Vision and the Visual System

Davida Y. Teller

Edited by John Palmer
December 2011

Editor's Forward

Davida Teller worked on this book until 2009. At that time her health prevented her from finishing and she passed away in 2011. The book was well along but incomplete. I am editing it to make it accessible to others.

This printing includes all of the text and draft figures. More to come. Many of the figures are hand drawn or modified from previously published figures with citation of the source. Permission to reproduce these materials will be requested when the manuscript is more complete.

The editing of this book was not done alone. The first round of copyediting, gathering of figures, and finding references was done by my student Serap Yigit-Elliott. Thanks Serap! It was supported by the Department of Psychology at the University of Washington at the direction of its chair, Professor Sheri Mizumuri. also thank my colleagues for their comments. And I thank my wife Zelda Zabinsky for her suggestions and eduring support. In addition, I thank Davida's husband, Tony Young, for his encouragement and help assembling the materials. Most importantly, I thank Davida Teller for her trust in allowing me to edit her work.

John Palmer
Seattle, Washington
December 2011

Preface

Vision science can be defined as the study of vision, the visual system, and the relations between the two. When we study vision, we use psychophysical and perceptual techniques to describe what and how well we see: how good are our spatial and temporal resolution? What is our color vision like? What are the properties of motion perception, form perception, object recognition? When we study the visual system, we use the techniques of neuroscience to describe the properties of the neural machinery – the optics, photochemistry, anatomy and/or physiology of the visual system – that makes seeing possible. And when we study the relationships between the two, we try to answer the question: How do the properties of the neural machinery leave their marks on the properties of perception?

In the terminology of this book, a Causal Story is an attempt to answer a question of this kind. For example, the statement that the shape of the scotopic spectral sensitivity curve is caused by the absorption spectrum of the photopigment rhodopsin is a (correct) Causal Story. So would be the statement that visual acuity is limited by the optical quality of the eye. (It would be incorrect, but a Causal Story nonetheless). I argue below that the most fundamental goal of vision science lies in the discovering and testing of Causal Stories.

Almost no one starts his or her intellectual life as a vision scientist. Most of us are trained in one or another of the classic disciplines: physiology, psychology, medicine, philosophy, physics, engineering, computer science, and so on. But eventually we may find ourselves working on a problem defined in the parent discipline, but whose answer impinges on the properties of vision and/or the visual system. One day our hearts suddenly beat faster with the insight that we might be able to contribute to the understanding of how people actually *see*. We're drawn to thinking of the visual system as a *system*, with vision as one of its high-level properties. We suddenly want to know how the topic we are studying fits into an understanding of the visual system as a whole. We're hooked – we've just become a vision scientist.

The conceptual base of vision science is remarkably varied and remarkably rich. This is partly because each new vision scientist brings along facts and concepts from his or her parent disciplines. Every few years new ways of thinking arrive from the parent disciplines. With a lot of intellectual work and many graduate seminars, the new concepts are eventually either found to enhance the old or are weeded out, and the discipline is the richer for it. An addiction to new concepts can keep a person in vision science for a lifetime.

But by the same token, vision science can be difficult for the beginning student to penetrate. This is partly because the conceptual base is so broad, and the factual base so extensive. But it's also because, as I mean to convince you, Causal Stories – explaining perceptual facts on the basis of neural facts – are a philosophically tricky matter.

In writing this book, I have had two goals: an initial goal which I chose, and a later one that forced itself upon me. The initial goal was to write a standard textbook – an introduction to vision science. I particularly wanted to weave together the concepts from the different parent disciplines. I wanted to make them mutually consistent, and accessible to beginners migrating into vision science from other disciplines. In particular, whenever a concept came up with which I had initially had particular trouble, I have tried to explain it in detail, with particular attention to the aspect with which I initially had trouble.

But as time and drafts went on, a second goal forced its way into the book. Again, the vision scientist's question is the question of Causal Stories: How do the properties of neural signals at the various levels of the visual system cause the properties of our perceptions? More deeply, where do hypotheses about Causal Stories come from, how are they tested, and by what criteria are they judged? How can we tell a good one from a bad one? I had done some prior work in this area (refs), and found myself inevitably drawn back to it.

As it turns out, there has been remarkably little explicit analysis of these questions in the vision literature. Consequently, Causal Stories can be difficult for the newly arriving student to evaluate. The second goal of the book, then, is to attempt to provide an extended, consistent analysis of the logic and the forms of argument used in vision science in general, and in Causal Stories in particular.

The format and content of this book are as follows. We begin with an explicit treatment of some of the kinds of propositions that enter into arguments about Causal Stories. Then, after an introduction to psychophysical techniques, we step through the visual system in the usual order – optics, photochemistry, photoreceptors, retinal processing, and so on. In each case, I provide at least a thumbnail sketch (and often a more extended treatment) of the properties and workings of the particular stage of processing. I then build upon this material to tell and evaluate one or more Causal Stories about how each level of visual processing leaves its marks on our perception.

Finally, a note about DT: As all of its practitioners know, science and philosophy are intensely personal passions. I find I can communicate that passion best to students by including personal anecdotes and making personal appearances in the book. But use of the first person in written work was beaten out of me in the third grade, and makes me uncomfortable still. Since the students in my lab have called me ”DT” for many years, DT has become my professional alter ego. She makes her presence known throughout the book. She feels free to express her opinions, and to suggest that the reader stop and think at certain points. Also, she feels free to just stop and *wonder* about things. Of course I do not claim that all the questions DT wonders about are original – surely most of them have been treated better by others. The goal is not to claim originality, but to model the sense of wonder that science engenders, and expose the students to the siren song of the next question down the road.

Davida Y. Teller (aka DT)
Seattle, Washington
September 2007

Acknowledgements

In 1970, Tom Cornsweet published an introductory text on visual science, entitled *Visual Perception.* (Academic Press). It is still read for its lucid accounts of the relationships between physics, physiology and perception. Although I have had to depart from his leisurely style of explanation because so much more is known by now, Tom's writing has nonetheless provided a model for this book.

Since 1995, several new books on vision science have been published by friends and colleagues. I have used them shamelessly as reality checks, and to educate myself on the parts of visual science that I knew the least about, and I think the authors for their contributions to my education:

Brian Wandell's *Foundations of Vision* (1995), Sinauer Associates,

D. Milner and M. Goodale's *The Visual Brain in Action* (1995), Oxford University Press,

Bob Rodieck's *The First Steps in Seeing* (1998), Sinauer Associates,

Bruce Goldstein's *Sensation and Perception* (1999), Brooks/Cole,

Clyde Oyster's *The Human Eye: Structure and Function* (1999), Sinauer Associates,

Stephen Palmer's *Vision Science: Photons to Phenomenology* (1999), MIT Press,

Mike Levine's *Fundamentals of Sensation and Perception* (2000), Oxford University Press ,

Martin Regan's *Human Perception of Objects* (2000), Sinauer Associates.

Special mention should also be made of the collections of papers on vision science in *The Cognitive Neurosciences* (1995) and *The New Cognitive Neurosciences* (2000), Michael Gazzaniga (ed), MIT Press.

Personal thanks must begin with Dr. Maureen (Mo) Powers, who started this book with me, and who produced early drafts of some of the chapters. Unfortunately, changes in her life led her to withdraw from the book early in the writing process. Mo brought activation energy and enthusiasm to the project, along with the conviction that writing a book was actually possible. Well begun is half done. (Well, not really, but it made all the difference.) Without Mo's enthusiasm, this book wouldn't have happened. Thanks, Mo.

I also especially thank my long-time colleague and friend, John Palmer, for challenging my thinking at many junctures over the years.

I also thank several colleagues for discussions and email conversations: Bill Newsome, Mike Landy (and others to be added....) and Tom Cornsweet, Temy Kennedy, John Palmer and Fred Rieke (and others......) for reading chapters of the ms. I also thank the students in my vision class who have read whatever chapters were available, and who annually rekindled my will to continue.

In my career I have had four mentors who most effectively challenged my intellect. They are the Gestalt psychologist Hans Wallach; the philosopher Michael Scriven; the engineer-turned-vision-scientist Tom Cornsweet, and the optometrist-turned-vision-scientist Gerald Westheimer. All communicated to me their passion for ideas. And their collective wisdom can be summarized

in two words: Think harder.

Finally, I thank my husband, Tony Young, for making the usual sacrifices an author demands of her family. I also thank him for using his skills as a photographer to provide some of the illustrations included in this book. We have had happy times searching the world for just the right image to illustrate one perceptual concept or another. The series of pictures on the development of vision in infants exists because of his creative skills and imagination.

Contents

Chapter 1

Introduction: The Domain of Visual Science

1.1 What is vision science?

Vision science is the study of vision, the visual system, and the relations between the two.

When vision scientists study *vision*, we study what and how well people see. The scientific goal is to describe and quantify our sensory and perceptual capacities – our capacities to respond to physical stimuli by using our eyes. What is the dimmest light we can detect? How fine are the finest details we can resolve? What is our color vision, or our stereovision, or our perception of motion like? How accurate are our perceptions of the sizes, shapes, and locations of objects? How well can we recognize objects? What can we see, and what can we *not* see?

When we study the *visual system*, we study the neural machinery – the optics, photochemistry, anatomy, and physiology of the eye and the parts of our brains that serve vision. How fine an image is made by the optics of the eye? How is light absorbed? Through what processing stages does incoming information pass? How is information recoded – what computations are performed – as we go from one stage of neural processing to the next? To what physical stimuli or aspects of the visual scene are neurons at different levels of the visual system tuned to respond?

What about the relations between the two? For many vision scientists, studying vision and the visual system separately is not the ultimate goal. Rather, the ultimate goal is to *explain why we see as we do, on the basis of the properties of the neural machinery that makes seeing possible.* Why can we detect some lights and not others; resolve some spatial details and not others; see objects accurately under some conditions but not others? Vision scientists want to understand how the computations carried out by the visual system both enable and limit our visual capacities, and how they leave their marks on our visual perception. We will call these attempted explanations *causal stories.*

1.1.1 Entities and mapping rules

Let us now put these questions in a slightly broader context. As shown in Figure 1.1, vision science is concerned with three kinds of entities, and three kinds of mapping rules.

Let us start with the three kinds of entities. The first kind is *physical objects*, or *physical stimuli* – the physical objects and light sources that send light to our eyes. The second is *physiological*

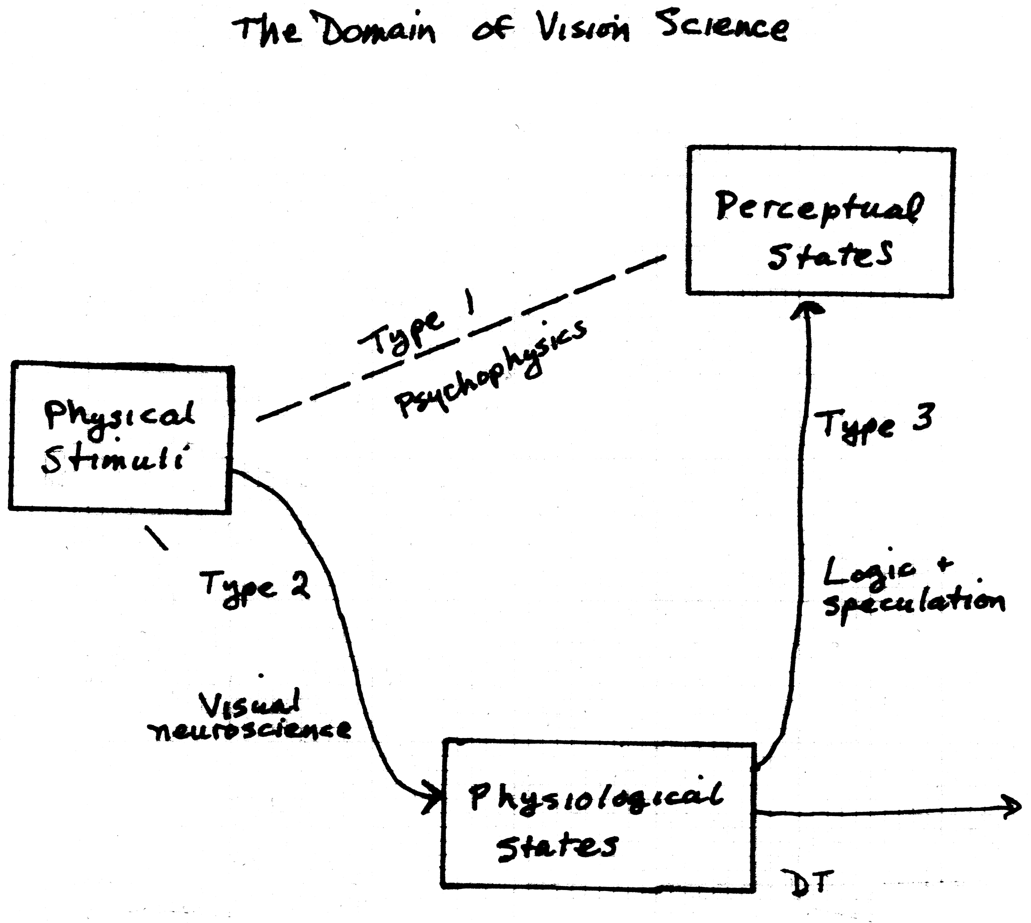


Figure 1.1: The domain of vision science. Three types of entities and three types of mapping rules make up the domain of vision science. The entities are physical stimuli, physiological (neural) states, and perceptual (phenomenal) states. The mapping rules are: Type 1, from physical stimuli to perceptual states; Type 2, from physical stimuli to neural states; and Type 3, from neural states to perceptual states.

states – the states of the many varieties of neurons in the visual system, occurring in response to the physical objects and light sources that lie in front of us. And the third is *perceptual* or *phenomenal states* – conscious states that usually correspond remarkably well to the objects and other stimuli in the physical world. These three kinds of entities are shown by the boxes in Figure 1.1.

Between each pair of entities there is a set of *mapping rules.* We are looking for rules of correspondence of the form, entities X1, X2 . in the physical domain occur in conjunction with entities Y1, Y2. in the neural domain, and with entities Z1, Z2. in the perceptual domain. The fundamental goal of vision science is to determine the mapping rules between each pair of entities, by whatever techniques are required, and however simple or complex these mapping rules might be. The three types of mapping rules are shown by the three arrows in Figure 1.1.

The three types of mappings are studied by three very different kinds of techniques. We study *Type 1 mappings* – mappings between physical stimuli and perceptual states – by means of the discipline of *psychophysics*, using sophisticated behavioral techniques to ask human subjects what they see when they view particular stimuli. We study *Type 2 mappings* – mappings between physical stimuli and neural states – with the techniques of visual neuroscience, such as presenting particular stimuli and recording the activities of particular neurons at various levels of the visual system.

What about *Type 3 mappings* – mappings between neural states and perceptual states? Suffice it to say that at this point the techniques used for exploring Type 3 mappings are much harder to define. At the same time, for many vision scientists, these are the heart of the matter, because as stated earlier, the ultimate goal of many vision scientists is to explain why we see as we do, on the basis of the properties of the neural machinery that makes seeing possible.

The world as perceived is a strikingly accurate representation of the physical world, allowing us both to perceive objects and to carry out appropriate motor activity with respect to them. But the existence of psychophysics notwithstanding, there are no direct causal mappings between physical and perceptual entities. The perceptual representations we have of the physical world are created by passing through the other two legs of the triangle: first through physical/physiological and then through physiological/perceptual mappings.

In the next few sections of this chapter, we will expand on each of these types of mappings. Then, in later sections of the chapter, we will expand at length on Type 3 mappings – mappings between neural states and perceptual states – because of their philosophical complexity and because of the fascination they carry for DT, the author of this book.

1.1.2 An example: Grating acuity

Let's take a concrete example of a physical/perceptual mapping. Figure 1.2 shows seven sets of regular black and white stripes, called *square-wave gratings*, and one homogeneous gray field. The stripes in each grating are half as wide as the stripes in the next coarser grating. At normal reading distance, you can readily see the spatial variation of light level across the coarser gratings (e.g., A-D). But what is the finest grating that you can see?

Somewhere between gratings E and G, your perception of black and white stripes probably fades perceptually into a uniform gray, and cannot be distinguished, or *discriminated*, from the homogeneous field H. Your *grating acuity* is defined as the finest stripes that you can just barely perceive as stripes, or discriminate from the homogeneous field on the basis of their spatial pattern. Grating acuity is a measure of the limit of detail you can resolve in space; it defines the limit of the *spatial resolution* capacity of your visual system.

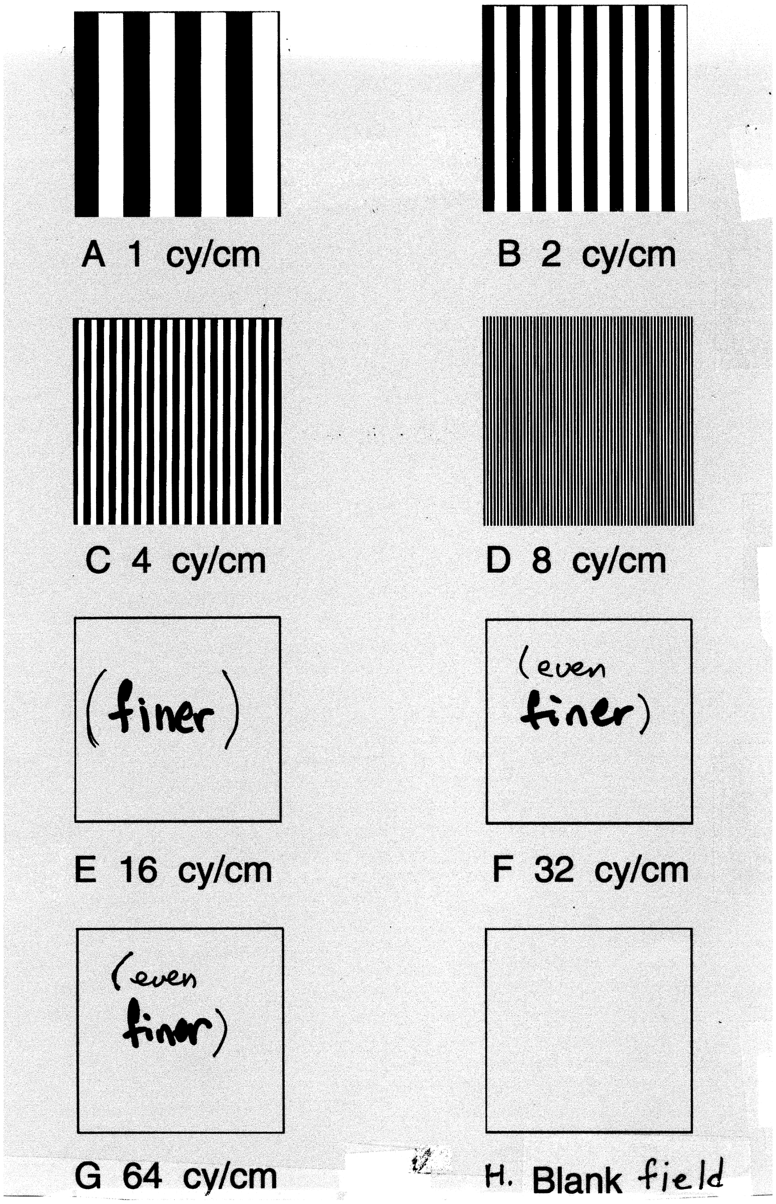


Figure 1.2: Seven square wave gratings and a blank (homogeneous) field. The gratins are labeled A, B, C, D, E, F, G. The blank field, H, approximately matches the gratings in average light intensity. If you have normal vision, at normal reading distance you will probably be able to resolve gratings A-xx, but not gratings xx-G. The finest grating you can discriminate from the blank field, H, defines your grating acuity.

We now elaborate on the three sets of entities and the three sets of mapping rules that make up the domain of vision science, using grating acuity as our example.

1.2 Three kinds of questions, three kinds of mapping rules

1.2.1 Type 1 – From physics to perception

What are Type 1, or physical/perceptual, mappings? Vision scientists want to know how and how well people see – to measure and quantify human sensory and perceptual capacities. To find out, we bring people (usually called *subjects* or *observers*) into the laboratory, and use well-controlled physical stimuli and sophisticated behavioral, or *psychophysical*, techniques to measure their visual capacities. The results of such experiments yield objective descriptions of the facts about visual acuity, color vision, distance perception, object recognition, and so on. We will look in detail at psychophysical techniques in Chapter 2 and 3.

The most immediately interesting fact about grating acuity is that it is so readily definable. There is a range of coarse gratings that you can resolve, an abrupt transition, and then a set of finer gratings that you can't resolve. Notice the first of many mismatches of properties between the physical and the perceived. The physical variation is continuous – there is nothing in the stimulus continuum that suggests a basis for any break of perceptual properties – but the change of perception from seeing to not seeing is abrupt.

At the perceptual level, grating acuity has several additional interesting properties. The finest grating you can see on the page varies with the viewing distance, the light level, and the part of your field of view in which the grating is presented. Try these experiments. First, prop up the book across the room, perhaps 20 feet away. From this distance, you will probably be able to resolve only the one or two coarsest gratings. Now walk toward the book. Every time you cut the distance of the book in half, you should be able to resolve one more grating.

Second, turn down the lights in the room in progressive stages, or prop the book up in the light of a window in the evening, when the outdoor light is steadily decreasing. As the light level decreases, your grating acuity will decrease, and you will need to move closer to resolve gratings that were easily resolvable at your original distance in full daylight. And third, instead of looking directly at the gratings, look above them by various amounts while still trying to resolve the stripes. The greater the *eccentricity* of the grating – the greater the displacement from your center of vision – the lower will be your grating acuity.

By picking out the finest grating you can see under a variety of conditions, you have just been a subject in an (informal) psychophysical experiment. You have made a series of measurements of the mappings from physical to perceptual entities. You have encountered the perceptual phenomenon of grating acuity, and you have learned about several important parameters – distance, light level, and eccentricity – that influence it.

1.2.2 System Properties: Bumblebees can fly!

Grating acuity and its variations with distance, light, and eccentricity can be called *system properties* of vision. Such system properties are interesting in and of themselves. But they become more interesting when we realize that system properties provide us with logically compelling information about some of the physiological properties of the visual system, without a single physiological experiment having been done. Oddly, we are arguing that perceptual results imply physiological conclusions. A fancier way of saying this is, system properties place important constraints on models of the visual system. The constraints depend on whether you (the subject) resolve the grating or not.

If you resolve the grating, information physically present in the stimuli must have been retained.

When you discriminate a grating from the homogeneous field, it follows that information that the two stimuli differ is retained from the physical stimulus, all the way through every one of the series of anatomical/physiological stages that make up your visual system. It is retained through every link of a *causal chain*, right up to your conscious perception, and right out through whatever motor system you use to tell the experimenter you resolve the grating. The question then becomes: How is the information carried, or *coded*, at each anatomical stage?

If you don't resolve the grating, information physically present in the stimuli must have been lost.

When a grating is present but you don't discriminate it from the homogeneous field, it follows that information that the two stimuli differ must be lost somewhere, at one or more stages of processing in your visual system. Like information retention, information loss implies a physiological conclusion – there exists a stage or stages of visual processing at which the information is lost. The question then becomes: at what anatomical stage (or anatomical *locus*) is the information lost? Questions of this kind are sometimes called *locus questions.*

Because they specify the limits of actual visual function, system properties like spatial resolution exert a great deal of power over mathematical and physiological models. A classic joke based on the importance of system properties concerns some early aerodynamic engineers who tried to make a mathematical model of the bumblebee, to find out how it flies. They tried and tried, but the model bumblebee fell out of the air every time. Eventually, the engineers concluded that bumblebees can't fly! But in fact, bumblebees can fly, so immediately we know that something about the model had to be wrong. Similarly, if a vision scientist makes a model of the visual system, and that model predicts that human beings can't resolve gratings as fine as the ones you could resolve in Figure 1.1, we know immediately that something is wrong with the model.

In more abstract terms: The system properties of vision are "black box" measurements made on a highly complex, multi-stage anatomical/physiological system, and rarely reveal the details of the machinery inside the box. But even in the absence of knowledge about the inside of the box, system properties put fundamental constraints on models of how it works. If the person can do X, then the underlying visual system must be such as to allow X to occur. Any model that claims that the person can't do X must be wrong. DT refers to arguments of this form as *"bumblebees can fly"* arguments. On the other hand, if the person can't do X, then information is lost, and there must exist a locus (or loci) of information loss.

In addition to such logically compelling implications, system properties often play a less formal but very important theoretical role in vision science. That is, system properties can encourage speculation and theory-building concerning information processing within the visual system. Such speculation, and the computational models it generates, can provide the motivation for visual neuroscientists to go and look for particular kinds of elements or processing circuits within the eye and visual system. We will see many examples of this pattern of argumentation, from psychophysics to physiology, as we go along.

1.2.3 Type 2 – From physics to physiology

In this section we provide a brief preliminary encounter with Type 2 mappings. We begin with a simplified cartoon of the anatomy of the eye and the early parts of the visual system.

Figure 1.3 and Figure1.4 show an overview of the early parts of the human visual system. As shown in Figure 1.3A, the optics of the eye create an optical image of the visual scene, called the *retinal image*, at the back of the eyeball. The major optical elements that form the image – the *cornea*, the *iris* (and the hole in its center, the *pupil*), and the *lens* – are shown in Figure 1.3B. The *retinal image* is focused on the *retina*, a thin sheet of neural tissue that lines the back of the eyeball. The *fovea* is a small, central region of the retina, specialized for high acuity. When you looked at each acuity grating in Figure 1.2, you turned your eyes to place your fovea under the retinal image of that grating.

Figure 1.4A shows a cartoon of a small piece of the retina, and some of the types of neurons it contains. The *photoreceptors* at the back surface of the retina catch the incoming light, and initiate a set of neural signals. Several other types of cells within the retina process these signals before they arrive at the *retinal ganglion cells.* The output processes (axons) of the ganglion cells make up the *optic nerve.* As shown in Figure 1.4A, the optic nerve passes out through the retina and eyeball, and leaves the eye [1].

As shown in Figure 1.4B, the optic nerve projects to a way station – *the lateral geniculate nucleus (LGN)*, deep within the brain – before signals are sent on to the *primary visual cortex*, a cortical region located at the very back of the brain. From there, signals project outward and forward to many higher levels of cortical processing (see Figures xx and xx in Chapter 18 for an intimidating preview).

When we ask about Type 2, or physical/physiological mappings, we are trying to trace the features of each physical stimulus through the visual system. We want to understand and quantify the characteristics of the eye's optics, and the means whereby light is transformed into physiological signals. We want to know the anatomy and physiology of each stage of processing in the early visual system and the cortex, and the computations – information losses, retentions, and recodings – that take place at each anatomical locus within this multistage information processing system.

1.2.4 Type 3 – From physiology to perception

But what about Type 3 mappings? From a vision scientist's perspective, the answers to Type 2 questions are made more interesting because the visual anatomy and physiology form the physical substrate of visual perception – these are the structures and processes that make seeing possible. But – here is a Type 3 question – which anatomical structures and which physiological computations

[1] Notice that, oddly, the retina is "in backwards" – the photoreceptors lie at the back (or outer) surface of the retina. In consequence, the light must pass through all of the other neurons in the retina before getting to the photoreceptors for absorption; and the optic nerve must pass through the retina to get out of the eye, creating a blind spot in your visual field. The probable reason for having the retina "in backwards" is that (as we will see) the photoreceptors are highly active metabolically, and need to be near a blood supply; and a blood supply that traversed across the front of the retina would itself get in the way of the retinal image.

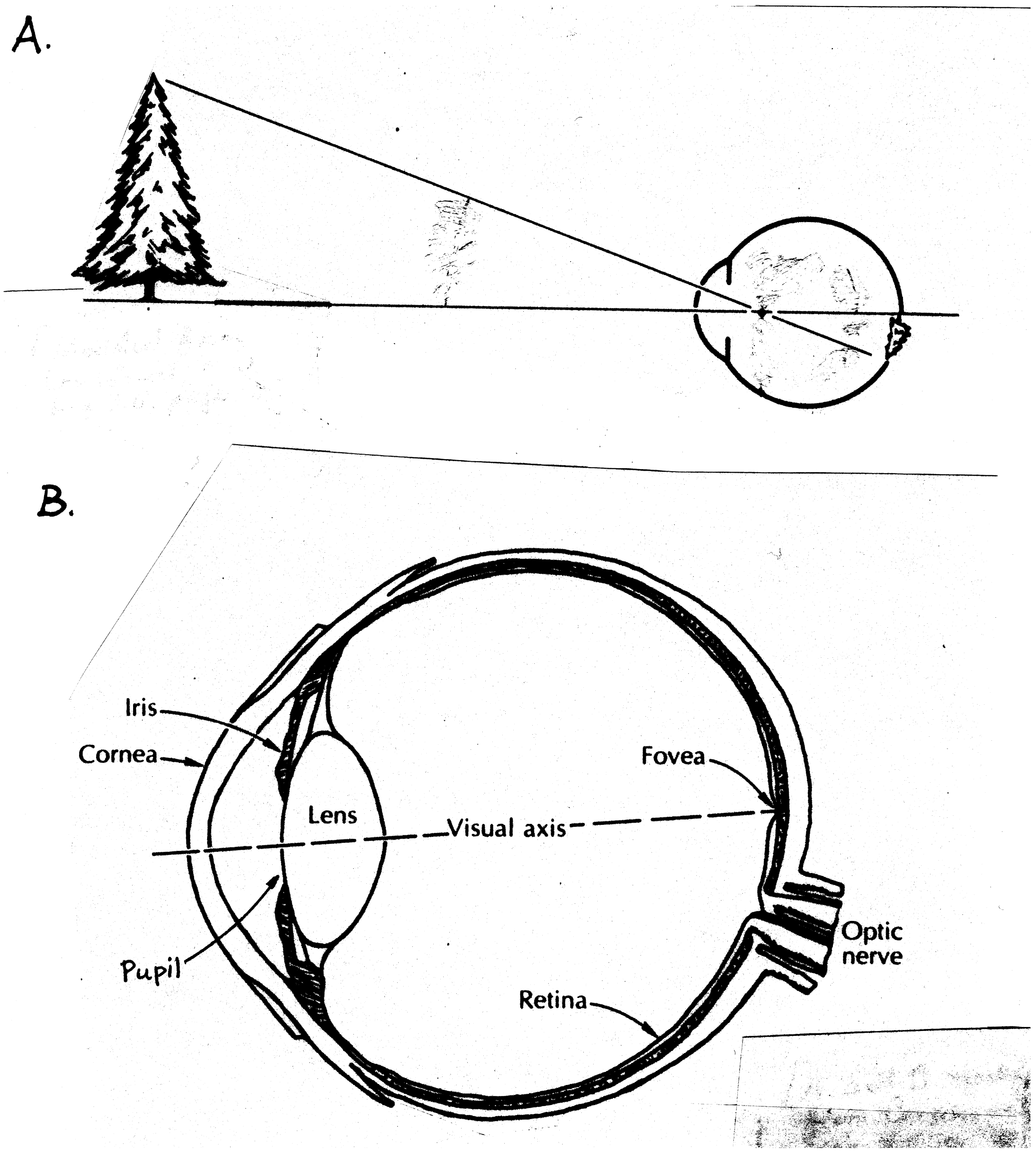


Figure 1.3: Overview of the eye and its optics. A. The optics of the eye form an image of the physical world on the back of the eyeball. B. This sketch shows a horizontal section through the right eye, labeling the major optical elements of the eye that work together to form the retinal image: the cornea, pupil, and lens. It also shows the retina, a thin sheet of neural tissue that lines the inside of the eyeball; the fovea, a central region of the retina specialized for high acuity; and the optic nerve, the nerve bundle that leaves the eye and carries visual signals toward the brain. [Modified from Cornsweet, 1970, Fig. 3.11, p.40]

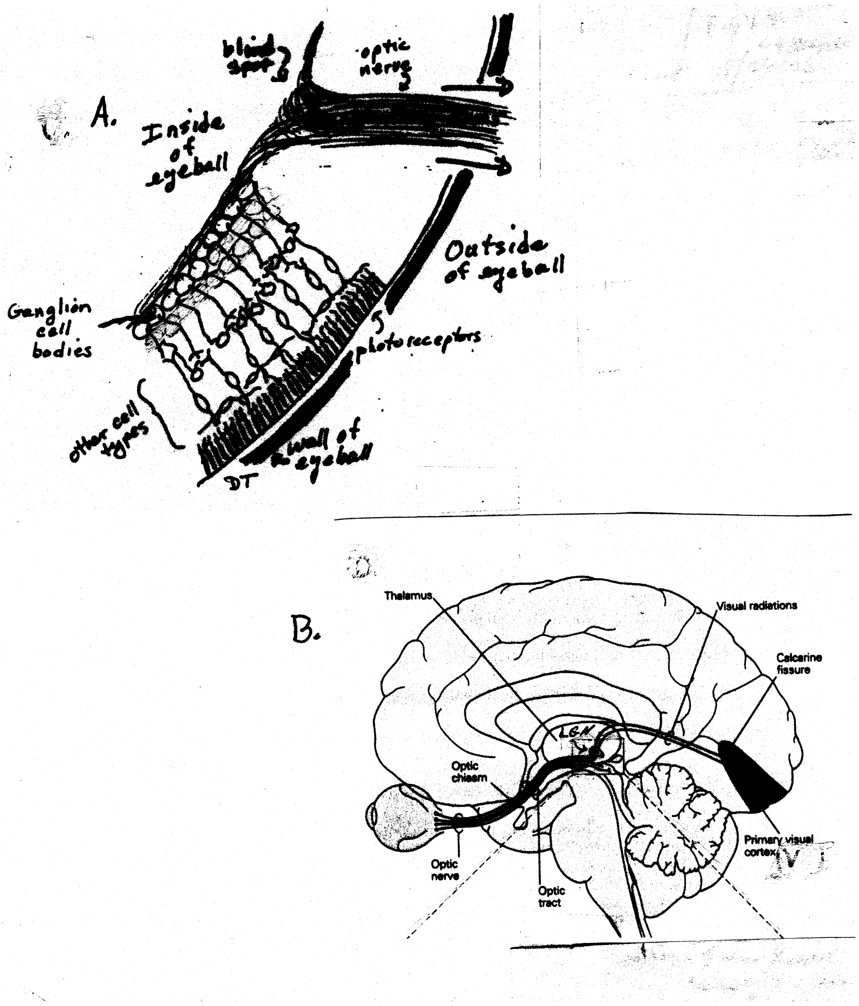


Figure 1.4: Overview of the visual system. A. A small section of the retina (schematic), showing its layers of neurons. The photoreceptors capture the light, and pass on neural signals (via several other types of neurons) to the ganglion cells. The output processes (axons) of the ganglion cells travel across the inside surface of the retina to join in a bundle, and plunge through the retina to form the optic nerve. B. A sketch of the pathway from the eye to the visual cortex. The eye is at the lower left. The optic nerve and the optic tract carry the signals sent by the ganglion cells to the lateral geniculate nucleus (LGN). Axons from the LGN project to the primary visual cortex and from there to other cortical areas (not shown) for further processing. [B modified from Kandel, Schwartz, and Jessell, 2000, Fig. 27.4, p. 527]

are most critical to producing any particular system property of perception? And how, exactly, do they produce it?

In order to give an example of exploring a Type 3 question in depth, we now ask: What limits grating acuity? At what stage or stages of visual processing is the spatial resolution limit, revealed by the system property of grating acuity, imposed on the incoming sensory signal? At which anatomical *locus* is the information lost? From Figure 1.3 and Figure 1.4 you can already intuit that the limit could be imposed at any of several stages.

In this section, we will raise four specific possibilities. We emphasize that the treatment of these possibilities is qualitative and intuitive at this stage – we just want you to end up believing that the answer is not obvious, and there are several very different and plausible candidate explanations. In later chapters, we will take up each of these possibilities in much more detail. In the meantime, just imagine you are looking out at Figure 1.2 through the neurons in your visual system, and trying to use the incoming spatial pattern of light or neural activity to tell the difference between a grating and a homogeneous field. Which level or levels of the processing system is the one that limits your spatial resolution, and what features of processing impose that limit?

1. The optics of the eye?

The black and white gratings in Figure 1.2 are patterns of light that exist in the physical world. When you look at a grating, the optics of the eye make an image of that grating on your retina. The process is straightforward, as shown in Figure 1.5A. Rays of light coming from a particular point on the grating leave that point and travel in straight lines in all directions. An image is formed because the optical system captures a subset of those rays, and (ideally) bends each ray just enough so that all of the rays that start at a point on the object are reunited at a point in the image. Neighboring points on the object are represented at neighboring points in the image, with the result that an image of the grating is formed at the back of the eye.

But real optical systems are not perfect. As shown in Figure 1.5B, in reality the rays from a single point on the object do not converge perfectly to a single point in the image. They are slightly spread out in the image, forming a small irregular blob. Technically, the optical image of a point of light is called a *point spread function* because it describes how much the rays from a single point in the object are spread over the image. Rays from neighboring points form neighboring point spread functions, and these imperfect blobs of light will soften the boundaries of the stripes of the grating in the retinal image.

Intuitively, what consequences would such optical imperfections have for grating acuity? A coarse grating will be represented faithfully, still recognizable and resolvable, but with slightly fuzzy edges. But we can intuit that when the stripes in the image of the grating are about equal to the width of the point spread function, the blobs from neighboring stripes will begin to overlap, so the stripes will become less resolvable. As the stripes become even finer, the overlapping point spread functions could eventually produce a homogeneous wash of light, and we wouldn't be able to tell the striped field from the homogeneous field in Figure 1.2.

Now, remember the argument you already know: bumblebees can fly. If you can resolve grating D (say) in Figure 1.2, you know immediately that the optics of your eye must be of at least sufficient quality to make a perceptible image of grating D on your retina. However, if you can't resolve grating E (say), you only know that information from grating E is lost somewhere within your visual system. The optical imperfections argument suggests intuitively that the optics *could*

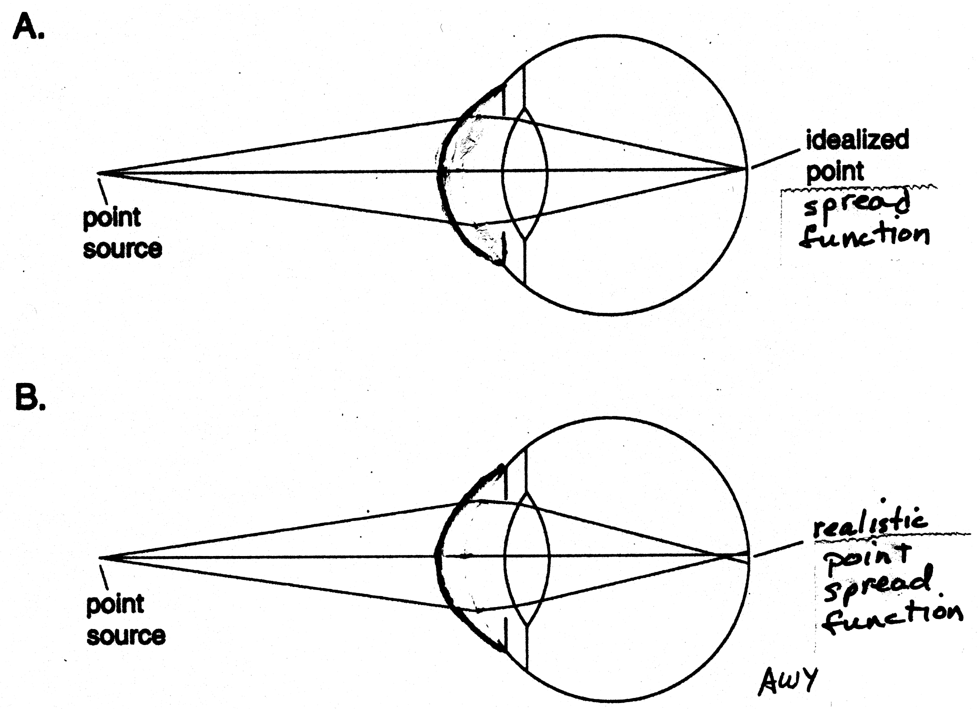

A.
idealized point spread function
point source
B.
realistic point spread function
point source
AWY

Figure 1.5: Idealized and realistic optical point spread functions. A. A point source and an idealized point image, drawn on the premise that all rays that leave the point source and enter the eye are perfectly bent by the optics, so as to end up at a single point on the retina. B. A more realistic, imperfect image – an extended "blob" of light – caused by optical imperfections in the eye. Some of the rays originating from the point source are bent too much or too little, so that they arrive near but not exactly on the idealized image of the point source. The distribution of light in the image of a point source is called a point spread function.

be the level that imposes the limit on your grating acuity, and prevents you from resolving grating E. To find out, we'd have to find a way to measure the actual quality of the optics of the human eye, and develop a quantitative theory of the effects of optical quality on vision. We will return to this task in Chapter 4 and 5.

2. Photoreceptor spacing?

Within the eye, at the back of the retina, lies a layer of tightly packed, highly specialized neurons called *photoreceptors.* The retinal image falls on the matrix of photoreceptors, and each photoreceptor captures the light from its particular region of the two-dimensional retinal image. However, the photoreceptor sums the light it catches over its whole extended region, and doesn't keep track of where each bit of light came from within that region. So the continuous retinal image is sampled piece-by-piece; that is, the photoreceptors perform a *discrete sampling* of the retinal image.

What consequences could discrete sampling have for grating acuity? As shown in Figure 1.6, each of the stripes in the optical image of a coarse grating covers many photoreceptors, so coarse gratings will yield variations of outputs across the *matrix* (or *mosaic*) of photoreceptors as a whole. By analyzing the spatial pattern of those outputs, we could readily tell that the input differed from that created by a homogeneous field. But a grating so fine that it puts more than one stripe on each photoreceptor is in danger of destruction, because it may yield no regular variation of output across the matrix of photoreceptors. In other words, a fine enough grating may produce the same spatial pattern of photoreceptor signals as does the homogeneous field of light, and thereby not be discriminable from it. At this point we might guess that the finest grating that gets through a discrete sampling matrix will have stripes just wide enough to put one stripe on each photoreceptor.

Again, bumblebees can fly. We already know that the limit imposed by discrete sampling at the photoreceptor mosaic can't be worse than the behaviorally measured acuity. But if the optics don't impose the resolution limit, the receptor matrix might. To find out, we'd need to know the sizes of the photoreceptors and their spacing, and make a quantitative model of the effects of discrete sampling. We return to this task in Chapter 5.

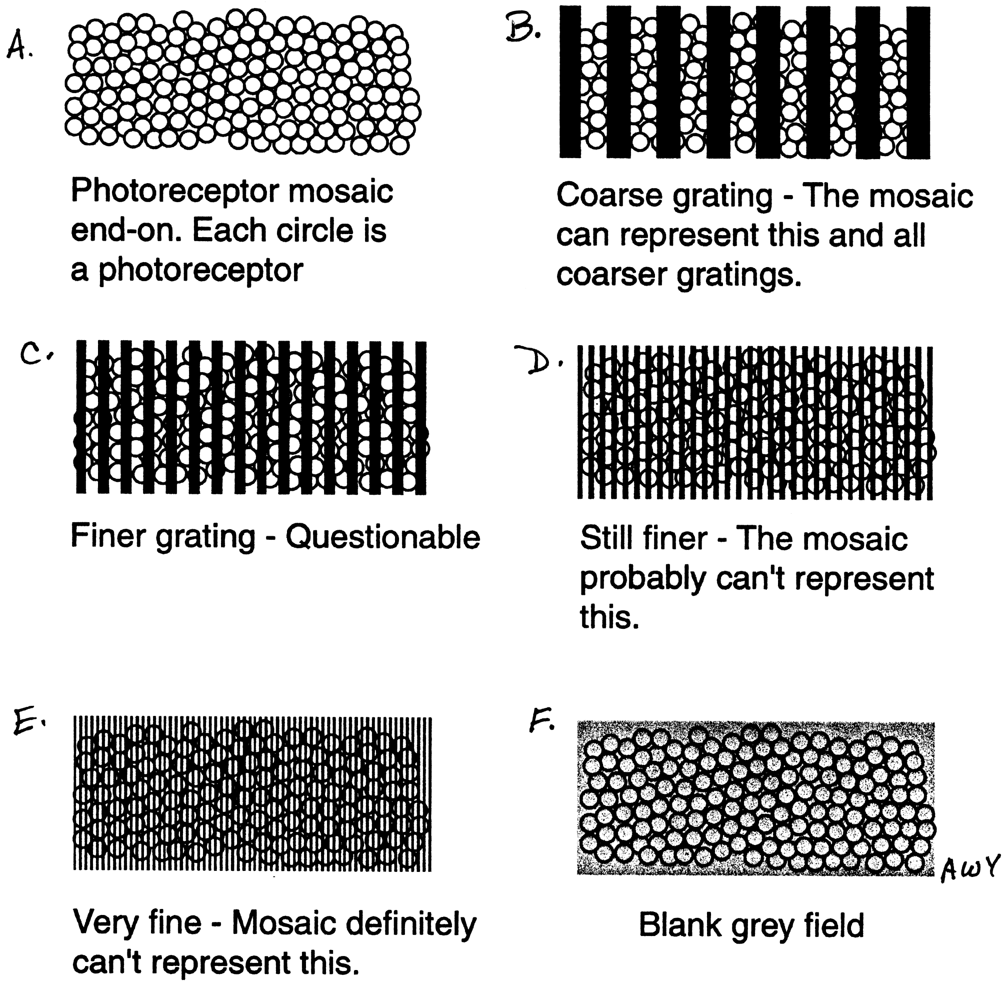
3. Neural convergence within the retina?

The concept of neural convergence has already been illustrated in Figure 1.4A: there are many more photoreceptors than ganglion cells. In fact, across the retina as a whole there are about 100 million photoreceptors but only about 1 million ganglion cells. That is, on average, over the retina as a whole, there is a 100:1 spatial *convergence* of neural signals. Intuitively, it is easy to imagine that unless special provisions are made, spatial resolution could be compromised here.

Bumblebees can fly. Even despite this average 100:1 convergence of photoreceptors onto ganglion cells, you already know that the ganglion cell layer, like all of the other layers of the visual system, must pass on information allowing us to resolve the finest gratings we do resolve. But if the optics and the photoreceptor spacing don't limit grating acuity, maybe neural convergence in the retina does. We'll look at this question again in Chapter 14.

4. Later levels of the visual system?

Beyond the retina, at the cortical level, there is a long series of anatomical stages and physiological recodings of visual information. In looking for the limits of grating acuity, our question about each



Photoreceptor mosaic end-on. Each circle is a photoreceptor

Coarse grating - The mosaic can represent this and all coarser gratings.

Finer grating - Questionable

Still finer - The mosaic probably can't represent this.

Very fine - Mosaic definitely can't represent this.

Blank grey field

Figure 1.6: Discrete sampling by the photoreceptor mosaic. A shows a schematic of the photoreceptor mosaic, viewed face on. Each circle is a photoreceptor. Each photoreceptor catches light from a small but spatially extended region of the retinal image, and sums the light over this region. B, C, D and E show images of four different square wave gratings, from coarse to fine. Intuitively, it seems likely that the grating in B will make a signal that varies systematically across the matrix of photoreceptors, but the finer gratings, especially the very fine grating schematized in E, might make only a homogeneous signal across the matrix. If so, the grating represented in E would not be discriminable from the blank field represented in F, and information about the spatial structure of the grating would be lost.

level would be the same. How does that level manage to preserve and pass on information about the finest grating we can resolve? How is information about the grating carried (or coded) at this level? And if the resolution limit is not imposed before this level, might it be imposed here, and if so, by what computational process? The second half of the book deals with these later levels of the visual system. [Before you go on, why not lay a bet as to which level imposes the resolution limit, and give the best justification you can at this stage for your answer.]

1.2.5 Causal Stories and Locus Questions

We now want to introduce two more terms that DT finds useful: *causal stories* and *locus questions. Causal stories* are proposed explanations of perceptual events on the basis of neural events. For example, the attribution of the grating acuity limit to the optics of the eye, or to the properties of the retinal mosaic, or to a combination of both, would all be causal stories. Causal stories can be speculative, or they can be argued on the basis of quantitative theory.

Similarly, *locus questions* take the form: where within the neural information processing system is information lost, or importantly recoded, in such a way as to bring about the correspondence between physical stimuli and perception? The four options for the locus of information loss in grating acuity represent four possible answers to a locus question. The usefulness of these concepts will become more meaningful through examples encountered throughout the book.

1.3 Design questions

Finally, a fourth type of question – design questions – is worthy of mention. Design questions are *why* questions. These questions are concerned with why human vision and the human visual system take the form they take. For example, why is our grating acuity as good as it is, and why is it not better? Design constraints are imposed from many sources: the laws of physics, the physiological properties of neurons, and the effectiveness of various visual coding schemes for various purposes. In addition, the design of the visual system is shaped by the competing evolutionary pressures that combined to shape the organism as a whole. Many competing pressures have acted upon the design of the visual system, and the current features of the visual system are doubtless historical compromises among these pressures. The answers to design questions are usually speculative, but often instructive and interesting as well.

The fundamental Design question about grating acuity is: What factor or combination of factors necessitates that visual resolution have the limits that it has? Is it that the optics can't be any better? Or the photoreceptors can't be any smaller? Or there can't be any more ganglion cells? Or that there is a constraint imposed on some later level of the system? Or might it be that no one level is to blame for the spatial resolution limit, but rather that the limit is imposed by conflicting design necessities, and many levels of the system conspire to impose this limit in a more complex way? What would have to be changed in order to improve our acuity by a factor of two, and what would be the cost? [Again, before you go on, write down your guesses about the answer to these Design questions.]

1.4 The mind/brain problem

We now return to Type 3 mappings – mappings between physiological and perceptual entities. To begin, we must turn briefly to the philosophy of mind. Vision science can be particularly perplexing from a philosophical perspective, because it seems as though with Type 3 questions, vision scientists hope to explain mental events (visual perception) on the basis of physiological events (neural activity). This hope brings us to close encounters with the mind/body or mind/brain problem.

For centuries philosophers have argued about the nature of the relationship between mind and brain. In particular they have argued about whether mind and brain are a single physical entity (a position called *materialism*), a single mental entity (a position called *idealism*), or two separate entities (a position called *dualism*); and if two, whether one of the two holds a causal priority over the other. Many variants of each of these positions have been formulated, and the debate continues to fascinate philosophical audiences across the centuries. (For recent treatments, see Chalmers, 1996 and Metzinger, 2000).

Most vision scientists are probably most comfortable with a materialist perspective. That is, most of us probably believe, implicitly if not explicitly, that between mind and brain, the brain is the primary causal agent. Moreover, perceptual events become less mysterious when they are viewed simply as high-level properties of the brain. To support this view, an analogy can be made between the properties of chemicals and chemical compounds, and the properties of brains and conscious states. Just as water can be viewed as a high-level property of hydrogen and oxygen, so a conscious perception can be viewed as a high-level property of a complex neural network. [Learning this argument scratched a huge and persistent itch for materialist DT. Does it do the same for other vision scientists, such as you?]

Taking the argument further, some philosophers make use of the concept of *emergent properties.* An emergent property of X can be defined as a high-level property of X that cannot be predicted, either in practice or in principle, from the characteristics of the lower- level elements from which X is made. Given this definition, some philosophers argue that consciousness is an emergent property of complex neural networks. But the concept of emergence is itself controversial, and many vision scientists and philosophers would sharply disagree with its usefulness in the mind/brain debate.

1.4.1 Finessing the mind/brain problem: Mapping rules

Given the inevitable continuation of these debates, it seems to DT that rather than developing a science that depends upon a single view of the mind/brain problem, vision science would be wise to finesse it. That is, we should try to find a formulation of the questions of vision science that will be robust, and survive across many or all of the different philosophical stances on the mind/brain problem.

In 1970, Ewald Hering (of whom you will hear more later), lecturing at the Imperial Academy of Sciences in Vienna, laid out the classic finesse:

”*If then, the student of neurophysiology takes his stand between the physicist and the psychologist, and if the first of these rightly makes the unbroken causitive continuity of all material processes an axiom of his system of investigation, the prudent psychologist, on the other hand, will investigate the laws of conscious life according to the inductive method, and will hence, as much as the physicist, make the existence of fixed laws his initial assumption.... it only remains for him to make* *one more assumption, viz., that* ***this mutual interdependence between the mental and the material is itself also dependent on law****, and he has discovered the bond by which the science of matter and the science of consciousness are united into a single whole....*(emphasis DT)

"*This, then, by no means implies that the two variables above mentioned – matter and consciousness – stand in the relation of cause and effect...to one another. For on this subject we know nothing. The materialist regards consciousness as a product or result of matter, while the idealist holds matter to be the result of consciousness, and a third maintains that mind and matter are identical; with all this the physiologist, as such, has nothing whatever to do; his sole concern is with the fact that matter and consciousness are functions one of the other.*"

In other words, Hering argues that we can set aside the philosophical problem, and get on with finding the lawful relationships (or *mapping rules*) that he assumes to hold between neural and perceptual states. We will adopt this perspective for the purposes of this book. [Does Hering's declaration provide a satisfactory finesse of the mind/brain problem? How would vision science be different if different vision scientists took different stances in regard to the mind/brain problem?]

1.5 Linking Propositions

We now turn to another of the major philosophical themes of this book: the topic of *linking propositions.* Let's take the next logical step beyond Hering's assertion that lawful relationships – mapping rules – exist between visual perception and visual neurophysiology. Can anything more be said about the properties of these mapping rules? The question isn't just, do neural states map to perceptual states? It's *which neural states map to which perceptual states?*

1.5.1 Mueller's axioms of psychophysical correspondence

Interestingly, an elaborate set of mapping rules was explicitly formulated right at the beginning of the discipline of psychophysics. The 19th century scientists who founded the discipline were motivated not just by an interest in sensations and perceptions, but also by a desire to use perceptual observations as a tool for drawing inferences about the workings of the brain. They argued that perceptual (mental) events and brain (material) events were of two different kinds, described by language from two different realms of discourse. Therefore, if conclusions about brain events were to be drawn from facts about perceptual events, some kind of special linking statements would be needed. Their attempts to specify the necessary arguments were concisely formulated by G.E. Mueller in 1896, and are known as *Mueller's axioms of psychophysical correspondence.*

Mueller's first three axioms have been translated as follows (Boring, 1942, p. 89):

"*1. The ground of every state of consciousness is a material process, a psychophysical*[2] *process so-called, to whose occurrence the presence of the conscious state is joined.*

2. To an equality, similarity, or difference in the constitution of sensations...there corresponds an equality, similarity, or difference in the constitution of the psychophysical process, and conversely. Moreover, to a greater or lesser similarity of sensations, there also corresponds respectively

[2] Notice that as used by Mueller (and his translator) the term *psychophysical process* was used to mean a special variant of a physiological process – a physiological process to which a conscious state is joined. But in more modern writings, the term *psychophysics* always refers solely to methods for describing and quantifying sensations and perceptions (Chapters 2 and 3). Other terms, such as *neural correlate of consciousness*, are used to refer to Mueller's so-called psychophysical process.

a greater or lesser similarity of the psychophysical process, and conversely.

3. If the changes through which a sensation passes have the same direction, or the differences which exist between series of sensations are of like direction, then the changes through which the psychophysical process passes, or the differences of the given psychophysical process, have like direction. Moreover, if a sensation is variable in ***n*** *directions, then the psychophysical process lying at the basis of it must also be variable in* ***n*** *directions, and conversely.*"

Mueller's axioms have several important properties. First, notice that Mueller called these linking statements *axioms* – statements that could not be proved, but that had to be assumed to be true if the discipline was to be pursued. With these axioms in place, one could use perceptual facts to deduce some aspects of the workings of the brain. Today we would be more likely to call these statements premises or assumptions. DT' argues that they are a special class of assumptions, and they enter into all claims about physiological/perceptual mappings in vision science. Moreover, they have a huge impact on the science we choose to do, and they govern the kinds of arguments that we entertain.

Second, notice that the first axiom differs fundamentally from the second and third. Mueller's first axiom states the very general premise that all perceptual processes arise from material processes, with the material process taking causal priority. However, it says nothing further about the forms these neural/perceptual correspondences might take. So far, the possibility is left open that any material process, or brain state, could give rise to any mental process, or perceptual state. In the mapping of physiological states to perceptual states, chaos could reign. Remember, however, that Hering earlier rejected this option, preferring to assume that " *this mutual interdependence between the mental and the material is itself also dependent on law*." Since the rest of the universe is lawful, it makes sense to DT to assume that mappings between neural and perceptual states are lawful too.

The second and third axioms, in contrast, are specific lawful relationships that might be assumed to hold between perceptual and neural states. Mueller recommends the assumptions that identical perceptual states imply identical neural states, and vice versa; similar perceptual states imply similar neural states, and vice versa; and so on through a longish list.

Statements like these are sometimes called *isomorphisms* – (assumed) *similarities of form* – between neural and perceptual states. More specifically, DT calls the mapping rules in Mueller's second and third axioms *relational isomorphisms.* Notice that in a relational isomorphism, perceptual states are compared to perceptual states, neural states to neural states; and the isomorphism is that the same *relationship* that holds between perceptual states is assumed to hold between neural states (cf. Coombs, Dawes, and Tversky, 1970). The propositions are that identical perceptual states imply identical neural states, and vice versa; similar perceptual states imply similar neural states, and vice versa; and so on.

The discussion of the possible limits on grating acuity (above) illustrates the use of relational isomorphisms. If you review each of the four cases, you will find in each case the assumption that due to a particular spatial processing imperfection within the visual system, as the grating gets finer and finer, the spatial distributions of signals produced by the grating and the homogeneous field become more and more similar, and eventually become *identical.* The linking proposition that enters each argument is that an identity of neural processes creates an identity of perceptions; and vice versa, the identity of the two perceptions implies that the neural identity has been reached.

In general, relational isomorphisms are not logical necessities. Two identical perceptions could in principle arise from different brain states if the mappings from brain states to perceptual states were chaotic, or if they were many:1 instead of 1:1 (Teller and Pugh, 1983). Other rules of relational isomorphism, such as those arising from similarity, are also easily challenged. The relational isomorphisms that enter into our beliefs about physiological/perceptual mappings are premises, not logical necessities. But they are certainly convenient!

1.5.2 Updating Mueller's first axiom: The Universal linking proposition

There has been surprisingly little work on the axioms of mental/material correspondence since 1896. Brindley (1960) treated the topic briefly, proposing the name *linking hypotheses* as a name for these rules of correspondence. Since they are rarely actually hypotheses, DT (Teller, 1984) suggested the name *linking propositions.* DT defines a linking proposition as *a claim that a particular mapping occurs, or a particular mapping principle applies, between neural and perceptual states.* She argues that linking propositions lie at the heart of vision science. Any causal story – any attempt to explain perceptual events on the basis of neural events, or vice versa – will necessarily include a linking proposition. Moreover, linking propositions are often implicit rather than explicit, and part of the fun of vision science is ferreting them out and examining them.

In thinking again about Mueller's first axiom, it makes sense to DT to reformulate it for modern times. The first axiom can be called the *Universal linking proposition: All perceptual states and processes are implemented in neural states and processes.* Most vision scientists would doubtless endorse this premise, because the alternative is to argue that perceptual states can exist without being accompanied by neural states. Notice, however, that the exact words that one uses to phrase the Universal linking proposition would vary with one's position on the mind/brain problem. Shall we say, perceptual states and processes are *implemented* in neural states and processes; or *arise* from them, or are *enabled* by them, or *emerge* from them, or that neural states and processes *cause* perceptual states and processes; or vice versa? [As you begin to read the vision literature, watch for these variations of meaning.]

1.5.3 Relational isomorphisms

Mueller's remaining axioms can be called the *Propositions of Relational Isomorphism.* The premise is that particular relational isomorphisms exist between perceptual states and their neural implementations.

Beyond Mueller's axioms, there are many other varieties of linking propositions. We briefly articulate two more – analogies and computational linking propositions – in the next two sections, before returning to a more general discussion.

1.5.4 Analogies.and Nothing mucks it up

The term *isomorphism* has also been used in a second way in the context of linking propositions (cf. Pessoa, Thompson, and Noe, 1998). DT calls this second category of isomorphisms *Analogies.* These are isomorphisms that bridge between perceptual and physiological domains, by making an analogy between some aspect(s) of perceptual and neural states. For example, think about your perception of a set of broad black and white stripes, such as those in Figure 1.2A. In speculating on the neural state that underlies this perceptual state, you might assume (or show) that there is a region of the retina across which the firing rates of neurons take on a similar pattern – a set of neurons firing more slowly, say, to provide the neural correlate of a black bar, and a similar set of neurons, displaced by the equivalent of one bar width, firing more rapidly to provide the neural correlate of a white bar. The plots of whiteness/blackness against space, and firing rate against space, would look very similar. In other words, there is a visual similarity, or *analogy* between the perceptually and neurally defined patterns.

An interesting feature of Analogies as linking propositions is that they often enter vision science as nothing more than two pictures that look alike. But typically, many pieces need to be added to the argument to make a compelling theory or explanation of the perception on the basis of the neural activity. In fact, DT argues that statements that she calls "*Nothing Mucks it Up*" provisos must be involved, implicitly if not explicitly. These assumptions are of the form that nothing within the visual system, between the neural pattern and the perception, interferes with the control of the particular neural pattern over the perception. In general, the earlier in the visual system the neurons on which the analogy is based occur, the more complicated and tenuous the required Nothing Mucks it Up provisos would seem likely to be.

1.5.5 Computational linking propositions

Another interesting set of linking premises might be called *computational linking propositions.* That is, what computations would you be willing to assume can take place within the mapping between neural and perceptual states? Must this mapping always be simple or 1:1, or might computations be parts of the mapping rules? And, how fancy are the complications we will allow?

Here's an example. Gustav Fechner, one of the very earliest psychophysicists, was struck with the idea that it might be possible to discover a mathematical formula that would specify the mapping of physical intensity to perception. Applying Fechner's argument to the case of brightness, suppose that brightness grew, not linearly with physical intensity, but with the logarithm of physical intensity. (This is approximately true in some cases, but not in others. But suppose it were true.) The question is, must the logarithmic transformation take place within the physiological system, so that the mapping from neural state to perceptual state is always linear; or might the growth of neural signals with intensity be linear throughout the visual system, and the logarithmic transformation come about in the mapping from neural to perceptual states? Vision scientists might well differ in their premises here.

1.5.6 From the sublime to the ridiculous

To DT, an interesting property of linking propositions is that they are often implicit. But once a linking proposition is made explicit, there is often a surprisingly good consensus among most vision scientists about its acceptability as a premise. At the one extreme lie some relational linking propositions, like Mueller's axiom of identity. Most vision scientists would probably regard this proposition as clearly true perhaps even analytically true or tautological (cf. Brindley, 1960). We will return to elaborate it further in Chapter 2.

At the other extreme, there are some candidate linking propositions that we would doubtless all reject. For example, we would probably be uncomfortable with the assumption that the neural code for seeing a three-dimensional object must be (literally) a neural circuit with the same three-dimensional shape as the object, somewhere within the brain. We would doubtless deny that a neuron that signals redness would have to be literally red (but then, what would a redness neuron have to be like?). And we would think it silly to argue that when we perceive a dance we must do so with dancing neurons, as shown fancifully in Figure 1.7.

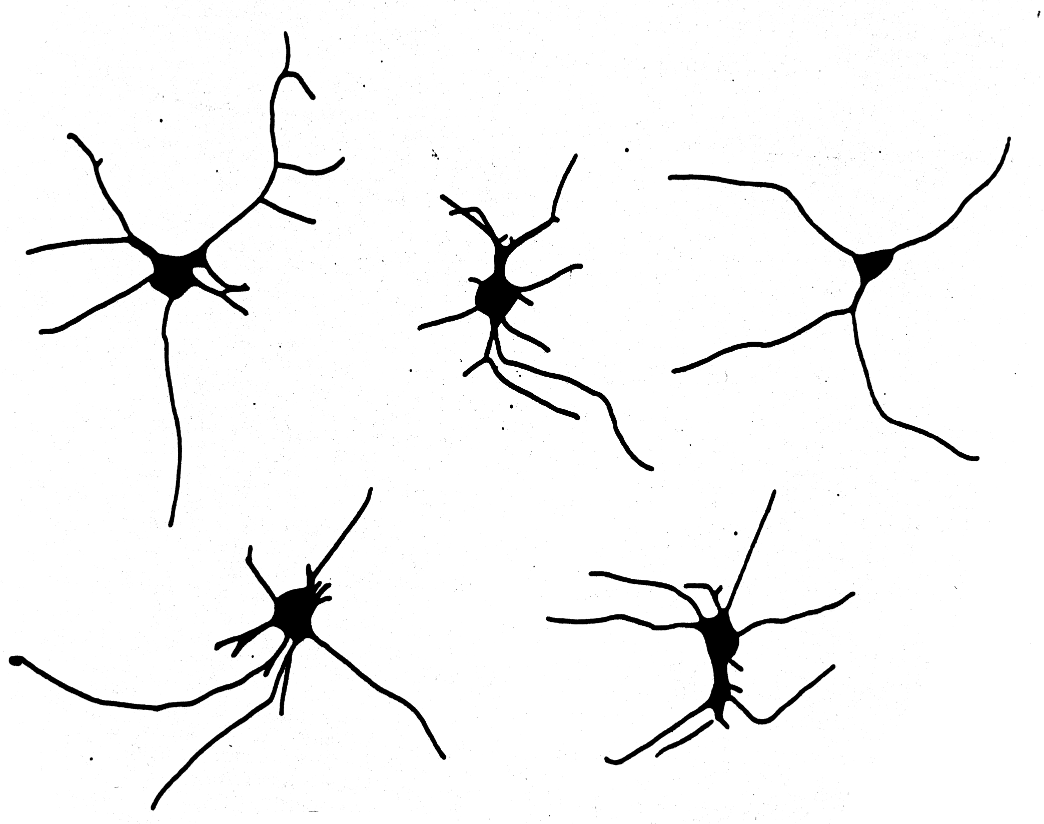


Figure 1.7: Dancing neurons? [From Gazzaniga, 1997, fronticepiece]

These examples are chosen because they define the ends of a continuum from high acceptability to silliness. But the credibilities of other kinds of linking propositions, such as analogies and computational linking propositions, fall between the two extremes, and are worth some thought.

The moral of the story is this. Whenever a causal story is proposed, a linking proposition lies within it. Is it a linking proposition that most vision scientists would readily accept, or readily discard? Would there be a consensus, or might different vision scientists differ in the kinds of linking propositions they are willing to incorporate as premises into their causal stories?

1.6 A more perceptual perspective

Historically, vision has been studied from two different perspectives. The first is a sensory perspective, in which vision scientists tend to emphasize the simpler aspects of vision, and (in general) attempt to account for them on the basis of the codings and recodings of information that take place within the early processing stages of the visual system. The second is a perceptual perspective, in which we tend to emphasize the more complex aspects of perception, and (in general) attempt to account for them on the basis of the more complex and higher level aspects of visual processing.

So far in this chapter we have been taking a classically sensory approach to vision. But let's switch briefly to a more perceptual approach. From the perceptual perspective the interesting parts of vision science lie not in the sensory details such as grating acuity, but rather in the complex system properties of perception. And the most fundamental phenomena are not illustrated by drawings on the pages of a book, but by looking at the world around you. [Look at the world around you!]

As you look around, you see a scene that contains three-dimensional objects of particular sizes, shapes and colors in particular three-dimensional locations. These objects may move, but their essential characteristics of size, shape and color tend to remain constant across viewing conditions (that is, we as perceivers have good size, shape, and color *constancy*). You are often able to recognize objects across time and across contexts. The perception scientist wants to describe and quantify these more complex system properties, and to understand the properties of the (presumably high level) neural processes that make these high level features of perception possible. We will argue later, for example, that incoming sensory information must be combined with stored information and processed with complex computational algorithms before it can provide the neural basis of high level visual perception.

Until relatively recently, physiological study of high level visual processing was still in its infancy. Since little relevant information was available, many perception scientists were not much interested in knowing the details of information processing within the visual system. In fact, some have denied the value of understanding anatomy and physiology for understanding perception.

But this situation is changing. By now, a great deal of information about the anatomy and physiology of cortical processing – Type 2 mappings – is well established, as we will show. Moreover, in the past few years a number of vision scientists have undertaken studies of single neurons, searching for the neural basis of particular perceptual processes and phenomena. As this knowledge comes in, perception scientists are beginning to invent causal stories about perceptual events and processes based on neural events and processes. Moreover, recent, more global analyses arising from neural imaging techniques such as functional magnetic resonanace imaging (fMRI), have also drawn the interest of perception scientists toward neuroanatomy and neurophysiology, and toward causal stories relating perception to neurophysiological activity.

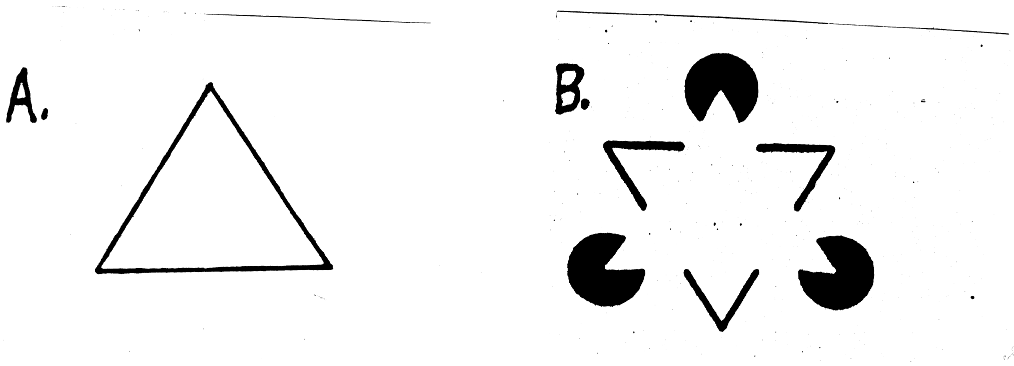


Figure 1.8: Illusory contours. The continuous contours at the left and the illusory contours at the right both give rise to similar perceptions of a triangle. Why?

1.6.1 Perception and linking propositions

Let us develop a couple of examples of linking propositions that might be involved in causal stories in high level perception. First, take the case of the triangles illustrated in Figure 1.8. This figure illustrates the phenomenon of *illusory contours.* The perception of a set of three borders can arise from at least two very different stimuli: three physical borders formed from solid lines; or a set of ”pacmen” at three corner locations. And in both cases, similar triangles can be perceived. Suppose we were to decide to search for the neural cause of this odd perceptual similarity. Where would we start, and why?

One linking proposition we could adopt would be a similarity (or identity) proposition: that within the visual system, there will exist neural elements that respond similarly (or maybe identically) to these two stimuli – the lines and the pacmen. Such a speculation will lead us to examine individual neurons at various levels of the visual system, and test them with both kinds of stimuli. The goal would be to try to find such neurons, and to locate the earliest level at which they occur. But the whole enterprise rests on a relational isomorphism: the premise that the two similar perceptual states – the two perceptions of a border – indicate the presence of two similar states of the same neurons arising from the two physical stimuli. [Is this an isomorphism you would endorse? Is it a logical necessity, a reasonable speculation, or just a silly argument?]

And as a complex example, we ask the same question about what DT calls the *Ultimate Code*: Is there a form that the neural code may or must take, in order to give rise to the conscious perception of an object or a scene? Some of you have undoubtedly already adopted the premise that every time we see the same object (famously, our grandmother, as we will see), there is a neural stage at which the same individual high-level neuron or small set of neurons must be active in the same way (cf. Barlow, 19xx). Others would reject this specific premise, but argue that some form of isomorphism must be involved in the mapping from neural to perceptual states, even though we cannot at present say what it will be. And yet others would argue that, much as we might prefer the universe to be otherwise, for complex perceptual phenomena there will turn out to be no consistent, definable isomorphism between perceptual and neural states.

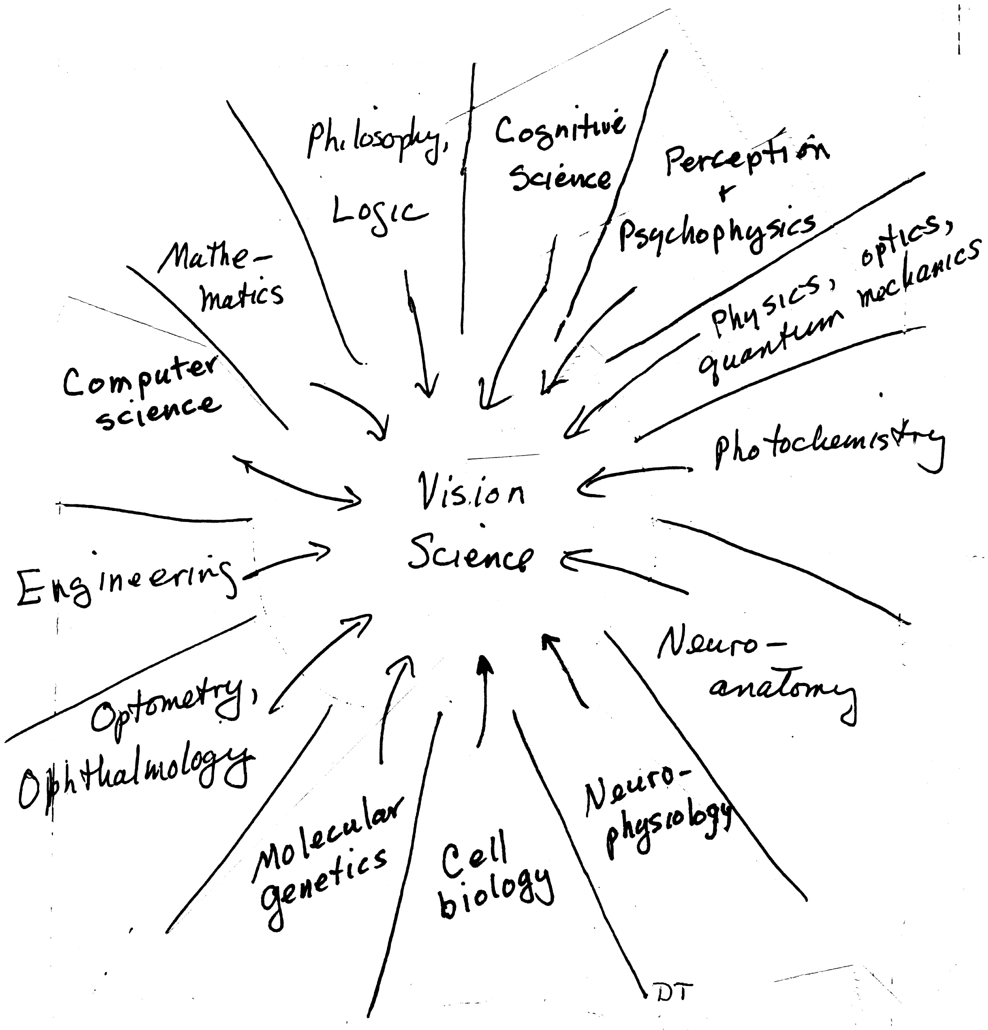


Figure 1.9: The major disciplines that contribute to vision science.

DT hastens to add that the problem of the Ultimate Code will not be solved in this book! And yet it provides a high-level example of the kinds of questions that initially attract many scientists and philosophers to the field of visual perception. It also illustrates the point that different implicit linking propositions lead to different physiological predictions and influence the choice of experiments a scientist undertakes.

1.7 An interdisciplinary field

It should be obvious by now that vision scientists cannot afford to respect the boundaries between classical scientific disciplines, much less be chauvinistic toward any of them. Instead, we first define the questions of interest, as we have done in this chapter. Then we look around to see what kinds of classical disciplines can provide us with the expertise we need to address our questions. Specialists of many different kinds are invited – we need all the help we can get! The disciplines that unite to form the field of vision science, and the expertise we need them for, include at least the following.

Psychophysics and perception:

To describe and quantify the system properties of vision.

Physics:

To describe the nature of light and the optical quality of the eye.

Photochemistry:

To describe the interaction of light with matter in the photoreceptors.

Neuroanatomy:

To describe the structure of the various parts of our visual systems.

Neurophysiology:

To describe the information processing characteristics of individual neurons and neural circuits at each stage of the visual system.

Cell biology:

To characterize the internal workings of the cells, and their mechanisms of communication.

Molecular genetics:

To describe the genetic control of the various parts of the visual system (at this stage, particularly the photopigments that capture light within the photoreceptors).

Optometry, ophthalmology:

To describe the disorders that can occur in vision and in the visual system, and use them to help us understand normal vision and the normal visual system. Also, to import the accumulating knowledge about vision and the visual system into medical practice, in order to help patients with vision problems.

Engineering:

To describe information processing systems; to provide conceptual and mathematical tools for describing complex systems and their properties.

Computer science:

To discover design principles and computational algorithms that might help us understand information processing within the human visual system.

Mathematics, statistics:

To provide tools for modeling the three kinds of mapping rules.

Philosophy, logic:

To provide logical analyses of our most basic scientific concepts.

Cognitive neuroscience:

To provide descriptions of the cognitive processes that affect perception, and models of the mechanisms that underlie it.

In fact, over the years specialists from all of these fields have been drawn into vision science. As a consequence, the field is conceptually very rich and sophisticated, and new ideas are always arriving from different sources. For DT, the excitement has lasted a lifetime. Everyone in the field comes from somewhere else; everyone has some areas of deep expertise, and some areas where he or she is an amateur.

1.8 Summary: The domain of vision science

In this chapter we defined vision science as the study of vision, the visual system, and the relations between the two. We argued that vision science spans three domains: physical stimuli, neuophysiological states, and perceptual states; and that vision scientists are interested in the mapping rules among these three domains.

The questions in which vision scientists are interested were illustrated by using the example of *grating acuity.* We defined grating acuity as a Type 1 phenomenon – a physical/perceptual mapping. We then provided a brief description of Type 2, or physical/neurophysiological mappings, mostly within the retina. Finally, we posed the Type 3 question: a locus question about neural/perceptual mapping. What stage or stages of visual processing limit grating acuity? We suggested four candidate answers. We will examine each of these possible answers in detail in subsequent chapters. The point is that these are the kinds of questions that vision scientists ask, and the ranges of answers that we find satisfying.

In pursuit of the essence of Type 3 questions, we then introduced the concept of a linking proposition: a claim that a particular mapping occurs, or a particular mapping principle applies, between neural and perceptual states. We updated Mueller's first axiom to define the Universal linking proposition – All perceptual states and processes are implemented in neural states and processes. We distinguished the Universal linking proposition from the propositions of relational isomorphism, which postulate specific, relational mapping rules between sets of perceptual states and sets of neural states. We also briefly mentioned several other kinds of linking propositions, including analogies and computational propositions, to which we will return throughout this book.

Finally, DT has had three broad goals in writhing this book. First, vision science is a complex interdisciplinary field, influenced by concepts from many kinds of science and by many kinds of scientists, and the beginning student is likely to have some difficulty with the parts that are the farthest from home. The first goal of this book is to provide a united, self-consistent set of tutorials, using simple examples, to make the science as a whole accessible. It is hoped that the tutorials in the various chapters of this book will slow down the moving train just enough for students with many different backgrounds to jump on.

Second, we have argued that vision science is a sophisticated discipline, spanning across physical, perceptual and physiological realms. To DT's knowledge there is no deliberate, consistent exposition of the forms of argumentation common in vision science. The arguments and premises are often implicit. It is hoped that making them explicit will demystify them, and thereby help students get on the train. And of all the implicit elements, linking propositions seem to DT to be the most consistently hidden in the shadows, and therefore the most fun to bring out into the light.

A final goal of this book is to encourage the seamless integration of sensory and perceptual approaches to vision science. Perhaps some successful causal stories from early processing levels will provide forms of argumentation that will be useful in evaluating causal stories at higher processing levels.

In Chapter 2 and 3 we examine the methodological tools with which vision scientists study physical/perceptual (Type 1) mappings, and a sample of the results they have found. In Chapter 4 and 5, we examine the optics of the eye, and the marks that the optical system leaves on our perceptions. Then in Chapter 6 and 7 we examine the workings of photoreceptors, both individually and in sets, and begin to analyse the code for color vision.

Chapter 2

Psychophysics: Class A Experiments

Psychophysics[1] is the study of quantitative relationships between physical stimuli and perceptions. Alternatively, we can define psychophysics as the science of quantifying the system properties of vision – the answers to Type 1 questions: What and how well do we see?

Somehow, information originating from physical stimuli arrives in our retinal images, is encoded and recoded in our visual systems, and eventually maps to our perception of those stimuli. These mappings are remarkably regular and lawful – the same stimulus usually brings about the same perception – as they must be if we are to respond properly and consistently to stimuli and objects in the physical world. Our first concern in this chapter is how scientists quantify the relationships between physical and perceptual realms. To that end, we introduce some of the measurement techniques used in psychophysics.

But measurement techniques become more interesting as they are used to discover the system properties of vision. As soon as we introduce a measurement technique, we want to put it right to work. In this chapter, after we discuss psychophysical techniques, we will use them to define a new set of system properties having to do with the effect on our visual perception of variations in the *wavelength and intensity of light* . As you will notice, many of the questions are the same as they were for grating acuity, but translated into the realm of wavelength. What wavelengths of light can we see, and at what intensities? Can we tell different wavelengths of light apart?

But first, how do we study the characteristics of human vision? Figure 2.1 shows two examples of subjects being tested in a psychophysics laboratory. In Figure 2.1A, the stimuli are being presented to the subject via a classical optical system. In Figure 2.1B, the stimuli are being presented on a video display system. In either case, the psychophysicist varies the physical characteristics of the stimuli, and the subject reports what she sees, either by turning a knob (as in A) or pressing a key (as in B). Now, more specifically, how do we quantify the lawful relationships between the physical stimulus and the subject's perception?

2.1 Class A vs. Class B experiments

In 1960, in his now-famous chapter on linking hypotheses, Giles Brindley also made a distinction between what he called *Class A* and *Class B* psychophysical observations. *Class A* observations

[1] DT was once trying to define psychophysics for a physicist. He listened attentively for about five minutes, and then said, "Oh, I get it. But why don't you just call it crazy physics?"!

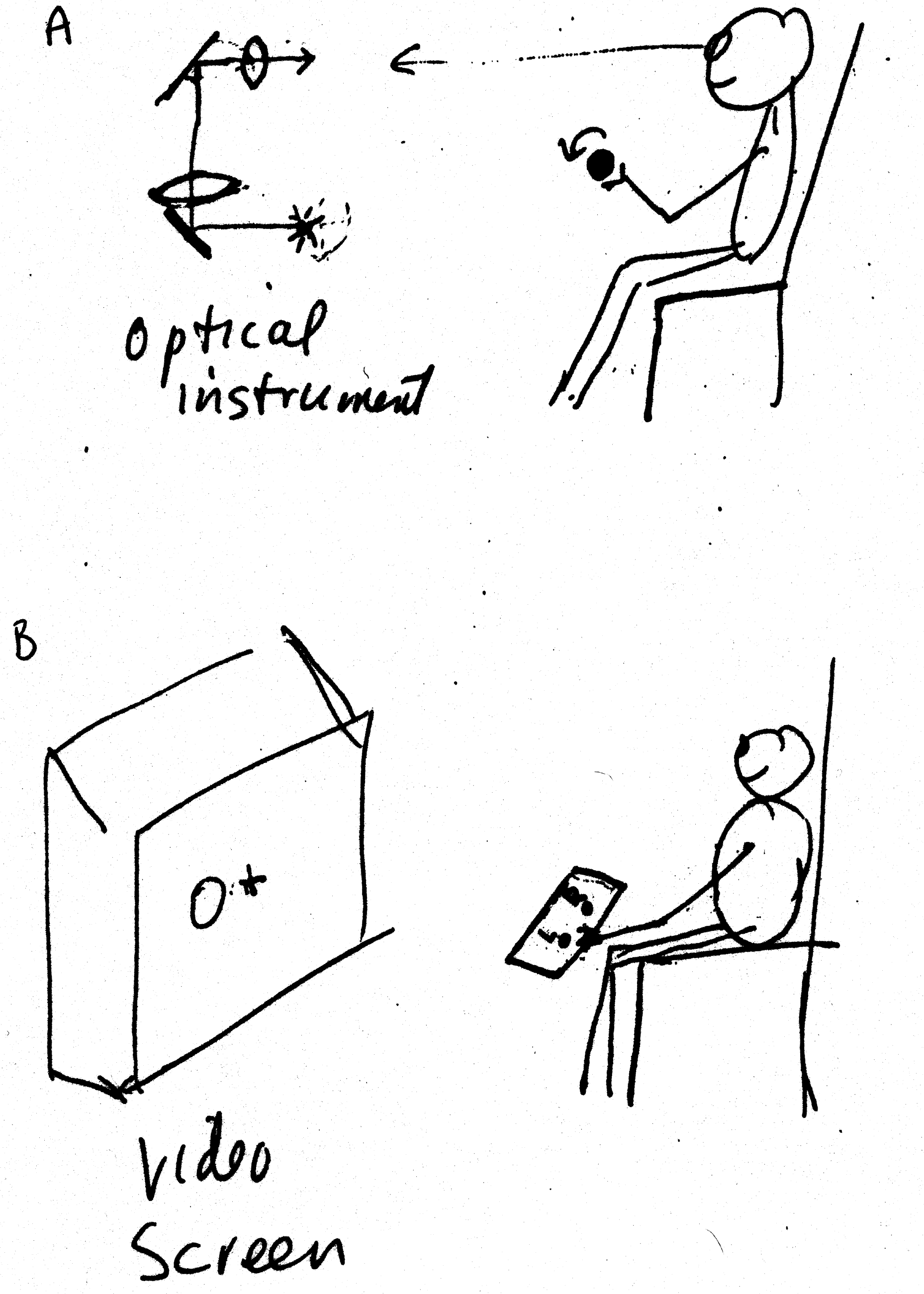


Figure 2.1: A subject being tested in a psychophysical experiment.

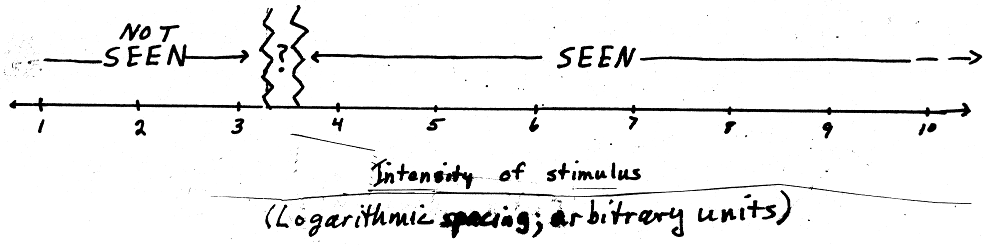


Figure 2.2: The threshold region. The horizontal line represents the intensities of a series of stimuli. The human eye can function over an enormous range of intensities, here represented as ten orders of magnitude (or 10 *log units* in vision science jargon). Under a given set of conditions, however, the threshold – the boundary between seeing and not seeing – is remarkably narrow, perhaps a factor of two or three. In the diagram the threshold region, marked with a question mark, occupies the range between about 3.3 and 3.6 log intensity units (its width is 0.3 log units, or a factor of two).

are those in which a subject is asked to detect the presence of a stimulus (*detection*), or to tell whether two stimuli differed from each other (*discrimination*). As we will see, both of these kinds of observations are measurements of what are called *thresholds.*

In contrast, *Class B* observations are those in which the subject is asked to look at *suprathreshold* stimuli – stimuli that are clearly detectable, and discriminable from each other. The subject's task is to report how his perceptions vary with variations in the physical properties of the stimulus – for example, he is asked to match two lights in perceived brightness when they differ in perceived color, or to make judgments about the perceived colors of the lights.

In the present chapter we confine our attention to Class A experiments – detection and discrimination thresholds. Class B experiments will be discussed in Chapter 3.

2.1.1 Class A experiments: What is a threshold?

In everyday English, a *threshold* is a boundary between one thing and another – between the inside and the outside of a house, for example. In psychophysics, a threshold is the boundary between conditions under which a stimulus is seen and conditions under which it is not seen. The term threshold captures the idea that the transition from seeing to not seeing, like the transition from inside to outside a house, is relatively abrupt. But in fact, the precise place at which we should say one enters the house is slightly ambiguous. Is it the porch steps, or the front door, or halfway between? A visual threshold is similar – relatively abrupt, but with a small region of ambiguity that requires further consideration.

Suppose that we arrange our laboratory equipment so that we can provide spots of light of many intensities, covering a range of (say) 1010, or 10,000,000,000 to 1 (a realistic estimate of the range of intensities that the human eye can handle)[2]. As shown in Figure 2.2, as we look at these

[2] Vision scientists generally plot the intensity of light in logarithmic units. This is because the range of intensities

different stimuli, letting our eyes adjust to changes of intensity as necessary, we will find informally that there is a large range at the low intensity end where we never see the test spot, and a large range at the high intensity end within which we always see it. In between there is a remarkably small region of uncertainty, within which we see the stimulus only some of the time. This region of uncertainty points to the location of the subject's threshold.

Notice, by the way, that this is another example of the departure of perceptual from physical properties. The intensity of the light – a physical variable – is continuous, and there is nothing in the physical stimulus to mark the range over which the subject's threshold will occur. The threshold is a perceptual variable, and marks a remarkably abrupt perceptual transition along the physical continuum.

2.2 Classical psychophysical methods

As psychophysicists, our first goal is to make quantitative estimates of thresholds. How shall we go about it? In the following paragraphs we give examples of three different, rather typical, classical psychophysical methods.

In choosing a psychophysical method, there are at lest three interrelated design factors: The *stimuli* to be used; the *task* the subject is asked to perform; ;and the *responses* the subject is allowed to use. As will be seen, these three factors all vary among the three psychophysical methods we will describe. Many more combinations, of course, are possible and have been used.

2.2.1 The method of adjustment

The first and most intuitively obvious method can be called the *method of adjustment.* In the method of adjustment, we present the subject with a stimulus such as a spot of light. We give him a knob to turn or a computer key to press, and ask him to adjust the physical intensity[3] of the spot until he can just barely see it. The subject is asked to repeat the adjustment some number of times; say, ten.

Hypothetical results of a method of adjustment experiment are shown in Figure 2.3A. In this figure, the abscissa shows the physical intensity of the spot of light, on an expanded, arbitrary intensity axis. The ordinate shows the frequency with which the subject sets the light to each of the different intensities in a set of ten trials. Subjects do this task quite reliably – the range of intensities might typically span only about a factor of about two or three. The mean or median of the set of intensities is the response measure, and is used to characterize the intensity required for threshold – that is, to specify quantitatively the location of the threshold region along the intensity axis.

The main advantage of the method of adjustment is that it is quick and efficient – each setting might take, say, 10 seconds. Thus, ten settings of a single threshold might take a couple of minutes, and a set of ten thresholds could be measured readily in a 20 minute session. Because of its speed,

over which the visual system operates is enormous – about 1010 from the dimmest star to a snowy ski slope when the sun is shining. The use of logarithmic units is a convenient way to compress the range into something that is manageable graphically. It also allows easy comparison between threshold curves and sensitivity curves, as will be discussed later.

[3] The term *intensity* has two uses in vision science. In this book, it is always used informally, to refer to variations in either the physical or the quasi-physical quantity of light (see Chapter 3). But it is also used technically, as part of a formal specification system for the physical intensity of light.

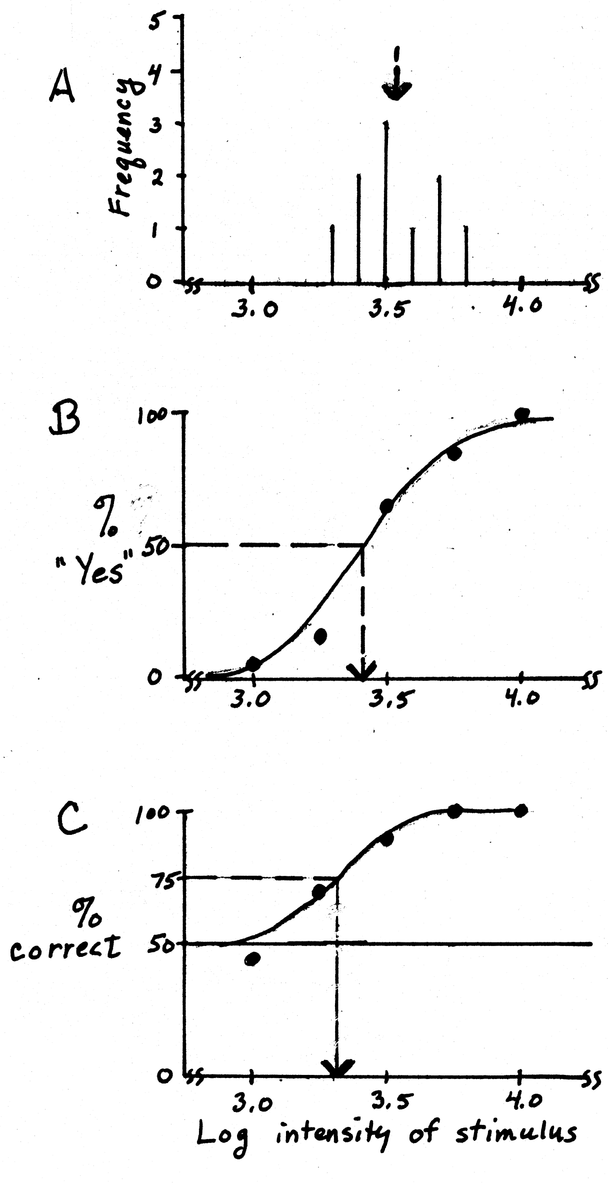


Figure 2.3: Illustrative data from three psychophysical methods. A: The method of adjustment. B: The Yes-No method of constant stimuli. C: The forced-choice method of constant stimuli. The arrows show the threshold estimates from each of the three methods.

the method of adjustment is extremely useful in preliminary work, or in cases in which large effects are being measured and/or only rough estimates of thresholds are required.

However, the method of adjustment has two major limitations. First, it leaves the definition of "seeing" up to the subject. That is, a liberal vs. conservative definition of "seeing" might well cause a difference in the measured threshold. One subject may set the threshold higher than another because the first subject will only say she "sees" the stimulus when it is clearly visible, while the second subject requires it to be only fleetingly so. And second, the subject can turn the intensity of the light up or down at will over any range she chooses. That is, the method of adjustment does not allow the experimenter to control the order of presentation of different stimulus intensities. If the immediate history of stimulation influences the detection threshold – and it does, as will be discussed in Chapter 10– these variations will increase the variability of the subject's individual threshold measurements.

2.2.2 The Yes/No method of constant stimuli

A second kind of approach, which allows the experimenter more control of the order of presentation of stimuli, can be called the *method of constant stimuli.* In this method, the experimenter pre-selects a set of stimulus intensities near where she thinks the subjects threshold will be. These stimuli are presented to the subject, one at a time, in random order, many times each. For example, the experimenter might decide to present the five stimuli 40 times each, for a total of 200 trials.

The experimenter's next decision concerns the choice of response measures and the subject's task. In what we will call the *Yes-No method of constant stimuli*, the experimenter asks the subject to use the responses 'Yes' and 'No'. The subject's task is to say 'Yes' (I saw the stimulus), or 'No' (I didn't see it) on each trial. The use of many trials at each of several different intensities allows the experimenter to plot the percentage of 'Yes' responses as a function of the intensity of the stimulus.

A hypothetical example of the kind of data one would obtain is shown in Figure 2.3B. A data set of this kind is called a *psychometric function.* In this example we have chosen our stimuli well, as the psychometric function spans the range from near zero to near 100% over the chosen stimulus range. By fitting an S-shaped curve to these data, we can estimate quantitatively the intensity at which the subject says 'Yes' on, say, 50% of the trials, and define that intensity as the threshold value. As with the method of adjustment, the threshold value defines the location of the psychometric function along the intensity axis[4].

The Yes/No method of constant stimuli has the major advantage of giving the experimenter control over the stimuli used. In particular, in case the prior history of stimuli viewed influences the threshold – and it often does – the method fo constant stimuli brings this history under control. It is not without its own problems, however. If each trial takes, say, 3 seconds, and 200 trials per threshold are required, measurement of a single threshold will take 10 minutes; and a set of ten thresholds will take two hours, compared to 20 minutes for the method of adjustment.

Moreover, even though the experimenter controls the order of stimuli, the subject is still in charge of deciding how he defines seeing. In fact, by changing the instructions to the subject, and

[4] Psychometric functions actually have four parameters: the threshold (or location along the abscissa), the slope (or steepness), and the upper and lower asymptotes (which are assumed to be 1 and 0 in Figure 2.3B, and 1 and 0.5 in Figure 2.3C). In contrast to the method of adjustment, the method of constant stimuli can also give estimates of these other parameters if enough trials are run. Mathematically, psychometric functions have traditionally been fitted with a variety of S-shaped functions, including the cumulative normal (probit), logit, or Weibull functions.

thereby changing the subject's *criterion* of what "seeing" is, it is easy to move the psychometric function by small amounts along the intensity axis. If the instruction is Be liberal – say 'Yes' if theres even just a tiny flicker of something the curve will shift toward lower intensities. If the instruction is Be conservative – say 'Yes' only if you see the stimulus very clearly the curve will shift toward higher intensities. The experimenter can also shift the curve by providing different incentives or payoffs for saying 'Yes', and in various other ways. In short, the experimenter controls the stimuli, but the experiment still confounds the sensory variable of threshold with the subject's cognitive criterion for saying 'Yes' or 'No'.

2.2.3 The forced-choice method of constant stimuli

A third approach can be called the *forced-choice method of constant stimuli.* In this method, the experimenter again modifies the pattern of stimulus presentation. Rather than presenting only a single stimulus on each trial, the experimenter presents the stimulus in either of two alternative spatial or temporal positions. For example, the stimulus can be presented either on the left or on the right, or in a first or second time interval. This stimulus format allows us to change the subjects task: instead of asking him to say whether or not he sees the stimulus, thereby leaving the criterion for seeing up to him, we can require the subject to make a judgment as to whether the stimulus occurred in one place or another, or in one time interval or another. For example, in a spatial forced-choice technique, the subject is asked to respond 'Left' if he judges that the stimulus was presented on the left, or 'Right' if he judges that the stimulus was presented on the right.

Notice that a critical change of task is being made here. In this case, the task is not, "Did you see it?" – a question about the subjects perceptions – but rather, "In which location did it occur?" – a question about the state of the physical world. This change of task has two other intertwined and important consequences. First, there is a right and a wrong answer on each trial – the stimulus is always presented in one or the other position (or time interval). And second, since the judgments can be either correct or wrong, trial by trial feedback can be given to the subject. The subjects overall task is to maximize the percent of trials on which his answer is correct, and providing trial-by-trial feedback allows him to learn, over time, to be correct on the maximum number of trials.

Tasks that involve judging the state of the physical world can be called *objective*, or *externally-referred*, or *physically-referred tasks.* Tasks that involve judging one's own perceptions can be called *subjective*, or *internally-referred*, or *perceptually-referred* tasks. We will use the terms physically-referred and perceptually-referred in this book.

Hypothetical results of such an experiment are shown in Figure 2.3C. As in Figure 2.3B, psychometric functions are plotted, but this time with the subject's *per cent correct* plotted on the ordinate. Since the subject can get 50% correct by guessing, the psychometric function spans the range from 50% to 100%. The threshold in such an experiment is typically defined as 75% correct – halfway between chance (50%) and 100%.

Among the three methods discussed here, the forced-choice method of constant stimuli is the most logically elegant, in the sense that it brings the stimuli, the subject's task and the subject's criterion under the tightest possible experimental control. However, it is also the least efficient, as many more trials (and therefore more time) must be invested to estimate the location of the psychometric function. For statistical reasons, it takes two to three times as many trials to locate the threshold to a given degree of accuracy when the lower asymptote is at 50% than when it is at zero. So a set of ten thresholds, held to the same criterion of accuracy, would take perhaps six hours, compared to 20 minutes for the method of adjustment and two hours for the Yes/No method of constant stimuli.

A final note on terminology: unfortunately, the term *forced-choice* is not consistently used in the psychophysics literature. Sometimes the term is used very broadly, to refer to any experiment in which the subject is allowed only two responses (e.g. 'Yes' and 'No'). With this definition, the Yes/No method of constant stimuli – the second method defined above – is also a forced-choice method. Other sources use "forced-choice less inclusively, to refer only to experiments in which two stimulus options are provided – e.g. the stimulus is present on some trials and absent on others. Within this mode of stimulus presentation, some sources use "forced-choice" regardless of whether the task is physically- or perceptually-referred, whereas others reserve "forced-choice" to refer only to the former. Finally, in the most stringent usage (and particularly in Signal Detection Theory, as discussed below), the term "forced-choice" is reserved for experiments in which, within each trial, the stimulus is presented in either of two spatial positions or temporal intervals *and* the subject's task is physically-referred. In this book we use the most stringent usage. But when an experimenter says that a forced-choice technique was used, the only way to find out what was actually done is to study the Methods section of the paper.

2.2.4 Staircases and other adaptive techniques

There's one more trick worth knowing. With the method of constant stimuli, we choose the set of stimuli before starting the experiment, and we're stuck with it. If it turns out not to sample the actual psychometric function well, we will do a poor job of estimating the threshold; we may even have to start over. There is another set of psychophysical methods, in which the trials that have been run before the current one are used to determine the optimal stimulus to use on the current trial. The earliest adaptive methods were called *staircase methods*, for reasons that become obvious in Figure 2.4.

In this figure, the experimenter is carrying out a Yes/No staircase experiment. The experimenter starts with a high intensity stimulus on the first trial. If the subject says "Yes", the experimenter *decreases* the intensity for the second trial. In this example, this pattern was repeated until on trial 5 the subject says "No". At this point the experimenter *increases* the intensity for trial 6. Depending on the subject's responses, the pattern of intensities goes up and down – hence the name staircase. [Could you use a staircase for a physically-referred forced-choice experiment? Why or why not?]

In more sophisticated versions of *adaptive techniques*, all of the trials up to trial N-1 can be collapsed into a psychometric function; the psychometric function can be fitted with a theoretical curve; and statistical rules can be used to select the optimal stimulus – the one whose use would yield the maximal information – to use on trial N. These adaptive techniques increase the efficiency of estimating the threshold, in comparison to the corresponding method of constant stimuli.

In sum, the choice of a psychophysical method depends upon one's needs for balancing speed and accuracy. For a "quick and dirty" estimate, the method of adjustment is the obvious choice; it's fast but open to serious stimulus artifacts and potential criterion problems. For a more controlled but still criterion-limited method, the Yes-No method of constant stimuli may be the best choice. For minimum influence of the subject's criterion, choose the forced-choice method of constant stimuli, but expect to spend a lot of time in the laboratory! And, with either of the techniques based on the

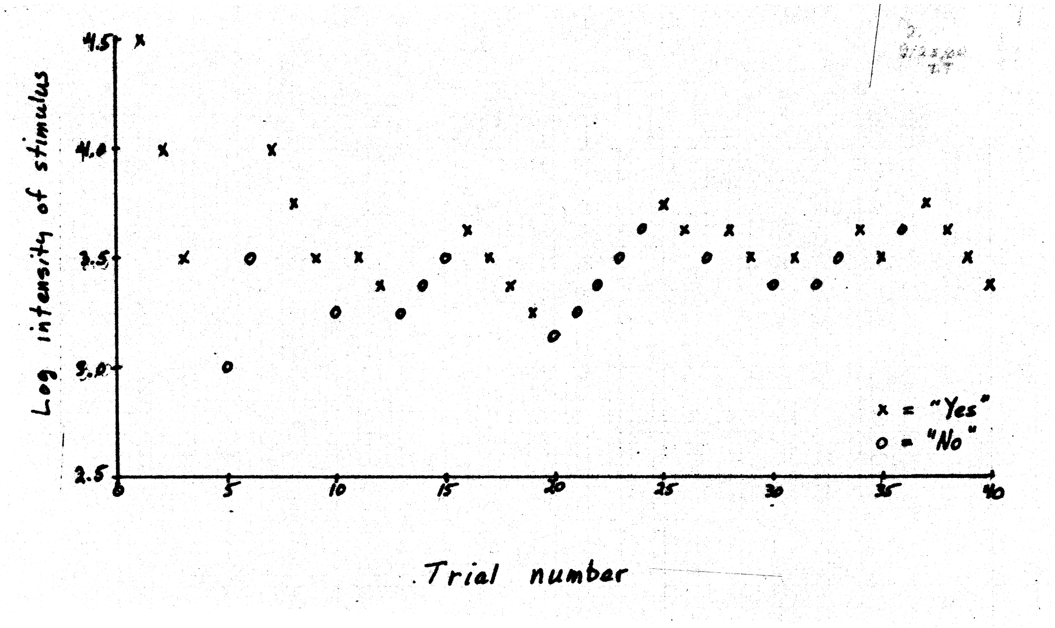


Figure 2.4: A "staircase". In this graph the abscissa shows the number of the experimental trial, and the ordinate shows the log intensity of the stimulus. The eexperimenter starts on trial 1 by presenting a stimulus she judges the subject will always be able to see. The subject replies "Yes", marked by the X over trial #1. The experimenter then presents an 0.5 log unit dimmer stimulus. This pattern continues until trial #5, when the subject says "No", whereupon the experimenter reverses direction and presents (say) an 0.5 log unit higher intensity stimulus on trial #6. When the subject again says "Yes" on trial #7, the experimenter reduces the step size, and presents (say) an 0.25 log unit dimmer stimulus on trial #8. The step size is reduced again at trial #11. Forty trials are run. The threshold can be estimated as the mean of the stimulus intensities used after the smallest step size is reached (although other scoring methods are also popular).

method of constant stimuli, efficiency can be improved somewhat by the use of adaptive techniques.

2.3 Discrimination thresholds

Up until now, we have confined our discussion to detection thresholds. Is a spot of light present? In which of two positions is it located? *Discrimination thresholds* are a closely related concept. To measure discrimination, two stimuli are presented, and the subject is asked to tell them apart. Is one of the spots of light of higher intensity than the other? In which of two locations is the higher intensity spot located?

Each of the methods used for detection experiments (adjustment, Yes/No method of constant stimuli, forced-choice method of constant stimuli, and staircase methods) can also be used for measuring discrimination thresholds. For example, one could use the method of adjustment by setting one of the two stimuli to a fixed value, and having the subject adjust the intensity of the second stimulus until it looks just barely different from the first. [How would you measure a discrimination threshold with the Yes/No method of constant stimuli? With the forced-choice method of constant stimuli? With a staircase?] We will return to both detection and discrimination experiments below.

2.4 Animal psychophysics

In Chapter 1, when we discussed three types of questions, you may have noticed a potentially major limitation. That is, Type 1 questions concern the system properties of visual perception – What and how well do we see? Obviously, Type 1 questions are usually studied by testing human subjects. In contrast, Type 2 questions concern the properties of the visual substrate – the optics, photochemistry, anatomy, and physiology of the visual system. Since we can't do invasive experiments on human subjects, we usually study the substrate – particularly the physiology of single cells – in animals. And Type 3 questions concern trying to explain system properties on the basis of substrate properties. But how can we assume that animals see as we do, or that our physiology works like that of the experimental animal we study? And if not, won't our causal stories always be potentially flawed?

Part of the answer to this question is yes – to some extent we have to live with this problem. But the other part is no. True, we can't do invasive experiments on human subjects; but we can in fact do psychophysics on animals. Many different species of animals have been tested successfully, including cats, fish, fruit flies, and (most importantly for our purposes) non-human primates.

Although other approaches are possible, the most straightforward approach to animal psychophysics is to use a physically-referred task – a task in which there is a right and a wrong answer. Suppose you train a monkey subject to sit in a primate chair facing a stimulus screen and two response buttons. On each trial of the experiment, the stimulus occurs either on the right or on the left; the monkey presses the right or left button, and he gets feedback – he is rewarded with a drop of water – for pressing the correct button. If he pushes the wrong button, he gets no reward. He may even get a ”time out” – say a ten-second period during which no trials are run. (If he's thirsty, this matters to him.)

Initially the experimenter presents high-intensity, easily visible stimuli, and the monkey is trained to use the buttons to respond right or left. Once the monkey gets, say, 90% correct or more over a series of trials, we use lower and lower intensity stimuli, backing up to higher intensities if his error rate goes up, and progressing toward lower intensities again when he does well. Eventually his performance will stabilize, and he will generate a forced-choice psychometric function of the kind shown in Figure 2.3C, just as a human subject does.

When old world monkey subjects are tested, the resulting psychophysical data often resemble the data from humans very closely. The parts of vision science that involve causal stories about human vision rely heavily on this similarity.

2.5 A theoretical account of thresholds: Theory of Signal Detection

In 1966, David Green and J.A. Swets proposed a theoretical analysis of detection thresholds, called the *theory of signal detection (TSD)* . Their analysis suggests that our previous discussion of the concept of threshold is incomplete, and that a full account of thresholds requires two parameters. The first parameter, which Green and Swets called *d'*, (d-prime) represents the *detectability* of the stimulus – a sensory variable. The second, which Green and Swets called β (beta), represents the subject's *criterion* – a more cognitive variable. Green and Swets argued that a traditional Yes/No experiment, as described above in Figure 2.3B, confounds these two variables. That is, when the subject chooses to be liberal or conservative, his psychometric function shifts. Is this to be considered a change of the sensory threshold, or just a change in the subject's criterion? How can we tease the two apart?

The essence of the theory of signal detection is shown in Figure 2.5. Green and Swets began by assuming that the psychophysical subject has access to an internal (perceptual) variable whose strength varies with the intensity of the stimulus. For the sake of concreteness, in the present case we can think of this variable as the perceived brightness of a spot of light. The abscissa of Figure 2.5 represents the possible values of this internal variable, and the ordinate represents the probability of occurrence of each of the possible values.

Now let's modify our Yes/No method of constant stimuli experiment in two ways. First, instead of presenting five different stimuli interleaved, let's present just one of these stimuli – say, the middle one of the former five. And second, let's present the stimulus on only half of the trials (*stimulus trials*), and not on the other half (*no stimulus trials*). The subject's task is to judge whether or not the stimulus was presented on each trial. (Notice that this is a physically-referred, Yes/No task – a different combination of experimental design factos than was used in any of our three examples.)

The fundamental argument made by Green and Swets is that a detection task should be viewed as a *signal/noise discrimination.* Even on trials on which no stimulus is presented, Green and Swets proposed that the internal perceptual variable will have a non-zero value due to sensory *noise.* The noise arises from many internal and external sources, and the value of the noise fluctuates randomly over time. The hypothetical distribution of the values that the noise will take is shown by the curve at the left in Figure 2.5, labeled *noise alone.* On trials on which a stimulus is presented, the value of the perceptual variable will be increased – say, a constant value will be added to the noise. As a result, on these trials, the values that the perceptual variable will take are shown by the right-hand curve in Figure 2.5, labelled *signal + noise.*

The key assumption is that these two distributions lie on a single perceptual dimension – noise and stimulus both contribute to the value of the same internal variable. Thus, there is no way

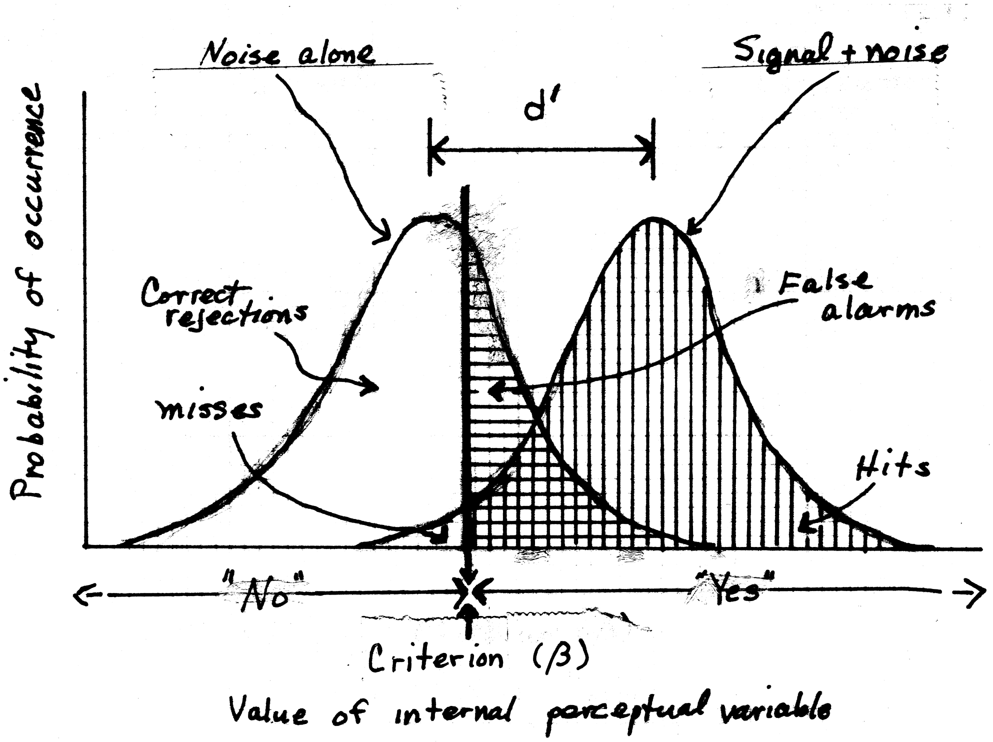
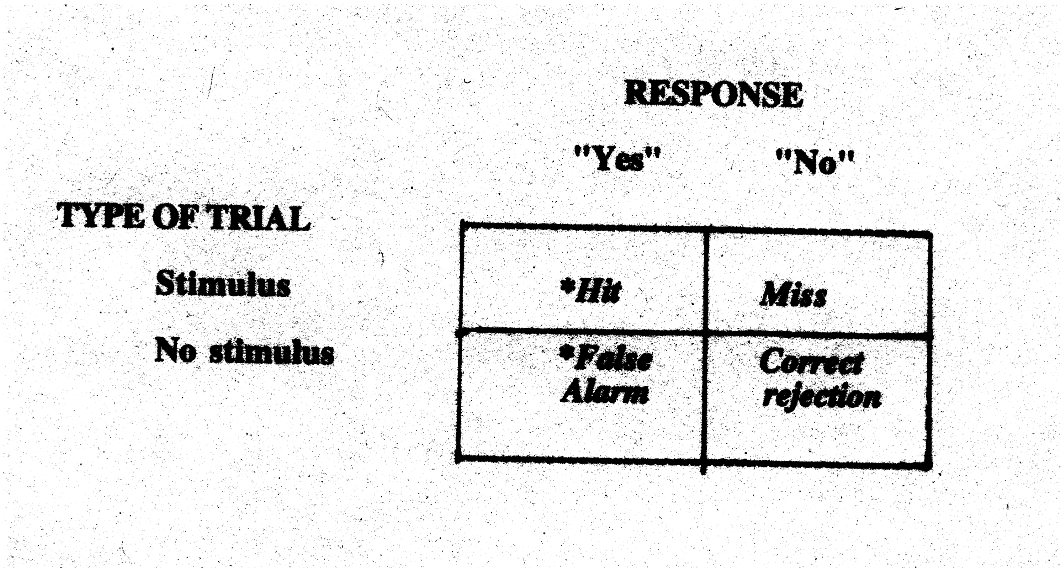


Figure 2.5: Signal Detection Theory. The abcissa shows the value of a hypothetical perceptual variable having to do with the strength of an internal signal. The ordinate shows the probability of occurrence of each of the possible values of the internal perceptual variable. A near-threshold stimulus is presented on half of the trials, and these trials generate the signal + noise distribution at the right. The other half of the trials contain no stimulus, and these trials generate the noise alone distribution at the left. The subject's task is to judge whether or not a stimulus occurred on each trial. The subject is free to vary his cutoff point, or criterion (β), along the abscissa. The number of Hits (saying "Yes" when the stimulus was present) and the number of False Alarms (saying "Yes" when no stimulus was presented) are tied together, and depend upon two variables: *d'*, the separation of the maxima of the two distributions, and β, the subject's criterion.



TYPE OF TRIAL	RESPONSE	
	"Yes"	"No"
Stimulus	*Hit	Miss
No stimulus	*False Alarm	Correct rejection

Figure 2.6: Contingency table for outcomes of trials in a signal detection experiment.

for the subject to know whether a given value arises from noise alone or from the presentation of the stimulus in the midst of the noise. Nonetheless, the subject’s task is to judge whether or not the stimulus was presented on each trial by saying, ’Yes’, the stimulus was presented, or ’No’, the stimulus was not presented. TSD suggests that the subject’s only possible strategy is to choose a *criterion* value, shown by the arrow on the abscissa in Figure 2.5, and to say ’Yes’ if the value of the internal variable is above the criterion value, and ’No’ if it is below the criterion value.

The four possible outcomes of each of the trials in a TSD experiment are shown in Figure 2.6. On each trial the stimulus is either present or absent, and the subject’s response is either ”Yes” or ”No”. On trials on which the stimulus was presented, a ”Yes” response yields a *hit*, and a ”No” response yields a *miss*. On trials in which no stimulus was presented, a ”Yes” response yields a *false alarm*, and a ”No” response yields a *correct rejection*. The probabilities of these four outcomes are related to areas under the two curves in Figure 2.5. The subject’s criterion forms the boundary between hits and misses on signal trials, and between false alarms and correct rejections on noise-alone trials. As the criterion is shifted leftward, the subject will say ’Yes’ on an increasing percentage of the trials, and generate more hits, but of necessity he will also generate more false alarms. By comparing the percentage of hits to the percentage of false alarms, and applying established formulas from TSD, the experimenter can estimate both d, the subject’s sensitivity to the signal, and β, the subject’s criterion.

What if we now go back to a more classical method of constant stimuli, and think about stimuli of several different intensities rather than just one? The result is shown in Figure 2.7. The location of the signal-plus-noise distribution will vary along the abscissa with the intensity of the stimulus. As shown in Figure 2.7A, the lowest intensity of our five stimuli will yield a signal distribution that overlaps nearly completely with the noise-alone distribution, so that the percentage of hits and the percentage of false alarms will have to be nearly equal no matter what the criterion. But as shown in Figures 2.7B-E, the higher the intensity of the stimulus, the more the signal plus noise distribution will shift to the right, and the more the percentage of hits can exceed the percentage of false alarms. Moreover, the parameter d' corresponds to the distance between the peaks of the noise-alone and signal-plus-noise distributions. The threshold region we first encountered in Figure 2.2 can now be seen to be the range of stimulus values that yield substantial (but not complete) overlap between the noise-alone and signal-plus-noise distributions.

In short, TSD is a useful mathematical model because it provides a theoretical account of several aspects of detection thresholds. It suggests a reason why observed psychometric functions are gradual rather than completely abrupt – because of the presence of noise, and the fluctuating value of the noise from trial to trial. It also allows us to separate the sensory variable *d'* from the cognitive variable β, and provides us with an honorable way of arguing that the subject's adoption of a liberal vs. conservative criterion does not change the incoming sensory signal.

Since Green and Swets first applied TSD to psychophysics in 1966, it has provided the basic conceptual foundation for our understanding of detection thresholds. Many quantitative accounts of phenomena we will see later have been formulated in terms of signal/noise models, with the noise arising from sources both outside and within the subject's visual system. Due to space considerations, we will not spend much time on such models, but it's important to know that TSD is one of the critical foundations of quantitative psychophysical theory.

2.6 Linking propositions for Class A experiments

We now return to the topic of linking propositions. In the same chapter in which Brindley (1960) introduced the Class A/Class B terminology to refer to differences in psychophysical tasks, he also introduced the concept of what he called a *linking hypothesis*. Brindley was at the time a visual physiologist, and his question was, what is the role of psychophysical data in elucidating the physiology of vision? He argued that mental terms (perceptual terms, hence psychophysical data) and physiological terms were from different realms of discourse, and could not be used in the same sentence without introducing some form of special statements by means of which their meanings could be linked. Rather than referring to these statements as axioms, Brindley called them linking hypotheses.

Brindley further argued that most of the linking hypotheses in use in the vision science of his day were quite arbitrary, and that a careful physiologist would not want to pay attention to arguments that depended on them. Hence, he was ready to exclude most of psychophysics and perception from vision science. However, he found one linking hypothesis which he felt was rigorous enough and safe enough to include among the premises of his science, and by allowing its use he allowed detection and discrimination experiments (Class A experiments) to slip into the science. The linking hypothesis Brindley found acceptable was as follows:

"Whenever two stimuli cause physically indistinguishable signals to be sent from the sense organs to the brain, the sensations produced by these stimuli...must be indistinguishable" (Brindley, 1960, p. 144). In other words, identical incoming physiological signals must yield identical perceptions. [What do you think of the logical status this statement? Is it a tautology? Could it be false?]

In 1984, DT wrote a paper elaborating on the concept of such linking statements (Teller, 1984). Since she felt that such statements are not usually hypotheses, she instead called them linking *propositions* – that is, statements that can potentially play many different roles in scientific argument, and whose truths and uses need to be individually evaluated. She defined a *linking proposition* as *a claim that a particular mapping occurs, or a particular mapping principle applies, between perceptual and physiological states.*

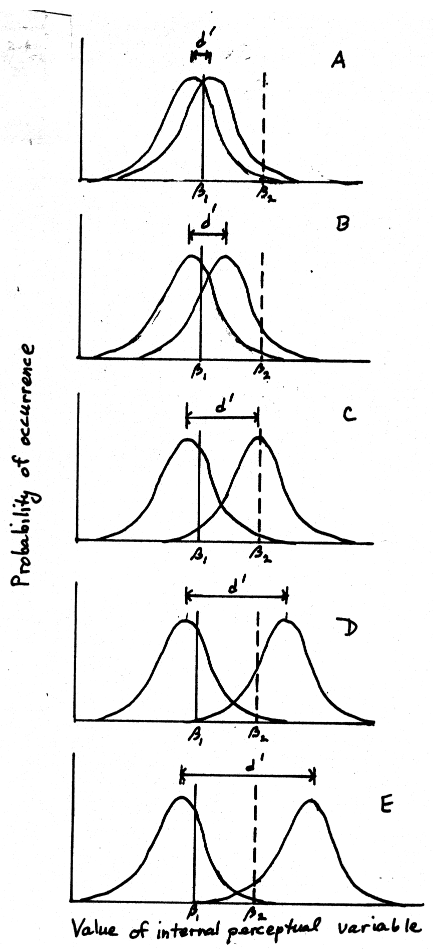
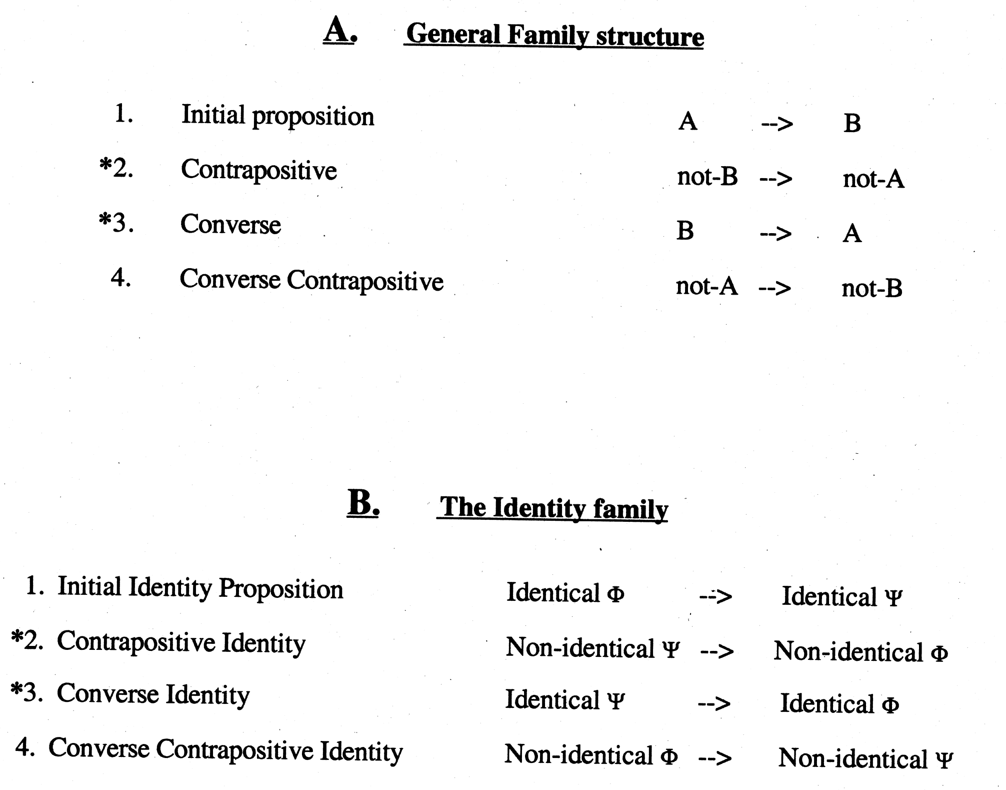

A
B
C
D
E
d'
β1
β2
Probability of occurrence
Value of internal perceptual variable

Figure 2.7: The effect of stimulus intensity on *d'*. The five panels show the effect of increasing the intensity of the stimulus, from a low (A) to a high (E) value within the threshold range. The noise alone distribution remains constant in all five panels. As the stimulus intensity increases, the signal + noise distribution shifts rightward, and *d'* increases. Two possible criteria, β_1 and β_2, are also shown. For β_1, the subject has chosen to hold the False Alarm rate constant at about 40%. For β_2, he is holding it constant at about 5%. The percents of Hits and False Alarms vary lawfully together with variations in *d'* and β. In a signal detection experiment, the numbers of Hits and False Alarms are the dependent variables. If the shapes of the two distributions are known, the pattern of covariation of Hits and False Alarms allows estimation of both *d'* and β.



A. General Family structure

1.	Initial proposition	A	-->	B
*2.	Contrapositive	not-B	-->	not-A
*3.	Converse	B	-->	A
4.	Converse Contrapositive	not-A	-->	not-B

B. The Identity family

1.	Initial Identity Proposition	Identical Φ	-->	Identical Ψ
*2.	Contrapositive Identity	Non-identical Ψ	-->	Non-identical Φ
*3.	Converse Identity	Identical Ψ	-->	Identical Φ
4.	Converse Contrapositive Identity	Non-identical Φ	-->	Non-identical Ψ

Figure 2.8: The family structure of relational linking propositions. A: The general family structure; B. The Identity family.

2.6.1 Family structure

Following the "and conversely" statements of Mueller (1896), DT also argued that relational linking propositions come in families, with different family members building on different experimental outcomes and allowing different directions of inference between psychophysics and neurophysiology. Specifically, she argued that a family of relational linking propositions has four members that relate to each other as shown in Figure 2.8A. The family is composed of (1) an *Initial proposition*; (2) its *Contrapositive*; (3) its *Converse*, and (4) its *Converse Contrapositive*. In abstract terms, the initial proposition (1) is: A implies B (if A is true, then B is true). The Contrapositive (2) is: not-B implies not-A (if B is not true, then A is not true). The Converse (3) is: B implies A (if B is true, then A is true). And the Converse Contrapositive (4) is: not-A implies not-B (if A is not true, then B is not true).

What are the relationships among these four statements? Students who have taken a course in logic will remember that the Initial statement and its Contrapositive – (1) and (2) –can be inferred from each other: if one is true, the other is also true. An example: if the Initial statement is, *if it rains the sidewalk will be wet*, the Contrapositive is, *if the sidewalk isn't wet, it didn't rain.* The same is true of the Converse (3) and the Converse Contrapositive (4). Importantly, however, the Converse (3) does not follow from the original statement (1): *if the sidewalk is wet, it rained* is not a logical inference, as the sidewalk could be wet for other reasons (perhaps the sprinklers are on)[5].

2.6.2 The Identity family

Following this format and the inspiration provided by Mueller's laws, Teller (1984) then formulated several families of relational linking propositions. The first family, with which we are concerned in this chapter, is called the *Identity family*, and is shown in Figure 2.8B. In order to use symbols that are more mnemonic for physiological vs. perceptual states, we will substitute the Greek letter ϕ (phi, for physiology) for the A's in Figure 2.8A, and the Greek letter ψ (psi, for psychology) for the B's.

With these symbol substitutions and formalizations, the Initial Identity proposition (1) becomes: *Identical physiological states imply identical perceptual states.* The Contrapositive Identity proposition (2) is: *Non-identical perceptual states imply non-identical physiological states.* The Converse Identity proposition (3) is: *Identical perceptual states imply identical physiological states.* And the Converse Contrapositive Identity proposition (4) is: *non-identical physiological states imply non-identical perceptual states.* As before, the Initial proposition and the Contrapositive can each be inferred from the other, as can the Converse and the Converse Contrapositive; but the Initial proposition and the Converse are logically independent.

Are the Identity propositions analytically true, or just highly likely; or are they relatively safe speculations, or risky ones? The Initial Identity proposition has the same content as the only linking hypothesis that Brindley found acceptable; that is, "Whenever two stimuli cause physiologically indistinguishable signals to be sent from the sense organs to the brain, the sensations produced by those two stimuli.must be indistinguishable." Moreover, Brindley argued (and DT would agree) that his acceptable linking hypothesis is probably analytically true and tautological (it follows from the definitions of the other concepts involved). The Contrapositive, being logically identical to the Initial proposition, would also have the same logical status – true and tautological.

The Converse and Converse Contrapositive, however, are not quite as necessarily true. For example, the Converse of Brindley's linking hypothesis would be something like this: Whenever the sensations produced by two stimuli are indistinguishable, these two stimuli must be sending identical signals from the sense organs to the brain. This statement is not necessarily true, because the signals could start out different in the retinal image but become identical at some later stage of processing; and because even if the neural states remain distinguishable, two different neural states could in principle map to the same perceptual state (a many:1 mapping between neural and perceptual states could occur). In fact, with only one possible exception (see later), Initial and Contrapositive Identity propositions seem to DT to be the only general linking propositions that are analytically true, and therefore completely safe to adopt as premises. All the rest are riskier.

We now return to causal stories. Notice that the second (2) and third (3) Identity linking propositions allow physiological conclusions to be drawn from psychophysical experiments. These two propositions are starred in Figure 2.8. Detection and discrimination experiments determine

[5] DT's all time favorite example of the fact that an initial statement doesn't imply its converse arose when she was a graduate student. A fellow graduate student remarked one day that he didn't mind being misunderstood, because to be great is to be misunderstood. DT undertook the delicious responsibility of pointing out to him that unfortunately, converses being what they are, to be misunderstood was not necessarily to be great.

whether the sensations produced by two different stimuli are discriminable (non-identical), or not discriminable (identical). If two stimuli are discriminable in a psychophysical experiment, Contrapositive Identity (2) insists that these two stimuli have sent non-identical signals from the sense organ to the brain. Most people find this proposition difficult to doubt, as Brindley did. On the other hand, if two stimuli are not discriminable, the Converse Identity proposition (3) suggests that they have sent identical signals, or more sensibly, that the signals differed initially but were rendered identical at some later stage of processing. Thus, any Class A psychophysical experiment allows us to draw a conclusion about the probable identity or non-identity of physiological states, depending only on the outcome of the experiment – which stimuli are discriminable and which are non-discriminable – and our willingness to employ either the Contrapositive (2) or the Converse (3) Identity proposition in our argument.

Finally, notice that these Identity arguments are the same as some arguments we introduced less formally in Chapter 1. We argued initially that if a subject can discriminate between a grating and a homogeneous field, information about the spatial structure of the grating must be retained through all levels of the visual system. There is a Contrapositive Identity proposition embedded in the premises of this argument. Similarly, we argued that if the subject could *not* discriminate between the grating and the homogeneous field, information about the spatial structure of the grating must have been lost somewhere within the visual system. There is a Converse Identity proposition embedded here.

2.6.3 Application of Identity propositions to Class A experiments

How, exactly, do we apply Identity propositions to the data from detection experiments, in which there is only one stimulus? Terminological issues can make this question confusing. The trick is that in this context, vision scientists think of the background alone as one ”stimulus”, and the background plus the test stimulus as the other ”stimulus”. Some intensities of the test stimulus are below the threshold region; these are marked NOT SEEN in Figure 2.2. In terms of signal detection theory, this means that the noise distribution is indistinguishable from the signal-plus-noise distribution. When such discrimination failures occur –when we are below threshold – a Converse Identity proposition will be included as a premise in any argument from perceptual data to physiological conclusions. We will conclude, slightly speculatively (as is always the case with Converse propositions) that the two physiological states arising from background and background-plus-test-stimulus are indistinguishable.

Other stimuli are above the threshold region; they are marked SEEN in Figure 2.2. In this intensity region, the noise distribution is very different from the signal-plus-noise distribution. In this case, a Contrapositive Identity proposition will be included as a premise in our arguments. Since the perceptual states are distinguishable, we can conclude with confidence that the physiological states are distinguishable. Thus, just as the psychometric function marks the transition from not seeing to seeing, it marks the transition from using Converse to using Contrapositive Identity propositions in drawing physiological conclusions from perceptual data.

For discrimination experiments, very similar arguments hold. When two suprathreshold stimuli are physically different but indiscriminable, we use Converse Identity any time we argue that the physiological states are indiscriminable. When the two stimuli are discriminable (as most pairs of stimuli are!) we use Contrapositive Identity in arguing that the physiological states are discriminable. The former is slightly speculative, whereas the latter is (as Brindley said) very difficult to doubt.

2.7 Two system properties of scotopic vision

Now that you have been introduced to psychophysical methods for measuring detection and discrimination thresholds, the next question is, what kinds of substantive questions can be investigated with measurements of thresholds?

As it turns out, an individual threshold value usually has little in the way of theoretical implications beyond the general one of information retention and information loss. However, experiments in which *sets* of thresholds are measured often give important hints about physiological processes. In such experiments thresholds are measured as a function of some stimulus parameter or parameters. As an example, we now turn to an important set of thresholds: detection thresholds as a function of the wavelength[6] of the stimulus.

What we experience as ”white” light usually contains a large range of wavelengths, the component wavelengths of which we experience as colors. As a student at Cambridge in the 1660s, Isaac Newton noticed a beam of sunlight coming through the shutters of his room. He passed the beam through a prism, and saw that the light now produced a rainbow of colors. That is, he had shown that sunlight can be broken down into its component wavelengths by passing it through the prism, which bends or *refracts* light differentially according to its wavelength. In the case of natural rainbows, internal reflections in water droplets act as the prism did for Newton.

As Newton showed, at moderate and higher light levels we can discriminate among lights of different wavelengths – different wavelengths map to different perceived colors. But you also may have noticed that if the light is sufficiently dim the colors fade away, and all that is left is shades of gray. [If you have not noticed this phenomenon, toss some shirts or towels of different colors around your room tonight, and see whether you can discern their colors when the room is nearly dark and your eyes have adjusted to darkness.]

In fact, vision turns out to have very different properties at low vs. high light levels. The term *scotopic vision* refers to vision at low light levels, at which no colors are perceived, and the term *photopic* refers to vision at higher light levels, at which colors are perceived. We will spend the remainder of this chapter discussing two of the major properties of scotopic vision.

2.7.1 The scotopic spectral sensitivity curve

The *absolute threshold* is the smallest amount of light a subject can detect when her eyes are fully adjusted to the dark. To begin our exploration of the effects of wavelength, we ask, does a person's absolute threshold vary with the wavelength of light?

The quickest way to answer this question is to use the method of adjustment with a series of stimuli that vary in wavelength. We use a calibrated light source so that the physical energy at any given wavelength is known. We put the subject in a dark room for an hour before we start. Then, we present each of the wavelengths in turn, and ask the subject to adjust the intensity of the light until he just barely sees the stimulus. The subject does this, say, ten times for each wavelength, and we take the mean of these ten settings. We then plot this threshold value as a function of the

[6] The wavelength of light is usually specified in *nanometers* (nm). One nm is 10^{-9} meters. We will say more about the nature and specification of light in Chapter 4.

wavelength at which it was measured. Of course, more elegant psychophysical techniques could also be used.

The results of the experiment, plotted in terms of thresholds, are shown in Figure 2.9A. This data set forms what can be called a *scotopic spectral threshold* curve. Alternatively (and more commonly), the data are plotted as *scotopic spectral sensitivity*, where sensitivity is defined as 1/threshold. With a logarithmically spaced ordinate, the conversion is particularly simple, because the threshold curve can simply be inverted to get the sensitivity curve. The same data plotted as sensitivity are shown in Figure 2.9C. We will use sensitivities rather than thresholds from now on.

This data set has several interesting features. First, the highest sensitivity is always around 500 nm (closer to 490 nm to be more exact). Second, sensitivity varies enormously with wavelength. Compared to the sensitivity at 500 nm, sensitivity declines by a factor of roughly 100 as we change the wavelength from 500 to 400 or to 600 nm; and by another factor of roughly 1000 as we change the wavelength to 700 nm. So the change in sensitivity for lights of 500 vs. 700 nm encompasses five orders of magnitude.

Third, the scotopic spectral sensitivity curve is a relatively simple U-shaped curve. And fourth, its shape is extremely stable. For example, it doesn't matter what psychophysical technique we use. All of the results reveal the same simple curve, possibly shifted up and down the ordinate, but of exactly the same shape. Similarly, backgrounds of various intensities and wavelengths shift the curve up and down; but over a broad range of conditions, the shape of the curve remains unchanged.

In sum, we have discovered a new system property of human vision. Absolute threshold experiments for lights of different wavelengths reveal a smooth, stable spectral sensitivity curve, with its maximum at about 500 nm, and with very large and characteristic losses of sensitivity with changes in wavelength.

2.7.2 Failures of wavelength discrimination: Metamer sets

Here's a second set of system properties for scotopic vision. Suppose that you set up a row of test lights of different wavelengths, each one set to its own threshold, and ask the subject to discriminate among them. The subject cannot do the task at absolute threshold, where the lights are barely visible, but this seems hardly fair. So lets set the intensity of each stimulus to twice its detection threshold. The new stimuli are indicated by the lowest dotted line, marked 2x for "two times threshold", in Figure 2.9B. The stimuli all look slightly brighter than they did at absolute threshold, but they all still look whitish, and remarkably, the subject still cannot discriminate among them. Similarly, the stimuli along the second dotted line (10 times the absolute threshold), or the third (100 times the absolute threshold), or any similar line in between, are indiscriminable, until we reach a limit (about to be described) for each wavelength of light[7]. (Stimuli from any two different dotted lines are discriminable because they vary in brightness.) In short, at scotopic light levels, wavelength information is lost.

Vision scientists are so impressed with the fact that very different physical stimuli can be indiscriminable, that we use several special terms to describe this phenomenon. Such sets of stimuli have been called *equivalence classes*, emphasizing the idea that the signals arising from

[7] In the early days of science, what a challenge it must have been to sort out the effects of physical variables from the effects of our own sensory systems. What visual sensations would Newton have seen if he had used his prism in moonlight instead of sunlight? (Moonlight, of course, is reflected sunlight.) If the whole spectrum looked white, what conclusion would he have drawn about the nature of light? Might he have decided that light has different properties when it is dim, or comes from the moon? Or would he have placed the cause correctly, within his own visual system?

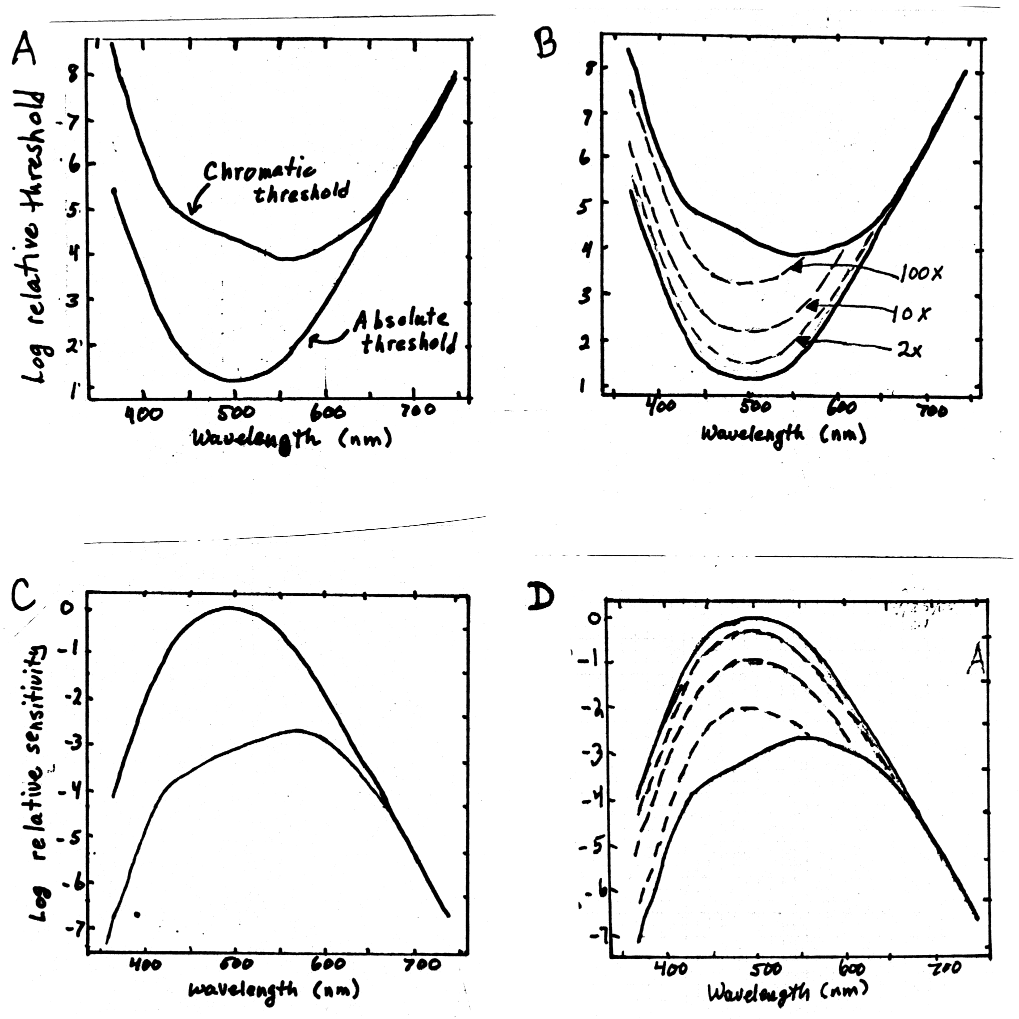


Figure 2.9: Spectral threshold and spectral sensitivity curves. A and B: Spectral threshold curves; C and D: spectral sensitivity curves. A: Absolute thresholds and chromatic thresholds as a function of wavelength. The abscissa shows the wavelength of light; the ordinate shows the relative intensities of the lights of different wavelengths needed for absolute thresholds (lower curve), or for chromatic thresholds (upper curve). B: Scotopic equivalence classes as equal multiples of thresholds. C: Absolute thresholds and chromatic thresholds plotted as spectral sensitivity curves. D: Scotopic equivalence classes plotted in terms of sensitivities. By convention, the maximum of a spectral sensitivity curve is labeled zero (0), with decreases in sensitivity away from the maximum labeled with negative log values (-1, -2, etc.).

them must be rendered equivalent within the visual system. They are also called *silent substitution sets*, emphasizing the idea that one could be substituted for the other with no change of the neural signal.

In addition, the term *metamers* is also used to describe lights of different wavelength compositions that are indiscriminable. The sets of lights indicated by each of the dashed lines of Figure 2.9 are *metamer sets.* The challenge, of course, is to explain why these losses of information occur.

2.7.3 A model of scotopic vision: The funnel analogy

Based on the universal linking proposition (Chapter 1), we posit that the system properties of scotopic vision arise from the physiological properties of the visual system. But specifically, how? As psychophysicists, we are entitled to use the system properties of vision as a basis of theory and speculation about how the visual physiology might work.

Here's a theory of how the scotopic spectral sensitivity curve might come about. Suppose that there were an anatomical stage of the visual system composed solely of a set of identical elements, and that each element had a spectral sensitivity curve matching that of the psychophysically defined scotopic curve in Figure 2.9. Under these assumptions, the system as a whole would necessarily show a spectral sensitivity curve that matches the scotopic curve. So we may choose to adopt the hypothesis that such elements, and such a processing stage, exist, and decide to go look for them.

But how shall we explain both the scotopic spectral sensitivity curve and the loss of wavelength information at the same time? It seems kind of odd – as though the system is influenced by wavelength yet loses wavelength information. But a gadget that would have the right properties can be created by combining a funnel with a counter (Figure 2.10). This analogy may seem a bit silly to those with some physical sophistication, but it will come in handy when things get more complicated later.

Let us assume we have an ordinary kitchen funnel. The funnel is a bit misshapen, being widest at the middle and narrowest at the bent corners. We add a counter to its output spout. To make the analogy, we metaphorically place the funnel under the wavelength scale of the scotopic spectral sensitivity curve, with the widest part at about 500 nm. Along the wavelength scale, we think of curtains of marbles of different colors, perpendicular to the page, raining down on the funnel at a specified rate. The probability that a marble of any given color will be caught is determined by to the width of the funnel at each particular wavelength. But once a marble is caught, it just rolls down to the spout of the funnel, and gets counted by the counter.

This analogy is also easy to express mathematically. Let R be the total number of marbles caught by the funnel – the count on the counter at the base of the funnel. For each wavelength λ, let Q_λ be the number of marbles that arrive per unit time in the curtain of marbles incident at the mouth of the funnel, and let r_λ be the width of the funnel at that wavelength. The catch of marbles at any wavelength, then, is just the number of incident marbles of the color corresponding to λ, multiplied by the probability that a marble of that color will be caught.

$$R = Q_\lambda r_\lambda$$

For several wavelengths – say, 450, 500, and 550 nm – presented together, we just add up the individual catches of marbles:

$$R = Q_{450}r_{450} + Q_{500}r_{500} + Q_{550}r_{550}$$

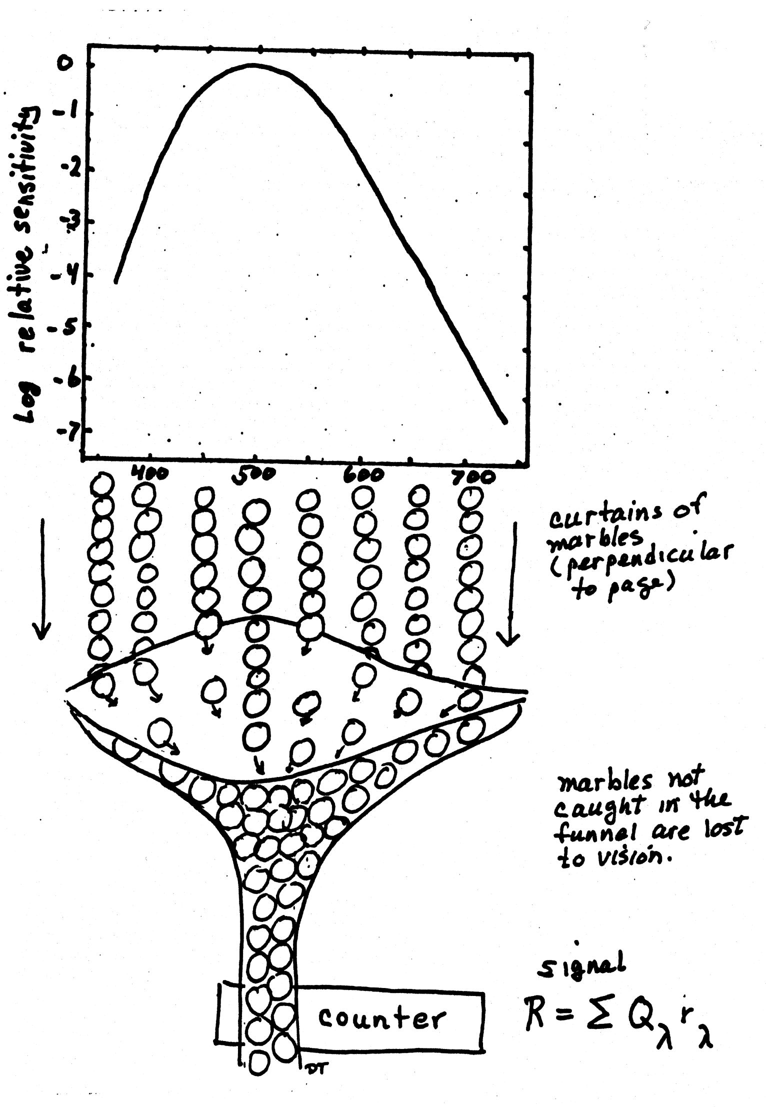

Log relative sensitivity
0
-1
-2
-3
-4
-5
-6
-7
400
500
600
700
curtains of marbles (perpendicular to page)
marbles not caught in the funnel are lost to vision.
signal
$R = \Sigma Q_\lambda r_\lambda$
counter
DT

Figure 2.10: The funnel analogy. In this analogy the visual system is modeled by a funnel with a counter attached. Curtains of marbles, perpendicular to the page, rain down upon the funnel. The funnel varies in width, providing an analogy for the fact that sensitivity varies with wavelength. But the counter only counts the marbles, providing an analogy for the loss of wavelength information that creates scotopic metamers.

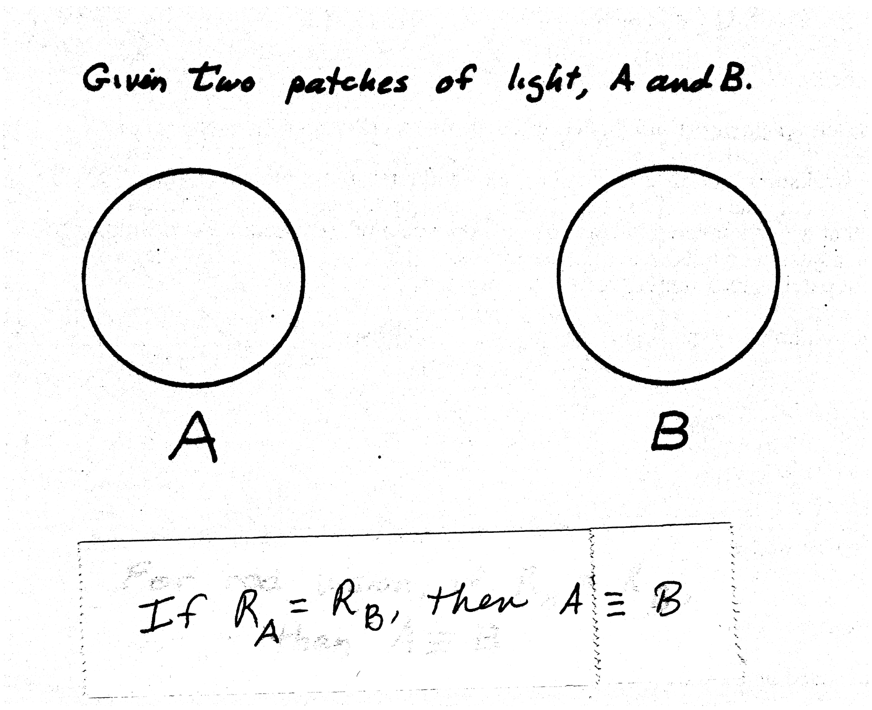

Given two patches of light, A and B.
A
B
If $R_A = R_B$, then $A \equiv B$

Figure 2.11: Scotopic metamers. A and B are lights of two different spectral compositions and intensities. If the intensities of A and B can be adjusted to make the two stimuli indiscriminable, A and B are called metamers. In the funnel analogy, lights A and B are indiscriminable because they make equal counts, R_A and R_B, on the counter.

For the whole spectrum of wavelengths, we keep adding to get a final expression:

$R = \sum Q_\lambda r_\lambda$, or for the continuous case, $R = \int Q_\lambda r_\lambda d\lambda$

The conditions that lead to metamerism in scotopic vision can now be formally stated. Suppose we have two patches of light, A and B, as shown in Figure 2.11. If the catches of marbles from the two patches are identical, the lights must be metamers. Why? Because the model system only has signals corresponding to R_A and R_B. There's no way for the system to encode the information that the two physical stimuli are different. So by necessity:

If $R_A = R_B$, *then* $A \equiv B$

Where = is the symbol for a metameric match. Hold the thought.

2.8 Chromatic thresholds and photopic vision

Now, let's relax our testing techniques for a while, and just let the subject tell us what things look like (or look at them ourselves). Let's present two dim stimuli, a spot of 500 nm and a spot of mixed wavelengths that looks white, and set them both to intensities just above their absolute thresholds. The subject tells us they both look white, and equally bright. Now we double the intensities of both lights together, and keep doubling them, each time asking the subject to tell us what he sees. What will happen?

At first, the subject continues to say that both spots look white, with the brightness of both spots increasing together as intensity is increased. But after the intensities have been increased about a thousandfold (3 orders of magnitude), the subject suddenly says, "OH! The one on the left just turned green!" The intensity at which the 500 nm light changes from looking white to looking green defines the subjects *chromatic threshold* – the threshold for the onset of color sensation – for 500 nm light. We have passed from the scotopic to the photopic realm.

This experiment can be repeated for each wavelength. Interestingly, the intensity range between the absolute threshold and the chromatic threshold varies with wavelength. The lowest absolute threshold occurs at about 500 nm. But, as shown in Figure 2.9, the lowest chromatic threshold occurs at about 555 nm, and chromatic thresholds increase as the wavelength increases or decreases from this value. Moreover, the shape of the chromatic threshold curve is not a simple U – it is asymmetrical, with a shallower rise at shorter wavelengths.

In sum, as we turn up the intensity of each light past a critical level, the system properties change. The *photopic spectral sensitivity curve* has its maximum at about 555 nm, and is more complex in shape than was the scotopic curve. It is also labile – its shape is affected by the method and conditions of measurement, as we will see in the next chapter. And above all, we become able to discriminate among different wavelengths of light – scotopic vision gives way to photopic vision: wavelength information is preserved, and color vision occurs.

2.9 Summary: Scotopic system properties in search of explanations

In Chapter 2 and 3 we introduce the field of psychophysics – behavioral methods for quantifying the mappings between physical stimuli and perceptions. In the present chapter we have concentrated on the kinds of experiments that Brindley called Class A – detection and discrimination. We described four examples of methods for measuring thresholds, and reviewed some of the advantages and limitations of each. We also introduced the theory of signal detection (TSD), which provides a mathematical model that allows the separation of sensory from cognitive parameters.

Next, we returned to the concept of a linking proposition. We reiterated the claim that arguments from psychophysics to physiology, or vice versa, will always involve linking propositions. We elaborated on a family of relational linking propositions – the Identity family – that deal with information retention and information loss. We argued that Identity propositions enter into physiological conclusions based on detection and/or discrimination data. The properties of the Identity family set the stage for examination of other, less intuitively obvious linking propositions in future chapters.

Finally, we used sets of threshold measurements to define two system properties of scotopic vision. Sets of detection thresholds were used to define a scotopic spectral sensitivity curve, and sets of discrimination failures were used to define metamer sets among supra-threshold but still scotopically detected stimuli. These two system properties may be seen as intuitively contradictory – the scotopic visual system is influenced by the wavelength of light, yet loses all information about it. In any case, these system properties are in need of physiological explanation.

We then invoked the Universal linking proposition to infer that elements that produce these system properties will be found within the visual system. We made use of our psychophysicist's speculation license to design a gadget, described by the funnel analogy, that would mimic scotopic vision. In fact, neural elements with the right characteristics will emerge without warning within the next several chapters. Keep the system properties of scotopic vision in mind, and when you think you spot the neural elements that explain them, make a note of them (but unless you are Archimedes, do not jump out of the bathtub and run up the street yelling ”eureka!”).

In the next chapter we turn to Class B psychophysical techniques – techniques for studying the supra-threshold characteristics of visual perception – using examples from photopic vision.

Chapter 3

Psychophysics: Class B Experiments

In Chapter 2 we began the discussion of psychophysics and psychophysical techniques. We introduced the distinction between Class A and Class B experiments, and discussed Class A experiments –experiments involving measurement of detection and discrimination thresholds – at length. Our examples had to do with measuring detection and discrimination thresholds for various wavelengths and intensities of light. In these experiments the perceived colors of the lights might or might not have been changing with light level and wavelength. But in order to stay within the domain of Class A experiments, we didn't ask the subject about these qualitative variations – a sort of "don't ask – don't tell" mentality applied to visual perception. As you may have noticed, one Class B observation did creep into the discussion, when the subject exclaimed, "Oh, the one on the left just turned green!" as we reached the chromatic threshold for a 500 nm light. In a strictly Class A experiment, the most we could do would be to show that above the chromatic threshold the subject could discriminate the 500 nm light from lights of other wavelength compositions, but we wouldn't ask him what color it looked, and he wouldn't tell.

But suppose we are specifically interested in the perceptual qualities of lights that are clearly detectable, and discriminable from each other. Suppose we want to map wavelength to perceived color, or set lights of different wavelengths to be perceived as equally bright, or characterize the supra-threshold similarities and differences among perceived colors. These questions simply cannot be addressed with Class A experiments. We will have to bite the bullet, and start asking subjects what they actually see.

Another way of highlighting the distinction between Class A and Class B experiments is to break down the difference between them into three interrelated parts. First, as discussed in Chapter 2, when forced-choice techniques are used, Class A experiments can be seen as externally referred – the subject's task concerns a judgment about the physical stimuli. Second, therefore, the subject's judgments can be either correct or wrong, depending on the state of the physical world. And third, since the judgments can be either correct or wrong, trial by trial feedback can be given to the subject. It can be argued that feedback provides a mechanism whereby the subject can learn more and more about exactly how the experimenter is defining the task. Overall, the experimenter exerts a great deal of control in a Class A experiment.

In contrast, in a Class B experiment the judgments are internally referred – the subject's task concerns a judgment about how things appear to her. What color do lights of 450, 500, 550, and 600 nm look? Second, therefore, there are no wrong answers; provided that the subject is telling the truth, we have to take her responses as correct at face value. And third, since her judgments cannot be wrong, no meaningful feedback can be given. It wouldn't make sense to say "correct" after every trial. We can give the subject descriptions and instructions, but we can't use feedback to further specify the task.

Viewed in this light, we can see that Class A and Class B experiments have tradeoffs of advantages and disadvantages. The most logically elegant techniques for Class A experiments – forced-choice techniques with feedback – embody questions about the physical world, and have an objectivity to which Class B experiments can never aspire. But all that Class A experiments can reveal is the discriminability of stimuli – for some of us, an impoverished topic at best! On the other hand, Class B experiments can be attacked for their subjectivity. But we get to set aside thresholds, and study what many of us came to study in the first place – the qualities of the subject's perceptions. Of course, given that we choose to do Class B experiments, we will want to develop and use the most rigorous possible techniques, as one would in any branch of science, and some examples of these techniques will also be introduced in this chapter.

There is one other major difference between Class A and Class B experiments: different linking propositions will necessarily be involved in the causal stories that arise from them. When we were studying the discriminabilities of stimuli, all we needed was the Identity family. But if we are to study perceptually richer phenomena, we will need more complex linking propositions, with names like Similarity and Mutual Exclusiveness. These new linking propositions will be discussed below.

In the present chapter, we discuss two major cases of Class B phenomena in the context of color vision. Case 1 deals with photopic spectral sensitivity, and Case 2 with the qualitative characteristics of perceived colors.

3.1 Case 1: Photopic spectral sensitivity

In physics, the intensity of a spot of light of a single wavelength is specified in terms of its radiant energy, E, in *watts*. But suppose that the physicist wants to specify the total physical energy in a spot of light made up of many different wavelengths. He can specify the radiant energy of each individual wavelength, E(λ), but how do they combine? If things were complicated, different wavelengths could be like apples and oranges, and the energies might not follow a simple combination rule. But in fact the physicist is lucky, because the measurement system works if the total energy, E, in a light of mixed wavelengths is specified to be just the sum of the energies at each of the individual wavelengths; that is,

$$E = \sum E(\lambda).$$

In other words, radiant energy is *additive* across wavelength. Additivity, of course, is a necessary condition for well-behaved units of measurement – if 2 inches plus 3 inches doesn't yield 5 inches, we are all in trouble.

But now suppose we want to specify the amount of light coming from a light source in terms of its *effectiveness for human vision.* For example, as a practical problem, suppose we want to design an airport runway system, using lights of different colors to mark different runways. What intensities shall we use? A physical specification, such as the energy of each bulb in watts, is not useful, because radiant energy near the middle of the visible spectrum is orders of magnitude more effective for vision than is radiant energy at the spectral extremes (see Figure 2.9). And radiant energy outside the visible spectrum, by definition, has no visual effectiveness at all.

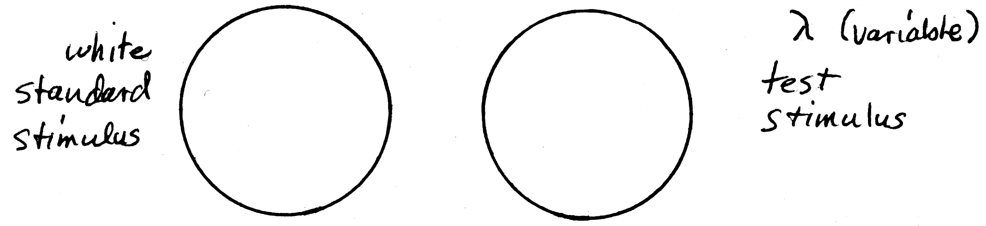


Figure 3.1: Heterochromatic brightness matching. The experimenter sets the radiance of the white standard light, and sets the test light to each of a series of different wavelengths in turn. The subject's task is to vary the radiance of the test light, to match the perceived brightness of each test light to the brightness of the fixed white standard light.

In such cases it is useful to specify stimuli in *quasi-physical* units; that is, in units based on their effectiveness for human vision. The appropriate quasi-physical units would weight the radiant energies of the different wavelengths, E(λ) by some psychophysical measure of the "visibility" – call it V(λ)– of the different wavelengths for the human eye. And ideally, the units would be additive: the total visibility, V, of a light of mixed wavelengths would be just the sum of the energy at each wavelength weighted by the visibility of light at that wavelength. Ideally,

$$V = \sum V(\lambda)E(\lambda).$$

????????shall????specify the "visibilities" of lights of different wavelengths, to allow these computations ? ??????????additivity prevail? Or are different wavelengths of light like apples and oranges when human vision is involved?

3.1.1 Heterochromatic brightness matching

The most obvious approach to the problem is simply to ask the subject to look at lights of different wavelength compositions, and match them in brightness. This task underlies the technique of *heterochromatic brightness matching.* The experimental set-up is shown schematically in Figure 3.1 A. We set up two spots of light: a *standard* light – say, a patch of white light at a fixed radiance – and a *test* light of the same size. To use the method of adjustment, we would set the test light to each of a series of different wavelengths, and for each wavelength have the subject vary the radiance of the test light to make brightness matches between the two. This is a Class B experiment because the white and chromatic lights are discriminable throughout the experiment. We are asking the subject to attend to one perceptual property – brightness – while ignoring another property – color.

A spectral sensitivity curve resulting from a heterochromatic brightness matching experiment is shown by the squares in Figure 3.2. Notice that the maximum sensitivity – the minimum energy required for a brightness match to the fixed white standard – is no longer near 500 nm, the sensitivity maximum we saw for scotopic vision. Instead, it shifts to about 555 nm, and the curve falls off sharply toward both shorter and longer wavelengths.

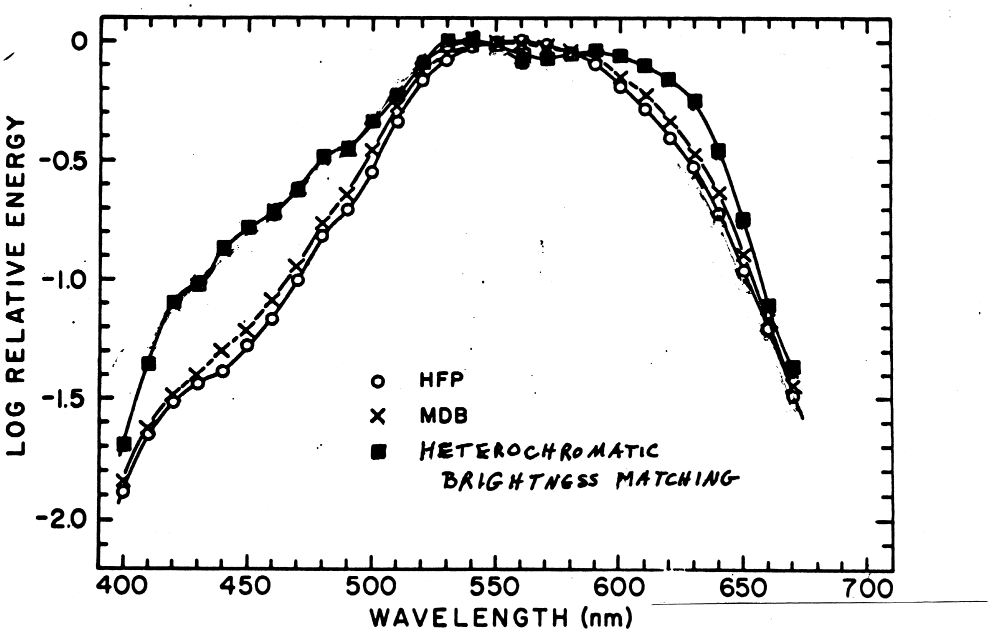


Figure 3.2: Photopic spectral sensitivity curves. The ordinate shows the relative radiance (physical energy) required for a given perceptual match. The results of heterochromatic brightness matching are shown by the closed squares. The open circles and the crosses show results from two photometric techniques, heterochromatic flicker photometry (HFP) and minimally distinct border judgments (MDB). The data from HFP and MDB agree well, whereas the data from heterochromatic brightness matching yield a broader curve. [Data from Wagner and Boynton, 1972. Graph adapted from Pokorny et al, 1979, p. 25, Fig. 2.1.]

3.1.2 Problems with brightness matches: Variability and non-additivity

Unfortunately, heterochromatic brightness matching is plagued by two problems. First, subjects find the task difficult, because the two lights that are to be matched in brightness differ so much in color. There is considerable variability from trial to trial within a session and between sessions for a single subject, and also considerable variability among subjects. We interpret these problems to mean that there is no *natural perceptual criterion* – no perceptually striking event that happens right at the brightness match – to guide the subject's performance. Lights of higher radiance look too bright; lights of lower radiance look too dim; nothing looks quite right; and the subject is on her own to interpret the instructions and do the best she can.

A second and more devastating problem is that the resulting values are non-additive. The non-additivity is shown schematically in Figure 3.3. Suppose that the subject sets a 650 nm light and then a 500 nm light to match the white standard. Now, we divide the intensities of the 650 and 500 nm lights in half, and superimpose the two lights. Subjects report that now the test and standard lights no longer match in brightness; the superimposed 500 and 650 nm lights look markedly dimmer than the white standard. These failures of additivity can be as much as a factor of five. The radiant energies of lights of different wavelengths sum additively, but their perceived brightnesses do not. Crazy physics indeed!

The non-additivity of heterochromatic brightness matching data makes this technique unacceptable as a basis for equating lights for visual effectiveness. We are left with our original practical problem: how to develop a set of quasi-physical units that specify lights in terms of their effectiveness in human photopic vision, yet are additive across wavelength.

3.2 Photometric methods

In about 1900, vision scientists had the insight that perhaps a solution might come from varying the psychophysical *task*. That is, perhaps a different task would yield additive values. In pursuit of this goal, several other methods, called *photometric methods*, have been developed. We will discuss three of these – heterochromatic flicker photometry, minimally distinct borders, and motion photometry – because they provide interesting examples of variations of the subject's task; and because, as it turns out, they each yield an additive system. Notice that in each case, the subject's task is to set two intensities to yield a perceptual minimum – in perceived flicker, in the perceived distinctness of a border, or in the amount of perceived motion – at a pair of relative intensities at which there is no physical reason to expect one.

3.2.1 Heterochromatic flicker photometry (HFP)

The oldest of the photometric methods is called *heterochromatic flicker photometry*, or *HFP*. In HFP, as shown in Figure 3.4A, we again choose a white light as the standard, and use each of a series of test lights of different wavelengths. We arrange to alternate the stimulus in time between the standard and test lights, at a rate of, say, 15 cycles per second (15 Hertz, or Hz). Perceptually, the light appears to flicker between white and the color of the test light. But as we vary the radiance of the test light over an appropriate range, remarkably, the sensation of flicker diminishes, passes through a minimum, and then increases again. The subjects task in HFP is to vary the radiance of the test light until the sensation of perceived flicker is minimal. We repeat the experiment for

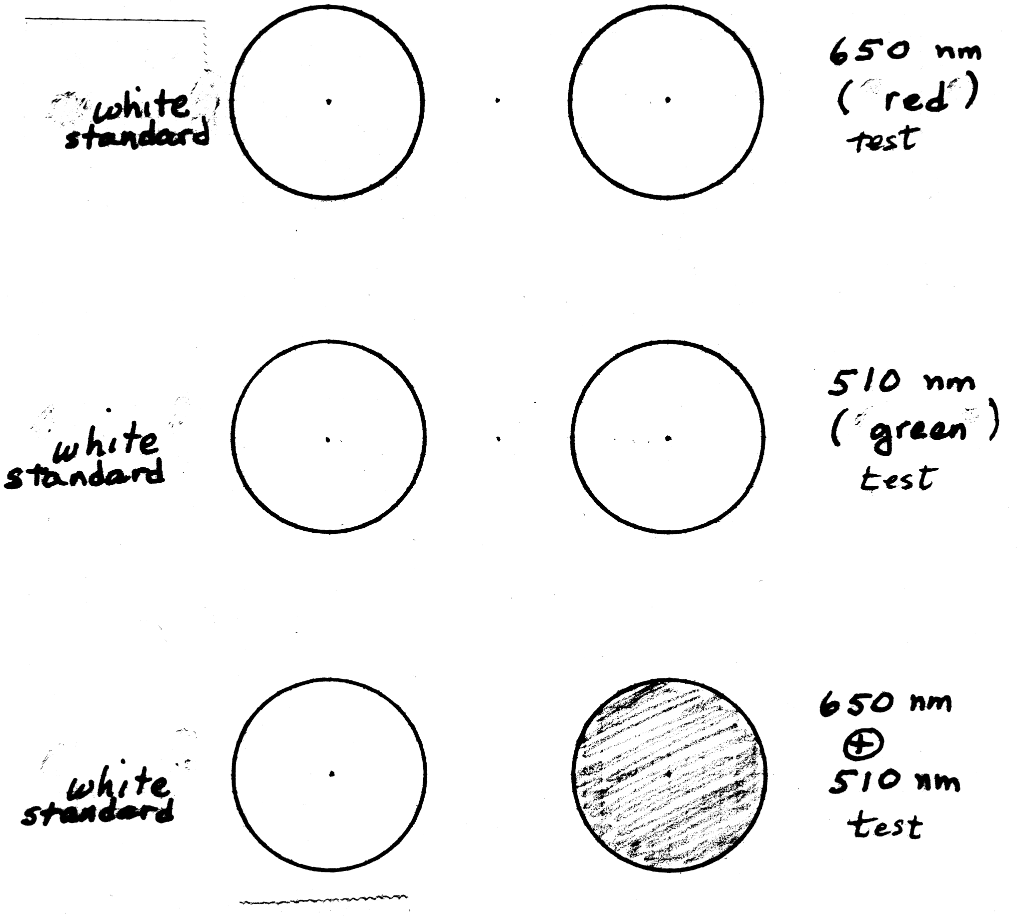


Figure 3.3: Non-additivity of heterochromatic brightness matches. The subject first matches the brightness of a 650 nm light to that of a white standard light, to yield a matching value – call it M650. He then matches the brightness of a 500 nm light to the same white standard light to yield a matching value M500. The experimenter then superimposes 1/2M650 and 1/2 M500, (the symbol @ stands for superposition), and asks the subject if the combination matches the same white standard. The subject's answer is no – the combined test field looks dimmer than the standard. Brightnesses are non-additive.

each different wavelength in turn.

It turns out that with a little practice, subjects can do this task remarkably consistently, both from one day to the next and from one subject to the next. We take this consistency to suggest the presence of a natural perceptual criterion in this task. A distinct perceptual event – a minimum in perceived flicker – is seen reliably at a particular radiance of the test light, and guides the judgments, allowing a high degree of consistency both within and between subjects.

The results of using the HFP technique are shown along with the heterochromatic brightness matching data in Figure 3.2. As was the case with heterochromatic matching, the maximum of the HFP curve falls near 555 nm. The two curves are similar in shape, but they differ in detail, with sensitivity at both extremes of the spectrum being higher with heterochromatic brightness matching than with HFP. And the matches made with HFP are additive.

3.2.2 Minimally distinct border (MDB) judgments

A second method of photometry is called the *minimally distinct border (MDB)* method. The MDB paradigm is shown in Figure 3.4B. The stimuli are a white standard and a series of test lights of different wavelengths, juxtaposed at a very sharp border. For each wavelength, subjects report that at a particular narrow range of relative radiances, the border between the two fields loses its perceptual sharpness – it becomes relatively indistinct. For some wavelengths, the border even appears to melt completely away, so that perceptually the colors of the two fields blend or smear together across the center of the stimulus field. The subject's task is to set the radiance of each test field to yield a minimally distinct border against the fixed standard light.

A spectral sensitivity curve generated by MDB judgments is also shown in Figure 3.2. The curve generated by MDB is highly similar to the curve found with flicker photometry, and again the values are additive. In addition the results are highly consistent, suggesting that a distinctive natural criterion – the minimal sharpness of the border – occurs reliably at one particular radiance of each wavelength.

3.2.3 Motion photometry

A third method of photometry, called *motion photometry*, is shown in Figure 3.4C. In motion photometry, a subject is shown a set of stripes of two different wavelength compositions (say, stripes of a white standard alternating with red, blue or green stripes produced by one of the three phosphors on a video monitor)[1]. The grating is set in motion across the face of the monitor. The radiance of the white bars is fixed, while the radiance of the chromatic bars is varied. As it varies, surprisingly, the perceived velocity of motion slows or even stops, and then speeds up again. The subject's task is to vary the radiance of the chromatic bars until the perception of motion is minimized. Again, the perceptual event is very distinct, and again the relative radiances required for motion minimization resemble those of flicker photometry and MDB methods, and again the values found are additive.

In summary, the problem of equating photopic lights of different wavelength compositions in visual effectiveness has been attacked with several tasks: match the brightnesses, minimize the

[1] Motion photometry is usually done on a video monitor because this is by far the easiest way to create the pattern of moving stripes. Since the phosphors of a video monitor produce only broadband stimuli, spectral sensitivity curves cannot be measured directly. However, the motion minima are predictable from HFP or MDB data, by adding up the visual effectiveness of the different wavelengths emitted by the phosphor, as determined with HFP or MDB.

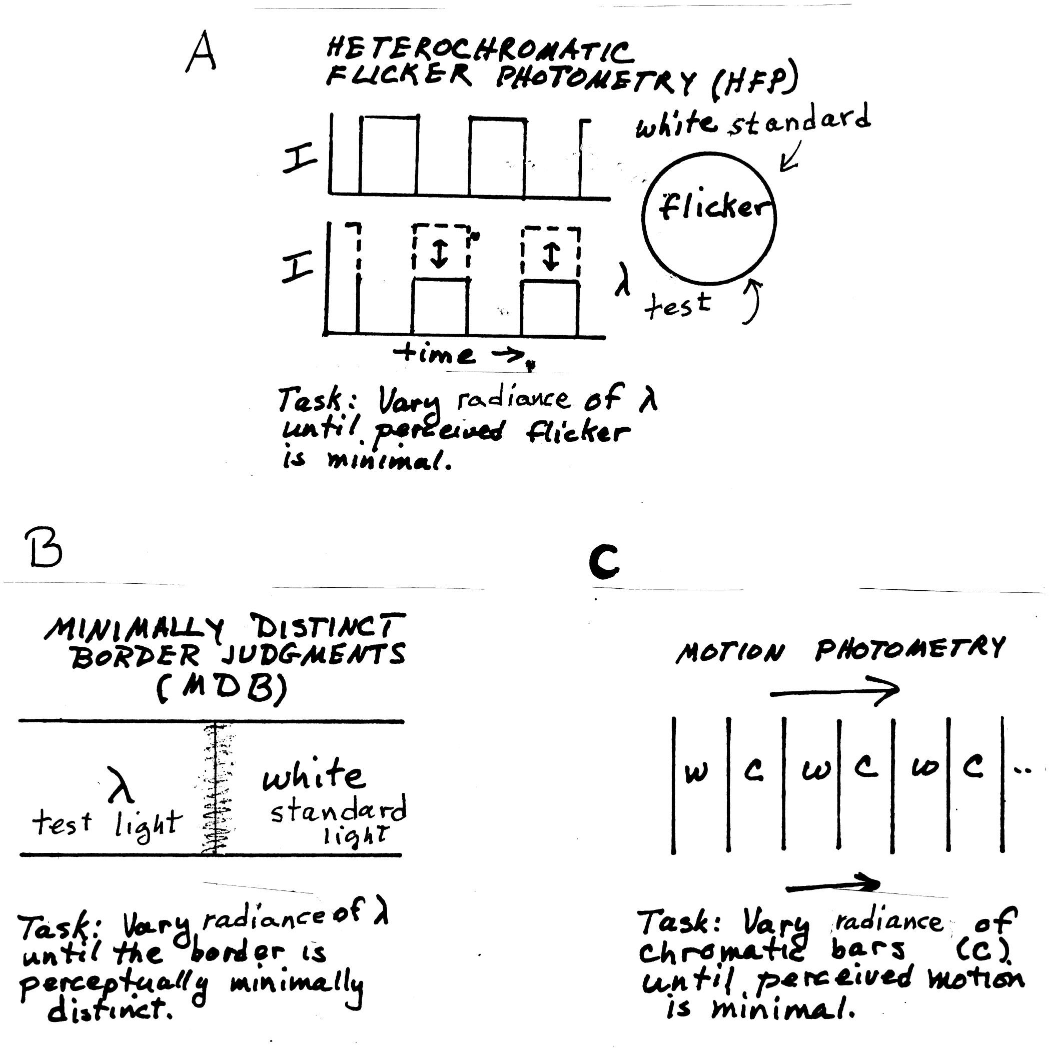

A
HETEROCHROMATIC FLICKER PHOTOMETRY (HFP)
I
I
white standard
flicker
λ
test
time →
Task: Vary radiance of λ until perceived flicker is minimal.
B
MINIMALLY DISTINCT BORDER JUDGMENTS (MDB)
λ
test light
white
standard light
Task: Vary radiance of λ until the border is perceptually minimally distinct.
C
MOTION PHOTOMETRY
w
c
w
c
w
c
..
Task: Vary radiance of chromatic bars (c) until perceived motion is minimal.

Figure 3.4: Three photometric methods. A: Flicker photometry. B: Minimally distinct border judgments. C: Motion photometry. In each case, the subject's task is to vary the radiance of the test light to set some aspect of her perception – perceived flicker, or the perceived distinctness of a border, or perceived motion – to a minimum.

perceived flicker, minimize the perceived distinctness of a border, or minimize the perception of motion. The last three produce estimates of visual effectiveness that are very similar. Moreover, all three are closely additive. These results are important because they have allowed vision scientists to develop a quasi-physical stimulus specification system that is both reasonably similar to the results of heterochromatic brightness judgments (compare the data in Figure 3.2), and at the same time, closely additive.

At the same time, from a purely physical point of view, the reason these tasks work is mysterious. There is no physical reason why perceived flicker, or perceived border sharpness, or perceived velocity of motion, should go through a distinctive minimum at some particular relative intensities of the test and standard lights. Moreover, there is no obvious reason why the quasi-physical specification systems they generate should be additive. More system properties in search of explanation.

3.3 The specification of light: physical vs. quasi-physical units

3.3.1 The standard spectral luminous efficiency functions V(λ) and V'(λ)

The additivity of photometric matches has allowed the development of an additive quasi-physical stimulus specification system for photopic vision. Figure 3.5 shows the now-standard photopic spectral sensitivity curve, or more formally the *photopic spectral luminous efficiency curve, V(λ)*, together with the corresponding *scotopic spectral luminous efficiency curve V'(λ).* These curves were adopted as standards in 1924 and 1951 respectively by the International Committee on Illumination, or more properly the Committee Internationale de L'Eclairage (CIE). They are widely used in industry, to approximate the radiances of lights of different wavelengths needed to produce lights of equal effectiveness for human vision.

Since the measurements that enter into the CIE photopic spectral luminosity curve V(λ) are additive, the total photopic effectiveness of a light of mixed wavelengths can be specified in this system by multiplying the energy at each wavelength by the visibility at that wavelength, and adding up the visibilities of its components, as we had hoped. That is, for photopic vision, we can now legitimately define the *photopic luminance*, L, of a light of mixed wavelength composition by:

$$L = \sum V(\lambda) E(\lambda).$$

The *luminance* of a light, then, is its intensity specified in units of its effectiveness for human photopic vision. A similar equation can be written for scotopic vision.

3.3.2 Radiometric vs. photometric units

Implicit in the above discussions is a distinction between two different ways of specifying the intensities of lights. *Radiometric units* are purely physical units – they specify light of any specific wavelength or wavelength mixture in units of physical energy (such as watts). In contrast, *photometric units* are quasi-physical units that weight the energy at each wavelength by the sensitivity of the human eye at that wavelength. Terminologically, words with the root *rad* (*radiance, irradiance, radient flux*, etc.) are used in specifying light levels in radiometric units. Terms with the root *lum* (*luminance, illuminance, luminous flux*, etc) are used in specifying light levels in photometric units. The curve V(λ) shows the multiplicative factors needed at each wavelength to convert radiometric to photometric units at photopic light levels.

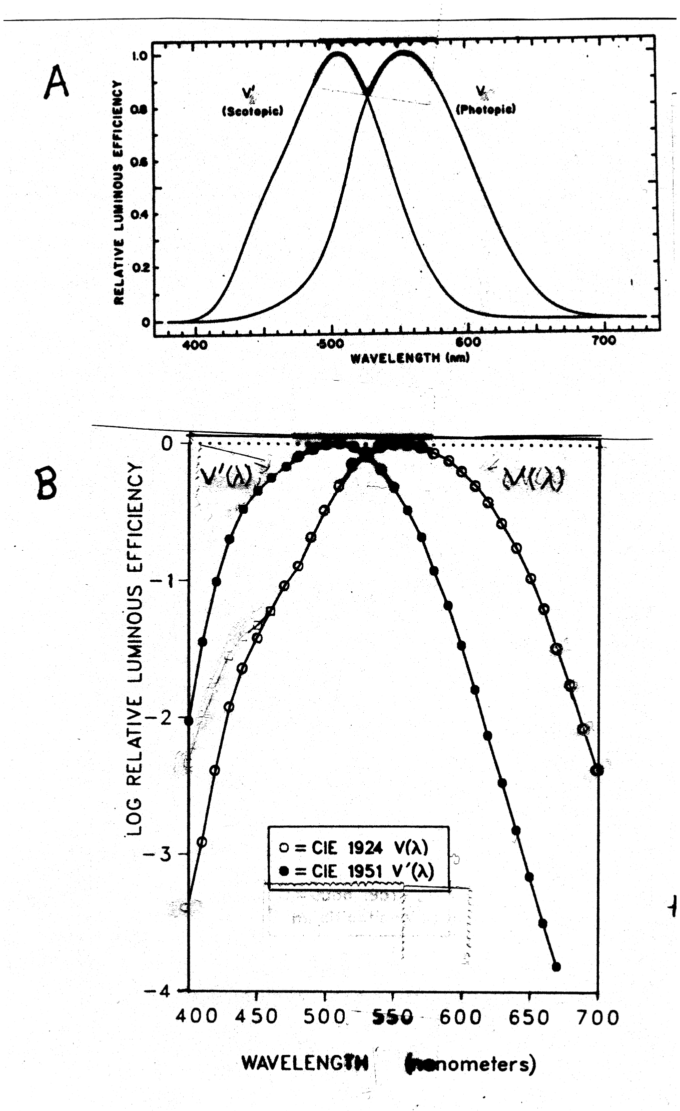


Figure 3.5: Standard photopic and scotopic spectral luminosity curves. V(λ) is the photopic curve, adopted by the CIE in 1924. V'(λ) is the scotopic curve, adopted in 1951. A. Plotted on a linear ordinate. B. Plotted on a logarithmic ordinate. Notice how the linear ordinate shows the detail near the maximum of each curve, but obscures it in the tails of the distribution; whereas the logarithmic ordinate compresses the values near the sensitivity maximum, but displays the continuing falloff of sensitivity in the tails. [A adapted from Pokorny et al, 1979, p. 28; B adapted from Pokorny and Smith, 1986, Fig. 8.9, p. 8-13.]

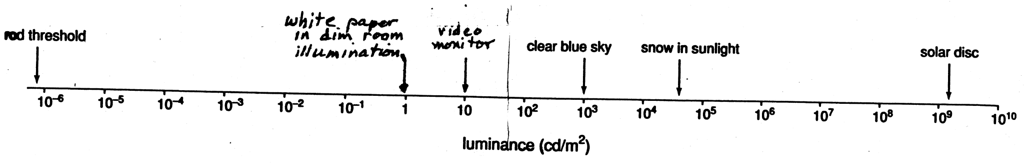


Figure 3.6: The luminances of some familiar objects, specified in candelas per meter squared (cd/m^2). Luminance is a quasi-physical specification of light level, in which the radiance at each wavelength is multiplied by the sensitivity of the human visual system at that wavelength. [Adapted from Rodieck, 1998, p. 152 (no figure number).]

Over the years many different sets of photometric units, with different names, have been used, but these need not concern us here. Suffice it to say that luminance is now usually specified in units of *candelas per meter squared (cd/m^2).* The luminances of a few familiar surfaces in cd/m^2 are shown for reference in Figure 3.6.

Finally, a brief mention of one other aspect of the specification of lights. It is important for various reasons to distinguish between the amount of light *falling on* a surface, and the amount of light *emanating* from the surface; these are specified in different kinds of units. Terms with the prefix *ir* or *il* (*irradiance, illuminance*) refer to the former, whereas terms without this prefix (*radiance, luminance*) refer to the latter. If the surface reflects, say, half of the light falling on it, then its luminance will be half of the illuminance falling on it, assuming that properly corresponding units are used.

We can now return to another practical problem – that of specifying the visual effectiveness of a light bulb. Light bulbs are traditionally specified in *watts* – a physical unit. That is, a 500 watt light bulb is being specified in terms of the energy it uses, but not directly in terms of its effectiveness for human vision. If it is specified in *lumens* it is being specified in photometric units, and its effectiveness for human vision is built into this specification. [Would it make sense to specify light in photometric units when testing the visual sensitivity of a goldfish? Why or why not? What would you do instead?]

3.4 Summary: System properties of scotopic vs. photopic vision

With our photopic spectral sensitivity curve in hand, we are finally ready to compare the system properties of scotopic and photopic vision. These properties are summarized in Table 3.1. The main features are these: the scotopic spectral sensitivity curve has a maximum sensitivity at about 500 nm, whereas the photopic maximum is at about 555 nm. The scotopic curve is a simple, smooth and relatively symmetrical U, and is remarkably stable over variations of measurement techniques and conditions, whereas the photopic curve is more complex and more labile. And wavelength information is lost at scotopic light levels, but preserved at photopic levels. Finally, with spectral sensitivities specified in accord with V(λ) and V'(λ), systems of units are additive for both scotopic

	λ max	Smooth over λ?	Stable over tasks?	Discrimination among λ's?
Scotopic vision	500 nm	Yes	Yes	No
Photopic vision	555 nm	No	No	Yes

Table 3.1: Summary of the properties of scotopic and photopic vision.

and photopic vision.

In Chapter 2, we argued that the funnel analogy captures the essence of two system properties of scotopic vision: sensitivity varies with wavelength, yet wavelength information is lost. In a similar vein, what can we infer, or what speculations might we be drawn to, by the differences in properties between scotopic and photopic vision summarized in Table 3.1?

3.5 Physiological implications of photometry

3.5.1 Causal stories: Bumblebees can fly

In G.E. Mueller's time, before neurobiological techniques were available, one of the goals of psychophysics was to use the system properties to deduce or guess the nature of the underlying physiological processes. Having read only the first few sections of this book so far, you are in the unique position of being like the early psychophysicists – you know quite a lot about the system properties of scotopic and photopic vision, but little about the physiological processing that gives rise to them. So think about the following "bumblebees can fly" arguments through your soon-no-longer-to-be nave eyes. Take the following paragraphs as exercises, and try to sort out strong from weak causal stories, and speculations from guesses.

The differences in spectral sensitivity lead to some new constraints on models. For example, the difference in the wavelength for maximum sensitivity for V(λ) and V'(λ) implies that scotopic and photopic vision arise from at least partially different physiological processes with different spectral characteristics. The simplicity and stability of the scotopic curve V'(λ) encourages the speculation that the neural basis of scotopic vision consists of a single process with a fixed spectral sensitivity curve that is not affected by task variables. In contrast, the relative complexity and lability of photopic spectral sensitivity suggests that it might be produced by combining inputs from several different physiological processes with different spectral sensitivity curves.

The differences in the loss vs. preservation of wavelength information also give us clues about physiological processes. The failure of wavelength discrimination in scotopic vision suggested a neural process that doesn't preserve wavelength information, like the counter on the funnel in the funnel analogy. The success of wavelength discrimination in photopic vision implies a set of neural processes that do preserve wavelength information – it's a strong inference that a single funnel with a counter will not do. What mechanisms might be involved? [Might several funnels do the trick? Think about this.]

What about additivity? We did not mention it previously, but scotopic vision is additive, as is implicitly assumed as a feature of the counter in the funnel analogy. But we know that the single funnel model must be discarded for the photopic system, since wavelength information is preserved. So how do photometric methods remain additive? One can speculate that both the inputs to each of the putative processes underlying the photopic spectral sensitivity curve (see above), and the combination of signals across these processes, are additive. In contrast, the non-additivity of heterochromatic brightness matching is consistent with a more complicated signal combination rule.

Finally, it is a very interesting question why the tasks that underlie the three photometric techniques – HFP, MDB, and motion photometry – are feasible psychophysical tasks at all. Like the color circle, the sources of flicker, border, and motion minima do not lie in the physics of the stimuli, but rather within the human visual system. It all seems very odd, but we might speculate that the physiological processes that support our perception of flicker, the perceived sharpness of borders, and the perception of motion, all somehow go through internally generated minima at particular relative intensities of lights of different wavelength compositions. The fact that these intensities are the same for all three tasks suggests that the three tasks all depend on neural processes with the same spectral sensitivity curve. Another way of saying this is that a single putative spectral process *leaves its signature* on our perception of flicker, of borders, and of motion.

As a general rule, when the same curve – the same signature – turns up repeatedly, vision scientists tend to speculate that this signature has its origins in the responses of individual neurons. That is, we speculate that there is a set of individual neurons within the visual system that also have this same spectral sensitivity curve. If we take these speculations seriously, we would predict that there will be neurons in the visual system with spectral sensitivities that correspond to V'_λ at scotopic light levels and to V_λ at photopic light levels, and we might choose to go and look for them.

3.5.2 Linking propositions

Finally, of course, each of the attempts at causal stories given above will have at least one linking proposition, implicit or explicit, among its premises. For example, arguments concerning simplicity and stability depend on premises concerning simplicity and stability, such as ”Simple perceptual facts suggest (or imply) simple neural correlates”, and ”stabile perceptual phenomena suggest (or imply) a stable underlying neural process”. Arguments concerning additivity depend on an additivity proposition: ”Additive perceptual properties suggest (or imply) additive neural properties”.

The argument on the tasks of photometry seems to rely on an Analogy linking proposition like, ”Perceptual minima imply corresponding physiological minima”. That is, a minimum of perceived flicker, or border sharpness, or motion, suggests a corresponding minimum in the signal that codes flicker, or borders, or motion respectively, somewhere within the visual system. And arguments from ”signatures” to neurons seem to involve a proposition to the effect that ”A recurring perceptual ”signature” suggests (or implies) the presence of neurons with an analogous signature ”.

These assumptions are all superficially innocuous, and are easy to ignore, or leave implicit in an argument. But they are probably not all true. Watch for processes that conform to the inferences and speculations that depend on them, or that reject them, in later chapters of the book.

3.6 Case 2: The properties of perceived colors

We now turn to a second example of the use of Class B methods. At photopic light levels, colors are perceived. How do psychophysicists specify and quantify the perceived colors of different wavelengths of light, and the relationships among the colors?

3.6.1 Physics vs. perception: The appearance of the spectrum

Imagine that you are looking at a rainbow, or at a spectrum made with a prism. The physical spectrum is a set of lights of different wavelengths, ordered by wavelength. If we ask a subject what he sees, he will say that perceptually the physical spectrum looks like an array of colors, with neighboring wavelengths taking on similar colors. If we give him color names and ask him to point to the location of each color within the spectrum, he will point approximately as follows:

Violets 400-450 nm Blues 460-480 nm Greens 500-530 nm Yellows 560-580 nm Reds 620-700 nm

with the intermediate colors – blue-violets, blue-greens, yellow-greens, and oranges – falling in between. In short, there is an orderly mapping between wavelength and perceived color[2], shown schematically in Figure 4.3 in the next chapter.

Beyond the orderly mapping between wavelength to color, there are a variety of striking differences between physical and perceptual realms in color vision. In particular, the physical spectrum varies continuously in wavelength, and short and long wavelength light have no special physical commonality. Yet most subjects report a perceived similarity – a common reddishness – between short wavelength lights (which appear violet) and long wavelength lights (which appear red). Moreover, if we mix long and short wavelength lights in varying proportions, we can generate a continuous variation in color, from violet through reddish violet to red – colors that never arise from individual wavelengths of light. Wavelength just varies from short to long, but the set of perceived colors makes a circle, as shown in Figure 3.7. And other mixtures of wavelengths are seen as whites and pastel (desaturated) colors.

Now imagine setting up a display with lights of many different wavelength compositions, arranged in haphazard order, but matched in brightness, and asking your subjects to arrange them in order of perceived similarity. If your subjects are color-normal, they will probably tell you that they cannot arrange the colors by similarity along a single line, but that all of the lights fit together naturally on a two-dimensional surface, with saturated colors around the outside (as in Fig. 3.7) and whites and desaturated colors in the center. If we now add variations in intensity to the display lights, the subjects will tell us that they require a third dimension, with the higher intensity lights occupying spaces above the lower intensity ones[3].

Notice that again, the regularities of the three-dimensional color solid correspond only partially to the physical similarities among the stimuli. Nothing in the physical stimulus explains why a straight line from red to green goes through mid-grey in the color circle; and nothing explains why red (associated with long wavelengths) and violet (associated with short wavelengths) are similar in our perceptions. [Invent a family of relational linking propositions – call it the Similarity family – that uses the psychophysical characteristics of the color solid to make predictions about the physiological coding of colors.]

[2] DT is reluctant to emphasize the mapping between wavelength and color, because doing so tends to reinforce the common but erroneous belief that perceived color depends only on wavelength. In fact, perceived color is influenced by many other factors. Except when we are viewing a physical spectrum, the perceived color of a light tells us remarkably little about its wavelength composition. We revisit this topic in Chapter 7.

[3] Like lights, paint chips also fit into a three-dimensional color solid, as can often be seen by the arrangements of chips on swinging leaves in a paint store. The differences between lights and surface (paint) colors, however, is beyond our scope.

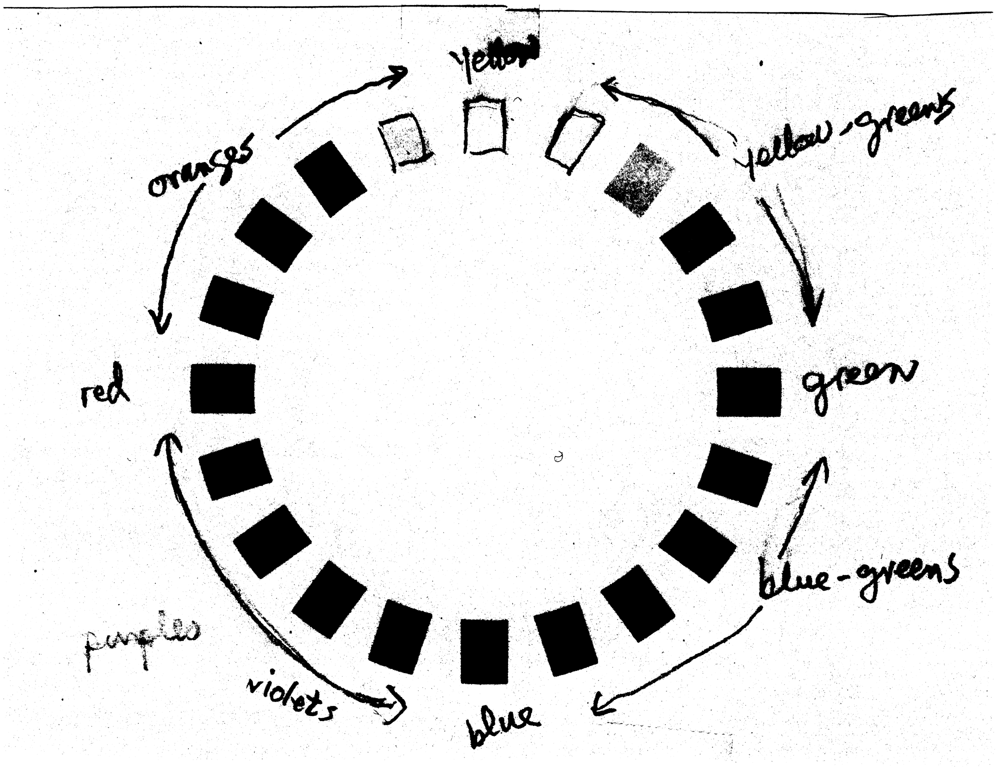


Figure 3.7: COLOR PLATE. A set of saturated colors arranged in a hue circle. [From Boynton, 1979, p. 35, Fig. 2.2.]

3.6.2 Unique vs. binary hues and mutually exclusive hue pairs

In 1878, the German scientist Ewald Hering drew attention to another set of facts about color appearances. First, he argued that perceived hues come in two distinctly different kinds: *unique* vs. *binary*. That is, some hues seem to be made up of perceptual combinations of other hues. For example, if I ask you what color is reddish yellow, you would probably readily come up with orange; purple can be described as a reddish blue, and so on. These are the binary hues. Other hues – red, yellow, green, and blue – seem to be perceptually simple, or unitary; they are not perceptually analyzable into components. For example, if you were asked to imagine a purplish orange, it would take you a while to figure out what hue fall between purple and orange, and then you would probably reject the original request. No, you would probably say, red is not a purplish orange. Red is red, and that's all there is to it! You have just cast a vote for red as a unique hue, and orange and purple as binary hues.

Hering further observed that the unique hues come in two pairs – red vs. green and yellow vs. blue – such that the colors in each pair are perceptually *mutually exclusive*. The claim is, perceptually there is no such thing as a reddish-green or a yellowish-blue. The mutually exclusive hue pairs were also called *opponent hue pairs*, and color theories that arise from these observations are called *opponent process theories* of color vision. A third dimension, lightness, was also included, with white and black as the defining sensations. This dimension isn't like the two color dimensions, particularly in that subjects usually find it possible to imagine blackish whites (as grays).

But shall we believe Hering's observations? Are there really unique and binary hues, and mutually exclusive hue pairs? Not everyone agrees, at least at first. Especially, many people say that for them green is perceptually "made up of" yellow and blue. We suspect that this common report comes from a knowledge of mixing paints; yellow paint mixed with blue paint often does yield a paint that looks greenish. But this is not what we're asking. Rather, we're asking, what do the colors themselves *look like*? To address this question, we need to develop new psychophysical techniques.

3.7 Quantitative color naming techniques

A more quantitative technique for describing the appearances of colors is called *hue naming* or *color naming*. In a color naming experiment, a subject is given a set of color words to use, and asked to use them quantitatively to describe the perceived hues of spectral lights. For example, the subject might be given the names corresponding to Hering's four unique hues – red, yellow, green, and blue – and asked to name the color of each wavelength. If more than one color is perceived, the subject is asked to assign a percentage of the perceived hue to each of the color names. He might describe a 610 nm light (which typically looks orange) as 40% yellow, 60% red; and a 575 nm light as 100% (unique) yellow. The percentages are averaged across presentations to yield an overall percentage score for each color name at each wavelength. Remarkably, subjects do this task very consistently, and agreement among subjects is strong, especially given the seemingly subjective nature of the task. [Try color naming on yourself and a friend with the colors in Figure 3.7. Make sure neither of you is color-blind!]

The results of a color-naming experiment using the four color names red, yellow, green, and blue, over the wavelength range 450 to 660 nm, are shown in Figure 3.8. Under the conditions of this particular experiment, the subjects very consistently used the color names blue, green and

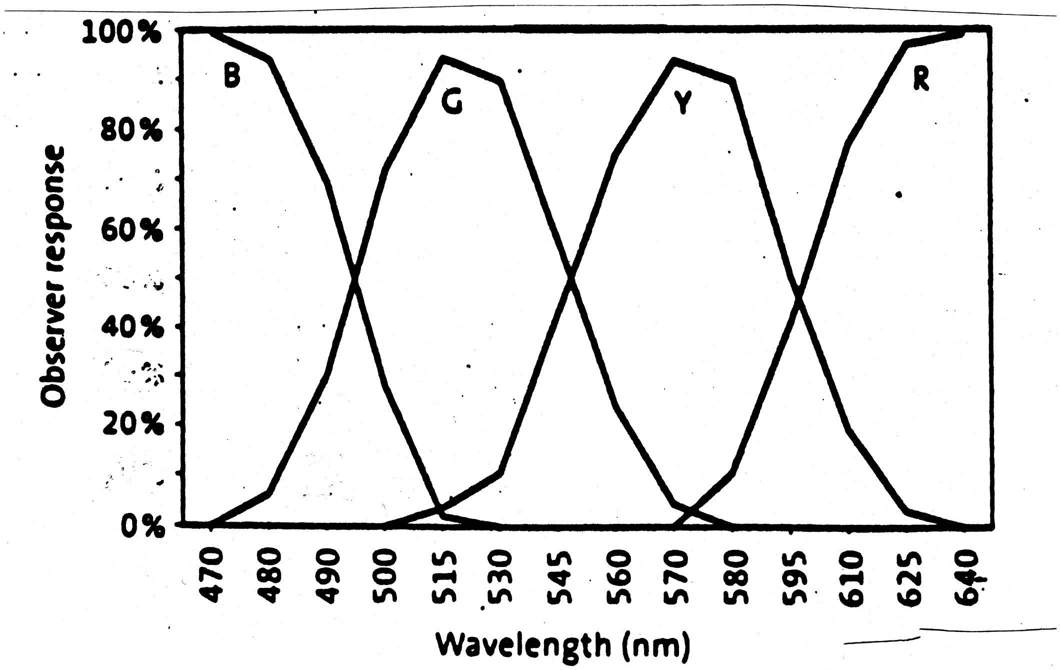


Figure 3.8: Results of a color-naming experiment over a large spectral range, from 470 to 640 nm. Subjects were allowed to use the color names blue, green, yellow, and red. [Adapted from Wooten and Miller (1997), Fig. 3.7, p. 77.]

yellow over the appropriate wavelength ranges. The consistency of use of the color names verifies the orderly mapping of wavelengths to colors. Unfortunately wavelengths below 470 nm were not used in this experiment; but other experiments show that in addition to the color name blue, the color name red is used at wavelengths below 450 nm. This result validates the claim (Figure 3.7) that perceptually the colors form a circle rather than just an ordered line.

Experiments like these provide some vindication of Hering's original claims. At 470, 515, and 570 nm, the appropriate unique hue names were used, almost to the exclusion of the other color names (e.g. the subject reported 98% yellow at 570 nm); and the color name red predominated beyond 625 nm. Moreover, the intermediate wavelengths were readily described with pairs of names of the neighboring unique hues. And the color names of the mutually exclusive hue pairs – red and green, and yellow and blue – were virtually never used to describe the same wavelength, confirming the mutual exclusivity of particular hue pairs. This experiment shows that subjects perform reliably, and that these four color names are *sufficient* to describe all of the hues; and it bolsters the claim of uniqueness vs. binariness as well as the perceptual mutual exclusivity of the members of the two mutually exclusive hue pairs.

But to what extent are the results determined just by the experimenter's original choice of color names? What would subjects do if we gave them fewer color names to use, or a different set? Charles Sternheim and Robert Boynton (1966) tested this question in the wavelength range 530-620 nm, asking subjects to use several different sets of color names on different runs of the experiment. Subjects were told that their response categories need not add up to 100% – if the available color names were insufficient to describe the perceived hues, the subject could just leave out some of the percentage points. These leftover percentage points were combined to generate a *computed function* (CF) that is plotted with each data set.

Some of Sternheim and Boynton's data are shown in Figure 3.9. As we would expect by now, the three color names green, yellow, and red were sufficient for the color-naming task across this wavelength range (Figure 3.9A) – virtually no percentage points went unused. But the two color names green and red were not sufficient (Figure 3.9B), nor were the three color names green, orange and red (Figure 3.9C). In both cases, the computed function CF closely resembled the curve generated by the term yellow when it was allowed (Figure 3.9C). Finally, all four color names – green, yellow, orange and red – were also sufficient to describe the hues in this wavelength range (Figure 3.9D).

The implication of Sternheim and Boynton's experiment is that the color name yellow is necessary for describing the colors in this wavelength range, whereas the color name orange is not necessary. Orange can be described satisfactorily as a reddish yellow, but yellow cannot be described as an orangish green. The results thus support the idea that yellow is a unique hue, whereas orange is a binary hue. Similar experiments have been done in other regions of the spectrum and along the purple line. In each case, the data support the uniqueness of Hering's four unique hues, and the binariness of the hues in between. These data thus provide a modern, quantitative vindication of Hering's original observations.

3.8 Physiological implications of color appearance

Again, bumblebees can fly. As was the case with photometry, the system properties associated with color appearance led color theorists to propose several physiological inferences and speculations.

3.8.1 The appearance of the spectrum

The fact that different wavelengths of light look different colors, in conjunction with the Universal linking proposition, implies that different wavelengths of light set up different neural codes. The fact that neighboring wavelengths look similar in color suggests that neighboring wavelengths cause similar values in the neural code for color. The perceived similarity between short and long wavelength lights (with violet having a reddish tinge) suggests that, surprisingly, the neural codes arising from short and long wavelength lights have some internally generated feature in common. The existence of the extra-spectral purples suggests that mixtures of lights set up novel neural codes that are not set up by any individual wavelength. And the circularity of the perceptual hue circle suggests that the values of the neural code form a continuous variation from short to long wavelengths, through the purples, and back to the code for short wavelengths again.

The existence of the three-dimensional color solid implies that three dimensions of variation are sufficient to encode all of the perceived colors (under a given set of conditions). These and other data have suggested strongly to vision scientists that the neural color code is confined to three variables; for example, that we will find three and only three classes of neurons that carry the perceived color code at any given level of processing. We return to this theme in Chapter 7.

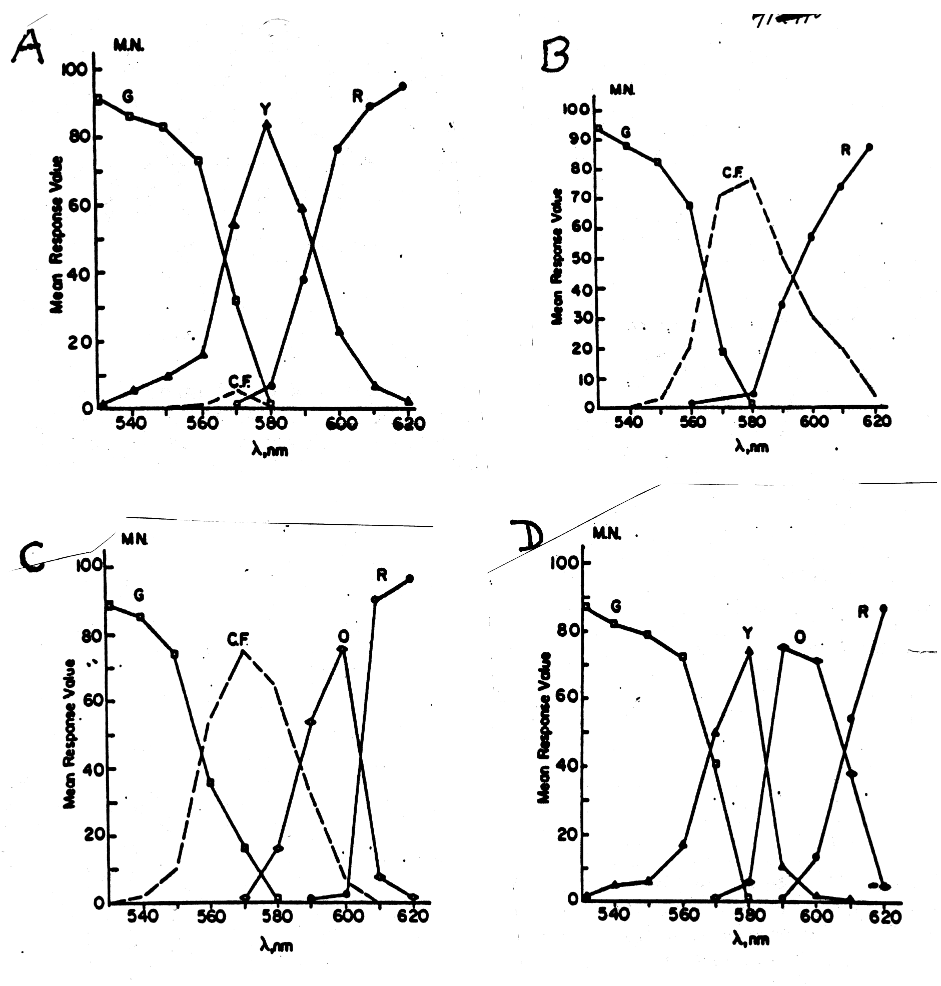


Figure 3.9: Color naming with four different sets of color names. Wavelengths over the range 530-620 nm were used. Subjects were allowed to use four different combinations of color names. The subjects were told that if there was a color name they needed, but were not allowed to use, they should assign less than the sum of 100 points to that wavelength. When points were left out, the authors calculated and plotted the missing points, labeled *C.F.* for *computed function.* The color names allowed were: A: Green, yellow, and red; B: Green and red; C. Green, orange, and red; D. Green, orange, yellow and red. The computed functions in B and C resemble the use of the color term yellow in A. The authors argue that the color term yellow is necessary to describe the colors in this wavelength region, whereas the color term orange is not necessary. [Adapted from Sternheim and Boynton (1966), pp. 772-773.]

3.8.2 A neo-Heringian opponent process theory

What about unique vs. binary hues, and mutually exclusive hue pairs? Ewald Hering not only called attention to thesesystem properties, but also proposed a theoretical account of why they should occur. His suggestion was that we could use the perceptually mutually exclusive properties to infer the existence of physiologically mutually exclusive processes. In fact, he argued that the same physiological process – a so-called opponent process, which changes in two opposite ways from a neutral state – provides a neural code for a pair of opponent colors.

Of course Hering lived before recordings had been made from single neurons, and his opponent processes could not have been framed in modern terms. But today, we can invent a neo-Heringian theory by thinking of a neural mechanism that (say) increases its output to signal redness, and decreases its output to signal greenness. Opponent process theory posits that the two perceived colors are mutually exclusive precisely because the two states of the neural mechanism are necessarily mutually exclusive: the physiological opponent process can't both increase and decrease its output at the same time. To complete the theory, we would add a second kind of opponent process to code the perception of yellow and blue; say, increasing its output to signal yellowness, and decreasing its output to signal blueness. (The polarity on each dimension is assigned arbitrarily in both cases).

This neo-Heringian theory is shown schematically in Figure 3.10. The theory posits the existence of two opponent processes, each of which can be at its resting state, or deviate from its resting state in either of two directions. The states of the two processes are plotted on the two axes. The unique hues come about when one of the opponent processes is active in a particular direction, while the other channel is at its resting level. The binary hues come about when both channels are active. For example, the unique hue yellow occurs when opponent process #1 is at its resting state and opponent process #2 deviates from its resting state in the positive direction. The non-unique hue orange occurs when both the redness/greenness process and the yellowness/blueness process deviate from their resting states in the positive direction; and so on.

3.9 New families of linking propositions: Similarity and Mutual Exclusiveness

When lights are arranged by wavelength, they go from short (say, 400 nm) to long (say, 700 nm). But when colors are arranged by perceptual similarity, they make a circle, with long and short wavelengths joined together by virtue of their perceptually reddish component hue (Figure 3.7). This system property led color theorists to suggest that the neural code for perceived color could be similar for long and short wavelength lights. The linking proposition involved in this speculation is that similar perceived colors suggest (or imply) similar neural states. This proposition is part of a new family of relational linking propositions – the Similarity family. [Spell out the Initial, Contrapositive, Converse, and Converse Contrapositive Similarity propositions].

Similarly, the neo-Heringian model uses system properties – the perceived mutual exclusiveness or mutual compatibility of particular hue pairs – to make predictions about visual physiology. In fact, it brings us to yet another family of relational linking propositions, which we can call the Mutual Exclusivity family. This family is shown in Figure 3.11.

Since the original observations were psychophysical, Hering's reasoning had to be from psychophysics to physiology, and had to start from the Contrapositive and Converse propositions. Hering's argument was that mutually exclusive perceptual states imply mutually exclusive physi-

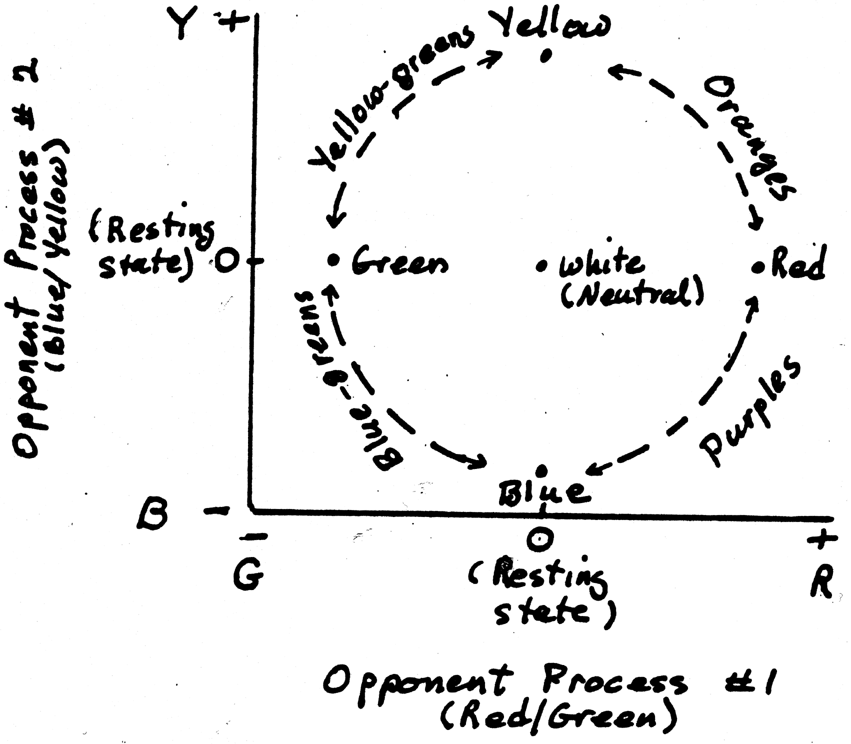
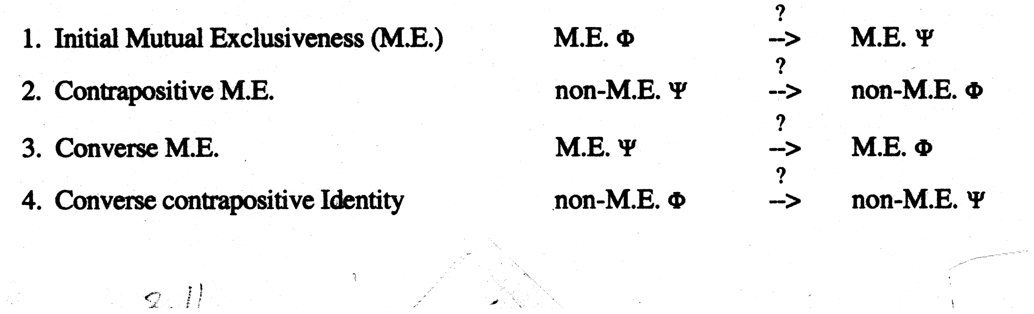


Figure 3.10: A neo-Heringian opponent process theory. According to this theory, the unique hues come about when one of the two putative physiological opponent processes deviates from its resting state in a particular direction, and the other is at its resting state. The binary hues come about when both channels deviate from their resting states. The perceptual mutual exclusivity of the mutually exclusive hue pairs comes about because of the literal mutual exclusivity of the signals: a single physiological process cannot both increase and decrease from its resting state at the same time.



1. Initial Mutual Exclusiveness (M.E.)	**M.E. Φ**	$\overset{?}{\text{-->}}$	**M.E. Ψ**
2. Contrapositive M.E.	**non-M.E. Ψ**	$\overset{?}{\text{-->}}$	**non-M.E. Φ**
3. Converse M.E.	**M.E. Ψ**	$\overset{?}{\text{-->}}$	**M.E. Φ**
4. Converse contrapositive Identity	**non-M.E. Φ**	$\overset{?}{\text{-->}}$	**non-M.E. Ψ**

2.11

Figure 3.11: The Mutual Exclusiveness family of linking propositions. Just as the Identity family, the Mutual Exclusiveness family has four members: the Initial proposition; its Contrapositive; its Converse; and its Converse Contrapositive. The Initial proposition is, Mutually exclusive physiological states imply mutually exclusive perceptual states. Its Contrapositive is, non-mutually-exclusive (mutually compatible) perceptual states imply non-mutually-exclusive (mutually compatible) physiological states. The Converse is, mutually exclusive perceptual states imply mutually exclusive physiological states; and the Converse Contrapositive is, mutually compatible physiological states imply mutually compatible perceptual states. The Contrapositive and the Converse are used in reasoning from psychophysical data to neurophysiological conclusions. The question marks indicate that the "inferences" may be speculations rather than logical necessities. [Adapted from Teller (1984).]

ological states – hence the axes with their mutually exclusive ends in Figure 3.10. And mutually compatible perceptual states imply mutually compatible physiological states – hence the off-axis regions and the binary hues. But of course, logically the Contrapositive implies the Initial proposition, and the Converse implies the Converse Contrapositive, so an opponent process theorist must be prepared to endorse the whole Mutual Exclusiveness family of linking propositions.

3.10 Summary: Class B experiments and color vision

In this chapter we have explored Class B experiments. In the course of the chapter, we addressed three goals: to develop two examples of the use of Class B experiments in color vision; to examine the inferences and speculations about neural coding that arise from them; and to ferret out some novel linking propositions implicit in these arguments.

The first set of Class B experiments centered on photometry. Photometry arose from the need for a set of quasi-physical units for the intensity of light. In searching for an additive system, physicists and psychophysicists tried several tasks, including brightness matching, as well as the minimization of perceived flicker, border distinctness, or motion. Of these, brightness matching has the greatest face validity, but the task is difficult and the measured values are not additive across the spectrum. The three minimization techniques yield more consistent and less variable data, and the values obtained closely obey additivity.

The second set of experiments dealt with color appearance, including the perceived hues of lights of different wavelengths. The psychophysics of color appearance reveals some properties not present in the physical nature of light. For example, ordered by similarity, the hues of the perceptual spectrum converge at the spectral extremes, and perceived hues form a circle when the extraspectral purples are included. We also examined claims that the color circle has an additional finer structure, in that certain hues (red, yellow, green and blue) are perceptually unique, whereas others are perceptually binary; and that the unique hues come in mutually exclusive pairs.

In terms of visual theory, we argued that vision scientists have tried to explain system properties in terms of inferences or speculations about the properties of the underlying neural codes. We examined several such arguments, and tried to develop an intuitive feel for the strength of the logic involved, from tight inferences down to cruder speculations. In regard to photometry, perhaps the most interesting argument is that the internally generated flicker, border distinctness, and motion minima that allow photometry to succeed, point to the existence of corresponding internally generated minima in the neural codes for flicker, border distinctness, and motion. In regard to color appearance, the most interesting argument is that the existence of unique and binary hues implies an opponent hue code at the neural level.

And what of linking propositions? In the case of photometry, the most novel linking proposition – an analogy – is that psychophysical minima suggest corresponding neural minima. In the color appearance cases, the most novel – a relational proposition – is that mutually exclusive hues imply the existence of mutually exclusive neural states.

Again, how do the system properties of vision come about? Why do we see as we do? What is it about the physiology of the visual system that creates the observed variations of perceived brightness with wavelength? Why are minimization matches additive, while brightness matches are not? Why are some colors unique and others binary, and some pairs mutually exclusive? All of these questions have been food for speculation and theory. We will return to them all more concretely in later chapters. For now, they wait on our list of system properties in search of explanations.

In the next two chapters we finally enter the visual system, and turn our attention to the optics of the eye.

Further Readings

For a technical treatment of the units of light and photopic spectral luminosity functions, see Pokorny, J. and V.C. Smith, 1986, Colorimetry and Color Discrimination. In K.R. Boff, L. Kaufman, and J.P. Thomas, Eds, Handbook of Perception and Human Performance. Vol. 1: Sensory Processes and Perception.

Chapter 4

Optics and the Eye

In Chapter 4, we leave the province of psychophysics and enter the province of optics, one of the classical branches of physics. Our topic also includes physiological optics – the optics of eyes built by biological systems. Students with backgrounds in physics and biology, of course, will be more comfortable with these topics than they were with psychophysics, whereas students with backgrounds in perception will be less so. Hard core neuroscientists will have to wait their turn!

In Chapter 1 we raised the question of spatial resolution: What limits grating acuity? We laid out four possibilities: the optics of the eye; the photoreceptor matrix; the spatial convergence of signals within the retina; and other factors at higher levels of the visual system. In the present chapter we return to the first of these options. The goal of the present chapter is to fill in the background we need, in order to evaluate the possibility that the optics of the eye are the major factor that limits grating acuity.

To begin, we first define the units used to specify the spatial frequency of square wave gratings, both in physical and in quasi-physical terms. We then introduce the basic properties of light, including its dual nature as both waves and quanta. We outline four ways that light interacts with matter – reflection, refraction, diffraction, and absorption. We expand on the property of refraction in order to explain how lenses make images, and on the property of diffraction because of the remarkable role that interference fringes produced by diffraction patterns play in defining optical quality (discussed in Chapter 5).

We then turn to physiological optics, and examine the human eye as an optical system. The optical elements of the eye – the cornea, pupil and lens – form an image of the physical world inside the eyeball, at the back, on the sheet of neural tissue called the retina. We spell out the consequences for vision of optical errors within the eye. We describe the possible sources of information loss within the human optical system, and outline how to specify its quality. Finally, we introduce adaptive optics, a technique with which, remarkably, it is possible to improve the quality of the retinal image within the living human eye.

In sum, the eye is a remarkable physical and physiological structure. It captures a bundle of rays of light coming from the three-dimensional world, and focuses the rays to make the retinal image. The retina in turn provides the initial encoding of the information contained in the retinal image, that eventually allows us to judge the shapes, colors, motions, and distances of physical objects. But how much do the eyes optics affect the information available in the retinal image? The background provided in this chapter will allow us to attack this question again at greater depth in Chapter 5, in which psychophysical, optical, and neural themes combine.

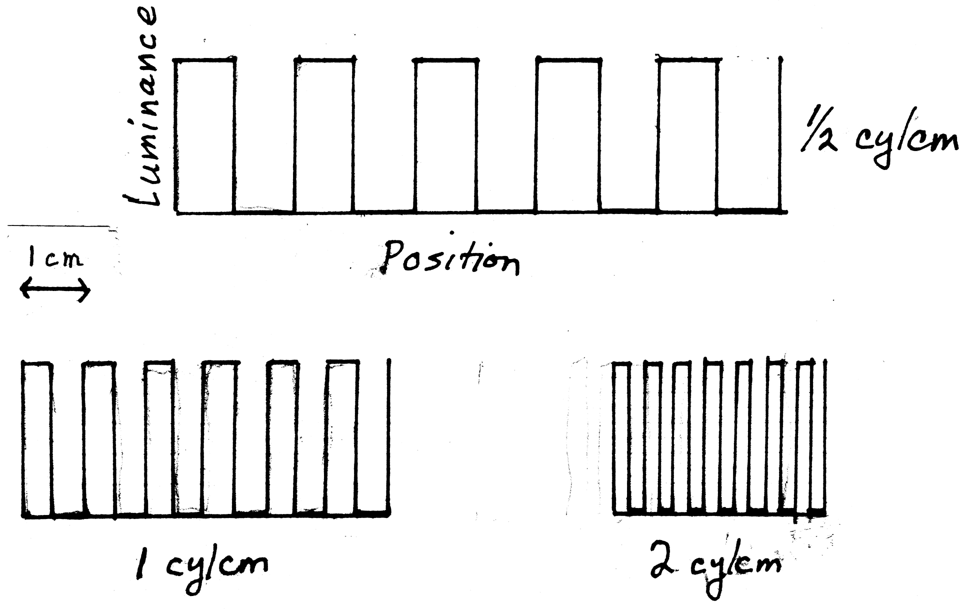


Figure 4.1: Physical specification of a square wave grating. The sketches show variations in luminance across spatial position. One cycle of a grating consists of one high-luminance and one low-luminance region. The spatial frequency of each grating is specified by the number of cycles per centimeter (cy/cm). Not to scale.

4.1 Square wave gratings

We begin with a digression on the question of units. In Chapter 1, we specified the acuity gratings in Figure 1.2 only by letters: A, B, C, and so forth. Obviously we need to adopt more formal units. Vision scientists specify gratings in two different kinds of units – one physical, and the other quasi-physical.

4.1.1 Physical specification: Spatial frequency in cycles per centimeter

Figure 4.1 shows luminance profiles of three square wave gratings, corresponding to three of the seven gratings from Figure 1.2. Each white stripe of the grating gives a relatively high luminance, and each black stripe gives a relatively low luminance. The name *square wave grating* comes about because the transitions between black and white are abrupt, or "square".

Physically, these gratings alternate between black and white stripes at regular intervals across space. Such cyclical patterns can be specified in terms of the number of *cycles* per unit distance. By this convention, one black and one white stripe of the grating constitute a cycle. The number of cycles per unit distance is called the *spatial frequency* of the grating (we will refine this definition later). Thus, the gratings in Figures 1.2 and 4.1 can be specified in terms of *cycles per centimeter*[1]

[1] Specifying the grating in cycles per unit distance yields units that are counterintuitive for some people, since

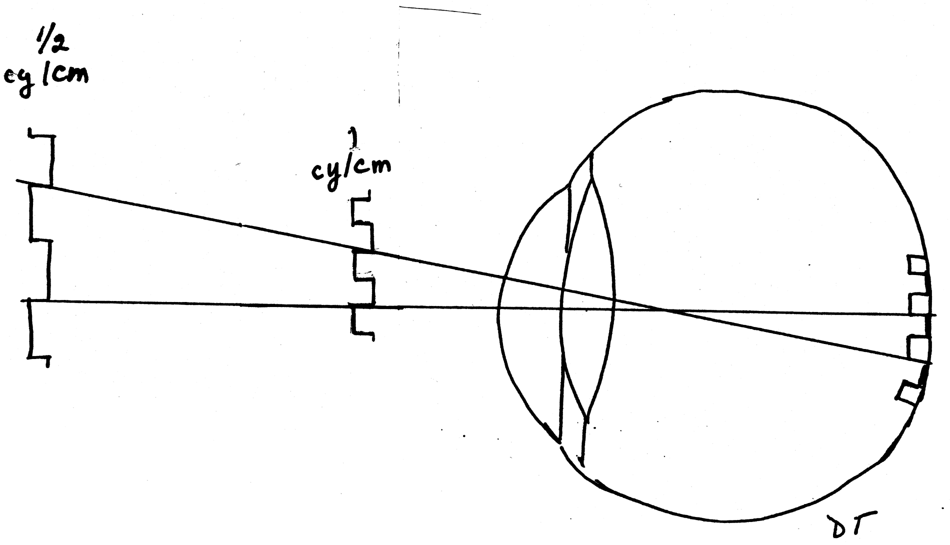


Figure 4.2: Quasi-physical specification of a square wave grating. The spatial frequency of the grating is specified in units of the angle subtended by one cycle at the eye. The two gratings pictured differ in spatial frequency, specified in cy/cm. However, the 1/2 cy/cm grating is twice as far from the eye as the 1 cy/cm grating, such that a single cycle of each grating occupies the same angular size at the eye. At their respective distances, these two gratings produce nearly identical retinal images of the same spatial frequency when specified in cycles per degree.

on the page.

4.1.2 Quasi-physical specification: Spatial frequency in cycles per degree

But second, as we saw in the case of luminance, vision scientists often develop quasi-physical units; that is, special units that specify the properties of the physical stimulus in terms of its probable effectiveness for human vision. Figure 4.2 shows another schematic view of a human eyeball. As stated previously, the eye forms an optical image of the physical objects in the visual field. For an object of a fixed size at a fixed distance, we can estimate the size of the image to a first approximation by drawing lines from the edges of the object, crossing (as it turns out) just behind the lens, and diverging again to hit the retina. For a fixed distance, the larger the physical object, the larger the retinal image. Similarly, the higher the spatial frequency of a physical grating, the higher the spatial frequency in its retinal image.

Things become more complicated when we vary the distance of the object. For an object of a fixed size, if we double the distance we will cut the image size in half; or, to keep the image the same size, we will have to double the size of the object. Similarly, for a square wave grating with a fixed spatial frequency, if we double the distance we will cut the width of each stripe in half, and so double the number of stripes per unit distance. To keep the same spatial frequency in the retinal

coarser stripes are designated by smaller numbers. Just remember – the *finer* the stripes, the *more* of them will fit in a unit distance – so the *higher* the spatial frequency.

image, as we double the distance we will have to double the widths of the stripes in the grating (*decrease* its spatial frequency).

How do we derive quasi-physical units for the sizes or spatial frequencies of objects in retinal image terms? We can think of the eye as occupying the center of a 360^o (degree) circle. So, we can specify a stimulus in units of its *angular size at the eye*; that is, in terms of the *visual angle* it occupies (*subtends*). For example, if three cycles of a grating fit into a single degree of visual angle, it is a three cycles/degree (cy/deg) grating. (Other common abbreviations for cycles per degree are *c/deg* and *c/deg.*)

There are two rules of thumb that will give you a better intuitive feel for specifying stimuli in terms of visual angle. The first is wonderfully literal: your thumbnail at arm's length subtends about 1^o of visual angle. The second is that the sun and the moon, at their respective distances, each subtend about $1/2^o$. You can check that these two rules of thumb are consistent by measuring the angular size of the moon (do not try this with the sun!) with your thumbnail held at arm's length. You will find that in terms of visual angle, the moon is about half as large as your thumbnail at arm's length.

Quasi-physical units are useful in specifying spatial resolution, for both empirical and theoretical reasons. Empirically you already have evidence, from viewing Figure 1.2 at different distances, that the physical grating you could just barely resolve varied with distance. When the grating is specified in physical units, every doubling of distance requires approximately a halving of spatial frequency for resolution. But specified in quasi-physical units, spatial resolution turns out to be virtually constant across changes in the distance of the grating, allowing us to separate the parameters of spatial resolution and distance.

More theoretically, it makes sense to guess that the spatial resolution capacities of the retina will be related to size or spatial frequency in the retinal image rather than in the world. A retinal image of a fixed spatial frequency will always make the same pattern on a fixed patch of retina. This pattern will be processed by the same patch of photoreceptors, interneurons, and ganglion cells both times, and it makes sense to assume that the same retinal image will be processed the same way each time. Thus, from this point on, gratings will always be specified in terms of cy/deg.

4.1.3 How good is grating acuity?

Armed with units of measurement, we can now ask, how good is grating acuity? The answer is that under the best conditions, in the best young eyes, grating acuity is just *about 60 cy/cm.* Alternatively, since there are 60 minutes of arc in one degree, the best visual acuity can also be stated as about *1 cycle per minute of arc.* That is, the finest grating you can resolve (if you have excellent eyes) contains about 60 black and 60 white stripes across your thumbnail, at arms length. In Figure 1.2, this is the spatial frequency of grating [E at xx cm, or grating F at xx cm]. Use Figure 1.2 to recheck your visual acuity with these numbers in mind. We will come back to these numbers several times in the next few chapters. Grating acuity of 60 cy/deg joins the scotopic and photopic spectral sensitivity curves as system properties in search of explanations.

4.2 Light and optics

4.2.1 The electromagnetic spectrum and the visible spectrum

Electromagnetic energy (also called *radiant energy*) is one of the basic forms of energy in the universe. Electromagnetic energy varies in its *wavelength.* As shown in Figure 4.3, the electromagnetic spectrum encompasses many orders of magnitude of variation in wavelength.

But what is *light*? The term light is used to refer to the part of the electromagnetic spectrum to which human eyes are sensitive. As you already learned in Chapter 2 and 3, the visible spectrum covers a range of wavelengths of just less than a factor of two, from about 400 to about 700 nm. These limits are not absolute, but represent practical extremes based on the rapid fall-off of the eye's sensitivity at the ends of this range (as shown, for example, in Figure 3.5Bxx).

Notice that, interestingly, psychophysical measurements such as those involved in defining the spectral sensitivity curves you saw in Chapter 2 and 3 even enter into the fundamental definition of the concept of light. In fact, the distinction between electromagnetic energy and light provides the quintessential example of quasi-physical specification. If it's a wavelength humans can see, it's light; if not, it isn't.

The other immediately striking thing about light is that at photopic levels, the different wavelengths of light take on characteristic colors, as we saw in the color naming experiment in Chapter 3. This fact is so universally appreciated that expectations about the perceived colors of lights of different wavelengths are often included in diagrams like that in Figure 4.3A. Vision scientists, however, usually avoid this practice, in order to avoid conflating physical with perceptual entities. Instead, we reserve one set of terms (say, wavelength) to describe the physical characteristics of the stimulus, and another set (say, color names) to describe the perceptual characteristics of the stimulus.

There are two important reasons for making such terminological distinctions. First, if we are initially clear in separating physical and perceptual realms, we are set up to ask how one realm maps to the other, without any initial presumptions. And second, the mapping between physical and perceptual realms is often complex, and (as we will see) many factors other than wavelength influence perceived colors. Figure 4.3B is separated from Figure 4.3A in order to make a clear distinction between perceived colors and the wavelength of light.

In terms of design, why is the visible range restricted to between 400 and 700 nm? Part of the answer is that it makes evolutionary sense to match the visual system to the wavelengths that are available at the earths surface and are therefore available to our eyes in our natural environment. Electromagnetic radiation at the long wavelength end of the visible spectrum, toward the infrared, gets increasingly absorbed in the earth's atmosphere; and electromagnetic radiation toward the short end, in the ultraviolet, gets increasingly scattered by the atmosphere[2]. The remaining radiation is the radiation available to our eyes for seeing.

A second part of the answer is that vision in either the ultraviolet or the infrared has practical disadvantages. Ultraviolet light can destroy biological structures in the eye – remember what it can do to the skin! Ultraviolet radiation also encourages the yellowing of the lens and the formation of cataracts; and there is some evidence that it can even damage the short-wavelength-sensitive

[2] On this view it is not surprising that different animals have different spectral ranges for vision, depending on their environments. Different species of fishes, for example, tend to match their spectral sensitivity curves to the wavelength composition of the light that penetrates water to the particular depth at which they live.

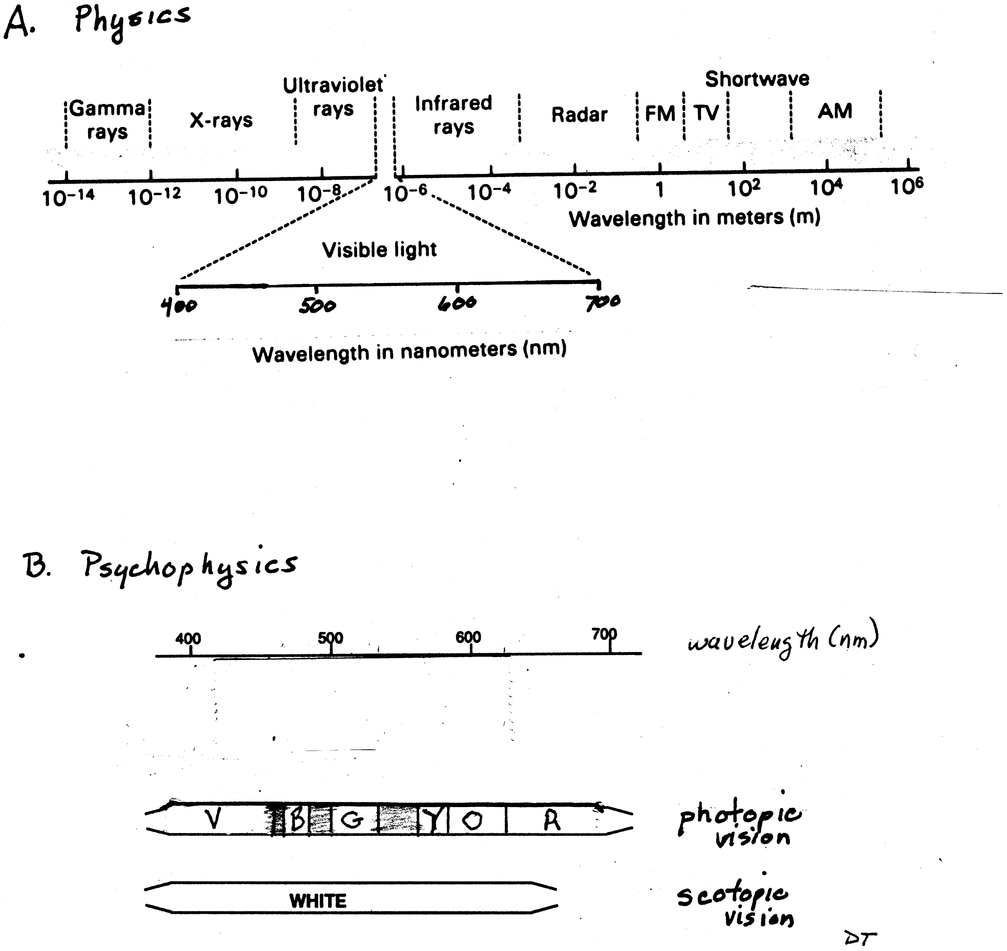


Figure 4.3: The electromagnetic spectrum and the visible spectrum. A: The electromagnetic spectrum, with wavelength specified in meters. The visible portion of the electromagnetic spectrum, which we call light, occupies only a narrow range of wavelengths, from about 400 to about 700 nanometers (nm; 1 nm = 10^{-9} meters). B. A typical mapping between wavelengths of light and perceived colors, based on psychophysical studies. In photopic vision, different wavelength ranges are typically perceived as different colors (V = violets, B = blues, G = greens, Y = yellows, O = oranges, R = reds; shaded areas show intermediate colors). In scotopic vision, all wavelengths of light look whitish, and no colors are seen. [A. Modified from Levine and Shefner (1991), Fig. 4.1, p. 66. B. DT]

photoreceptors – the cells that capture short wavelengths of light within the retina.

Infrared wavelengths, at the opposite end of the spectrum, are similar to our bodily radiations due to heat. Snakes can sense infrared radiation, to help them locate prey. But for us, seeing our own body heat within our own eyes would add noise that would tend to mask the images of objects in the real world. Detection of our own body heat would be especially deleterious at absolute threshold, where every bit of energy in the visible range counts.

In support of this latter argument, it can be shown that our absolute threshold varies as body temperature changes over the course of the day. At night, when the human body temperature is low, due to normal circadian fluctuations, absolute threshold is low too. When body temperature rises during the day, so does the absolute threshold[3].

4.2.2 Quanta vs. waves

Electromagnetic energy sometimes behaves like waves and sometimes behaves like particles, or discrete packets of energy. For a long time physicists argued about whether electromagnetic energy was ”really” waves or ”really ” particles. We now know that both the wave-like and the particle-like properties of light, or any electromagnetic radiation, can be fully and consistently described mathematically. For the non-mathematician, the conceptual problem is that there is no single entity at the level of things we can observe directly that has both kinds of properties, so its hard to imagine light as having them both. The way around this conceptual blockade is to be willing to use different analogies to elucidate different properties of light. We will do this below.

In fact, the wave-like and particle-like properties of light manifest themselves under different conditions. Light behaves like waves when traveling through air or another transparent substance (*medium*) – for example, from the sun to the earth, or from a physical object to your eye, or within the eyeball. It behaves like particles of energy when interacting with matter – for example, when it is absorbed by a physical object, or by your retinal cells to start the visual process. Both the particle-like and the wave-like properties of light are important to understanding vision, and vision scientists use the two different concepts interchangeably at different times, depending on what the light is doing.

Physicists use the term *quantum* (plural *quanta*) to refer to a particle of light when its particle-like properties are being emphasized. Quanta of light (within the visible range) are sometimes called *photons*. However, in other contexts the terms quantum and photon are used interchangeably.

4.2.3 Quantal fluctuations

When considered as particles, light has another important property for vision. It turns out that the emission of a quantum is a probabilistic occurrence. Thus, the output of any given source of light is not precisely constant in terms of quanta/sec, but varies over time. Moreover, the quantal fluctuations increase (the magnitude of the noise increases) as the radiance of the light increases.

[3] DT has a friend who was interested in the influence of body temperature on absolute thresholds. His lab was in an old house, so he happened to have a bathtub in the lab. He sat in a very hot bath while he dark adapted for an hour or so. He then wrapped himself up in a sleeping bag and measured his absolute threshold (we're not sure how he got from the bathtub to the apparatus), and showed that his absolute threshold was elevated while his body temperature was elevated. We suggested that he put the bathtub outside in the winter and see if he could lower his absolute thresholds, but he declined. There's a limit to what even he would do for science!

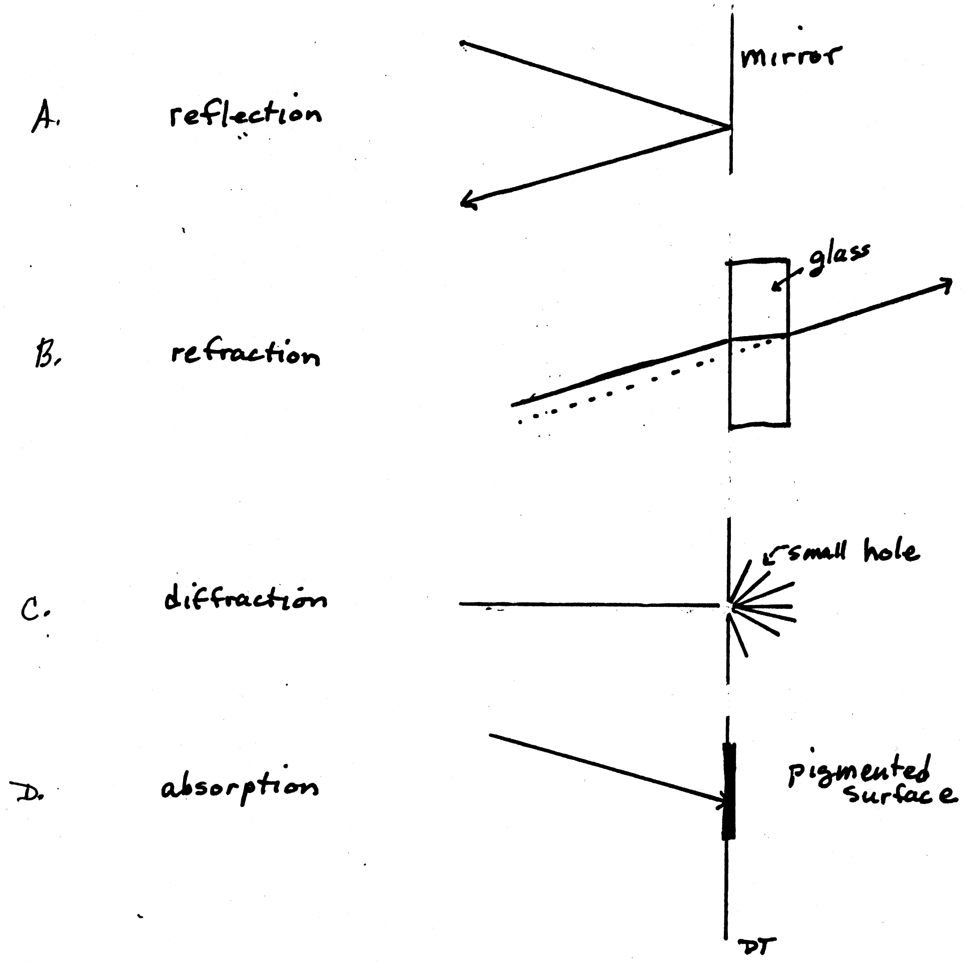


Figure 4.4: Four ways that light can interact with matter. A: reflection; B: refraction; C: diffraction, and D: absorption.

Thus, the physical variability of the light source itself is one of the major factors that limits detection thresholds in human vision. This topic is elegantly discussed in Cornsweet (1970).

In Chapter 2, in the context of signal detection theory, we introduced the idea that a threshold can be considered as a signal/noise discrimination. We can now add to that discussion by noting that quantal fluctuations are a classic example of noise. In this case, the source of the noise is external to the observer, or *extrinsic*. *Intrinsic* noise – noise generated within the observer – will also influence visual thresholds.

4.3 Optics

4.3.1 Interactions of light with matter

A beam of light is traveling along as a wave on a straight path through the universe, minding its own business, when suddenly it encounters a bit of matter. What happens? There are four possibilities, as shown in Figure 4.4. First, *reflection* occurs when light, acting briefly in its particle mode, bounces off the surface of the matter. The wave now changes its direction of travel in a precise way. A useful analogy for reflection is a billiard ball bouncing off the side of the table. The angle at which the light hits the edge of the table (the *angle of incidence*) determines the angle at which it bounces off (the *angle of reflection*). In fact, all things being equal (e.g., no spin on the ball) the angle of reflection will be exactly equal to the angle of incidence.

The second possibility is *refraction.* Refraction occurs when light enters (but is not absorbed by) a new medium – for example, in passing from air to glass. If the medium is more dense (has a higher *index of refraction*) the wave is slowed down. As a result, it changes its direction of travel. A useful analogy here is a heavy vehicle going from *a*sphalt to *g*ravel at an angle (the *a*sphalt is the *a*ir, and the *g*ravel is the *g*lass). As the first wheel (say the right front wheel) hits the gravel, it is slowed down, and the vehicle tends to turn toward a line normal (perpendicular) to the boundary, changing its direction of travel. As it goes from gravel to asphalt again, the right front wheel hits the asphalt first, and speeds up again, and the vehicle turns the opposite way, once more changing its direction of travel. Any skier has experienced a similar phenomenon when going from ice to snow or vice versa. Note that by manipulating the boundaries between various transparent materials, we can manipulate the direction of travel of a beam of light. [What would happen in Figure 4.4B (refraction) if the pane of glass were triangular in cross-section?]

The third possibility is *diffraction.* Diffraction occurs when a ray of light passes very close to the edge of a piece of matter. The ray is bent in proportion to how close it is to the edge – the closer, the more bent. Think of water in a fast-moving stream as it courses around a rock. The bits of the stream that are close to the rock bend around it, while those sufficiently far away are not affected. Analogously, when light passes through a very small hole, it is bent outward in all directions at the edge of the hole, and it will make a blurry spot (not a sharp one) on a piece of paper placed on the far side of the hole. As the hole gets larger, only the rays very near the edge are bent, so the light will make a concentrated spot with only a slightly blurry edge. Another reasonable analogy is an adjustable hose nozzle. With a small opening youll get a spray, but with a larger one you'll get a stream.

The final possibility is *absorption.* Acting as particles, individual quanta of light are absorbed by the individual molecules that make up the absorbing substance. They then cease to exist as electromagnetic energy, and become part of the energy state of the molecules that absorb them.

4.3.2 Lenses and image formation

The formation of an optical image by a lens depends upon the property of refraction. As shown in Figure 4.5A, light rays leaving a point on an object (or a point source of light) diverge from that point in straight lines in all directions. Suppose that a cone-shaped group of those rays encounters a glass lens. As the light passes from the air to the lens, it bends; and the greater the angle at which it strikes the air-glass interface, the more it will bend.

If we design the lens cleverly, with a surface that varies in curvature, we can bend each ray by a different amount; say, so that all of the rays are parallel to each other within the lens. If the second surface of the lens is equally cleverly shaped, we can bend each ray again, say just enough so that all of the original rays will converge to a single point on the far side of the lens. Rays from neighboring points on the object will converge at neighboring points in the image, and voila! – an optical image of the object. As shown in Figure 4.5B, the farther the source is from the lens, the closer to the lens the image will be.

The power of a lens is defined by its *focal length.* Rays from one point on a very distant object[4] (say, a point on the surface of the sun) arrive at the lens virtually parallel to each other. This case is illustrated in Figure 4.5C. When these parallel rays pass through the lens, they will converge at a point on the far side of the lens. The distance from the lens at which the parallel rays converge is called the focal length, f, of the lens. The shorter the focal length, the greater the *power* of the lens. We express power in *diopters*, which are units of one over the focal length (1/f), where f is in meters. So, if f = 1 meter, the power of the lens is 1 diopter. If f = 1/2 meter, the power is 2 diopters, and so on.

The lenses shown in Figure 4.5A-C are all convex, or *positive*, lenses – they *converge* the incoming rays of light and form an image. Concave, or *negative*, lenses, on the other hand, *diverge* the light and do not form images. The more the divergence, the higher the power of the lens. A negative lens is shown in Figure 4.5D. Both positive and negative lenses are used in fitting glasses to correct focusing errors of the eye, as will be discussed below.

4.3.3 Interference

The phenomenon of *interference* is illustrated in Figure 4.6. Interference patterns are a manifestation of the property of diffraction. If beams of light from the same source pass through two small neighboring slits in, say, a metal plate, each of the slits will diffract the light. The two diffracted beams will spread out, and can overlap beyond the slits. If the two overlapping beams then fall on a screen, they will form a set of fuzzy light and dark stripes called *interference fringes.* The analogy here is to the overlapping ripple patterns produced when you drop two stones into a pool of water[5] – each set of waves produces high and low points, and where they overlap, the highs add to produce super highs and the lows add to produce super lows.

4.4 Physiological optics: The eye as an optical system

How do the properties of light and optics manifest themselves in the human eye? All of the interactions that light can have with matter are of importance to vision. Reflection is important in that some of the light reaching the front surface of the eye is reflected, never enters the eye, and therefore cannot contribute to vision. Refraction is important because the eye contains an optical

[4] The term *optical infinity* is used to refer to a distance beyond which further variations in distance have only negligible effects. For most purposes, objects more than 30 feet or so away are considered to be at optical infinity, and the rays from a point on an object at 30 feet or more are considered functionally parallel.

[5] Heres a puzzle that illustrates the paradoxical nature of light. Suppose that you set up a double slit experiment. You shine a light from a laser onto the two slits, but you make the light so dim that on average, only one quantum per day will reach the two slits. Since a quantum is indivisible, one might think it would have to go through only one of the slits; and since interference is a property of the two beams together, one might think that no interference fringes could be made. Now you put a sensitive photographic film where the screen was, and go away for a year. The question is, when you come back and develop the film, will you see interference fringes? The answer is yes.

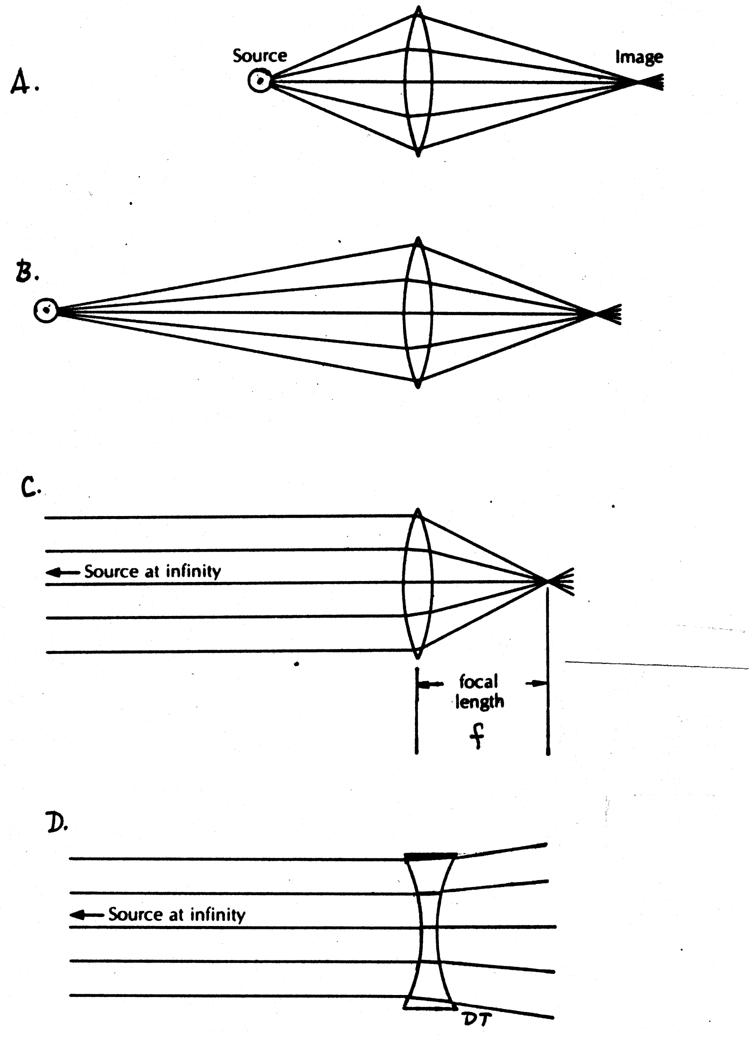


Figure 4.5: Lenses and image formation. A: A convex, or positive lens. forming an image of a point source. B. The farther the source is from the lens, the closer to the lens the image will be. C. When the source is at optical infinity, the rays from the source are parallel. The image is formed at a distance *f* behind the lens. The distance f is the focal length of the lens. D. A concave, or negative, lens diverges the light. [Modified from Cornsweet (1970), Fig. 3.9, p. 37]

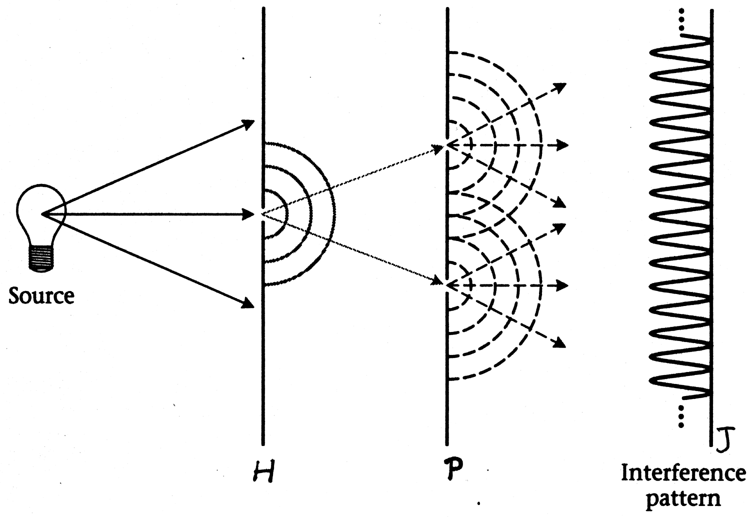


Figure 4.6: Interference. Light from a single source is diffracted by a slit in plate H, and diffracted again by a pair of slits in plate P. The interference patterns made by the two slits in P overlap, and produce an interference pattern on the screen J. [Modified from Wandell (1995), Fig. 3.8, p. 55.]

system, and forms the retinal image. Diffraction is important because the pupil of the eye is a small hole, and all of the light that reaches the retina must pass through it. When the pupil is very small, many of the rays will be diffracted and will not reach the proper point on the retina. And, of course, absorption is critical, because the absorption of light by the photoreceptors within the eye changes electromagnetic energy into the first stage of the physiological signal, as we will see in Chapter 6.

4.4.1 Major optical elements: Cornea, lens, pupil

The optics of the human eye are shown in Figure 4.7. The light entering the eye passes through the transparent window, the *cornea*, which forms the external surface of the eye. It then passes through a thin liquid called the *aqueous humor*; then through the *pupil*, a small hole in the *iris* (the colored part of the eye); then through the *lens*; and finally through a viscous material called the *vitreous humor* and on to the retinal surface, where it passes through the inner retinal layers before being absorbed by the photoreceptors (see Figure 1.4).

The cornea and lens together serve to form an image of the physical world on the retina. Every time light encounters a change of refractive index, it changes its direction of travel. Interestingly, since the largest change in refractive index occurs between the air and the cornea, most of the focusing or bending of light rays is actually done by the cornea, and not by the lens. Then the lens fine tunes the focus onto the retina. The total refractive power of the eye is about 60 to 70 diopters; of the total, about 40 diopters is due to the cornea, and 20 to 30 diopters to the lens (see below).

You can demonstrate to yourself the importance of the cornea for focusing by opening your eyes underwater. The corneas index of refraction is very close to that of water. Thus, if the light

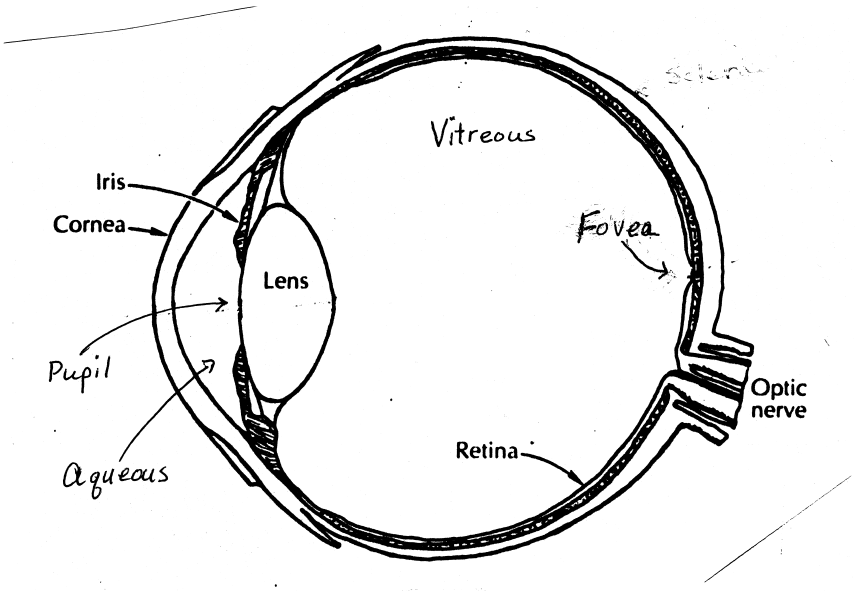


Figure 4.7: Optical elements of the eye. [Modified from Cornsweet (1970), Fig. 3.11, p. 40.]

travels from water to the cornea, it will not be bent much at all, and the cornea is essentially ineffective. You cant change focus enough with the lens to compensate, so your vision is vastly degraded. However, if you put on a pair of goggles or a dive mask, vision is restored because you have provided the air interface required by the cornea.

4.4.2 Spectral transmissivity of the eye's optics

The optical elements of the eye do not transmit all wavelengths of light equally well. The most important contributor to this effect is the lens, which absorbs light more strongly at short than at middle or long wavelengths. The differential absorption of different wavelengths by the lens is shown in Figure 4.8[6]. A second entity, the macular pigment, also absorbs at short wavelengths. So the optics of the eye act as a yellowish filter, letting through middle and long wavelengths but reducing the radiance of the light at short wavelengths. [Why might the optics be designed this way?] The density of the lens pigmentation increases with age, causing older people to lose more and more of the incoming short wavelength light.

4.4.3 Accommodation

A glass lens such as that shown in Figure 4.5A-C has a fixed focal length. However, the lens of the eye can change its focal length in order to focus objects at different distances on the retina at different times. Changes in the focal length of the lens are called *accommodation.* To demonstrate

[6] The absorption of light by a filter (or other medium, such as the lens) is specified quantitatively in terms of *optical density*, D. Optical density is defined as log10 of the ratio of incident light, I, to transmitted light, T. That is, D = log_{10} (I/T). A filter with a density of 1 transmits 1/10 of the incident light; with a density of 2, $1/100^{th}$ of the light; and so on. Optical density is plotted on the ordinate of Figure 4.8.

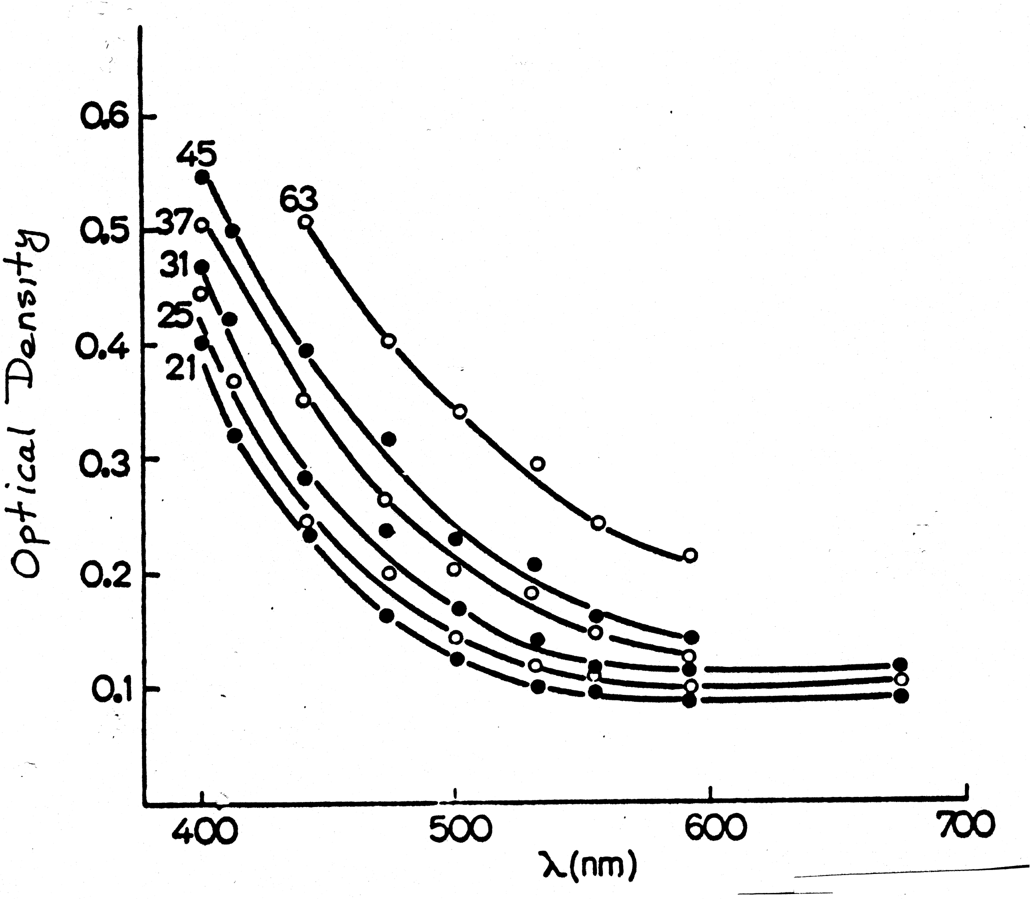


Figure 4.8: Optical density of the eye's optics as a function of wavelength and age. Notice that the density of the optics varies a great deal with wavelength. Most of the absorption of light is done by the lens. At 400 nm, the optics absorb all but 1/10th to 1/100th of the incoming light. The lens "yellows" with age, absorbing more and more of the short wavelength light. [Modified from Ruddock (1972), Fig. 3, p. 458. Original data from Said and Weale, 1959.]

accommodation, hold a finger as close to your eye as you can and still keep it in focus. Now concentrate on the finger and notice the blur of distant objects. Now reverse the process – look at the distant object and notice the blur of the finger.

Accommodation is shown schematically in Figure 4.9. These changes in focal distance are brought about by changes in the shape of the lens. For far away objects a thin, flat lens is sufficient to bring the image to focus on the retina; whereas for near objects a thicker, more curved lens will be required. When you focus at a near distance, special muscles within your eye (the *ciliary muscles*) contract to make your lens thick – increase its power. When you relax your accommodation to focus far away, these muscles relax and allow your lens to become thin again – decrease its power. A young person with normal optics has a *range of accommodation* – the range of distances that can be brought into focus by accommodation – that covers 10 or more diopters, and goes from about optical infinity to within a few cm of the nose. [How close can you bring your finger to your nose and still keep it in focus?]

4.4.4 Common optical problems and their corrections

The human eye is susceptible to a variety of focusing problems, known collectively as *refractive errors*. An eye with a normal range of focus and no other optical problems is called *emmetropic*. Common refractive errors include *myopia*, or nearsightedness, *hyperopia*, or farsightedness, and *presbyopia*, or ”old eyes”. Myopia, hyperopia, and presbyopia involve changes in the range of accommodation away from the normal range typical of emmetropia. These four types of eyes are illustrated in Figure 4.10, in terms of a set of distances important to a child (not to scale).

Myopia occurs when the whole accommodative range is moved in toward the eyeball. In consequence, close objects can still be focused readily, but distant objects cannot be brought into focus; the myopic child can't see the blackboard in class. Myopia often appears in adolescence and may be the result of long periods of focusing at near, or of continued growth of the eyeball while the eye socket (*orbit*) slows in its growth. This mismatch results in an eye that is too long for the available focusing power – the lens can't be made thin enough to focus objects at far distances. Myopia can be corrected by putting negative lenses – glasses or contact lenses – in front of the eyes. The negative lens diverges the incoming light, moving the accommodative range away from the eyes and back toward the normal range. (For example, a glasses prescription of -5.25 indicates that the myope needs a -5.25 diopter lens to move her accommodative range back to normal.)

Hyperopia, in contrast, results when the accommodative range is moved too far away from the eyeball, usually because the accommodative power of the lens is too limited. A person with hyperopia can focus far away objects fine, but cannot make his lens thick enough to focus close objects. A hyperopic child often has difficulty learning to read because she can't focus the type on the book page on her retina. Like myopia, hyperopia can be corrected with a properly chosen external lens. In this case a positive lens is needed to bring the accommodative range in closer to the eye. (For example, a prescription of +5 indicates the need for a 5 diopter positive lens to bring the hyperope's accommodative range back to normal.)

Presbyopia (”old eyes”) refers to a condition that most people encounter after about age 40. Throughout the lifespan, from childhood on, the lens continues to grow, adding layers to its structure like an onion. As the lens grows it becomes less flexible, and harder for the muscle that controls accommodation to change its shape. So starting at about age 17, the range of accommodation narrows, and the near point of accommodation gradually moves out away from the nose. [Can you

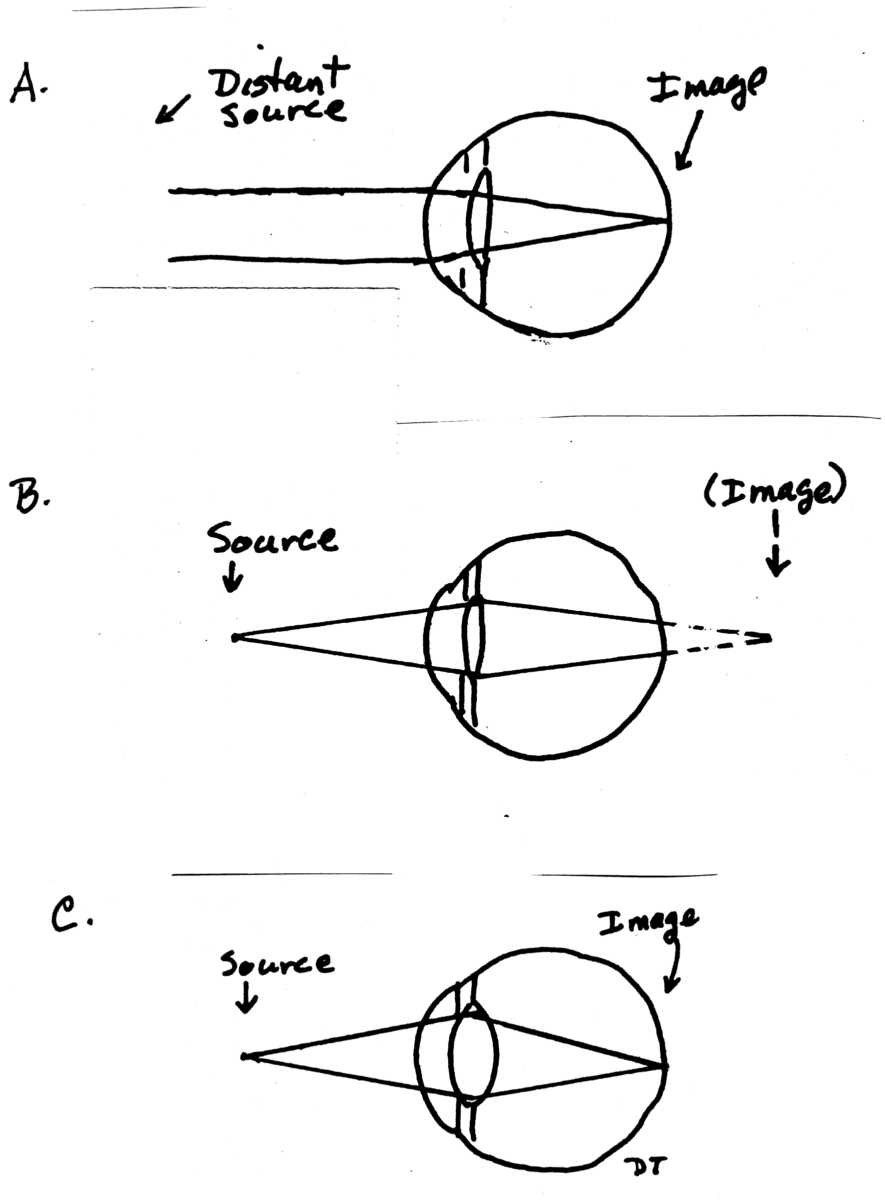


Figure 4.9: Accommodation in the normal (emmetropic) eye. A. A parallel beam of light is coming from a distant source. The lens is relaxed and thin, and the beam is focused on the retina. B. If we bring the source closer but keep the lens thin, the image will fall "behind" the retina, and will be out of focus on the retina. C. The solution: Increase the curvature of the lens surfaces. This increases the power of the lens, bringing the source back into focus on the retina.

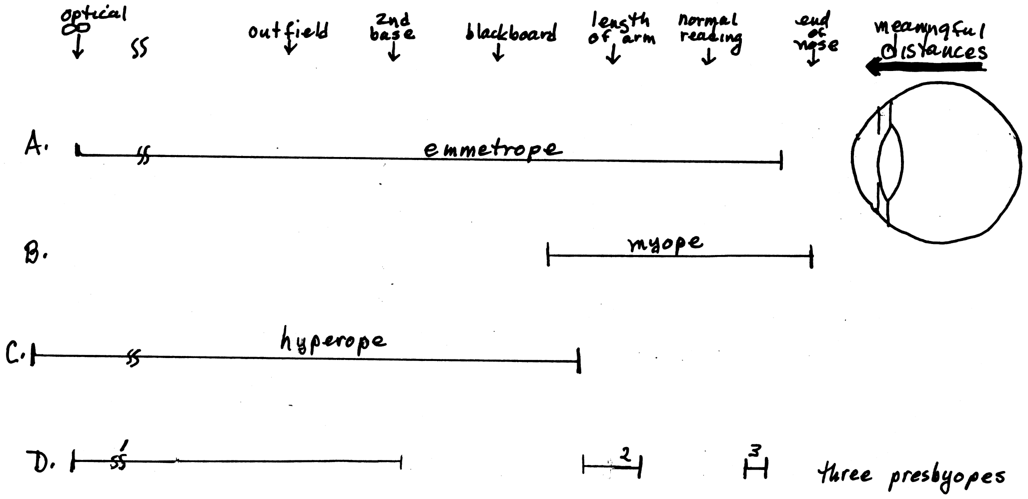


Figure 4.10: A normal eye and three common optical problems. The top row shows a child's view of the world – a set of important distances, from the end of his nose to the outfield in a baseball diamond (not to scale). The lines A-D show the accommodative ranges of eyes of various types. A. An emmetrope (person with normal eyes) can focus at all of these distances. B. A myope (nearsighted person) has a range of accommodatation restricted to near distances. C. A hyperope (farsighted person) has a range of accommodation restricted to far distances. D. A presbyope (a person with "old eyes") has only a very narrow range of accommodation. The diagram shows three different presbyopes with different residual ranges of accommodation.

still focus as close as you used to?]

As humans age we compensate for a receding near point of accommodation by holding our books farther from our noses, playing the trombone to put the book in focus. The trend finally catches up with us at about age 40, when the near point of accommodation becomes farther away than the length of our arms, and the print is too small to resolve at that distance anyway! When this happens we get bifocals or ”granny glasses” to provide artificial accommodation for reading and viewing close objects. The glasses prescription for bifocals has two parts – one for optical infinity and one for reading distance. But bifocals still give us only two distances at which objects are in focus, and are a poor and frustrating substitute for our natural accommodation. Eventually the lens becomes virtually rigid, and we are left with our accommodation frozen permanently at a single distance. The person who figures out a solution to this problem will be a millionaire!

Astigmatism is another common refractive problem. People with astigmatism have an optical system that has one power for focusing lines or gratings at one orientation (say vertical) and a different power at the opposite orientation (say horizontal). In consequence lines at one orientation will be in focus with one level of accommodation, while lines at the opposite orientation will be in focus with a different level of accommodation. The astigmat can't ever focus both orientations in the visual scene at once, and always sees images with one kind of blur or the other. Astigmatism can be corrected by fitting the patient with an astigmatic (cylindrical) lens that compensates for the differential focusing power of the eye in the two orientations. (A prescription that looks like -5.25 +1.00 x 180 describes the correction needed by a myope who is also astigmatic. The first number gives the spherical correction; the second, the additional cylindrical correction to counter the astigmatism; and the third, the angle at which the cylinder axis is to be placed.)

Still other optical problems have to do with losses of transparency of the eye's optics. *Cataracts* are opacities in the lens which can greatly interfere with vision. They become increasingly common in old age. In addition, the lens develops more and more of the yellow pigment that gives it its selective absorption of light at short wavelengths.

Relatively routine surgical procedures have been developed to remove the cataractous lens and replace it with a plastic lens. Unfortunately, no plastic zoom lens is available. However, a person who was so myopic that he wore ”coke bottles” all his life can have a myopic correction built into the implanted lens, and be left delighted with needing only light weight bifocals. Also, a person who has just had a cataractous lens removed will often marvel at how the colors of things are restored – the blues look like they used to before the aging lens began to steal the short wavelengths of light away.

4.5 Optical information loss

Imperfections in the eyes optics limit the flow of information from the physical world to the visual neurons. Imperfections arise not only from refractive errors and opacities (discussed above), but also from additional factors we have yet to consider. How much degradation is there, what are the reasons for it, and how can we quantify it?

4.5.1 Why can't retinal images be perfect?

Even when the eye is in perfect focus, retinal images are not perfect, because the optics of the eye degrade the retinal image in several more complex ways. The cornea and lens refract the incoming light to form the retinal image. Insofar as their surfaces are not perfectly shaped, they will make blur circles instead of point images, leading to degradation of the overall image. The pupil, when it is very small, can degrade the image by diffraction. The vitreous too plays a role in image quality because it is not completely clear; the vitreous can contain imperfections including floaters, pieces of cellular debris that make a wash of scattered light and degrade the contrast of the image. Finally, internal reflections from various structures within the eye scatter the incoming light and further degrade the image. In the following sections we treat some of these problems in greater detail.

The refractive system of the cornea and lens produces four kinds of complex distortions: *chromatic, monochromatic, spherical* and *higher order aberrations. Chromatic aberration* occurs because different wavelengths of light are refracted differently as they pass from one medium into another. When light passes through a traditional lens system the shorter wavelengths are bent more than the longer ones. The result is that if the long wavelengths are focused on the retina, the short wavelengths will be focused in front of the retina, and out of focus at the retina. Conversely, if the short wavelengths are focused on the retina, the long wavelengths will be aimed at a focus behind the retina, and out of focus at the retina. In short, it is impossible for the optical system of the eye to focus all wavelengths of light at the same time. In the laboratory, we often eliminate chromatic aberrations by using light of a single wavelength (*monochromatic light*).

Monochromatic aberrations occur because the surfaces of lenses arent perfect. If the curvature of a lens (or the surfaces of a lens system) do not bend each ray of light by exactly the right amount, the image will not be perfect – it will be spread out due to the inaccuracies or irregularities in the lens surfaces. In *spherical aberration*, the center of the lens has one focal length and the periphery of the lens has another. With a large pupil both are used together, making a fuzzy image. More idiosyncratic irregularities in the lens and cornea can also produce *higher order aberrations* that contribute to the imperfection of the image.

What pupil size makes the sharpest image?

The diameter of the human pupil ranges from about 1.5 mm to about 8 mm, as the person moves from bright to dim light. In general, optical quality degrades slightly from the center to the periphery of the lens. In addition, the power of the lens changes slightly from center to periphery, so optical degradation is worst overall when the pupil is largest and the whole lens contributes to the image. One might conjecture, then, that the way to optimize the image is to make the pupil small. But when the pupil is small, diffraction can make a significant contribution to imperfections in the image. How should one balance these two opposing demands? Is there an optimal pupil size?

4.6 Line-spread functions

How can the quality of the human retinal image be measured? Conceptually, one would want to form an image of a simple target – say, a point or a line – on the retina, with optimal focus. Then one would measure the diameter of the blur circle produced by the point, or the width of the image of the line, on the retina. These measurements would yield quantitative values for how far the image of the point or line was *spread* across the retina, due to optical errors of all kinds. That is, we could measure the *point spread function* or *line spread function* of the eye.

The classical measurements of the line spread function were made by Campbell and Gubisch in 1966. They used a special optical system similar to the system used in an ophthalmoscope – an instrument that allows another person to peer into your eye and examine your retina. This

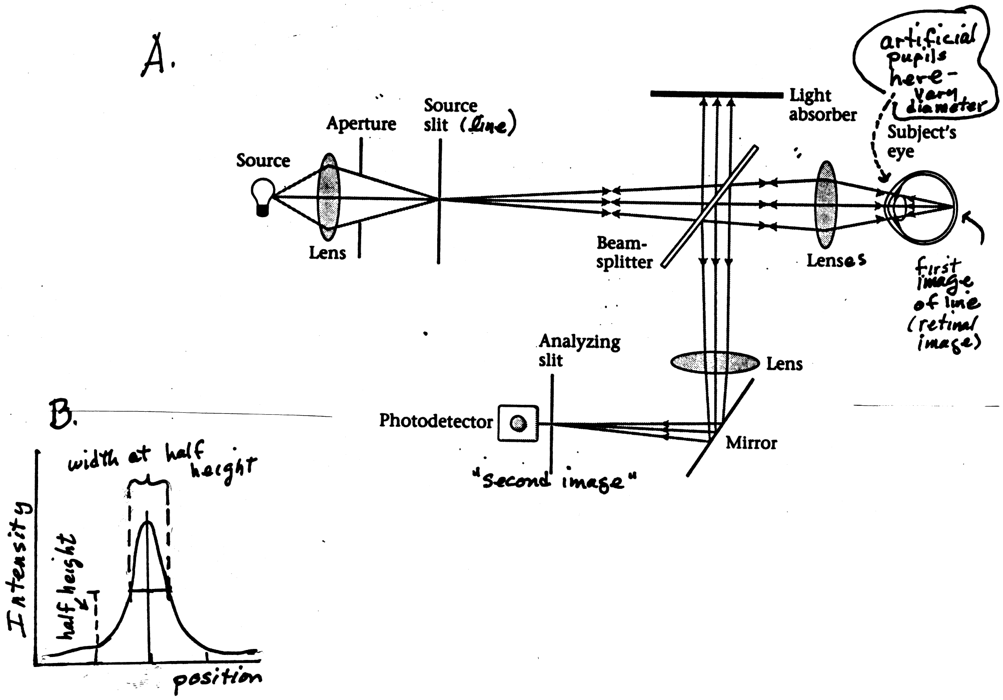


Figure 4.11: Double-pass method for measuring line spread functions. Light from the source (the light bulb at left) is imaged by a lens onto a slit. The slit makes the line that will be imaged on the retina. Light leaves the slit and passes through a beam splitter. The light reflected by the beam splitter is lost to the experiment at the light absorber at the top. The light transmitted by the beam splitter passes through another lens, enters the eye and is imaged on the retina (the *first* or *retinal image*). Some of the light from the retinal image is reflected back out of the eye. The returning light is again divided by the beam splitter, and the light transmitted is lost to the experiment. The light reflected by the beam splitter passes through another lens, is reflected by a mirror, and forms a *second image* in space. The second image is scanned by a photodetector, and the amount of light in the second image is plotted as a function of spatial position. The experimenter back calculates from the second image to estimate the line spread function in the first (retinal) image. B. The distribution of light in the retinal image can be characterized by its *width at half height.* [After Wandell (1995), Fig. 2.3, p. 16.]

approach is known as the *double pass technique*, for reasons that will be clear below. The essential elements of their apparatus are shown in Figure 4.11. (Notice that the subject makes no judgments in this experiment. The measurements are entirely physical, not psychophysical. The subject need only hold still and fixate a fixation point.)

First, Campbell and Gubisch used drugs to paralyze the subjects pupillary and accommodative systems temporarily – the pupil was dilated to its largest possible diameter, and the lens flattened to be focused at optical infinity. They set up a light source to project a line of light onto a subjects retina, to form the retinal image. The light was monochromatic in order to avoid the problem of chromatic aberration. Then they captured the light reflected back from the retina out of the eye, to make a *second image* (an image of the retinal image) in physical space outside the eye. Finally they optimized focus by placing glass lenses in front of the subjects eye until they produced the sharpest possible line spread function in the second image.

Campbell and Gubisch next scanned the second image with a photocell to quantify its luminance across space. They repeated the experiment with a series of artificial pupils – small holes in thin metal plates placed just in front of the eye – of different diameters, to simulate variations in pupil size. Of course the second image has passed through the optics twice, not once as it would for normal vision. Campbell and Gubisch overcame this problem by factoring out the double pass mathematically. The calculations produced an estimate of the line spread function in the retinal image.

Figure 4.12. shows both theoretical predictions and empirical measurements from Campbell and Gubisch's experiment. The dotted lines show the theoretical predictions based on diffraction alone. If diffraction were the only source of light dispersion in the retinal image, the line spread functions should follow these predictions. As expected, the image produced by diffraction varies with pupil diameter – the larger the pupil, the narrower the diffraction limited image. For large pupils the predictions based on diffraction have a width at half height of only about 0.2 minutes of arc – about five times narrower than the diffraction limit for small pupils.

The solid lines in Figure 4.12 show the measured line spread functions. For small pupils (1.5 to 2 mm), the data come quite close to the limit set by diffraction. The width of the distribution at half height is about 1.3 minutes of arc. Thus, for small pupils, the optical quality of the retinal image is said to be *diffraction limited*, and not further degraded by the optics of the eye. Since diffraction is an absolute limit imposed by the laws of physics, the optics of our eyes do remarkably well in matching the physiological limit to the physical limit. The measured line spread function is actually narrowest at a pupil diameter of 2 to 2.4 mm. The distribution at its best is very tight – the width at half height is only about 1 minute of arc, and the central part of the distribution still approaches the diffraction limit. So for small and intermediate pupil sizes, what we have is just about as good as it gets.

But for large pupils, the observed line spread functions are much broader than the predictions based on diffraction, with a width at half height of about 2 minutes of arc. That is, for large pupils the quality of the retinal image is about ten times worse than the diffraction limit. When the pupil is large, other aspects of optical quality – spherical and higher order aberrations, and scattered light – take over to limit the quality of the retinal image.

In short, given the properties of diffraction, the line spread function is potentially narrowest with a large pupil. For a large pupil, a more perfect optical system – one without such marked spherical and higher order aberrations – could in principle yield a retinal image with a much narrower line spread function than we in fact have. But our visual systems do not take advantage

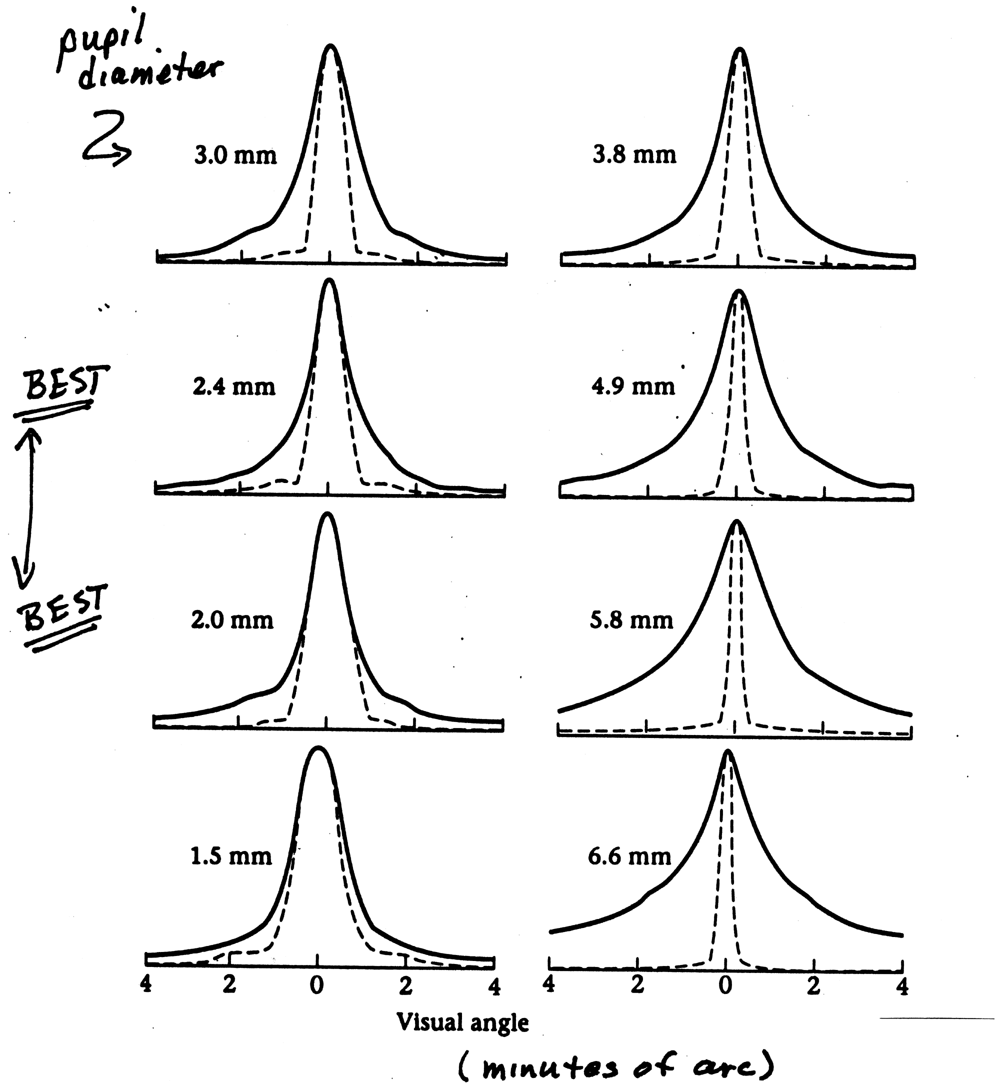


Figure 4.12: Campbell and Gubisch's results. The numbers beside each distribution show the diameter of the artificial pupil. The dotted lines show the diffraction limits for each pupil size. The larger the pupil, the narrower the distribution predicted from diffraction. The solid lines show the measured line spread functions. The measured distributions approach their respective diffraction limits for small and midsize pupils, but not for large pupils. [After Wandell (1995), Fig. 2.11, p. 29.]

of this opportunity. Viewed from this perspective, the optics of the human eye are remarkably poor. And it's interesting to wonder why better optics haven't evolved for the human eye. We will return to this question in Chapter 5.

4.7 Adaptive optics: Improving on nature

Meantime, science is improving on nature In the late 1990s an exciting new chapter was added to the story of optical quality: the use of adaptive optics. An adaptive optical system is one in which measurements of the optical quality of an individual eye can be made, and then fed back to correct the path of each ray in the incoming bundle of light rays, in such a way as to improve the quality of the retinal image for that particular eye. The system consists of two parts: a *wave front sensor* that allows measurement of the optical aberrations of a given eye, and a *deformable mirror* that allows correction of these aberrations.

In 1997, Liang and Williams (Liang and Williams, 1997; Liang, Williams, and Miller, 1997) built the first successful adaptive optical system for use with the human eye. Simplified optical diagrams of the system are shown in Figure 4.13 and 4.14. The first part of the adaptive optical system, the wave front sensor, is shown in Figure 4.13A. First, in a double-pass optical design similar to that used by Campbell and Gubisch, light from a laser (with its radiance carefully controlled so as not to do damage) is shined into the eye, and forms a tiny spot of light on the retina. Light from this spot is reflected back, and a cone of light – a wave front – emerges from the pupil. The trick is that different subparts of the wave front have passed through different parts of the eye's optics – different locations within the pupil, corresponding to different parts of the cornea and lens. By analyzing this beam of light region by region, we can evaluate the eye's optics region by region as well.

The light emerging from the eye eventually falls on a tightly packed array of 217 tiny lenses (lenslets). Each lenslet makes an individual image of the retinal spot, so the output of the device is an array of 217 tiny dots of light. But the different dots have been processed by different regions of the optics. As it happens, any irregularity in one region of the optics will shift the corresponding dot out of its place in the matrix of dots. The matrix of dots falls on a charge coupled device (CCD), which creates a record of the location of each dot in the matrix and passes it on to a computer for the next step. By analyzing the spatial irregularities in the whole pattern of dots, a description of the overall irregularities in the optics can be derived.

Two dot matrices produced by the wave front sensor with a 3 mm artificial pupil are shown in Figure 4.13B. The first matrix was made by analyzing an "ideal" eye rather than a real one, and the regularity of the matrix is apparent. The second matrix was made by analyzing a real eye (subject DRW). When the pupil is small, as it is in this case, the matrix of dots remains regular – as we said earlier, the retinal image is usually diffraction limited for this pupil size, and spherical and other aberrations have little effect. In contrast, two dot matrices produced with 7.3 mm pupils are shown in Figure 4.13C. In this case, there are obvious irregularities in the matrices of dots, showing again that spherical and higher order aberrations degrade the retinal image when the pupil is large.

The complete adaptive optics system, including the deformable mirror, is shown in Figure 4.14. How does it work? A subject and his eye are aligned in the apparatus. Starting with the deformable mirror set to be flat, the apparatus is activated, and a matrix of dots is made by the lenslets. The spacing of the matrix of dots is analyzed by a computer. The computer makes an educated guess as

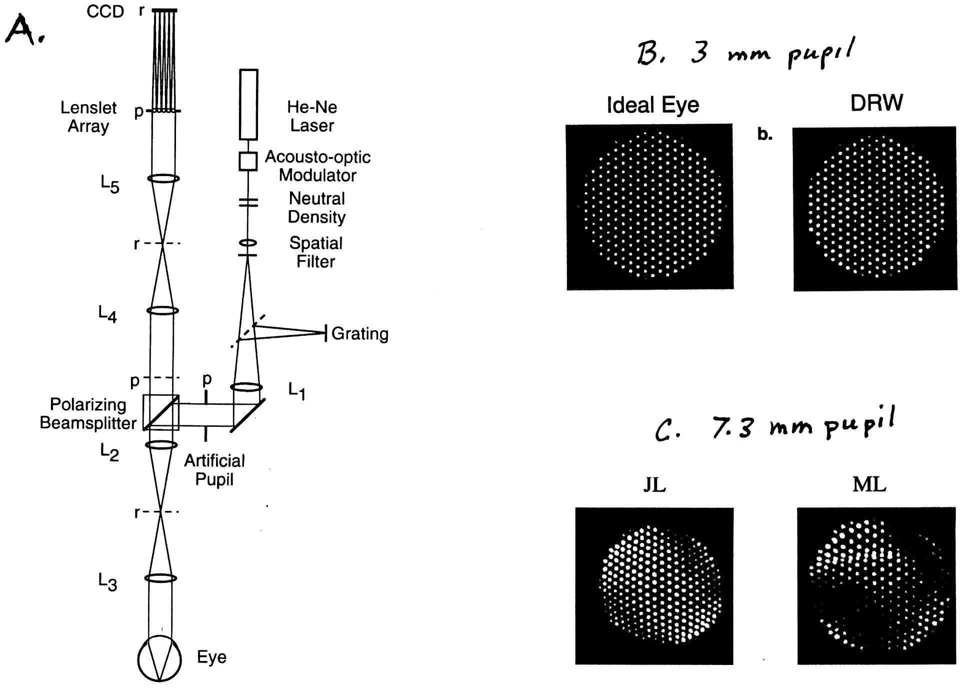


Figure 4.13: Adaptive optics: the wave front sensor. A tiny dot of light from the laser is imaged on the retina, and reflected back to the lenslet array. The lenslet array makes a matrix of tiny images of the retinal image. Irregularities in this matrix indicate aberrations in the optics of the eye. B. For a 3 mm pupil, the matrix of images is regular, both for an ideal eye (left) and for a real eye (right). C. For a 7.3 mm pupil, the matrix at the left shows a closer spacing of the dots at the edge of the array, indicating the presence of spherical aberration. The matrix at the right shows a set of irregular aberrations in the region where the eyelid normally rests against the cornea. [From Liang and Williams (1997), Figs. 1 (p. 2874), 2 (p. 2875) and 3 (p. 2876).]

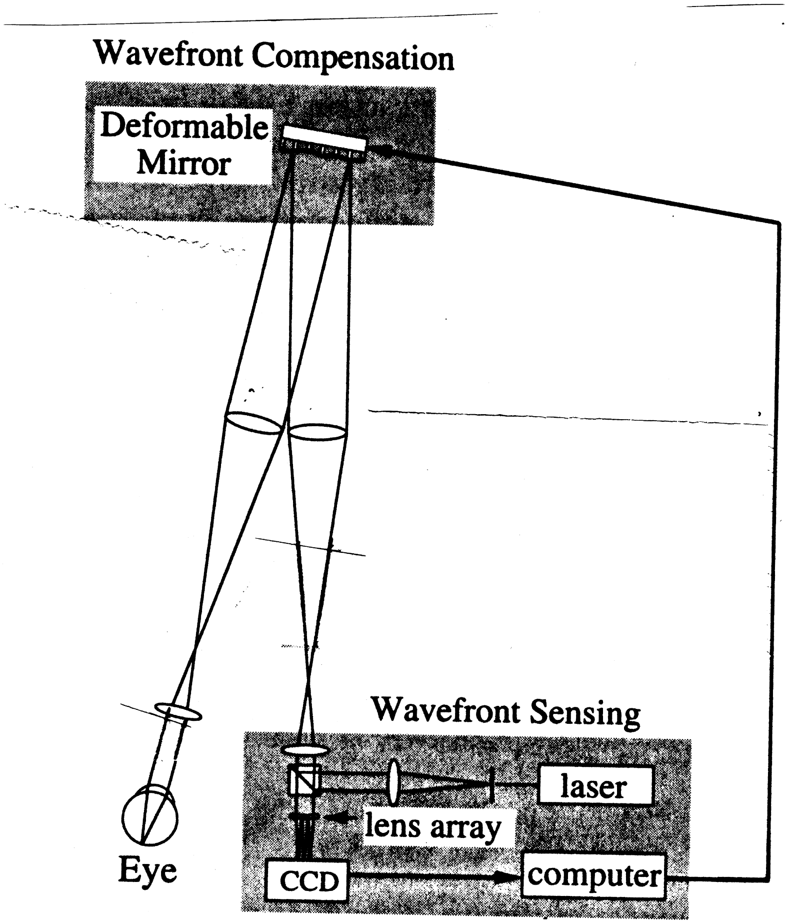


Figure 4.14: Adaptive optics: the complete optical system. In this figure the wave front sensor, including both the laser source and the array of lenslets, has been folded up into the box at the lower right. The parts of the wave front sensor closest to the eye have been spread apart by adding lenses, so that the light from the laser can be bounced off the deformable mirror on the way to and from the eye. The deformable mirror is mounted on a set of tiny pistons. The surface of the mirror can be deformed by advancing some of the pistons and retracting others, in order to compensate for the particular aberrations of the eye being studied. [Adapted from Liang, Williams, and Miller (1997), Fig. 2, p. 2885.]

to how the mirror should be deformed to make the matrix of dots more nearly regular. This guess is implemented by deforming the mirror. New measurements are then taken, and a new deformation is tried. After 10 to 20 iterations, a highly regular matrix is usually produced. The deformations required to produce the regular matrix provide a description of the aberrations introduced by the particular eye being studied.

To what accuracy can the eye be corrected with adaptive optics? As of the late 1990s, with the pupil fully dilated, line spread functions could be made about a factor of two narrower than the best line spread function Campbell and Gubisch saw with a 3 mm pupil. The ultimate goal is to use a large pupil, for which the line spread function is potentially narrowest, and to reduce the line spread function to the diffraction limit calculated for the large pupil. If that goal were achieved, gratings of frequencies much higher than 60 cy/deg could be imaged on our retinas! [But could we see them? Think about it. We return to this question in Chapter 5]

Adaptive optics have generated much excitement, because they can be applied in at least three important ways. First, some people have major optical aberrations that are not readily corrected with currently available glasses and contact lenses. Adaptive optics may eventually allow optometrists and ophthalmologists to analyze the optical aberrations in these eyes, and fit patients with specialized contact lenses designed to correct them. Second, by creating higher quality retinal images in normal eyes, adaptive optics can be useful for laboratory studies of the post-optical limits of acuity and spatial resolution.

And third, adaptive optics can also be used to look *into* the eye. Currently, when an opthalmologist or optometrist dilates your pupil and looks into your eye, she looks in through your imperfect optics. If adaptive optics could be incorporated into ophthalmic instruments, she could see your retina more clearly. Moreover, clearer picture of the retina could be taken, for both basic science and clinical purposes. We will return to these applications in later chapters.

4.8 Summary: Optical properties of the eye

We began this chapter by returning to square wave gratings, and introducing the specification of spatial frequency in both physical and quasi-physical units. We continued with a brief review of the nature of light and the properties of physical optical systems.

We then examined the properties of the human eye as a physiological optical system. The eye has two major optical components: the cornea and the lens. The cornea has about 40 diopters of optical power, and the lens a variable power between about 20 and 30 diopters. The variable power of the lens – accommodation – allows us to focus objects at different distances on the retina. Many common clinical vision problems, including myopia, hyperopia, presbyopia, and astigmatism, are due to problems in focusing, and can be corrected or ameliorated by placing lenses (glasses or contact lenses) in front of the eye.

Beyond questions of focus, the optical quality of the eye is limited by two major factors: the diameter of the pupil (which diffracts the light substantially when the pupil is small), and optical aberrations (which result in poor image quality when the pupil is large).

We then introduced the double-pass method developed by Campbell and Green for making *in vivo* measurements of the optical quality of the human eye. Campbell and Green showed that the optimal pupil diameter is in the range of 2 to 2.5 mm. At that pupil diameter the line spread function has a width at half height of about 1 minute of arc. But spreading the lines of a 60 cy/deg grating this much should lead to a major loss of contrast in the retinal image. Thus, optical quality could indeed be the major factor that limits grating acuity to about 60 cy/deg. We return to this question in the next chapter.

Finally, we introduced a more recent development in studies of the optics of the eye: adaptive optics. With adaptive optics we can measure the specific pattern of aberrations present in an individual eye *in vivo*, and use external instrumentation to shape the incoming light, in order to improve the quality of the individual's retinal image. Thus, we can potentially form images on the retina that are finer than those allowed by the eye's optics; and the question is, what will we see?

In the next chapter, we look at an alternative method of specifying the quality of an optical system: the modulation transfer function, or MTF. We also consider the effects of discrete sampling of the retinal image by the photoreceptors, and reconsider the question of whether the optics of the human eye limit our acuity.

Chapter 5

Optics and Vision

In Chapter 1 of this book, we introduced grating acuity as a fundamental psychophysical measure of spatial resolution. We argued intuitively that in terms of a causal story, spatial resolution is potentially limited at each of four different anatomical stages: the quality of the optics: the discrete sampling imposed by the photoreceptor mosaic; or higher factors in the retina and/or the brain. We then posed the fundamental locus question that provides a unifying theme for the early chapters of this book : which of these stages actually limits grating acuity?

To answer this question, we need to re-address all of the alternatives introduced intuitively in Chapter 1, but at more sophisticated levels. In Chapter 4 , we began this journey with the optics. We posed another of the most fundamental questions in vision science: how might one specify the physical quality of a lens or an optical system, and how good is our physiological optical system when such measurements are carried out?

Before we can continue, however, we need to introduce a considerable amount of background material, largely deriving from optical and electrical engineering. Toward that end, we begin the present chapter by addressing the concept of *linearity.* We then describe a new kind of visual stimulus called a *sinusoidal* or *sine wave grating*, and try to explain why vision scientists (crazy physicists?) use sinusoidal gratings as stimuli in vision experiments. We end the section with an intuitive introduction to *linear systems theory.* In the second section, we introduce a new system property called the *contrast sensitivity function*, or *CSF*. The CSF, and its cousin the *optical modulation transfer function*, or *MTF*, extend the topic of spatial resolution to include sensitivity for large as well as small features of the visual scene.

In the third section, armed with the MTF and the CSF, we return to the goal of specifying optical quality. We describe three different techniques for measuring optical MTFs in the human eye. The first and third of these techniques depend on physical measurements. However, the second technique, *interferometry* – which is arguably the most accurate of the three – depends on psychophysical rather than on physical measurements. Moreover, interferometry serves our broader goals because, remarkably, it allows us to separate the limits of visual processing combined in the ordinary CSF into optical vs. neural components. The story of interferometry is told in detail because it illustrates the rich interplay among disciplines that for DT makes up the essential charm of vision science.

In the fourth section of the chapter we turn to the second of the factors that might limit acuity: discrete spatial sampling by the photoreceptor mosaic. We will see that discrete sampling also imposes some remarkable and unexpected system properties on our vision. Armed with this information, we return with increased sophistication to our original locus question – what limits grating acuity?

We end the chapter with a summary of the code transformations imposed by the optics of the eye and the sampling properties of the photoreceptors. These concepts serve to emphasize the idea that visual processing can be seen as a series of recodings of the visual signal.

5.1 Some background: Why sinusoidal gratings?

Suppose you agree to be a subject in a vision laboratory. You walk in on the first day, and on a video screen you see a set of very fuzzy-looking stripes. Say hello! You have just been introduced to sinusoidal gratings, which are used very often as stimuli in vision science. But why? It takes a while to explain.

5.1.1 Linearity

We begin with the fundamental mathematical concept of linearity (Wandell, 1995). Most generally, linearity has to do with the way two signals combine. At the simplest level, a linear system is one that does ordinary arithmetic: it just adds and subtracts, and multiplies and divides. In consequence, when a linear system is provided with two inputs simultaneously, the system's output (response) to one input is not affected by the level of response to another input. The output to the combination of inputs is just the sum of the outputs to the two separate inputs when each is provided alone. Figure 5.1 treats the topic of linearity, and Figure 5.1A shows an example of linearity operating in an optical system.

In symbols, suppose you have a system such that when you put in a signal S you get out a response R. Further, let the signal S_1 yield the response R_1, and the signal S_2 yield the response R_2 when each is presented alone:

$$S_1 \rightarrow R_1$$

$$S_2 \rightarrow R_2$$

Then, if the system is linear, when you put in S_1 and S_2 simultaneously, you get out the sum of R_1 and R_2:

$$S_1 + S_2 \rightarrow R_1 + R_2.$$

From a vision science perspective, it is important to note that the stimuli and the responses can be highly dissimilar. Suppose that the input is a spot of light. The output can be the light distribution in the retinal image; or it can be a change in the state of some particular neuron along the visual pathway; or the perceptual report of the whole human subject. As long as the response to two stimuli is the sum of the responses to the individual stimuli, the system is linear (for the conditions tested).

An important special case of linearity arises when $S_1 = S_2$. In that case the input is $2S_1$; and if the system is linear, the output will be $2R_1$. Doubling the input doubles the output; and generalizing the argument, multiplying the input by any factor will also multiply the output by the same factor. In other words, in a linear system the relation between input magnitude and output magnitude is a straight line through the origin.

Several possible linear relationships between inputs and outputs are shown in Figure 5.1B. In a linear system, the slope of the line – the change in output per unit change in input – is constant across the whole range of inputs and outputs, and is called the *gain* of the system. The input-output relationships of several linear systems with different gains are shown.

Now that we've defined linearity, let's ask, what is a *non-linearity*? This question can't be answered in general, because there are many different ways in which linearity could fail[1]. Three examples of non-linearities are shown in Figure 5.1C-E. Panel C shows a *thresholding non-linearity*: the system does not respond to very small inputs, and the output starts to rise only after a threshold input level is reached. Panel D shows a *saturating non-linearity*: at high levels of input, the output approaches an asymptotic value. Panel E shows a *compressive non-linearity*: the output continues to grow with the input, but it grows more and more slowly at higher and higher input values. In all three cases, the consequence of the non-linearity is that the same change in input, across different ranges of input, leads to different changes in output. We have seen thresholding and saturating non-linearities in the psychometric functions of Figs. XX and XX, and we will see compressive non-linearities in the photoreceptors in Chapter 6.

5.1.2 Sinusoidal gratings

We turn next to sinusoidal gratings. A set of six vertical sinusoidal gratings is shown in Figure 5.2. Figure 5.2A shows three graphs of luminance as a function of position across the face of the video monitor. Each of these graphs traces out an undulating mathematical function called a *sine wave* or *sinusoid*. The horizontal lines through the graphs show the average luminance values. Figure 5.2B shows three simulated pictures of the video monitor, each displaying a sinusoidal grating corresponding to the one graphed just above it. Speaking non-mathematically, it is immediately obvious that sinusoidal gratings are a lot like the square-wave gratings introduced in Chapter 1, but their transitions are smoothed rather than sharp. All three of the gratings in Figure 5.2A and B have the same mean (space-average) luminance, and all have (nominally) 100% contrast, with the actual contrast of the pictures in B depending on the properties of the reproduction process used in printing the figure.

In addition to mean luminance, sinusoidal gratings have three interesting parameters: *spatial frequency, contrast*, and *phase*. As in the case of square wave gratings, the *spatial frequency* of a sinusoidal grating describes the number of repeats of the spatial variation per unit distance. Spatial frequency can be expressed as the number of cycles of the sinusoidal grating per cm on the video screen (cy/cm) – here, 0.25, 0.5 and 1.0 cy/cm – or more typically in vision science, in terms of cy/deg.

The *contrast*, or *modulation*, of a sine-wave grating describes the amount of variation of luminance around the mean luminance level. The peaks and troughs of a sine wave are symmetrical about its mean value, L_o. The minimum luminance that the troughs can take (L_{min}) is zero. Thus, by symmetry, if the mean luminance is fixed at Lo, the maximum luminance that the peaks of the sine wave can take (L_{max}) is $2L_o$. The contrast (often called Michelson contrast) of a grating is

[1] DT, with her minimal background in mathematics, was confused on this point for many years. Every time a colleague spoke about non-linearity, he seemed to be talking about something different! She finally figured out that this is because linearity and non-linearity are asymmetrical opposites. There's only one kind of linearity, but many different kinds of non-linearity. It's like a person being normal vs. eccentric – there are fewer ways to be normal, and more ways to be eccentric!

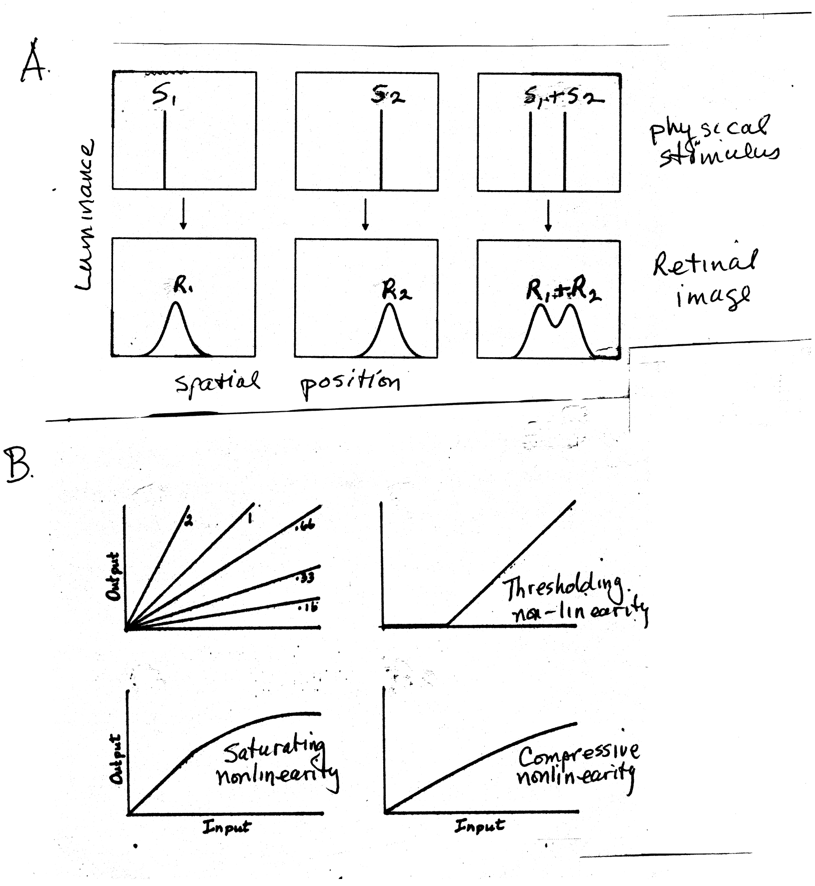


Figure 5.1: Linearity. A. The upper row shows three input stimuli – S_1, S_2,, and the combined input S_1+ S_2. S_1 and S_1 are in different positions (e.g. on a video screen). The lower row shows the outputs (e.g. retinal images) R_1, R_2, and the combined response to S_1 and S_2. Because the response to S_1 + S_2 equals R_1 + R_2, the system is said to be linear. Any other output would reveal a non-linearity. B. Now imagine that S_1 and S_2 are spatially superimposed, for a stimulus of $2S_1$. In a linear system, multiplying the input by a constant, k, multiplies the output by the same constant. Thus, the output magnitude plotted against the input magnitude yields a straight line. Several linear input-output mappings with different gains (slopes) are shown. The other three panels show examples of specific nonlinearities: thresholding (C), saturating (D) and compressive (E). [A modified from Wandell, 1995, Fig. 2.8, p. 21. B-E: DT]

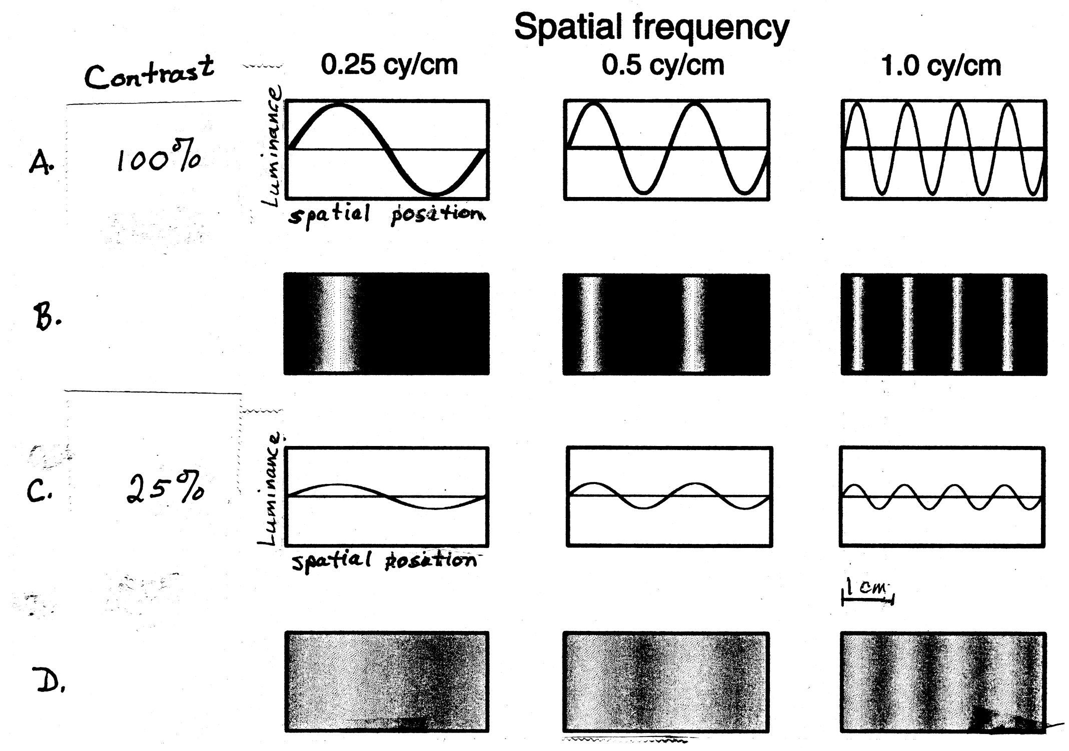


Figure 5.2: Sinusoidal gratings: variations of spatial frequency and contrast. A. Graphs of the variation of luminance with spatial position across the video screen, for sinusoidal gratings of three different spatial frequencies: 0.25, 0.5, and 1.0 cy/cm, all at 100% contrast. B. Simulations of what these three sinusoidal gratings would look like on a video screen. C. Graphs of the same three gratings at 25% contrast. D. Simulations of what the 25% contrast gratings would look like. [AWY]

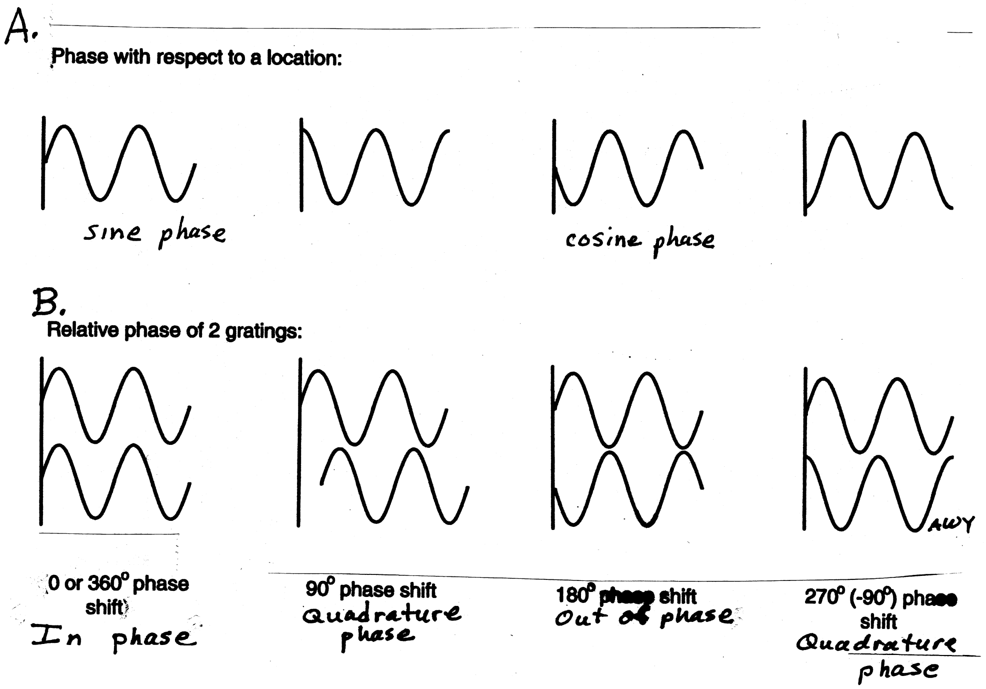


Figure 5.3: Sinusoidal gratings: Variations of phase. A: A sinusoidal grating in various phases with respect to a fixed line. B. Two gratings in various phase relationships to each other. [AWY]

formally defined as

$$Contrast = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

and it takes values between zero (a homogeneous field) and 100%. The gratings in Figure 5.2A and B, whose luminances nominally vary from zero to $2L_o$, have nominal 100% contrast. Figure 5.2C and D show three sinusoidal gratings of the same spatial frequencies shown in Figure 5.2A and B, but at a lower nominal contrast level, 25%. Contrasts lower than 25% cannot be represented readily in print. However, as we shall see, contrasts as low as 1%, and even less, are visible to human subjects under the right conditions.

The third parameter of the sinusoid is its *phase.* The phase of a sinusoidal grating describes its spatial displacement relative to a fixed reference point or to another grating. Figure 5.3A shows a series of locations of a sinusoidal grating with respect to a fixed reference point, and Figure 5.3B shows two gratings varying in phase relative to each other.

One of the historically earliest mathematical uses of sinusoidal functions is to describe some of the properties of circular motion. From this history comes the custom of designating phase in degrees in the same way that circles are designated – 360 degrees is a complete circle, or a complete cycle of the grating. As shown in Figure 5.3B, two gratings of the same spatial frequency that are aligned in space are said to be *in phase* ; i.e. shifted with respect to each other by 0 or 360 degrees. Two gratings shifted by one-quarter of their period (cycle length) are said to be shifted by 90 degrees, or in *quadrature phase.* A half-period shift is a shift of 180 degrees, and two gratings in this relationship are said to be *out of phase* or in *counterphase.* And a three-quarters of a period shift is a 270 degree shift – again, a quadrature shift, but in the opposite direction. For the sake of simplicity, questions of phase will be largely ignored in this book.

5.1.3 Fourier analysis

Why are sinusoidal gratings interesting? In the late 1700s, the French mathematician Jean Baptiste Fourier described an important mathematical theorem. Fourier's theorem states that *any signal that varies in space or time can be described mathematically as the sum of a set of sinusoids that vary in frequency, amplitude, and phase.* By starting with sinusoids of the requisite frequencies, manipulating amplitudes and phases, and summing the sinusoids appropriately, any more complex function can be generated. Breaking down a complex pattern into its frequency components is called *Fourier analysis*, and recombining them to make the original pattern is called *Fourier synthesis.*

An example of Fourier analysis – the analysis of a square wave grating into its sinusoidal components – is shown in Figure 5.4. Fourier's theorem asserts that Fourier analysis can similarly be carried out on any spatial scene (although the results are much more complex than the square-to-sine-wave example).

Why do we care about sinusoidal gratings and Fourier analysis? Most basically, vision scientists are in the business of trying to understand how the visual system codes and recodes visual stimuli. Natural visual stimuli – objects and scenes – vary in an infinite number of ways, and before the arrival of Fourier analysis vision scientists had no common specification system in which to describe them all. But in principle, Fourier's theorem gives us such a descriptive system – we can in principle specify each stimulus in terms of its Fourier components. In addition, the existence of Fourier analysis led some vision scientists to the fascinating idea that, at some level of coding, the visual system might actually use a Fourier-like description to represent visual scenes and objects – but that is getting way ahead of our story[2] (see Chapter 17xx).

5.1.4 Linear systems theory

Linear systems theory is a set of concepts that originated in electrical and optical engineering, and were imported into vision science in the 1950s. Suppose that we are interested in a system such as an audio amplifier, or a lens, or the human visual system as a whole. Also suppose that our goal is to be able to predict the output of the system for any arbitrary input. Rather than measuring the response to each possible input, perhaps it is possible to measure the response of the system to just a few carefully selected inputs, and use these responses to derive a general description of the system. Perhaps this description together with a standard algorithm could then be used to predict the system's response to any arbitrary input.

Now, Fourier's theorem tells us that any pattern can be represented as the sum of a set of sinusoids. Perhaps if we knew the response of the system to a sinusoid of each spatial frequency, we could calculate its response to any arbitrary visual pattern! In pursuit of this goal, then, we

[2] When this approach to vision was first introduced in the 1960s, we irreverant young students were reminded of an old quip about Freudian theory: "You shouldn't criticize psychoanalysis until you've been psychoanalyzed." We provided an update: "You shouldn't criticise Fourier analysis until you've been Fourier analyzed!" We proceeded to amuse ourselves at meetings by Fourier analyzing the various senior scientists in our minds – the slim ones perhaps having high pass amplitude spectra, and the round ones being particularly well endowed with low spatial frequency components.

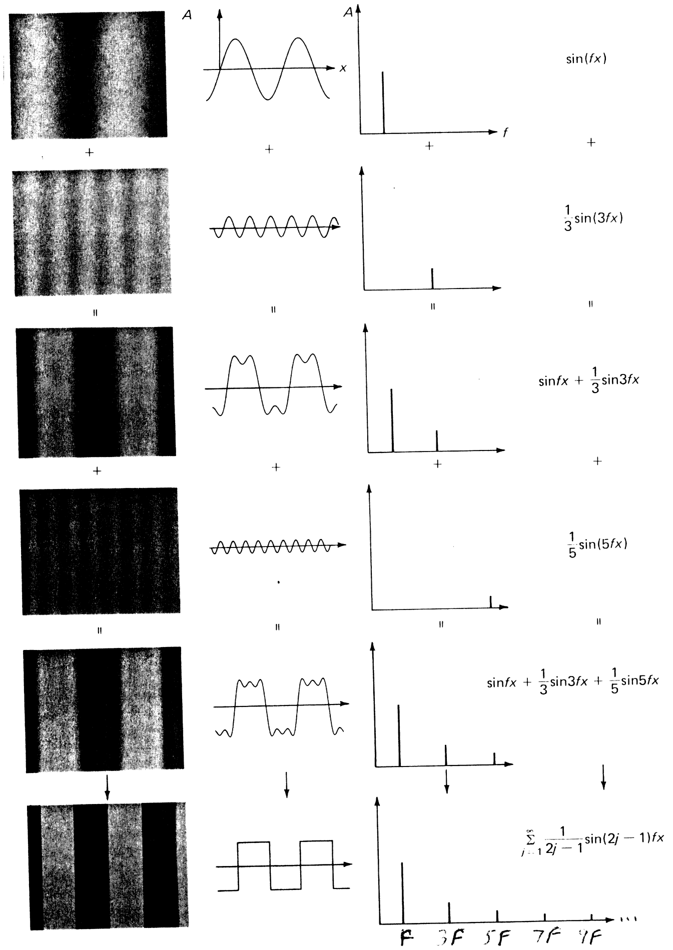

A
x
A
f
sin(fx)
+
$\frac{1}{3}\sin(3fx)$
=
$\sin fx + \frac{1}{3}\sin 3fx$
+
$\frac{1}{5}\sin(5fx)$
=
$\sin fx + \frac{1}{3}\sin 3fx + \frac{1}{5}\sin 5fx$
$\sum_{j=1}^{\infty} \frac{1}{2j-1}\sin(2j-1)fx$
...
F
3F
5F
7F
9F

Figure 5.4: Fourier components of a square wave grating. The top row shows a sinusoidal grating of frequency f. The second row shows a sinusoid of three times the frequency, 3f. The third row shows f and 3f superimposed, yielding a somewhat squared-off waveform. The fourth row shows a sinusoid of spatial frequency 5f. The fifth row shows f, 3f, and 5f superimposed, resulting in more squaring off. Finally, the sixth row shows all of the odd harmonics of f (f, 3f, 5f, 7f.). When superimposed in the appropriate phases and amplitudes they produce a square wave grating. The four columns show four ways of characterizing each grating: with a picture; with a trace of luminance across position; with an amplitude spectrum (showing the amplitudes of the sine waves at each spatial frequency); and with an equation. [Levine and Shefner, 1991, Fig. 10.6, p. 217.]

would need to begin by *establishing a function that specifies the response of the system to sinusoidal inputs of each different spatial frequency.*

Once this function is established, we could in principle Fourier analyse the pattern we are interested in, specifying it in terms of its frequency components; multiply each frequency component by the gain of the system at that frequency; and use Fourier synthesis to calculate the system's response to the pattern. This process is schematized in Figure 5.5 [3]. Of course, the results will be much more satisfying if the system is linear, but the conceptual framework could be interesting even if it is not.

To embark on this pathway, we could begin by measuring the response of a system – such as the optics of the eye, or the visual system as a whole – to sinusoidal gratings of different spatial frequencies. The response is specified in terms of the *contrast ratio*, or *gain* – the output contrast for a given input contrast – at each spatial frequency. If the system can be assumed to be linear, as optical systems are, the resulting function is called a *modulation transfer function* (*MTF*). The term MTF is a particularly meaningful one, as the MTF specifies the fraction of *modulation* (contrast) that the *system* transfers from the input to the output. For example, if at a particular spatial frequency the input contrast is 100% and the output contrast is 25%, the contrast ratio, or gain, is 0.25 at that spatial frequency.

On the other hand, if the linearity of the system being measured is not known, or the system is known to be non-linear, the corresponding function is called a *contrast sensitivity function* (*CSF*). CSFs are most widely studied psychophysically, as system properties of human vision.

Obviously, the next thing we would like to do is determine both MTFs and CSFs for the human visual system. In particular, the MTF of the optics of the eye would move us toward the goal of the chapter by providing us with our much-desired second method for defining human optical quality. But as it turns out, the simplest and cleanest measurements of the optical MTF are derived from measurements of the psychophysically measured CSF! So let us move back to the psychophysics laboratory.

5.2 A new system property: Contrast sensitivity functions (CSFs)

To measure a CSF, we ask a subject to sit in front of a video monitor, and present him with sinusoidal gratings of different spatial frequencies in turn. For each spatial frequency, we measure the subject's *contrast threshold.* That is, we measure the contrast on the video screen required for the subject to just barely detect the grating. We then take the reciprocals of the contrast thresholds to generate sensitivity values, and plot the subject's sensitivity as a function of spatial frequency.

Figure 5.6 shows an early, classical CSF measured on a single human subject. Psychophysical CSFs like this one have three interesting features. First, in a log/log plot such as that of Figure 5.6,

[3] The above description is intended to convey the conceptual basis for using Fourier analysis and linear systems theory at an intuitive level. However, as a matter of honesty, we need to point out two complications. First, a stimulus cannot be specified unambiguously by its spatial frequency content alone – the phases of the components must also be specified. And second, grating patterns are a particularly simple class of stimuli. They are *one-dimensional*, in the sense that luminance varies along only one spatial dimension, and is constant along the other. In contrast, a scene is *two-dimensional* in the sense that its luminance varies along both vertical and horizontal dimensions. As it turns out, Fourier analysis of two-dimensional spatial patterns reveals spatial frequency components in an infinite number of different *orientations*: vertical, horizontal, left diagonal, right diagonal, and every orientation in between. Thus, the true amplitude spectrum of a scene would be much more complex than that shown in Figure 5.5, and would include spatial frequency components and their amplitudes at all possible orientations.

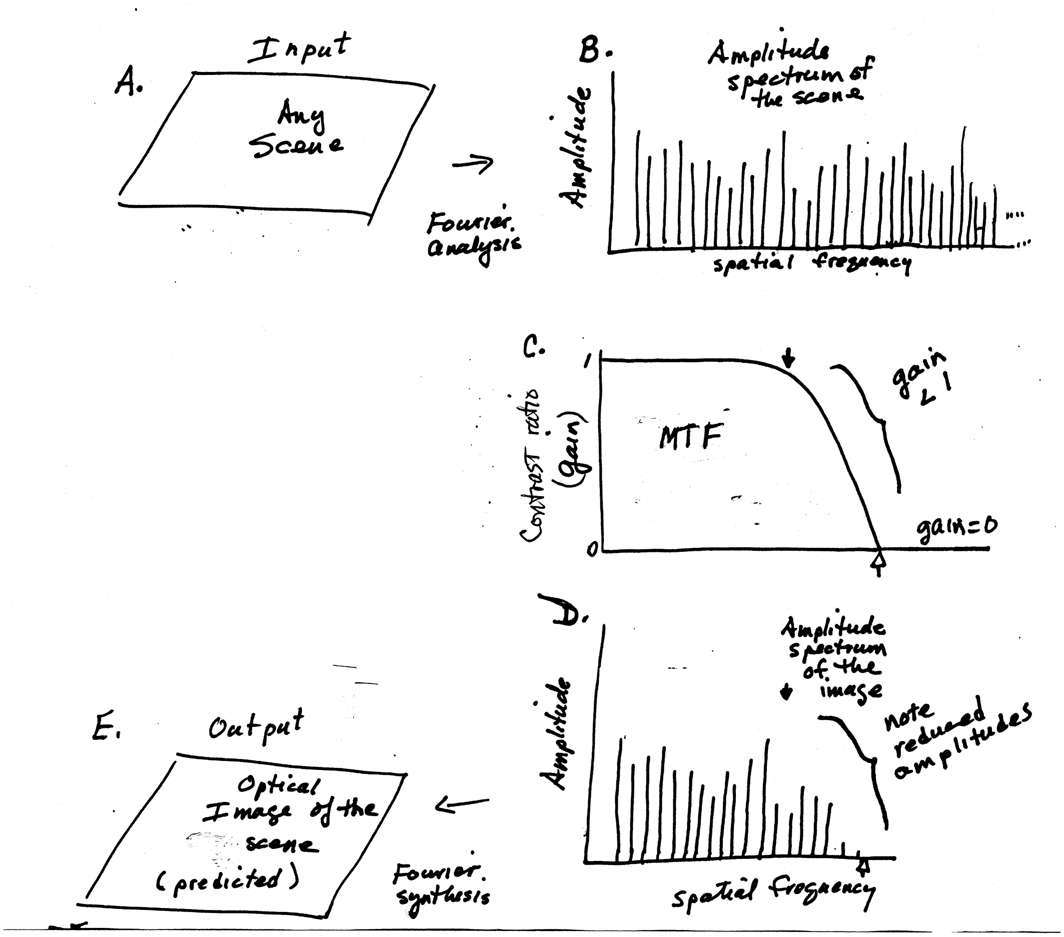


Figure 5.5: Linear systems theory. A schematic illustration of how Fourier analysis and synthesis, in combination with a modulation transfer function (MTF) for the system, allows prediction of the output of a linear system to any arbitrary input. Panel A represents any scene. Panel B represents analysis of the scene into its Fourier components (the amplitude spectrum of the scene). Panel C represents the MTF of the system. The solid arrow shows the spatial frequency at which the contrast ratio (gain) falls below 1; the system reduces the amplitudes of the spatial frequencies above that value. The open arrow shows the spatial frequency at which the contrast ratio falls to zero; the system eliminates all spatial frequencies above that value. Panel D shows the amplitude spectrum of the image formed by the system. Each of the spatial frequencies represented in Panel B is multiplied by the contrast ratio of the MTF at that spatial frequency as represented in Panel C. Finally, Panel E represents the recombination of components across spatial frequency, to produce the predicted output (optical image).

the CSF is band-pass: sensitivity is maximal in the vicinity of 5 cy/deg, and falls at both lower and higher spatial frequencies. Second, as shown on the right-hand ordinate, the minimum detectable contrast is less than 1% – about 0.2% to be exact – and correspondingly, the maximum contrast sensitivity is about 500. This is a remarkably high level of sensitivity – in the spatial frequency range around 5 cy/deg, the subject can detect a sinusoidal luminance wiggle of much less than 1% across the video monitor.

And third, sensitivity falls off sharply at high spatial frequencies. By definition, the highest available stimulus contrast is 100%. Correspondingly, the highest measurable threshold is a threshold that requires 100% contrast (a sensitivity of 1); and the *high frequency cut-off* is defined as the spatial frequency at which contrast sensitivity falls to 1. Between 5 and 60 cy/deg, the visual system shows a sensitivity loss of almost three orders of magnitude. If we extrapolate the smooth curve fitted to the data, it will fall to a sensitivity of 1 at just about 60 cy/deg. Thus, the high frequency cut-off of the CSF is numerically equal to our original estimate of grating acuity: 60 cy/deg.

Is this a coincidence? No. To measure grating acuity, we set the contrast of a square-wave grating to 100%, and vary the spatial frequency to find the highest visible spatial frequency at 100% contrast. To measure the high frequency cut-off, we vary the contrasts of high frequency gratings to find the spatial frequency with a contrast threshold of 100%. Other than the choice of the stimulus variable, the two are the same.

Contrast sensitivity functions are of interest for several reasons. First, CSFs are recognized as the fundamental threshold-level descriptors of spatial vision. They generalize the concept of grating acuity, and describe the sensitivity of our eyes to different spatial patterns in the visual scene. Second, the measurement and modeling of CSFs is fundamental to the multiple spatial frequency channels approach to vision, which we describe in detail in Chapter 17xx.

And third and most immediately relevant, CSFs can be used to derive a new, more functional estimate of optical quality. Now, instead of asking about line spread functions, we can ask the more sophisticated question: what is the optical MTF of the human eye? We next describe three very different techniques for measuring optical MTFs, and compare the results.

5.3 Optical modulation transfer functions (MTFs) for the human eye

5.3.1 The double pass technique

In 1963, Gerald Westheimer used a double-pass technique to measure human MTFs, using sinusoidal gratings rather than lines as the physical stimuli. For a grating of each spatial frequency, he compared the contrast in the physical stimulus and the contrast in what we have called the second image, and back calculated to estimate the contrast in the retinal image. The results are expressed in terms of the *gain*, or *contrast ratio*: the contrast in the retinal image as a fraction of the contrast in the physical stimulus.

Fig. 5.7 shows the estimated MTFs, on both linear/linear axes (Figure 5.7A) and log/log axes (Figure 5.7B). The overall appearance of the curves, of course, depends heavily on the choice of axes, due to the marked differential stretching of both axes at low values and compression at high values in the log/log plot. Our description applies to the log/log plot.

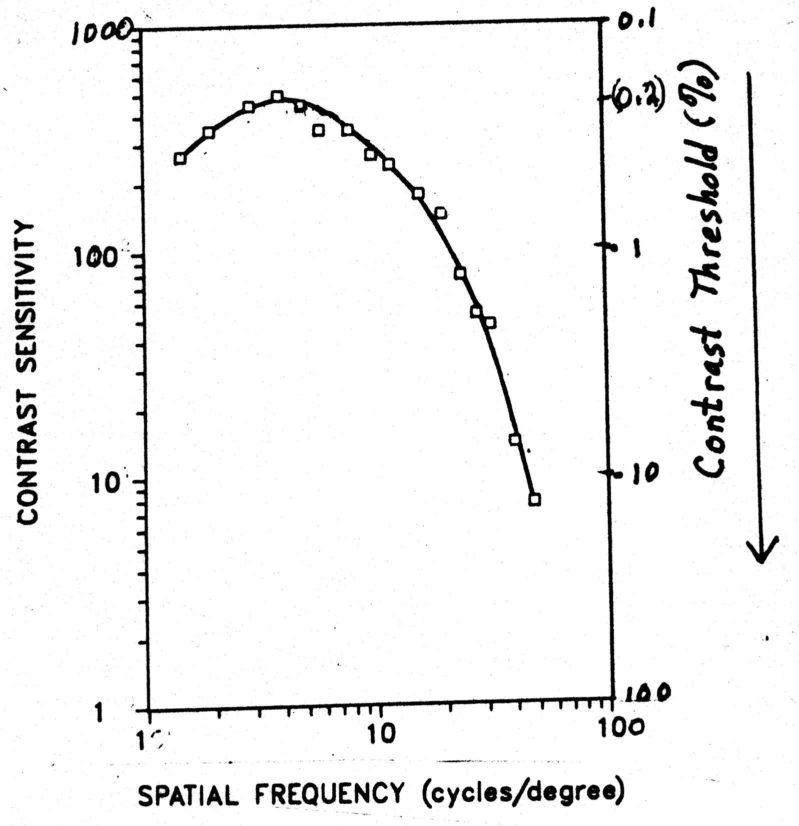


Figure 5.6: A psychophysical contrast sensitivity function, or CSF. Contrast sensitivity – the reciprocal of the contrast threshold – is plotted as a function of spatial frequency. Sensitivity is maximal in the middle of the function, at about 5 cy/deg, and falls off at both lower and higher frequencies. Contrast thresholds are plotted on the right-hand ordinate (with contrast increasing downward) to show the conversion from detection thresholds to sensitivity values. A curve of this shape, with the maximum in the middle, is called a *band-pass* function. [Modified from Van Nes and Bouman (1967), via Olzak and Thomas (1986), Fig. 7.11, p. 7-18.]

MTFs, like CSFs, have several interesting properties. First, unlike CSFs, optical MTFs show constant contrast ratios across the whole low-spatial frequency range, below about 5 cy/deg. Whereas the CSF is band pass, the MTF is low pass. Second, like CSFs, optical MTFs show very high values at low spatial frequencies. Below about 5 cy/deg for small pupils, the contrast in the retinal image is equal to the contrast in the physical stimulus, for a contrast ratio of 1. Remarkably, no modulation is lost in the optics of the eye at low spatial frequencies.

And third, above about 5 cy/deg, the contrast in the retinal image decreases, first slowly and then more and more rapidly, with increasing spatial frequency. The spatial frequency at which the MTF crosses the ordinate value of 0.01 (a contrast ratio of 0.01, or 1%) is called (somewhat arbitrarily) the *high-frequency cut-off* of the MTF[4].

5.3.2 Interferometry

Interestingly, through the technique of *interferometry*, psychophysically defined CSFs provide us with a second approach to measuring the optical MTF. It begins from the remarkable fact that we can produce sinusoidal gratings on the retina in two very different ways. The first way is, of course, by viewing a physical grating directly on a video monitor. Since the optics of the eye are linear, the retinal image will also be a sinusoidal grating, with its contrast degraded in accord with the optical MTF.

The second way of producing sinusoidal gratings on the retina makes use of the physical phenomenon of diffraction. Using a specialized optical system called an *interferometer*, two beams of light from a single (laser) source are focused at two different points in the plane of the subject's pupil (cf. Figure 4.4). The two beams then diverge to make overlapping fields of light, forming interference fringes on the subject's retina.

The interference patterns have two properties dear to the hearts of vision scientists. First, as it turns out, the variation of luminance across the interference fringes closely resembles the variation across a sinusoidal grating. And second, notice that formation of the grating pattern on the retina does not involve the focusing properties of the eye! Thus, remarkably, nature allows sinusoidal gratings to be produced on the retina in two different ways, one (with *direct*, or *ordinary viewing*) that makes use of the optical focusing properties of the eye, and the other (with interferometry) that does not.

The next step is to make psychophysical measurements of CSFs, using each of the two kinds of sinusoidal gratings in turn. For direct viewing, the subject views sinusoidal gratings on the usual video monitor, and adjusts the contrasts of the gratings to threshold. These measurements provide us with a *direct*, or *ordinary CSF*, that describes the transfer of contrast through the whole sequence of stages of the visual system, including both the optics and the neural visual system.

For interferometric measurements, the subject is positioned in a sophisticated optical device called an *interferometer*. He sees the interference fringes as sinusoidal gratings, and again adjusts the contrasts of the gratings to threshold. These measurements provide us with an *interferometric CSF* that, remarkably, describes the transfer of contrast through the neural visual system – retina and cortex – in isolation, but omitting any focusing by the optics of the eye! Since it is not

[4] Notice that there is a practically motivated difference in the definition of the high spatial frequency cut-off of the CSF vs. the MTF. CSFs are psychophysical measurements, and the limit is taken to be the spatial frequency at which a contrast of 100% is required for threshold (no higher contrasts are available). MTFs are physical measurements, and the limit is taken to be the spatial frequency required for a contrast ratio of 1 (measurements at lower contrast values would be difficult because of noise). We will ignore this difference.

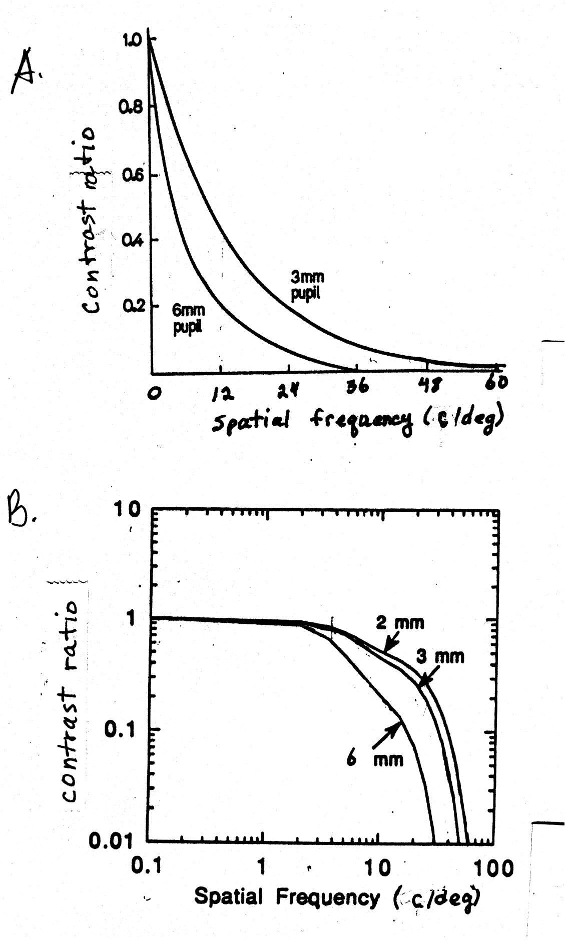


Figure 5.7: Optical modulation transfer functions, or MTFs, measured with the double pass method. A. Linear axes. The abscissa shows the spatial frequency of the physical grating. The ordinate shows the contrast ratio (or gain) of the optical system – the output contrast (the contrast in the retinal image), as a fraction of the input contrast (the contrast in the physical stimulus). B. Logarithmic axes. When logarithms are used to stretch the axes at small values and compress them at high values, the optical MTF is flat at low spatial frequencies. A curve of this shape is called a *low-pass* function. In both panels, MTFs are plotted for pupil diameters of both 3 and 6 mm. A 2 mm pupil is also included in B. As was the case for the line spread functions in Fig. 4.12, the smaller pupils yield better optical quality (higher contrast ratios) than does the 6 mm pupil. [A modified from Westheimer (1963), Fig. 2, p. 173. B modified from Banks and Crowell (1993), Fig. 6-3, p. 94.]

influenced by the optics of the eye, it must depend only on neural factors; and for this reason, the interoferometric CSF has also been called a *neural CSF*. We will return to neural CSFs below.

The results of a classic study by Campbell and Green (1965) are shown in Figure 5.8. Figure 5.8A shows both ordinary and interferometric CSFs. The same subject was used for both sets of measurements. Since the ordinary CSF has been degraded by the optics of the eye and the interoferometric CSF has not, it makes sense that the interoferometric CSF falls above the ordinary CSF throughout the measured spatial frequency range.

Campbell and Green then determined the ratio of contrast sensitivities between the ordinary CSF and the interferometric CSF, as shown in Figure 5.8B. These contrast ratios provide an estimate of the reduction in contrast of an ordinary grating caused by passing through the optics of the eye. As such, remarkably, they provide us with our desired second estimate of the optical MTF. (The two MTFs will be compared more directly below.)

5.3.3 Adaptive optics

More recently, as part of their development of adaptive optics (see Chapter 4), Liang and Williams (1997) have added a third technique for estimating optical MTFs. The wave front sensor of the adaptive optics apparatus allows the description of optical aberrations, and from them theoretical estimates of optical MTFs can be calculated. The details of the technique, however, are beyond the scope of this book.

5.3.4 Comparison of three estimates of optical MTFs

When quantitative comparisons of the data of Figures 5.7 and 5.8 are carried out, they reveal that estimates of the optical MTF based on interferometry show higher contrast ratios, indicating higher optical quality, than do estimates based on the double-pass method. However, it is impossible to tell whether these differences are real; or whether they are based on individual differences among the eyes of the subjects tested in the two different experiments, or on the different kinds of artifacts that potentially impact the different techniques.

To attack this question, Williams, Brainard, McMahon, and Navarro (1994) repeated measurements of optical MTFs with both the double-pass technique and the interferometric technique, under tightly parallel conditions, on the same subjects. Moreover, Liang and Williams (1997) used adaptive optics to derive a third estimate of MTFs on the same three subjects, keeping other factors as similar as possible.

The results from all three techniques, averaged across three subjects, are shown on linear axes in Figure 5.9. The measured optical MTFs are poorest with the double pass technique, medium with interferometry, and best with adaptive optics. However, individual differences are large, especially for the wave front sensor technique; and the differences among techniques are only about the size of the individual differences among subjects.

Which of the three techniques best characterizes the real MTF of the eye's optics? The double pass technique is suspect because it depends upon light scattered within the eye and reflected from the retina. Several different retinal structures in different depth planes could in principle contribute to the scattered and reflected light. These variations could artifactually reduce the contrast in the second image, and thereby reduce the estimate of contrast in the retinal image. In sum, the estimates of the MTF provided by the double pass technique are probably too low.

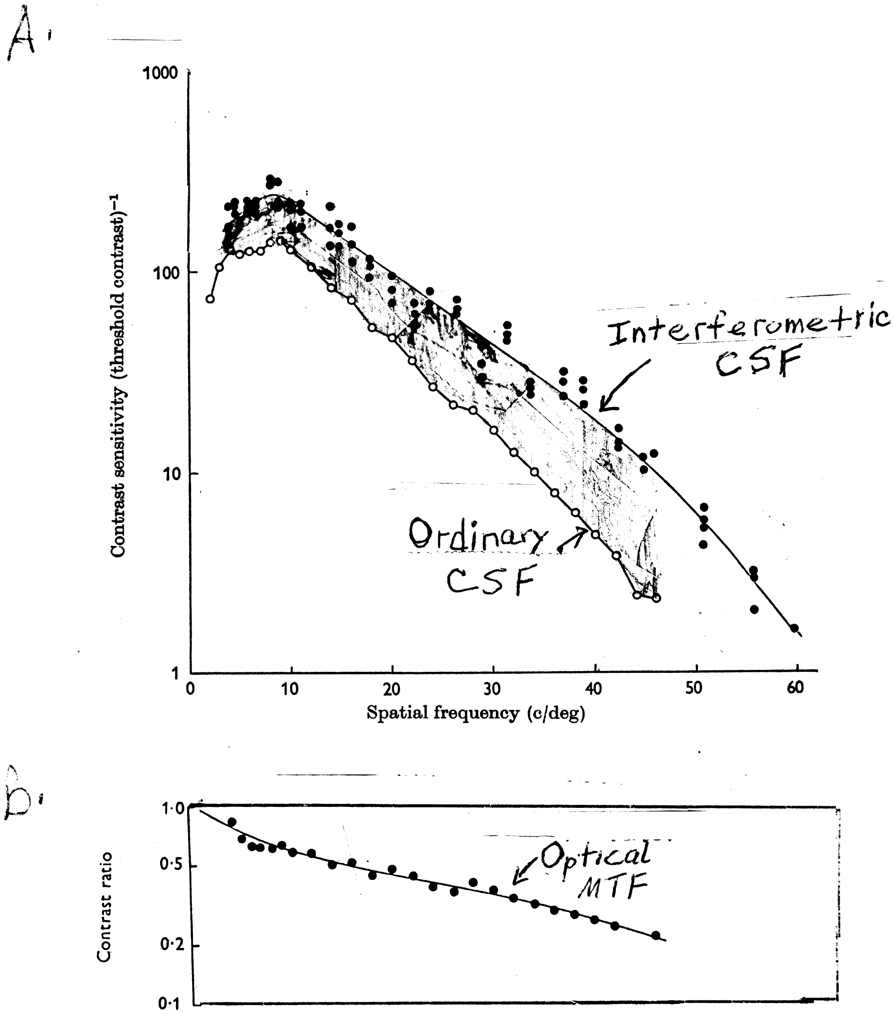


Figure 5.8: An optical MTF estimated from interferometry. A. The open symbols show an ordinary CSF measured by direct viewing of a video monitor, and therefore influenced by both optical and neural factors. The closed symbols show a CSF measured with interference fringes, and therefore influenced only by neural factors. B: The contrast ratio between the ordinary and interferometric CSFs (also shown by the shaded area in A), influenced only by optical factors. This function provides a new estimate of the optical MTF. [Modified from Campbell and Green, 1965, Fig. 9, p. 586.]

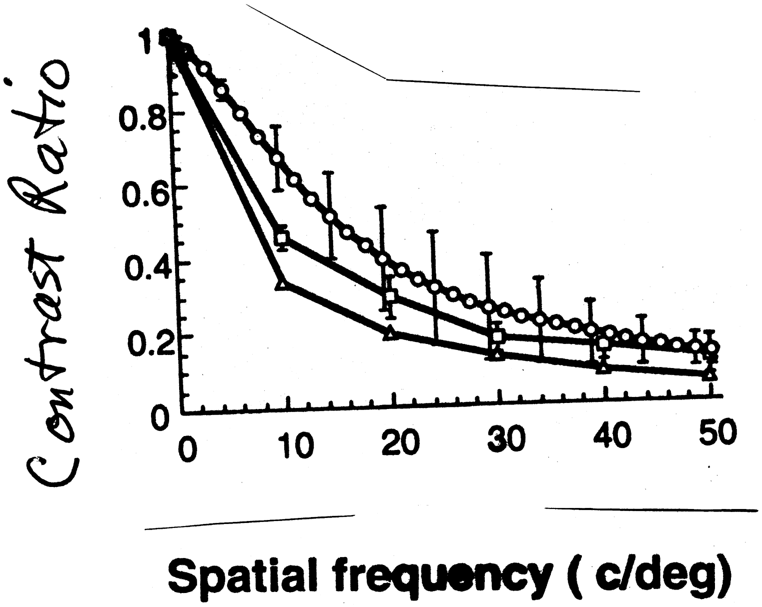


Figure 5.9: Optical MTFs measured with three techniques. The same three subjects were tested in each case, and the data are averaged across the three. The triangles, squares and circles show the results of the double-pass, interferometric, and adaptive optics techniques respectively. [Modified from Liang and Williams, 1997, Fig. 6, p. 2877.]

The interoferometric technique, on the other hand, is based on the actual visual performance of a human subject. That is, it provides a measure of the optical image in whatever plane within the eye is actually used for the quantal absorptions that initiate the visual signal. And as it turns out, the new adaptive optics measurements are much quicker than those from the other two methods. Liang and Williams suggest that the speed of the measurements may sharpen the MTF by allowing the subject to hold a more constant accommodative state over the whole measurement interval.

From a broader perspective, given the wide variation of measurement techniques, the most striking thing about these three sets of measurements is the agreement among them. According to Liang and Williams, the difference in MTFs between the interoferometric and adaptive optics measures would be brought about by a factor as small as a change of accommodation of only about 0.15 diopters. And all three techniques agree that the percentage of stimulus modulation transferred through the optics to the retina is about 50% at 10 to 20 cy/deg, and 10 to 20% at 50 cy/deg.

In sum, one of the fundamental questions we posed at the beginning of this chapter was, how good are the optics of the eye? Thanks to the work described above, the question of optical quality is now a solved problem in vision science. The answer is that the eye's optics are excellent for low spatial frequencies, transferring virtually 100% of the contrast from the physical stimulus to the retinal image. But for spatial frequencies above about 5 cy/deg, the optics clearly degrade the contrast in the retinal image. By 60 cy/deg, most of the contrast in the physical stimulus is lost in the optics, and does not make it to the retinal image.

5.4 Photoreceptor spacing

We now return to the second possible cause of the 60 cy/deg limit on grating resolution mentioned in Chapter 1: the anatomical layout of the photoreceptors. Beyond the optics, the next processing elements of the visual system are the photoreceptors – the entities that absorb light and start the neural signals in the visual system. As was shown schematically in Figure 1.6xx, the retina is paved with a mosaic of photoreceptors. Each photoreceptor absorbs only the quanta of light that arrive at that photoreceptor's location. Moreover, a photoreceptor sums the signals arising from all of the quanta it absorbs, without regard for the spatial location of each quantum within it. Thus, although the optical image is continuous in space, the photoreceptors sample the optical image *discretely*: that is, they sum the signal over small, separate local regions.

The photoreceptor layer is often referred to as the *receptor mosaic.* This terminology makes an analogy between the discrete sampling implemented by photoreceptors and the representation of a visual scene in a mosaic. In a mosaic, each local region of a scene is represented by a single tile of a homogeneous color. Just as the mosaic distorts the scene by representing each local region by the average color of that region, so too the photoreceptor mosaic distorts the incoming visual signal by summing the effects of quantal catches within local regions. As it turns out, this stage of discrete spatial sampling influences the visual signal.

5.4.1 The Nyquist limit

The effects of discrete sampling by a regularly spaced array of photoreceptors are shown schematically in Figure 5.10. This figure consists of five panels. In the top row of each panel is shown the luminance profile in the retinal image . In the second row is a set of schematic photoreceptors with a fixed inter-receptor spacing, d. And in the bottom row are the signals produced by each of the corresponding photoreceptors across the matrix.

As shown in Figure 5.10A, a homogeneous field of light produces an equal quantum catch in all of the photoreceptors (give or take a little noise). In Figure 5.10B, a coarse grating falls on the photoreceptor array. Several photoreceptors fall under each dark stripe, and several under each bright stripe. That is, there will be several neighboring photoreceptors with high quantum catches, and then several with low quantum catches. Think of yourself looking out at the world through this array of photoreceptors. Analysis of the regions of high and low quantum catches would allow you to deduce that there is a spatial pattern in the world, and to have veridical (accurate) information about its spatial frequency.

In Figure 5.10C the grating is matched to the spatial separations of the photoreceptors, so that the grating produces a dark stripe on one photoreceptor and a bright stripe on its neighbor. The pattern of photoreceptor outputs would be finer in this case – one photoreceptor per stripe of the grating – but again the pattern carries veridical information about both the presence and the spatial frequency of the grating.

But there's a limit to the fineness of the grating that can be represented unambiguously by such a set of sampling units. Consider the grating in Figure 5.10D. In this case the spatial frequency of the grating is high enough so that one period of the grating – one dark and one bright stripe – falls on each photoreceptor. The pattern of quantum catches across the set of photoreceptors will wash out, and be quite similar to the pattern made by the homogeneous field in Figure 5.10A. Based n this argument, you would no longer be able to tell the grating from the homogeneous field, and information about the spatial pattern would be lost.

Intuitively, to preserve the spatial variations in the grating, one needs to sample both the bright and the dark stripe of each period of the grating, and have not more than one stripe (1/2 cycle of the grating) per photoreceptor. That is, a single cycle of the grating must occupy at least twice the inter-receptor spacing, or 2d. 2d is defined as the *Nyquist limit* of the sampling array, stated in terms of the period of the grating. Since the spatial frequency of the grating is the reciprocal of the period, in terms of spatial frequency, F, the Nyquist limit is F = 1/2d. The grating in panel C is exactly at the Nyquist limit[5] of the receptor array.

5.4.2 Alias patterns in the primate fovea

But there's one more level of complication to this argument, because the Nyquist limit is not an absolute limit. Under the right conditions, some information about the presence of spatial frequencies above the Nyquist limit does get through a discrete sampling mosaic. The phenomenon is called *aliasing* because the pattern of photoreceptor responses across the sampling matrix, made by each supra-Nyquist frequency, closely resembles the pattern made by a sub-Nyquist frequency. Just as Mack is also called "The Knife", each frequency above the Nyquist limit potentially creates the same spatial pattern as a frequency below the Nyquist limit, and slips through the photoreceptor mosaic under an assumed identity – an alias.

Consider Figure 5.10 again. Figure 5.10C depicts a grating with a frequency at the Nyquist limit (call it grating N), and Figure 5.10E shows a frequency three times the Nyquist limit (call it 3N). For grating 3N, two bright stripes and one dark stripe fall on the first photoreceptor, two dark stripes and one bright stripe fall on the second photoreceptor, and so on. Even though this

[5] When we wrote Nyquist limit, the spell checker suggested that we might mean the nicest limit....

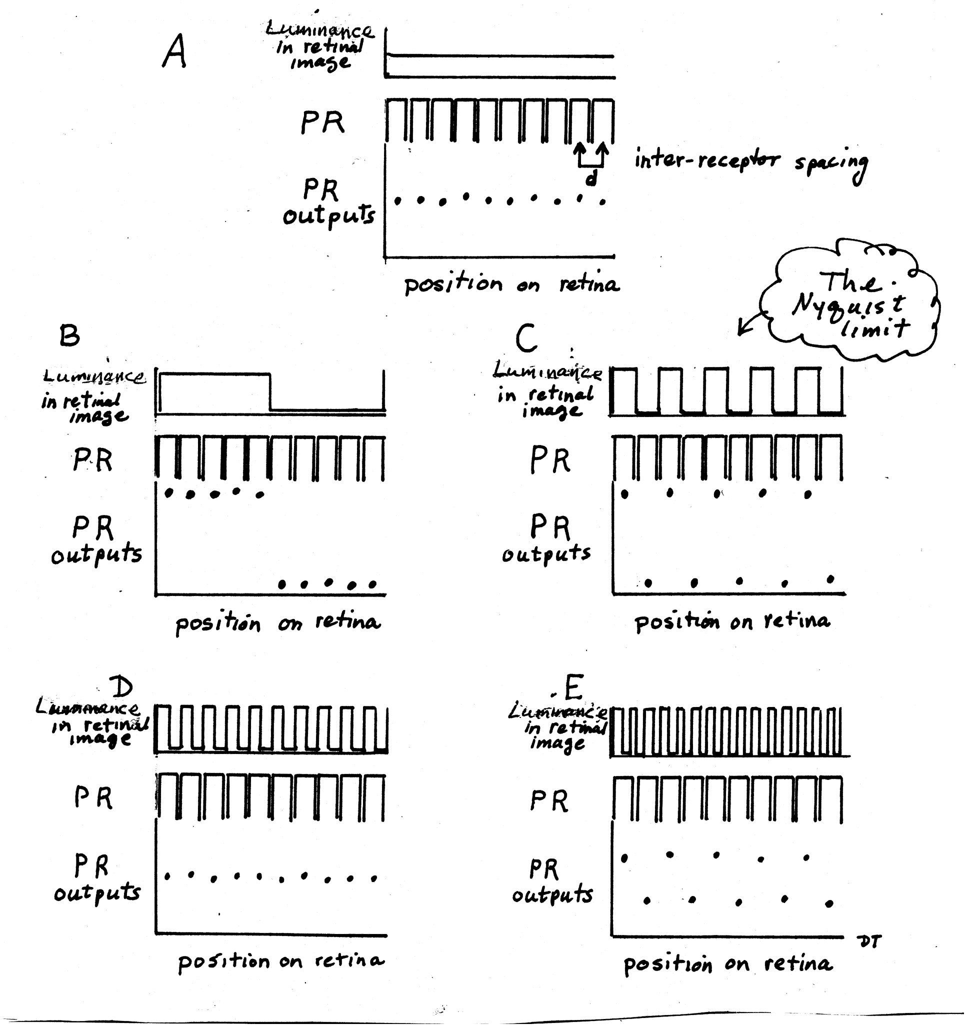


Figure 5.10: Discrete sampling and the Nyquist limit. A. A homogeneous field of light produces an approximately constant number of quantal catches in each photoreceptor (PR) across the matrix. B. A coarse grating creates a coarsely varying pattern of quantal catches, and is readily distinguished from the homogeneous field. C. A grating at the Nyquist limit creates alternating high and low quantal catches in neighboring photoreceptors. D. A grating at twice the Nyquist limit creates nearly equal quantal catches in each photoreceptor, and aliases to the homogeneous field (A). E. A grating at three times the Nyquist limit aliases to the pattern at the Nyquist limit (C).

spatial frequency is above the Nyquist limit, the pattern of quantum catches will vary across the row of photoreceptors, and the pattern of activity across the row of photoreceptors differs from that created by the homogeneous field.

In fact, for our assumed perfectly regular matrix, the grating 3N would give you the *same* pattern of quantum catches as does the grating N, although at a lower contrast. If we lower the contrast in grating N, then gratings N and 3N will yield the same spatial pattern; that is, they will alias to each other[6]. The same argument will hold true for many pairs of spatial frequencies. For each sub-Nyquist frequency, there will be a supra-Nyquist frequency that will alias to it. [Work out some examples of such *alias pairs* using diagrams like those in Figure 5.10. What pattern do you find?]

Now, how do these considerations apply to human vision? In the human fovea the interreceptor spacing, d, is about 0.5 minutes of arc, so (ignoring some complications) the Nyquist limit should be about 60 cy/deg. So, if we could bypass the optics of the human eye, and create gratings with frequencies above about 60 cy/deg directly on our retinas, we should see alias patterns!

Simulations of the alias patterns predicted for a primate retina are shown in Figure 5.11. Figure 5.11A shows a map of the locations of individual photoreceptors in the foveal region of a macaque monkey retina. The map shows irregular patches of regular arrays of photoreceptors. Figure 5.11B-D show patterns produced by physically laying horizontal square wave gratings of particular spatial frequencies on the matrix. The simulated spatial frequencies are 40, 80, and 110 cy/deg in panels B, C, and D respectively. In B, 40 cy/deg is below the Nyquist limit of the monkey's fovea, and the grating is represented veridically as a set of horizontal rows of dots. In C and D, 80 and 110 cy/deg are both above the Nyquist limit, and alias patterns appear. They are irregular because of the patchiness of the photoreceptor matrix.

5.4.3 Can we see our own alias patterns?

Conclusive experiments demonstrating the detection of alias patterns by human subjects were carried out by David Williams in 1985. Williams used interferometry to create supra-Nyquist gratings of up to 200 cy/deg directly on the retina. He then measured contrast sensitivity functions with a two-interval forced-choice experiment. Each trial of the experiment consisted of two time intervals, one of which contained a grating (notice that this is an externally referred, forced-choice, detection experiment). The spatial frequency and contrast of the gratings varied from trial to trial. The subject's task was to judge which time interval contained the grating.

The resulting interoferometric CSFs are shown in Figure 5.12A. As expected, the subjects detected gratings below the Nyquist limit, as shown by the left hand lobe of the CSF below 60 cy/deg. But they also detected gratings far above the Nyquist limit, as shown by the right hand lobe and the tail above 60 cy/deg. With the forced-choice technique, gratings were detected all the way out to at least 200 cy/deg, the maximum spatial frequency that could be produced by the interferometer.

Of course a forced-choice experiment such as this one only tells us that the gratings were detected, but not what they looked like (remember, for Class A experiments, don't ask; don't

[6] Alias patterns often occur on video systems. They are produced by aliasing between the spatial patterns in the scene and the spatial sampling characteristics of the video. The most exotic example of aliasing DT has ever seen occurs in the Apache dance in the movie Can-Can, viewed on a video system. The female dancer is wearing finely striped tights, and her alias patterns flash spectacularly every time she moves or changes her distance from the camera.

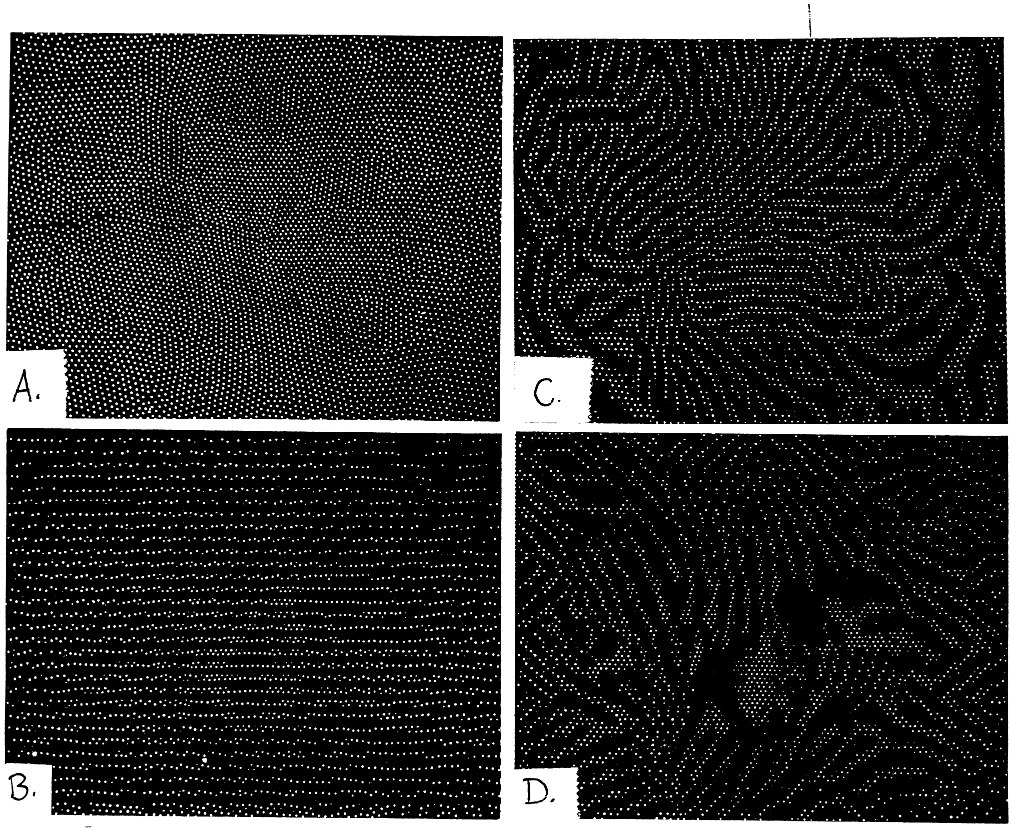


Figure 5.11: Simulated alias patterns for a monkey fovea. A. The monkey's retinal mosaic. Each white dot represents the location of a photoreceptor. Notice the hexagonal packing in local regions, and the irregularities among local regions. B. Simulation of the pattern arising from a horizontal 40 cy/deg grating. 40 cy/deg is below the Nyquist frequency of the matrix, and the grating is represented veridically, as horizontal rows of dots. C. An alias pattern arising from simulation of an 80 cy/deg grating. D. An alias pattern arising from simulation of a 110 cy/deg grating. [Williams (1985), Fig. 6, p. 202.]

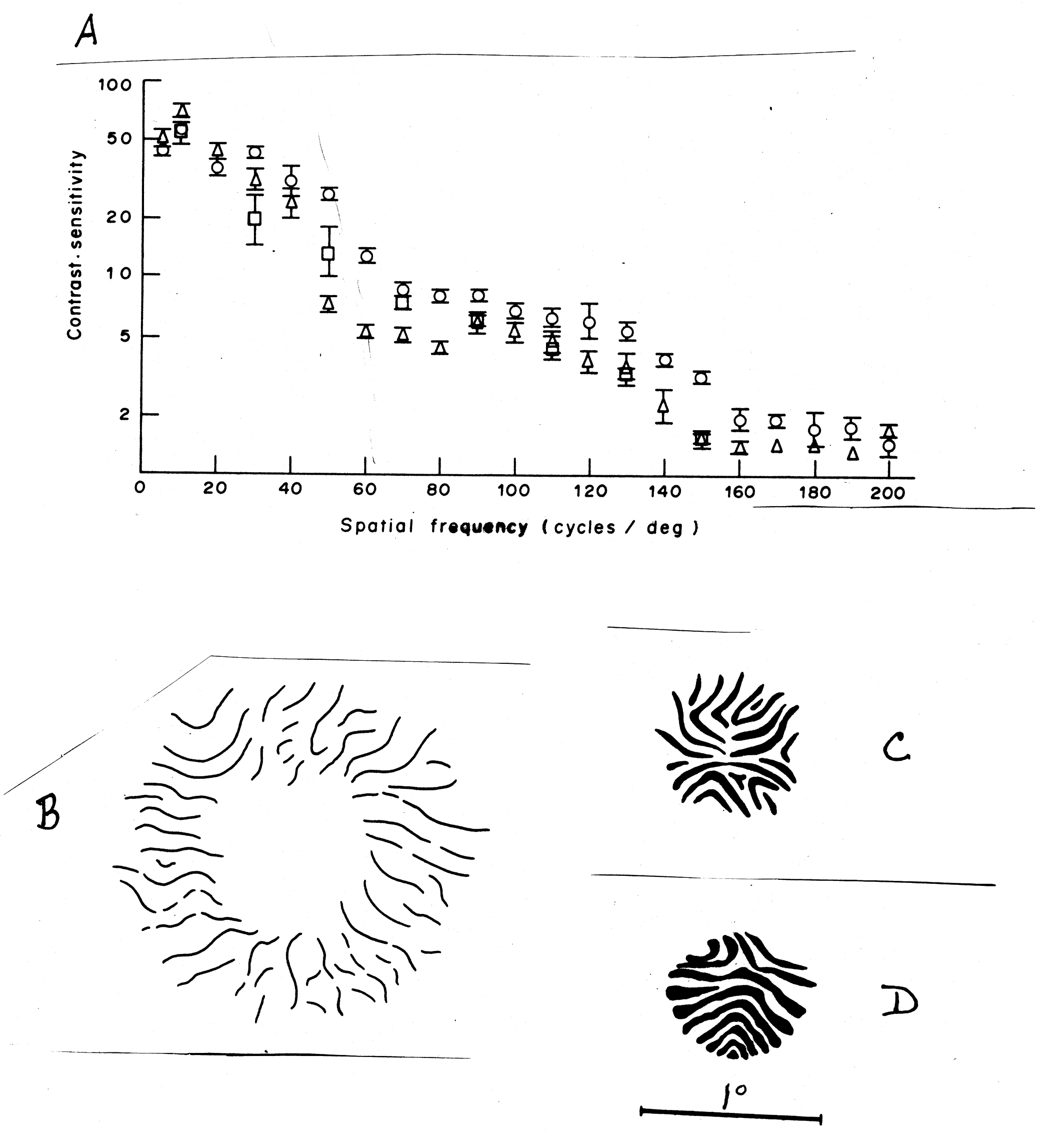


Figure 5.12: Detection of alias patterns. A. An interferometric CSF based on forced-choice detection thresholds. Gratings are detectable out to a spatial frequency of 200 cy/deg. Subjects report that gratings of spatial frequencies below 60 cy/deg are perceived veridically, whereas spatial frequencies above 60 cy/deg are visible as alias patterns. B-D. Drawings of alias patterns. The scale bar in D shows 1^o of visual angle, and applies to all three drawings. [From Williams (1985). A from Fig. 3, p. 199; B, C, D from Fig. 4, p. 200.]

tell). To address this question, the subjects also described and drew the patterns they saw in the interferometer (a Class B experiment). Three of these drawings are shown in Figure 5.12B-D. For fringes below about 60 cy/deg, subjects reported seeing regular gratings, with spatial frequencies corresponding to the actual spatial frequencies of the interference fringes. But above about 60 cy/deg, they saw coarse, wiggly ”zebra stripes” or ”worms”, resembling the drawings shown in Figure 5.12B-D (note the similarity to the alias patterns shown in Figure 5.11C-D).

5.4.4 Neural CSFs

Because of the differences in the perceived qualities of the test stimuli – regular gratings vs. alias patterns – the different parts of the CSF shown in Figure 5.12A are attributed to two different origins. The region below 60 cy/deg is attributed to the processes that underlie the detection of ordinary gratings, whereas the region above 60 cy/deg are attributed to the processes that underlie the detection of alias patterns. As noted previously, the region below 60 cy/deg is called the *neural CSF* ; remarkably, as discussed above, it reveals the CSF for the neural visual system in isolation, unaffected by optical factors.

More recently, Sheng He and Donald MacLeod (1996) have extended the analysis of the neural CSF. As shown in Figure 5.11C-D, and 5.12B-D, the stripes in the alias patterns formed by our retinal mosaics are wiggly and variable in orientation, and in general do not conform to the orientation of the fringes that produce them. He and MacLeod proposed that subjects be tested with a new forced-choice psychophysical task – *orientation discrimination.* They reasoned that subjects should be able to make orientation discriminations for spatial frequencies that are detected veridically, but not for gratings detected only by their alias patterns.

He and MacLeod's results are shown in Figure 5.13. Unexpectedly, their subjects could often do the orientation task above the Nyquist limit for large differences in orientation – vertical vs. horizontal, or ± 45 degrees – presumably because the two gratings produced two discriminably different alias patterns. But for smaller orientation differences – gratings oriented at ± 5 or ± 10 degrees from horizontal – the subjects failed above about 60 cy/deg, and no orientation thresholds were measurable, even at 100% contrast. Thus, the lower lobes of the interferometric CSFs for orientation discrimination in Figure 5.13 provide additional estimates of neural CSFs for the human visual system.

Although this chapter concerns optical rather than neural factors, notice that we seem to have received a bonus for the future – the neural CSF, which is a description of the CSF for the neural visual system, unaffected by the optics of the eye. Notice that in both Figure 5.12 and Figure 5.13, the neural CSF declines relatively slowly below about 40 cy/deg, but falls precipitously at high spatial frequencies – more than an order of magnitude in the spatial frequency range between 40 and 60 cy/deg. We return to the implications of this finding below, when we return to the limits on grating acuity.

In the meantime, He and MacLeod complicate the interpretation of the interferometric CSF one step further by proposing a three-factor model, shown in Figure 5.14. They argue that ordinary grating detection accounts for the left-hand lobe of the interferometric CSF – the neural CSF – below about 60 cy/deg. And alias patterns account for the second lobe that rises to the right of the Nyquist limit, above about 60 cy/deg. But if these were the only two factors, there should be a sharp minimum in the CSF at about 60 cy/deg, where veridical perception has cut off and aliasing is not yet available. In fact, no such minimum appears in the data, particularly in Figure 5.12 and

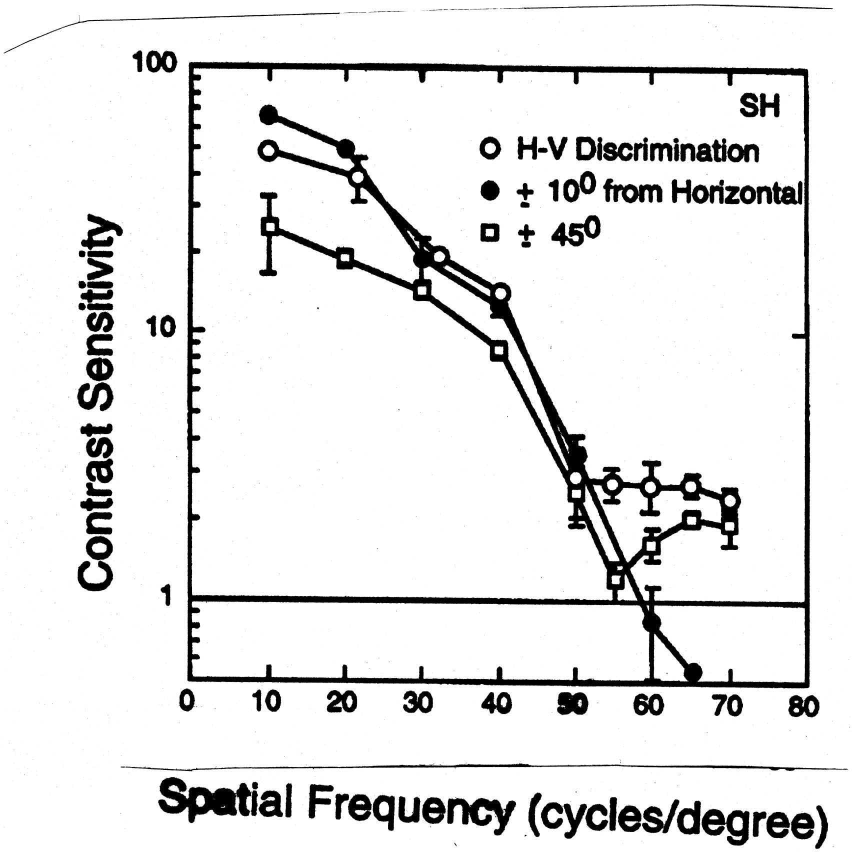


Figure 5.13: Orientation discrimination and the neural CSF. Using interferometry, subjects were asked to discriminate between gratings of two different orientations. For small orientation differences (closed symbols), discrimination became impossible above about 60 cy/deg, supporting the argument that detection above this value is mediated nonveridically by the detection of alias patterns. The lower lobe of the data, which arises from veridical grating perception, provides a new estimate of the neural CSF.

for one subject in Figure 5.13.

What is going on? Interestingly, He and MacLeod argue that detection in the spatial frequency range near 60 cy/deg is caused by yet a third process – a non-linearity arising beyond the optics of the eye, within the neural retina. We will return to this argument in Chapter 6, after we have filled in some background on photoreceptors.

5.5 Hyperacuities

The foveal photoreceptors are spaced about 30 seconds of arc apart, and as we have just seen, the spacing (the *spatial grain*) of the photoreceptors limits the high frequency cutoff of the optical MTF to about 60 cy/deg. By definition, a *hyperacuity* is a sensitivity to the locations of stimuli that is finer than the spatial grain of the visual system (Westheimer, 1979).

For example, the inset in Figure 5.15 shows a target for the task of *vernier acuity.* The vertical line that contains a small horizontal offset. Human subjects can see the offset and detect its direction when the offset is only a few seconds of arc – much smaller than the diameter of a photoreceptor. Similarly, suppose that a subject is looking at a line of light, and at a certain point in time it jumps either rightward or leftward. How far will it have to jump, in order for the subject to be able to tell the direction of the jump? The *displacement threshold*, like vernier acuity, is only a few seconds of arc.

How is it possible for the visual system to have such refined sensitivity to spatial location? Figure 5.15 provides an analysis. In this figure, two vertical lines, A and B, are physically separated by 12 seconds of arc. In the geometrical image, line B is slightly to the right of line A. But because of the line spread function of the optics, the retinal image of each line is a relatively broad distribution of light, and information about the relative locations of the lines is contained in the spatial pattern of quantal catches across each retinal image. By catching enough quanta and doing enough statistical analysis of the two patterns, the visual system can in principle sort out the relative locations of the two lines.

Aside from their counterintuitive nature, hyperacuities raise two interesting points. First, notice that the successful analysis of spatial location in Figure 5.15 depends on the fact that each line makes a broad line spread function. If the optics were perfect, the images of the two lines could fall on the same set of photoreceptors on at least some trials, and the two lines would then be indiscriminable in location on those trials. We can thus speculate that perhaps the driving force that makes the optics of the eye remain as imperfect as they are is the potential usefulness of optical blur for enabling spatial hyperacuities.

Second, what makes it worthwhile for the visual system to keep track of the small differences between the two quantum catch patterns in Figure 5.15? There must be something very important about the relative locations of lines or objects, to justify doing the fine analysis required by hyperacuities.

5.6 Reprise: What limits grating acuity?

So, finally, how shall we answer our initial apparently simple locus question: what limits grating acuity? We now have quantitative descriptions of three possible limiting stages – optics, photoreceptor spacing, and (collapsing retinal and central stages) neural processing. The limits imposed

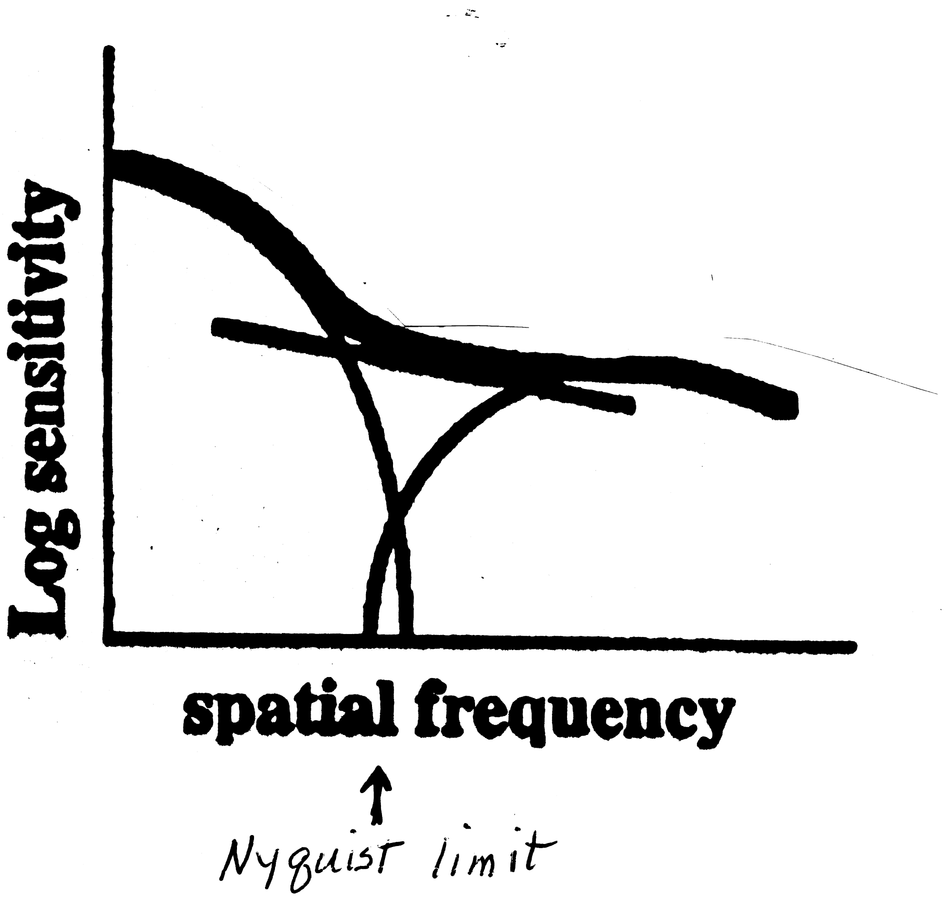


Figure 5.14: A three-factor model of the interferometric CSF. The lower lobe, below 60 cy/deg, is attributed to ordinary (veridical) grating detection; the upper lobe, above 60 cy/deg, to (non-veridical) detection of alias patterns; and the filling in of the gap between the two to a retinal non-linearity (see Chapter 6). (Modified from He and MacLeod, 1996, Fig. 8, p. 1145.)

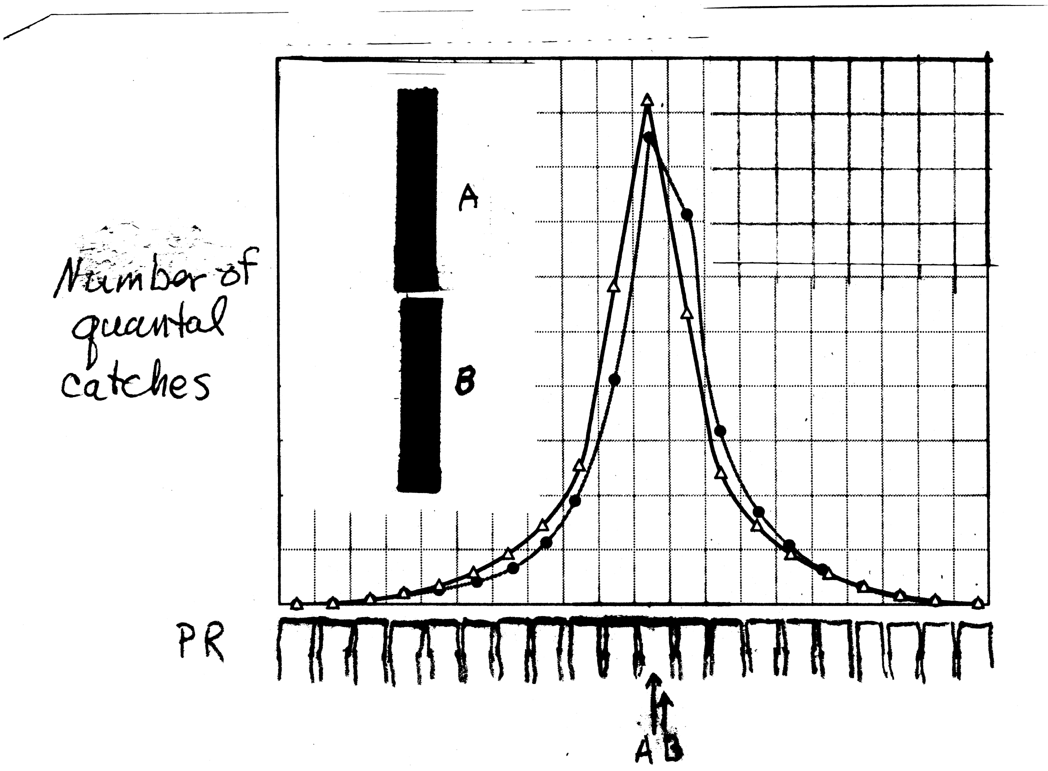


Figure 5.15: Hyperacuity. The inset shows a vernier acuity target. In the main figure, two lines A and B are shown at locations separated by 12 seconds of arc – about 1/3 the width of a foveal photoreceptor. Each line makes a line spread function in the retinal image, with the centers of the two functions separated by 12 seconds of arc. The row of boxes marked PR denotes a row of photoreceptors. The figure shows a hypothetical distribution of quantal catches arising from each of the two lines. The relative locations of the two lines can be determined by statistical analysis of the two patterns of quantal catches. [After Wandell, 1995, Fig. 7.27, p. 242.]

by all three loci – the high frequency cut-off of the optical MTF, the Nyquist limit of the photoreceptor matrix, and the high frequency cut-off of the neural CSF – all converge at the original grating acuity limit of about 60 cy/deg. What conclusion[7] shall we draw?

The simplest option, of course, is to stay with our original answer – the optics of the eye. Since the optics are the first stage of processing, the incoming light encounters the optics before it encounters the photoreceptors or later levels of neural processing. The optics remove high spatial frequencies from the retinal image, and thereby impose the initial limit on spatial resolution. They get the blame, even though if they had a higher cut-off frequency, the receptor mosaic and the neural CSF would still impose approximately the same resolution limit.

But, taking a slightly more sophisticated perspective, why are these three limits so closely matched? Of course, the matching of components in a serial processing system is just a sensible design feature, implemented by natural selection. There is no point in having the sampling limit or the neural limit better than the optical limit, or vice versa.

But there's still a deeper argument. Remember that if the optics weren't cutting out the spatial frequencies above the Nyquist limit , aliasing would occur. But aliasing can yield false perceptions, which could in principle be disadvantageous – we don't want to be seeing zebras when we look at picket fences or venetian blinds! Assuming there is no alternative to discrete sampling, how might one build a system that would still avoid the false perceptions that would arise from alias patterns? One solution would be to make the sampling matrix irregular, and indeed there are irregularities in the photoreceptor mosaic in Figure 5.7. However, we know that these irregularities are not sufficient to eliminate alias patterns, because we see alias patterns with interference fringes.

The solution to the alias problem is to limit the spatial resolution of the optics to below the Nyquist limit. An optical system that greatly reduces the contrast of gratings above 60 cy/deg would obviate the problem of alias patterns. On this argument, the optics of the eye may have evolved to be as poor as they are, in order to function as an *anti-aliasing device* for the photoreceptor mosaic. The important design principle here is, processing stages are not just selected to have similar limits, but also to cover for each others' deficiencies.

We can further speculate that the ultimate limit of fineness of the photoreceptor mosaic is very likely set by the minimum size that photoreceptors can be. Perhaps the organelles inside the photoreceptors can only be made so small, and no smaller. Perhaps human foveal photoreceptors are as small as they can be without a radically new design. Perhaps some limit on photoreceptor design sets the Nyquist limit at 60 cy/deg, and the Nyquist limit necessitates the optical limit at 60 cy/deg in order to protect the system from alias patterns.

If this argument were true, then what limits grating acuity to 60 cy/deg? At the functional level we might choose to blame the optics. But at the system level we might choose to blame the size and spacing of photoreceptors, which require the optics to be poor if aliasing is to be avoided[8].

The neural CSF also poses a puzzle. In evolutionary terms, it makes no sense for the neural CSF to process spatial frequencies beyond the cutoff frequency of the optics, so it is not surprising that it too cuts off near 60 cy/deg. However, why does the neural CSF fall off so rapidly at spatial frequencies between 40 and 60 cy/deg? In principle, in order to preserve all of the information transferred by the optics, the neural CSF should be low pass, with close to 100% modulation

[7] Other than, Beware of simple questions!

[8] We note that this conclusion is controversial. He and MacLeod argue that the optics are even poorer than would be needed to avoid aliasing. If so, there could be some other evolutionary factor that requires a limited optical quality. Perhaps this factor is the broad line spread function needed to support the hyperacuities for spatial location.

transferred right out to 60 cy/deg.

Why doesn't the neural CSF take this shape? One would have to argue that preserving the available contrast information above 40 cy/deg is too expensive to be worthwhile in terms of neural processing. But why discard information needed for spatial pattern, but preserve information about spatial location? Perhaps refined hyperacuities are more important than high contrast sensitivity. A quantitative design argument is needed if these issues are to be taken further.

5.7 Summary: Recoding the visual signal

In Chapters 4 and 5 we have talked about the first stage of information processing within the visual system – the optics of the eye. In this book we will call the mapping of the physical world to the retinal image, via the optics of the eye, the *First Transformation.*

The First Transformation imposes two major information losses on the incoming visual signal. First, the physical world and the objects within it are three dimensional, but the retinal image is only two dimensional. The optics of the eye collapse all points in the same line of sight to the same location in the retinal image. Moreover, the size of the retinal image of an object is not tied to the size of the object, but varies with its distance from the observer. At this stage it is hard to see how we will ever be able to perceive the sizes and distances of objects veridically. Yet, since we can, the information must be available somehow. We return to this question in Ch. xx [Perception of depth and distance].

Second, the physical world contains all spatial frequencies. But the optical system low-pass filters the incoming visual signal, reducing the contrasts of spatial frequencies above about 5 cy/deg, and imposing a cut-off at about 60 cy/deg. We see the world low-pass filtered, through the "window" of our optical MTF.

As we move from one stage of visual processing to the next in this book, we will try to capture the essential form of the code at each level with a slogan, "in 25 words or less". These slogans are intended to serve as mnemonics for remembering the effects of each stage of processing on the visual code. We are now ready to summarize the effects of optics. As summarized in Figure 5.16, the two most important transformations implemented by the optics render the incoming signal *two-dimensional*, and *low-pass filtered.*

What about the next stage – discrete sampling by the photoreceptors? The artistic style called *pointillism*, shown in Figure 5.17, provides a nice mnemonic for the form in which spatial information is coded by the photoreceptors. In pointillism, the artist represents a visual scene by using discrete dots of paint on the canvas. At the level of the photoreceptors, the visual scene is represented by sets of spatially discrete quantum catches. Notice that there is no signal that binds together the parts of a given object in the scene. (Of course there was no such signal in the retinal image either.) The work of assembling the image into meaningful parts remains to be done by higher levels of the visual system. In sum, discrete sampling makes the incoming signal *pointillistic.*

In Chapter 6 we turn to a more detailed analysis of the photoreceptors. How do they work, and what limits do they place on our vision?

OPTICS: THE FIRST TRANSFORMATION

Physical world ---> Retinal image

Three dimensional --->Two dimensional

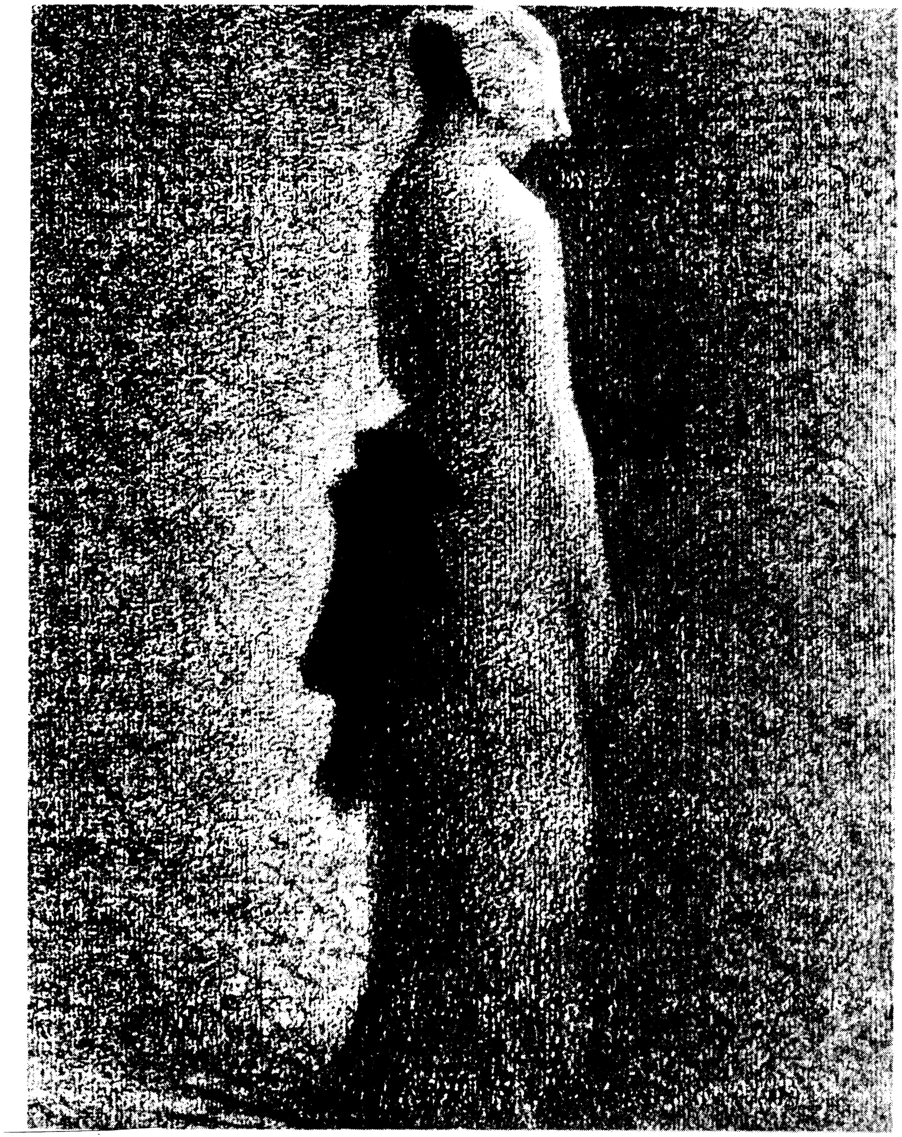
All spatial
frequencies --->Low pass filtered

DISCRETE SAMPLING

Retinal image ---> Photoreceptor inputs

Spatially
continuous --> Discretely sampled, "pointillistic"

Figure 5.16: Optics and discrete sampling recode the visual signal.



The Black Bow, *by Georges Seurat. From Rewald (1954).* © *SPADEM, 1954, by French Reproduction Rights, Inc.*

Figure 5.17: A pointillistic painting: Seurat, The Black Bow. [See copyright information on the painting; via Ratliff (1965), Fig. 7.27, p. 242]

Chapter 6

Photoreceptors and Transduction

In this chapter we leave behind the purely physical aspects of vision – light and optics – and begin a section on retinal physiology. In Chapters 6 and **??** we explore the anatomy, physiology, and function of photoreceptors. In Chapters 8-13 we explore the properties of other retinal neurons. In each case, we first describe the anatomy and physiology of the neurons themselves, and then introduce some of the causal stories of how the characteristics of these neurons leave their marks on the system properties of vision.

Logically speaking, photoreceptors have two major tasks to perform: *phototransduction* and *signal transmission*. The first goal of this chapter is to create at least a qualitative (if not quantitative) appreciation of these two processes.

In the *phototransduction* process, each individual photoreceptor – rod or cone – absorbs quanta of light. A quantum of light – a physical entity – ends its existence, and creates a neural signal. For many vision scientists, including DT, phototransduction has always held a special fascination, because it forms the immediate interface between the physical and physiological worlds. A part of the universe becomes a part of the individual.

The second task of photoreceptors is *signal transmission*. The absorption of quanta occurs in the outer segment of the photoreceptor. But in order to influence vision, the photoreceptor must transmit a neural signal all the way to its synaptic terminal, at which the photoreceptor communicates with later retinal neurons. The processes involved are complex, and the technical details are really available only to those students with a background in biochemistry and cell biology. However, even if you don't have the background to understand these processes, we hope to provide at least an intuitive appreciation of them.

Technical advances have allowed vision scientists to carry out physiological recordings from single living, functioning photoreceptors. At low light levels, these recordings show that, amazingly, the absorption of a single quantum in a rod outer segment creates a physiological signal that is sufficient to affect the output of the rod. At higher light levels, they show that the responses of rods increase with increasing light levels, but eventually saturate; that is, they provide evidence for a saturating non-linearity very early in visual processing.

The second goal of this chapter is to continue our analysis of causal stories. How do the properties of photoreceptors leave their marks on the system properties of vision? We will examine three examples. The first deals with the effects of transduction in the rods on the spectral characteristics and wavelength information losses of scotopic vision. The second deals with the effects of the exquisite sensitivity of rods on scotopic absolute thresholds. The third causal story depends upon cones, and concerns the psychophysical consequences of photoreceptor saturation.

Transduction and signal transmission processes in the cones also have profound consequences for color vision and for photopic spectral sensitivity. However, the color story is too long to tell within the present chapter, and is postponed to Chapter **??**.

6.1 Family portraits: The anatomy of photoreceptors

As shown in the schematic overview of Figure 1.4A, the photoreceptors lie in the *outer* portion of the retina, against the back wall of the eyeball. Quanta of light coming in through the lens traverse several other types of neurons before they arrive at the photoreceptors and are finally absorbed.

6.1.1 Rods and cones

There are two basic kinds of photoreceptors in the eye – *rods* and *cones*. Figure 6.1 shows some family portraits of rods and cones. Figure 6.1A shows a drawing of a primate cone and two primate rods, as seen through a light microscope. Figure 6.1B shows a scanning electron micrograph of two cones and several rods. Anatomists divide each photoreceptor into three basic parts: the *outer segment*, the *inner segment*, and the *synaptic terminal*.

6.1.2 Retinal distributions of rods and cones

The numbers of rods and cones vary across the retina in different ways, as shown in Figure 6.2. Figure 6.2A shows the concentration of rods and cones as a function of retinal eccentricity. The photomicrographs in Figure 6.2B show the varying sizes and densities of the two types of photoreceptors. In the central fovea (eccentricity 0.0), all of the photoreceptors are cones, and (as we already know) the outer segment diameters of foveal cones subtend only about 30 seconds of arc. At the other eccentricities both rods and cones are present, with the cones being increasingly larger in size, and the rods increasingly more numerous.

Figure 6.3 shows a high power electron micrograph of a rod outer segment. It shows a highly specialized structure of tightly packed membranes, called *disks*. There are about 1000 disks per rod outer segment. Each disk is composed of two membranes joined at the ends with a space between, like a stack of pita bread. The whole stack of disks is contained within the outer membrane of the rod outer segment. In cones these structures are slightly different (see the inset to Figure 6.3), in that the ”disk” membranes are actually continuous with the outer membrane of the cell (a cone is like a comb).

These highly organized structures hold the machinery for catching quanta and starting neural signals in the photoreceptors. Because of structural differences and other factors, the details of transduction differ slightly between rods and cones. Our descriptions will apply most strictly to rods, but the differences are small.

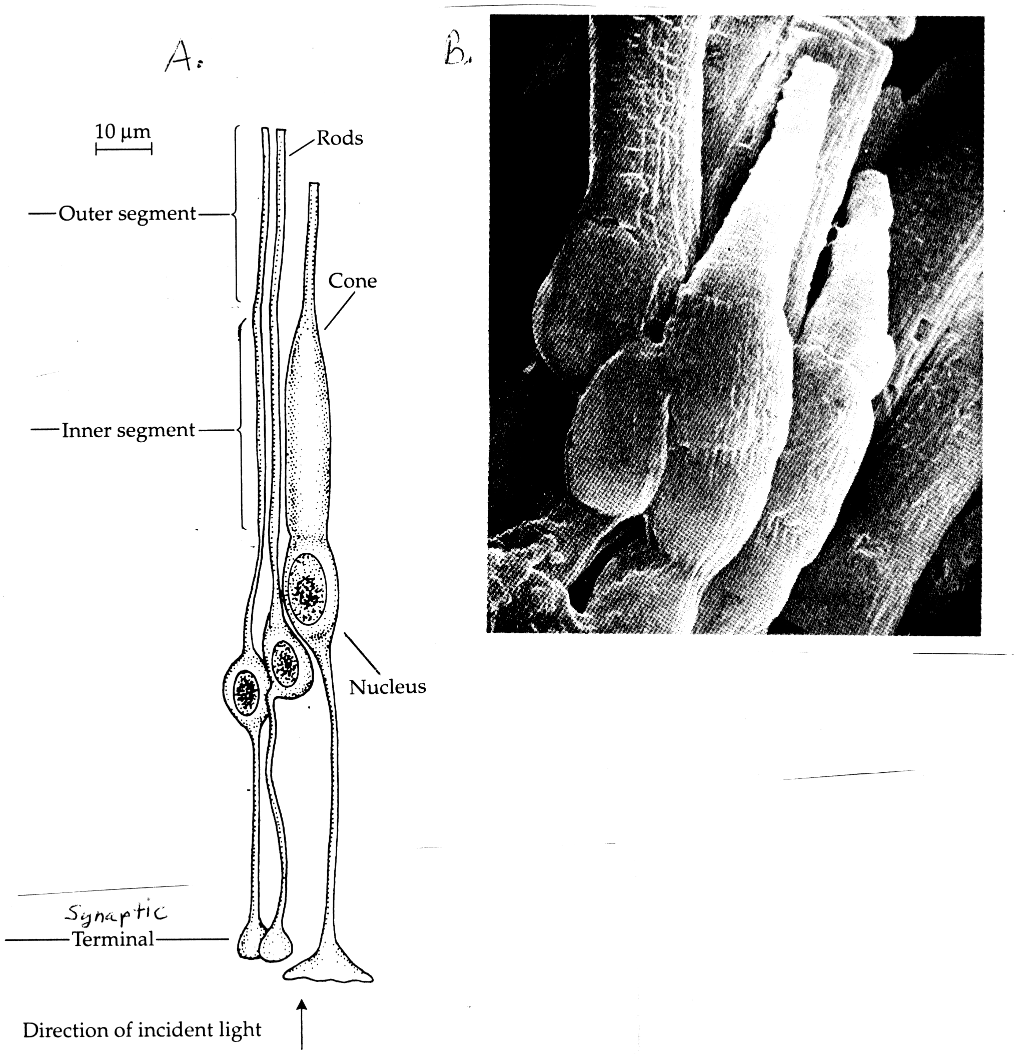


Figure 6.1: Rods and cones. A: Drawings of a rod and a cone as seen under a light microscope. The three basic parts of the photoreceptor – outer segment, inner segment, and synaptic terminal – are shown. The outer segments lie against the pigment epithelium, at the back of the eyeball, and the synaptic terminals lie closest to the center of the eyeball. Light entering the eye passes through the synaptic terminals and inner segments of the photoreceptors before being absorbed in the outer segment. B: A scanning electron micrograph of two cones and several rods. [A from Oyster (1999), Fig. 13.5, p. 550, after Polyak (1941). B from Lewis, Zeevi, and Werblin (1969), via Goldstein (1999), Fig. 2.12, p. 37.]

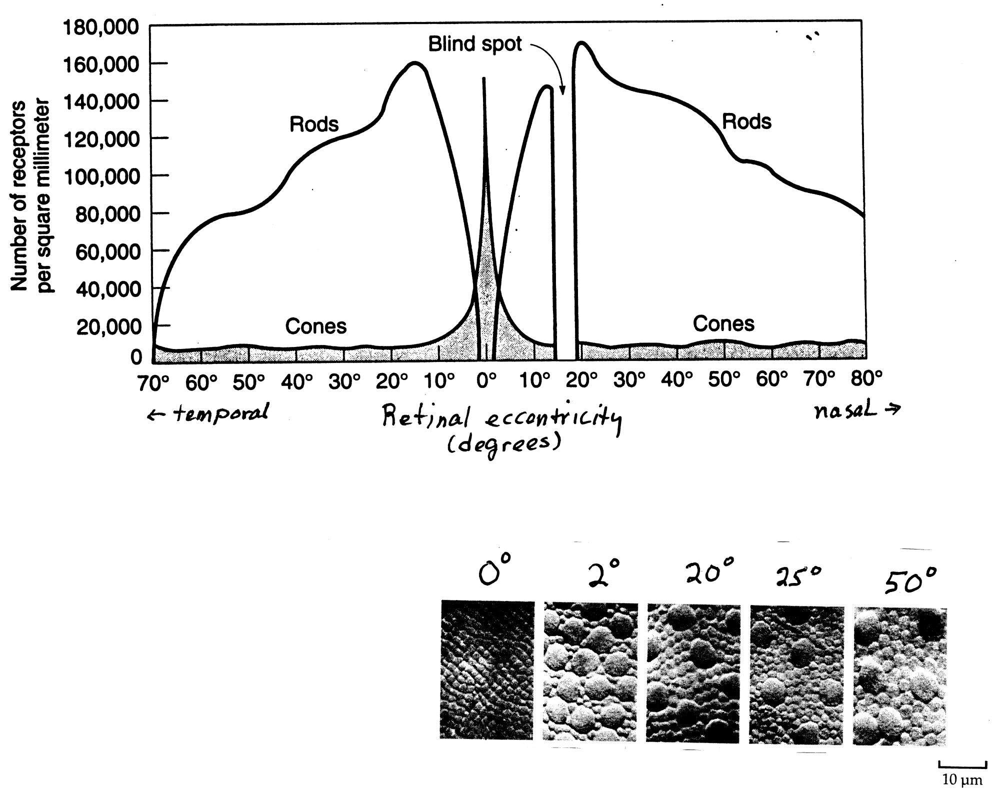


Figure 6.2: Distributions of rods and cones across the retina. A: The concentrations of rods and cones as a function of retinal eccentricity. The concentration of cones is highest at the fovea, drops off rapidly out to about 10^o eccentricity, and remains roughly constant across the rest of the peripheral retina. The rods are absent from the fovea, increase to a maximum concentration at about 15 to 20^o eccentricity, and then taper off toward the far periphery. The blind spot – the region at which the optic nerve exits the eye – contains no photoreceptors of either type. B. Photomicrographs at the level of the inner segment, showing the relative sizes and concentrations of rods and cones at five retinal eccentricities. At zero degrees, in the fovea, all of the photoreceptors are cones. Outside the fovea, the large circles are cones and the small ones are rods. [A adapted from Goldstein (1999), Fig. 2.13, p. 37. B Adapted from Oyster (1999), Fig. 15.12, p. 665. From Curcio et al (1990).]

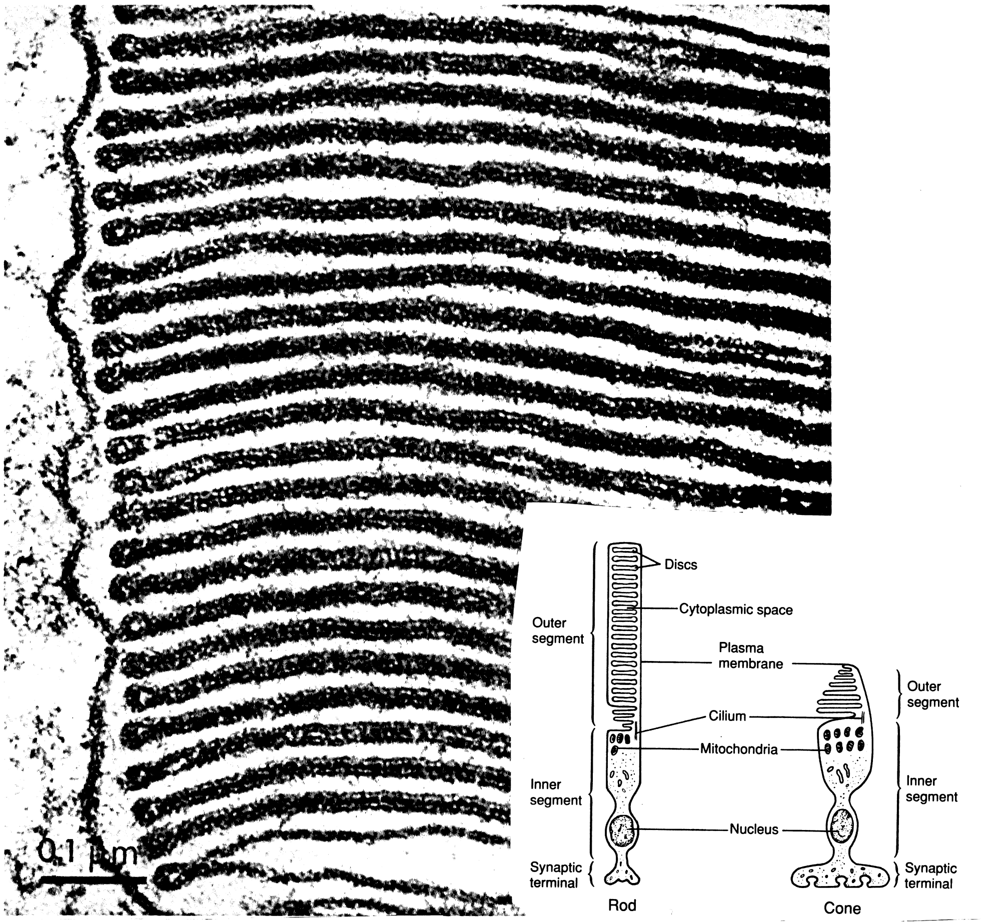


Figure 6.3: Distributions of rods and cones across the retina. A: The concentrations of rods and cones as a function of retinal eccentricity. The concentration of cones is highest at the fovea, drops off rapidly out to about 10^o eccentricity, and remains roughly constant across the rest of the peripheral retina. The rods are absent from the fovea, increase to a maximum concentration at about 15 to 20^o eccentricity, and then taper off toward the far periphery. The blind spot – the region at which the optic nerve exits the eye – contains no photoreceptors of either type. B. Photomicrographs at the level of the inner segment, showing the relative sizes and concentrations of rods and cones at five retinal eccentricities. At zero degrees, in the fovea, all of the photoreceptors are cones. Outside the fovea, the large circles are cones and the small ones are rods. [A adapted from Goldstein (1999), Fig. 2.13, p. 37. B Adapted from Oyster (1999), Fig. 15.12, p. 665. From Curcio et al (1990).]

6.2 Phototransduction

6.2.1 The rod photopigment: Rhodopsin

The substance that absorbs quanta of light in the photoreceptor outer segment is called a *photopigment*. In rods, the photopigment is called *rhodopsin*[1]. Rhodopsin molecules sit tightly packed in the disks of the outer segments of the photoreceptors, as well as in the surrounding membrane that contains them. In mammalian retina, each rod contains about 1000 disks, and each disk contains about 105 rhodopsin molecules, for a total of about 108 rhodopsin molecules per rod outer segment.

The structure of the rhodopsin molecule is shown in Figure 6.4. It has two parts – the *opsin molecule* and the *chromophore*. The opsin is by far the larger part – for those with a background in chemistry, it is a protein composed of 348 amino acids. The molecular weight of the opsin molecule is about 39,000.

As shown in Figure 6.4A, the opsin molecule twists into a helical structure, and loops back and forth through the disk membrane a total of seven times, crossing between the outside and the inside of the disk. As shown in Figure 6.4C, the loops are arranged in a cylindrical conformation, so that each molecule of rhodopsin forms a more or less barrel-shaped structure within the membrane.

The chromophore – also called *retinal* – is by far the smaller part of the molecule. Retinal is the form of Vitamin A commonly found in carrots and other vegetables. Its molecular weight is 285. As shown in Figure 6.4B, the chromophore commonly exists in either of two forms, called *11-cis* and *all-trans retinal*. These are terms used to describe the atomic connections, and hence the three-dimensional shape, of the chromophore. The two states of the chromophore are called *isomers* – the chemical composition of the chromophore remains unchanged, but its three-dimensional shape changes. In the 11-cis configuration, the carbon backbone of the molecule is bent at the 11th carbon atom; in the all-trans configuration, this bend is straightened. The chromophore is attached to the opsin within the membrane, in the middle of the seventh loop (as labeled in the rightmost loop in Figure 6.4A). It lies in wait in the 11-cis configuration in the middle of the barrel, primed for action when a quantum of light arrives.

Now, recall that a quantum is an indivisible packet of energy. It cannot be divided up among different photoreceptors, but can only enter a single photoreceptor and be absorbed by a single rhodopsin molecule. Moreover, absorption is an all-or-nothing event: either a quantum is absorbed or it is not. But it is also probabilistic: when the quantum arrives at the outer segment, there is a certain probability that it will be absorbed by a molecule of rhodopsin. That probability varies with the wavelength of the light (the energy in the quantum). For rhodopsin that probability is maximal at about 500 nm, and it falls off sharply at both shorter and longer wavelengths. If the quantum is not absorbed, it is lost to visual processing.

6.2.2 The mechanism of transduction: Cis-trans isomerization

So far, so good. But how does light act on the rhodopsin molecule? When a quantum is absorbed into the structure of a rhodopin molecule, the energy of the quantum is used to excite an electron, and the decay of the excited electron leads to a change in the conformation of the chromophore. In other words, the only thing light does in the entire visual process is to trigger a change in the

[1] In an eye that is fully adjusted to the dark, the retina has a rosy appearance – hence the name *rhodopsin* (= red vision substance) for the rod photopigment. When the eye is exposed to high levels of light, the retina changes color – it loses its color, or "bleaches". In vision jargon, light is said to *bleach* photopigments.

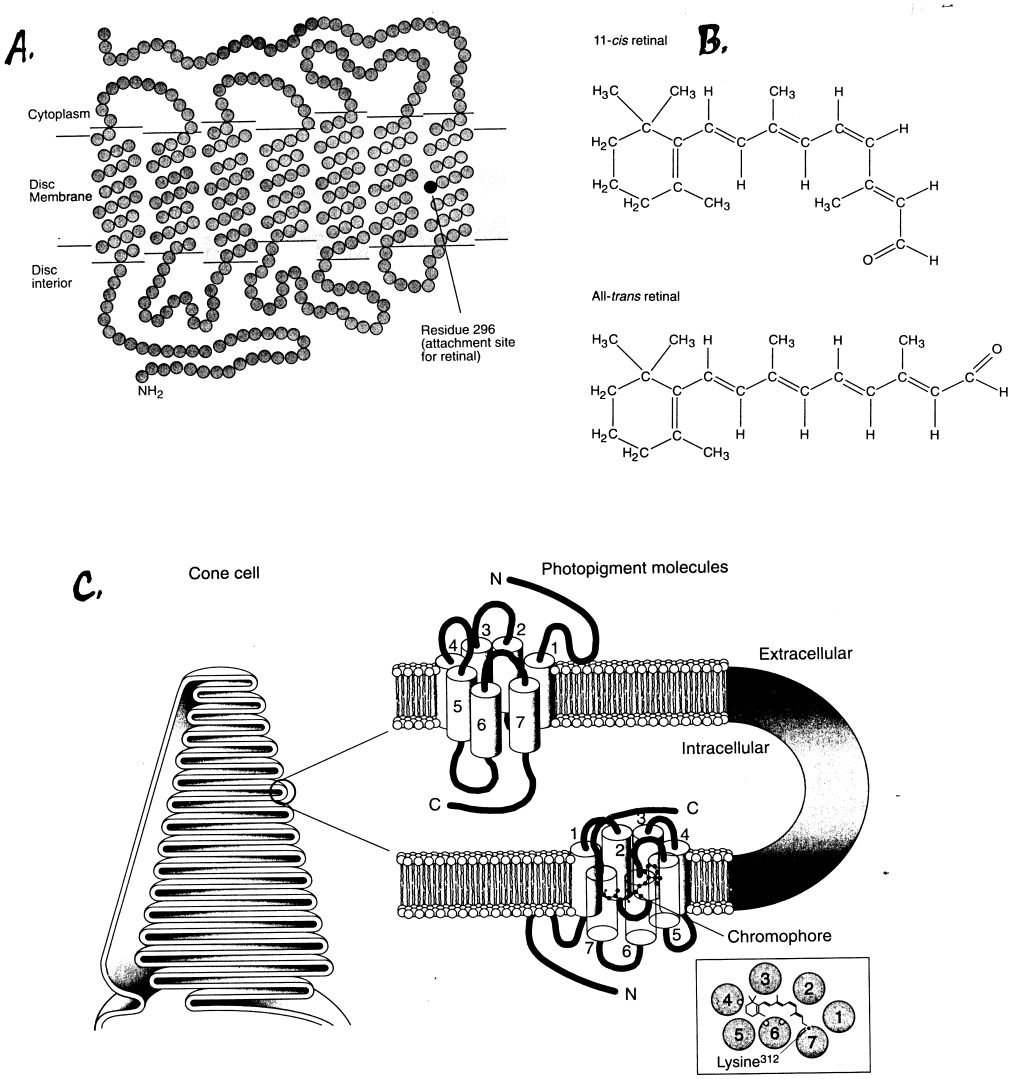


Figure 6.4: The rhodopsin molecule and the disk membrane. A: The structure of rod opsin. The opsin twists into a helical structure, and loops back and forth through the disk membrane seven times, with one end outside the disk and the other inside it. B: The chromophore in its 11-cis and all-trans configurations. C: The three-dimensional shape of a different opsin, this time in a cone membrane, showing the 11-cis chromophore inside the "barrel". [A and B modified from Kandel et al, 1991, Fig. 28-4, p. 405. C from Sharpe et al (1999), Fig. 1.2, p. 6.]

shape of the chromophore. Moreover, the change of shape is always the same, regardless of the wavelength of light that has been absorbed[2].

After the chromophore absorbs the quantum, the rhodopsin molecule goes through a series of very rapid changes in three-dimensional shape, finally including separation of the chromophore from the opsin. With the help of enzymes located in the pigment epithelium, the chromophore eventually gets changed back into 11-cis retinal, and rejoins the opsin in the ultimate recycling process. On average, it takes several minutes for a rhodopsin molecule to reform.

6.3 Signal transmission

As we said earlier, the transduction process is the first of two tasks that each photoreceptor needs to accomplish. The second task is signal transmission: the photoreceptor must create and transmit a signal, passing the information that the quantum has been absorbed, all the way from the rhodopsin molecule to the synaptic terminal.

Photoreceptors are neurons, but it turns out that they are very atypical neurons. In particular, the more typical neurons, of which students have often heard, have axons and fire action potentials (spikes). To work through the properties of photoreceptors, it will be useful to be able to compare them to the properties of typical neurons. In this section, we first review the properties of typical neurons, and then proceed to the unusual properties of photoreceptors.

6.3.1 A typical neuron in a nutshell

Figure 6.5 shows the anatomy and physiology of a typical neuron. A portrait of a typical neuron, with its *dendrites*, *cell body*, and *axon*, is shown in Figure 6.5A. The direction of signal transmission is in through the dendrites, across the cell body and out the axon. The axon ends in a set of structures called the *axon terminals*, and it contacts the dendrites of the next set of neurons across intercellular spaces called *synapses.*

How does a typical neuron work? As shown in Figure 6.5B, in its resting state a typical neuron sets up a small electrical voltage across its outer membrane. It does this by maintaining different concentrations of ions with different electrical charges inside vs. outside the cell membrane. Part of the reason for the charge difference is that nerve cell membranes are *semipermeable.* That is, like the filter in a coffee maker, the openings (channels) in the membrane pass particles of a certain size and electrical charge, and exclude others. So for each kind of channel in the membrane, some chemical ions can enter and leave whereas others cannot. We will be most concerned with the sodium permeable channels.

In addition to the semipermeable membrane, all neurons also have a *sodium-potassium pump* – an active transport mechanism that pumps sodium ions (Na^+) out of the cell and potassium ions (K^+) in. Because more sodium ions are pumped out than potassium ions are pumped in, the net charge is negative on the inside with respect to the outside; in the jargon, neurons in their resting state are *hyperpolarized.* For most neurons, the resting membrane potential – the magnitude of the

[2] Here's a way to get an intuitive feel for the rhodopsin molecule and the isomerization process. Hold your left hand vertically in front of you, with the fingers and thumb spread apart and shaped into a barrel (pretend you have seven digits). Put your right thumb inside the barrel, and bend it at the knuckle. This is the 11-cis configuration. To absorb a quantum, flip your thumb from bent to straight – the all-trans configuration.

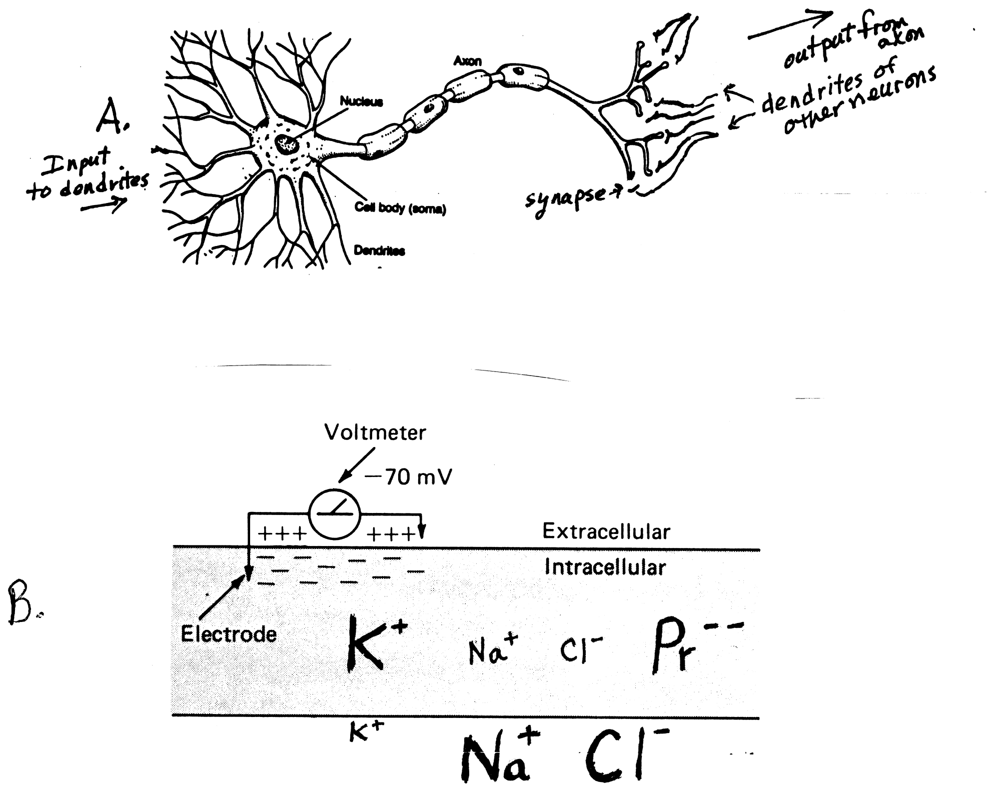


Figure 6.5: A typical neuron. A: Anatomy of a neuron: dendrites, cell body, axon, and synaptic terminals. B. A typical neuron maintains different concentrations of different ions (charged particles) on the inside vs. the outside of its cell membrane. The inside contains relatively high concentrations of potassium (K+) and proteins (Pr–); the outside contains relatively high concentrations of sodium (Na+) and Chloride (Cl-). These charge differences create a resting potential – an electrical voltage – of -70 mV between the inside and the outside of the cell. [A after Levine and Shefner (1991), Fig. 3-11, p. 36; B after Levine and Shefner (1991), Figs. 3-3 and 3-4, p. 38.]

charge – is about -70 mV, with the minus sign indicating that the *inside* of the cell is negative with respect to the *outside*.

How do typical neurons process incoming signals? When a neuron receives a signal across a synapse from a neuron earlier in the causal chain, the incoming signal perturbs the charge across the membrane in the vicinity of the input synapse. These perturbations, called *slow potentials*, can be either depolarizing (excitatory) or hyperpolarizing (inhibitory). They spread passively along the dendrites and the cell body, with decreasing effect the greater the distance from the input site. Slow potentials from many input sites combine their positive and negative effects across the dendrites and cell body of the neuron.

At the base of the axon there is a specialized location called the *axon hillock*. When the neuron is *depolarized*, so that the voltage across the membrane of the cell is sufficiently *decreased* at the axon hillock, the properties of the axon membrane suddenly change. The mechanisms for this change are well understood, but beyond the scope of this chapter. Suffice it to say that the end product is an *action potential* or *spike* – a brief, all-or-none wave of *depolarization* that travels rapidly along the axon, propagating itself all the way to the synaptic terminal.

A reasonable analogy to a spike traveling down an axon is a flame travelling down a long match. The act of striking the match starts a flame at one end. Each segment of the match that burns provides the energy to ignite the next segment, so that the flame travels all the way down the match to its end. The analogy would be even better if the segments of the match regenerated themselves after the flame had passed, to be ready for the next traveling flame.

At the synaptic terminal, the depolarization brought about by the arrival of the spike produces an increase in the release of a *neurotransmitter* – a chemical substance specialized to transmit signals across the synapse. The neurotransmitter in turn creates slow potentials in the dendrites and cell bodies of postsynaptic neurons. Any given postsynaptic neuron can receive and combine inputs from thousands of presynaptic cells. The process is repeated countless times throughout the complex neural network of the brain.

In sum, whenever a neural signal must travel long distances, the signal is carried by the patterns of spikes in the axons of typical neurons. We will return to typical neurons in Chapter 8 xx, when we explore the properties of ganglion cells – the neurons whose axons carry information from the eye to the brain.

6.3.2 Photoreceptors are not typical neurons

In the meantime, we return to photoreceptors. Vertebrate photoreceptors are extremely atypical neurons – in fact, as we will see, most of their properties differ from those of a typical neuron. Most importantly, photoreceptors do not have axons, and do not produce spikes; they transmit messages only with slow potentials.

6.3.3 Photocurrents

The unique physiological properties of photoreceptors are shown in Figure 6.6. The first notable difference between a photoreceptor and a typical neuron is that the resting potential of a photoreceptor is only about -40 mV rather than the -70mV seen in the typical neuron. Why is this so? As with typical neurons, there are sodium permeable channels in the outer membrane of the outer segment of the photoreceptor (shown in Figure 6.6B). In the dark, these channels are open, and they allow sodium to leak in. Meanwhile, the inner segment is permeable to potassium ions.

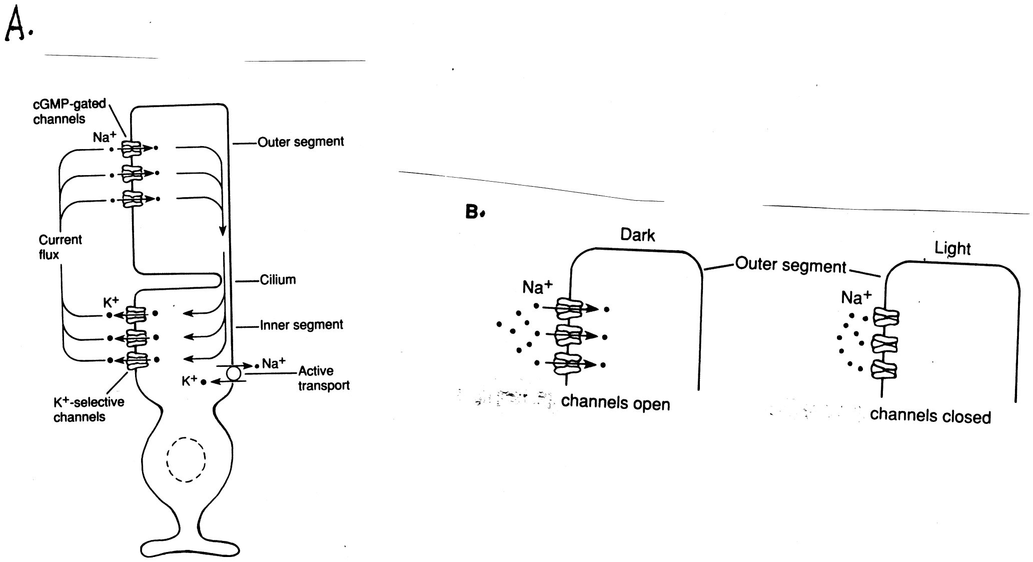


Figure 6.6: The dark current. A. The arrows show the current that flows in darkness, when the sodium permeable (cGMP gated) channels in the outer segment are open. B. In the light the channels close, and sodium ions are excluded from the outer segment. [Kandell, Schwartz, and Jessell (1995), Fig. 22-5, p. 415.]

This state of affairs produces an electrical current – a flow of ions – from the outer to the inner segment inside the cell, and back again on the outside of the cell (Figure 6.6A). This continuous depolarizing current, called the *dark current*, results in the membrane potential of about -40 mV (a *depolarization* compared to the -70 mV seen in a typical neuron).

In the dark, the constant depolarization produces a constant release of transmitter from the synaptic terminal. This property of photoreceptors is actually consistent with the corresponding property of typical neurons, in which depolarization (at the axon hillock) produces an increase in the release of transmitter (at the synaptic terminal, at the far end of the axon). In summary, in the dark, the resting membrane potential is about -40 mV, the photocurrent flows continuously, and the photoreceptor continuously releases transmitter from its synaptic terminal.

What happens when light is absorbed? It turns out that (for reasons discussed immediately below) sodium channels in the outer segment close, excluding sodium ions, and thus reducing the dark current. In consequence, the cell *hyperpolarizes* toward -70 mV; and the release of transmitter that occurred continuously in the dark is *reduced* by the action of light.

We should pause for a minute to consider this mechanism of action. Intuitively, most of us would probably have guessed that increased light on the photoreceptor would yield an increase in transmitter release from the photoreceptor. But in fact the opposite is the case – increased light absorption in the photoreceptor yields a *decreased* signal from the photoreceptor. Is this a logical problem? Not really. The fact that the signal for an increase in light level is a decrease of transmitter (and vice versa) is logically perfectly OK. It is the *change* of transmitter release with the change of light level that matters, not the absolute direction of change.

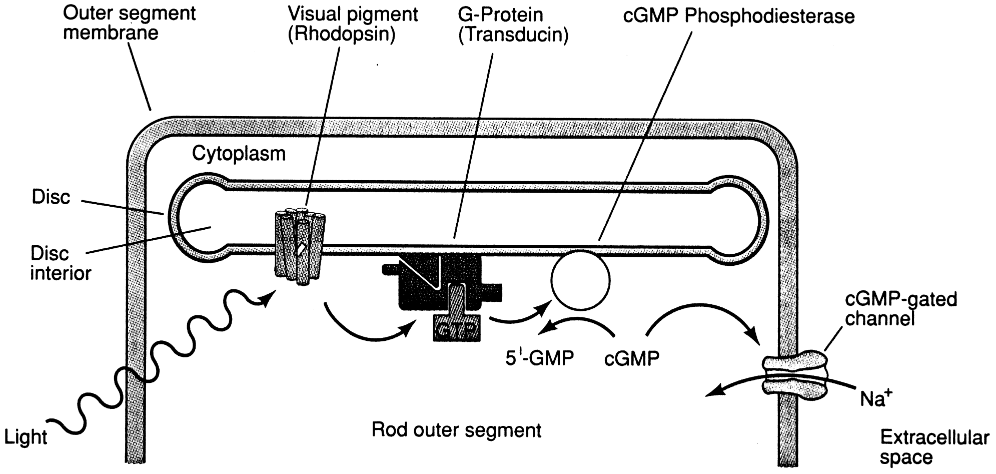


Figure 6.7: The chemical cascade. In the dark, cyclic GMP (cGMP) keeps the sodium permeable (cGMP-gated) channels open, so that sodium ions pass into the outer segment, and the dark current flows. When a quantum is absorbed, resulting in isomerization of a molecule of rhodopsin and a set of conformational changes in the opsin molecule, a protein called transducin becomes activated. Transducin in turn activates an enzyme called cGMP phosphodiesterase in the disk membrane, and cGMP phosphodiesterase hydrolizes cGMP to 5'GMP. 5'GMP cannot hold the sodium permeable channels open, so they close. Sodium ions are excluded from the outer segment, and the dark current slows. The cell hyperpolarizes and the release of transmitter is reduced, signaling to the next cell that a quantum has been absorbed. [Kandel, Schwartz, and Jessel (1995), Fig. 22-3, p. 412.]

6.3.4 The chemical cascade

Returning now to the outer segment: what happens after a photon is absorbed? In particular, how does the message that light is absorbed get from the rhodopsin molecule to the outer membrane of the cell, and bring about a closing of the sodium channels? To make a long and highly technical story short, the absorption of a quantum triggers a complex series of biochemical changes, called the *chemical cascade*, that results in the closing of sodium permeable channels in the outer membrane of the outer segment of the photoreceptor.

For those with some biochemical background, the details of the chemical cascade are shown in Figure 6.7. From our perspective, the bottom line is that the chemical cascade produces an enormous *amplification* of the signal. Absorption of a single quantum in the outer segment of a rod results in the closing of several hundred sodium permeable channels, and each closed channel blocks the entry of as many as 10,000 sodium ions per second into the rod's outer segment. As we will see, the change in sodium flux is so large that it causes a detectable change in the charge across the photoreceptor membrane, as well as in the output of the photoreceptor.

6.4 Physiological responses recorded from rods

In the 1970s, a remarkable new technique was developed: the *suction electrode*. A suction electrode preparation is shown in Figure 6.8. Using an excised retina, it is possible to draw the outer segment of a rod or a cone into a closely fitting hollow glass electrode. The membrane current that would ordinarily flow along the outside of the cell then flows inside the electrode, and changes in the current can be measured. By shining lights of various intensities and wavelengths on the outer segment of the photoreceptor within the electrode, one can record the changes in current flow in response to light, all the way down to the responses to absorption of an individual quantum. This work was extended to primate photoreceptors, including human photoreceptors, in the early 1980's.

6.4.1 Low light levels: Responses to single quanta

Suppose that a rod outer segment has been drawn into a suction electrode, and the current flowing through the photoreceptor is being recorded. Now suppose that the experimenter produces a series of flashes of light so dim that on average only a single quantum of light will be absorbed by the photoreceptor. Because of the quantal nature of light, the actual number of quanta absorbed will vary from one flash to the next – sometimes zero, sometimes one, sometimes two, and occasionally more than two. If the response of the photoreceptor to a quantal absorption is consistent and repeatable, then the change in current on each flash should take one of only a few stereotyped forms, corresponding to the absorption of zero, or one, or two, or (occasionally) higher numbers of quanta.

The results of such an experiment are shown in Figure 6.9A. The upper tracing shows the current recorded as a function of time. The tick marks under the tracing show the times at which the flashes were nominally delivered – a little more than one flash every 10 seconds. Notice that the response of the cell is variable from one flash to the next, as predicted. Changes in the membrane current occur on some but not all trials; and when they occur, they are usually of a stereotyped form, either small or large in size. In sum, and remarkably, rod photoreceptors can indeed initiate a measurable signal from the absorption of a single quantum of light[3].

Similar experiments show that the photocurrent produced in a rod by a single quantal absorption is the same regardless of the wavelength of the quantum. Figure 6.9B shows responses to flashes of 550 and 659 nm lights that each produced one quantal absorption.

6.4.2 Higher light levels: A saturating non-linearity

What about the photoreceptors' responses to multiple photons at higher light levels? Figure 6.10 shows photocurrents measured from dark adapted primate rods and cones, measured with suction electrodes. The traces in Figure 6.10A show current flow in response to lights of varying radiances. For both rods and cones, as light levels increase, the amplitude of the current decreases. The time course of the response also changes with the light level. For rods, as the quantal catch increases, latency decreases. The cone response is biphasic but shows the same trend. Eventually, at high

[3] In cones as in rods, single quanta are caught by single photopigment molecules, and single quantal absorptions must initiate functional physiological signals. But a rod produces a much larger signal than a cone does in response to the absorption of a single quantum. Some estimates suggest that the difference in the magnitude of current produced is as much as 100/1. Thus, the magnitudes of the cone responses to individual quanta are too small to measure, even with suction electrodes.

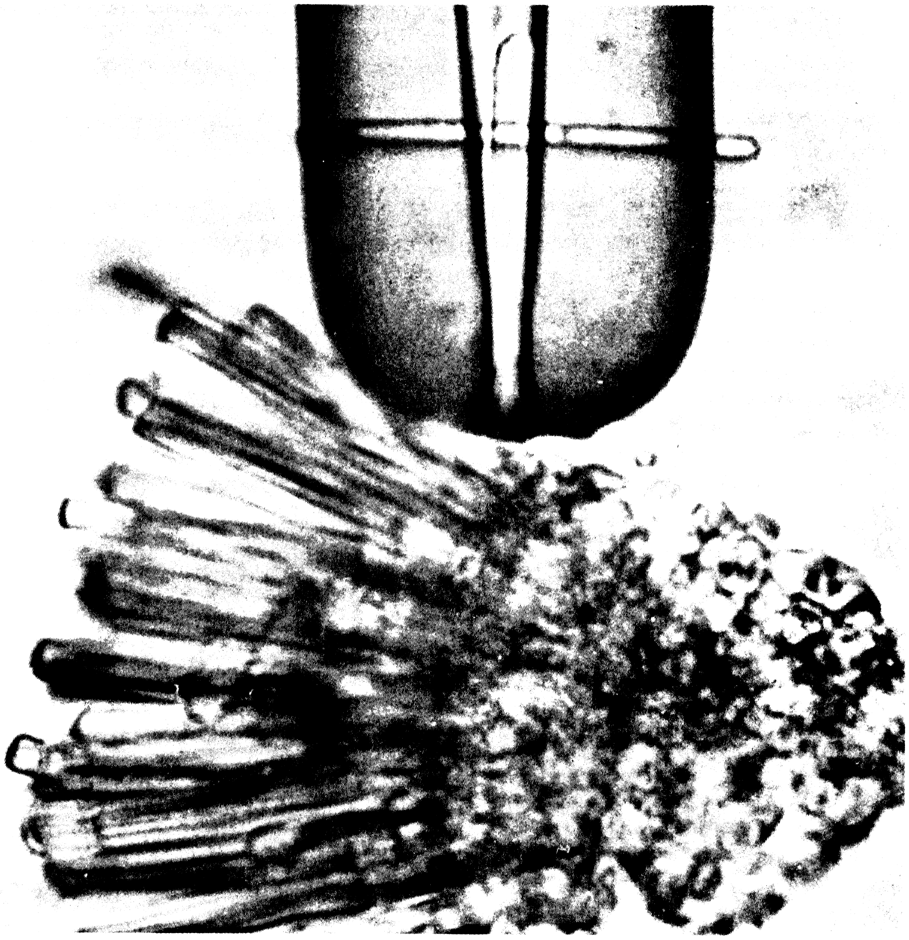


Figure 6.8: The suction electrode technique. The photomicrograph shows a piece of retina from a toad (*Bufo marinus*). The outer segments of the rods are at the left. One of the outer segments has been sucked into the electrode so that its membrane current can be recorded. The horizontal bar across the electrode is the stimulus – a bar of light used to deliver quanta to the rod. [Baylor (1987), Fig. 4, p. 36.]

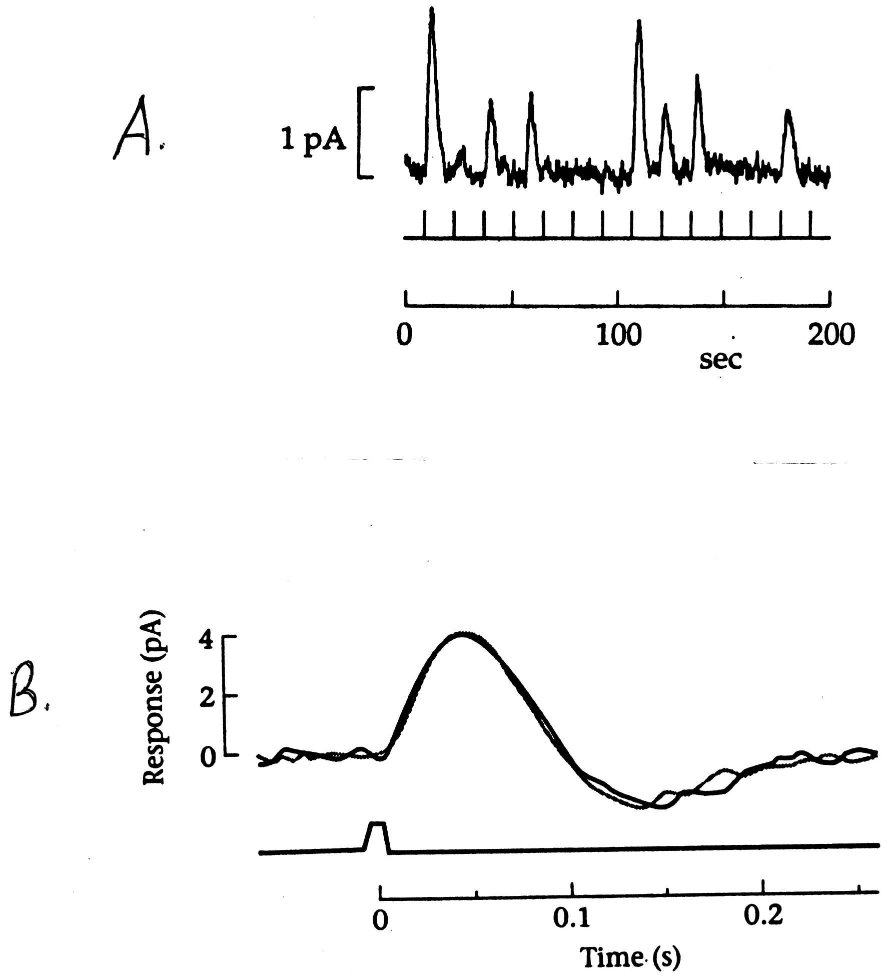


Figure 6.9: The responses of rods to very dim flashes of light. A: Responses to a series of very dim flashes spread over a 200 second period. The ordinate shows the change in the photoreceptor current in picoamps (pA; 1 picoamp = 10_{-12} amperes). The responses of the photoreceptor come in three sizes, corresponding to the absorption of zero, one, or two quanta from each particular flash. The response to one isomerization is about one pA. B: The response to capture of single quanta, on an expanded time scale. The responses to quanta of different wavelengths, such as the 559 and 659 nm cases shown here, are identical. [A from Rieke and Baylor (1998), Fig. 4, p. 1030; B from Baylor et al, (1987), via Wandell (1995), Fig. 4.17, p. 92.]

enough light levels all of the sodium channels are closed, the current is reduced to zero, and no further changes in current amplitude can occur.

Figure 6.10B shows the peak change in the membrane current of rods and cones as a function of the number of quantal absorptions. In each case, the size of the response increases with the number of quanta absorbed, but the response eventually saturates. That is, photoreceptors show a *saturating non-linearity.*

The *dynamic range* of the photoreceptor is defined as the range of inputs over which the photoreceptor's output changes. Dynamic ranges for both rods and cones are shown in Figure 6.11. Defined in terms of the peak change in current flow, the dynamic ranges of both the rod and the cone cover about a factor of 100, or two log units: from about 1 to about 100 quanta absorbed for rods, and from about 200 to 20,000 for cones. We will see in Chapter 10 that these descriptions strictly apply only to the dark-adapted rod and the dark-adapted cone, and things become more complex when light adaptation processes are included.

6.5 Three causal stories

In this section we discuss three causal stories concerning how photoreceptors leave their marks on perception. The three stories illustrate the effects of our three major properties of photoreceptors. The first rests on the properties of transduction; the second on signal transmission; and the third on the saturating nonlinearity. The first two rest on the properties of rods; the second, on the properties of cones.

6.5.1 Transduction and wavelength information

In Chapter 2, we introduced two major system properties of scotopic vision. First, scotopic spectral sensitivity varies with the wavelength of light, with a maximum at about 500 nm and a specified, sharp falloff of sensitivity to either side. And second, the ability to preserve wavelength information is lost in scotopic vision – lights of all wavelengths look whitish, and they can be matched psychophysically to one another – made indiscriminable – simply by adjusting relative light levels. Some people see these two properties as intuitively contradictory – how can sensitivity vary with wavelength if wavelength information is lost? In fact, the transduction process precisely accounts for both.

Figure 6.12 shows a comparison of the scotopic matching curve of human subjects to the absorption spectrum of rhodopsin. In this figure, the psychophysical data have been corrected for the differential absorption of light of different wavelengths by the optics of the eye (Figure 4.8). Both curves have their maxima at just about 500 nm, and fall off virtually identically, both at shorter and at longer wavelengths. The fit between the two curves is remarkable, particularly given that one data set is psychophysical and the other biochemical. In short, the absorption spectrum of rhodopsin – the probability of a cis-trans isomerization as a function of wavelength – perfectly predicts the shape of the psychophysical spectral sensitivity curve.

At the same time, the transduction process discards wavelength information. When a rod absorbs a quantum, an isomerization occurs, but the effect is exactly the same regardless of the wavelength of the quantum. It follows that the rod has *equivalence classes* – sets of stimuli which, even though they are physically different from each other, are rendered identical by the transduction process. Any set of stimuli that lead to equal numbers of quanta caught are in an equivalence class

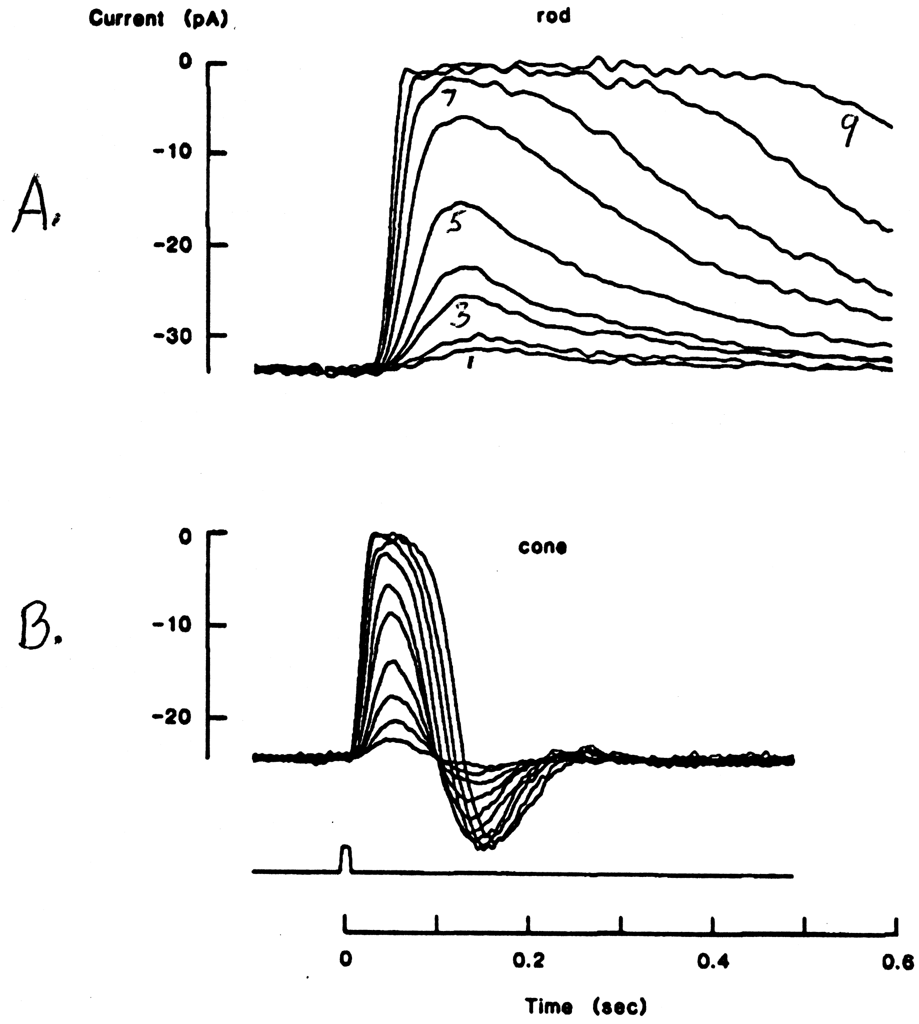


Figure 6.10: Saturation in rods and cones. A: Superimposed responses to flashes of light of increasing intensity, recorded from a monkey rod with a suction electrode. The average number of quanta absorbed per photoreceptor per flash increases by a factor of two from each trace to the next. In the top panel, for traces 1-7, the higher the intensity the higher the peak membrane current. For traces 7-9, the peak response shows little if any additional increase; these traces reveal physiological saturation in the rod. B: A similar trend can be seen in the responses of cones. [After Baylor (1987), Fig. 11, p. 42.]

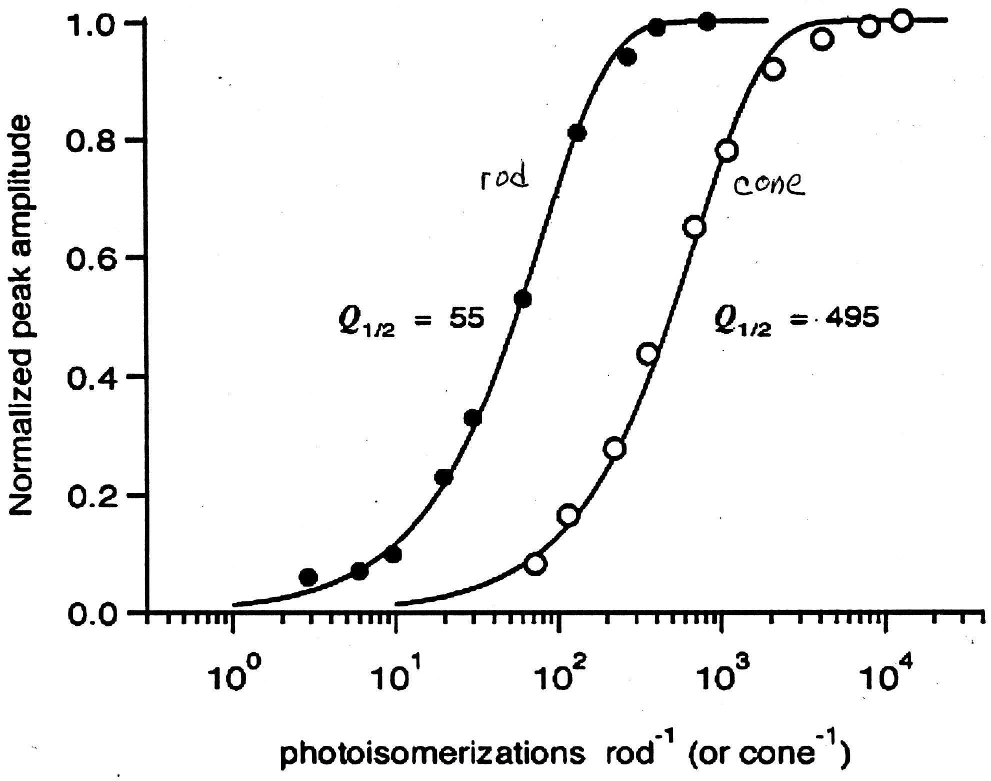


Figure 6.11: Dynamic ranges of rods and cones. Peak responses of a rod (left) and a cone (right) to flashes of light of increasing intensity. The abscissa shows the number of photoisomerizations per rod or per cone. $Q_{1/2}$ is the light level required to produce a half-maximal response; this is about 55 quanta for rods and 495 quanta for cones. Both rods and cones show saturating nonlinearities. The range over which the cell's response changes with changes in the light level is called the dynamic range of the cell. [Schnapf, personal communication.]

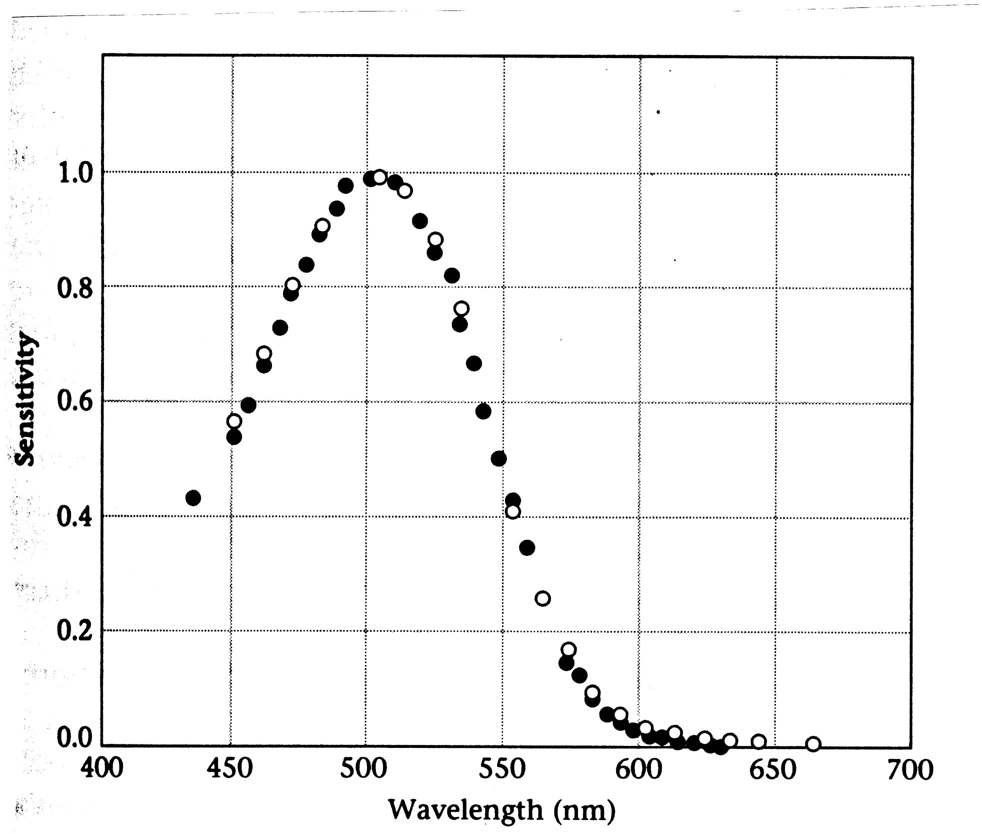


Figure 6.12: The dark adapted human spectral sensitivity curve and the absorption spectrum of rhodopsin. The psychophysical data have been corrected to compensate for differential absorption by the lens of light of short wavelengths. The match of these two curves, together with the theoretical account of how rods work, and the Nothing Mucks It Up proviso, provide a satisfying causal story, explaining a set of psychophysical facts on the basis of a set of facts about the visual system. [From Wandell (1995), Fig. 4.9, p. 79, after Wald and Brown (1956).]

for the rod. It is these quantal equivalences that lead to the suprathreshold discrimination failures seen in scotopic vision. So these two properties of transduction – the variation of the probability of quantal absorption with wavelength, and the loss of wavelength information at the instant of quantal absorption – are exactly sufficient to model the two system properties of scotopic vision – the shape of the spectral sensitivity curve, and the existence of equivalence classes.

By what criteria do we evaluate the quality of a causal story? The more fully established the facts at both levels, the better the match of details between the two sets of facts, the fewer the free parameters, and the fewer the reasonable alternative explanations, the more compelling the causal story. In this case, both the rhodopsin spectrum and the psychophysical spectral sensitivity curve are known from direct measurements, and equivalence classes exactly like the ones originally discovered psychophysically can be demonstrated physiologically by recording rod signals to light of different wavelengths (Figure 6.9B). The story fits together perfectly, with no questionable assumptions and no free parameters. In short, this is arguably the most compelling causal story in vision science.

6.5.2 Signal transmission and absolute thresholds

In 1942, Hecht, Schlaer, and Pirenne carried out a psychophysical experiment on absolute detection thresholds. The stimulus was a test spot that subtended 10' of visual angle. It was placed 20 degrees eccentric to the fovea, near the region of maximum density of rod photoreceptors (Figure 6.2). After the subject adjusted fully to the dark, detection thresholds were measured using the Yes/No method of constant stimuli.

Hecht and his colleagues then made careful calibrations of the light source, and combined these with estimates of the fractions of quanta that are lost within the eye vs. absorbed by the rod photoreceptors. Based on these calculations[4], Hecht et al concluded that the subjects could detect the test spot when only a total of 5-10 quanta were absorbed by the whole set of photoreceptors covered by the test spot. Going back to the information retention proposition we discussed in Chapter 1, this system property implies that the absorption of 5-10 quanta by a set of neighboring photoreceptors is sufficient to initiate a signal that traverses every stage of processing in the visual system.

Moreover, comparisons to retinal anatomy showed that the 10' test spot covered several hundred rod photoreceptors. Hecht et al calculated that with only 5-10 quanta required for detection, the probability was very low that any one rod would have absorbed two or more quanta. Thus, they also concluded that the absorption of a single quantum must be sufficient to make a detectable signal in an individual rod[5]. Half a century later, this prediction was confirmed by direct physiological measurements, as we know from Figure 6.9 above.

[4] More complicated arguments supporting the same conclusions were also made by comparing the slopes of the actual psychometric functions to slopes predicted on the basis of quantal fluctuations. These arguments are presented in elegant detail by Cornsweet (1970).

[5] The claim of such exquisite sensitivity was not immediately accepted by all vision scientists. DT and a friend both went to graduate school in the early 1960's, DT in psychology and the friend in biochemistry. It turned out that we were both assigned to read the Hecht, Shlaer, and Pirenne paper (for the friend it was part of an examination question on cell biology). The friend argued that the energy in a quantum was not sufficient to make a signal that could traverse the whole rod photoreceptor, and therefore that Hecht et al's conclusion must be wrong. DT maintained, on the basis of a bumblebees-can-fly argument, that a single quantum must be sufficient; and that new mechanisms of photoreceptor function, consistent with the psychophysical data, must remain to be discovered. The photocurrent and the amplification provided by the chemical cascade eventually resolved the argument.

The causal story, then, is that the detection of lights that yield a quantum catch of only 5-10 quanta over a 10' field is mediated by the exquisite sensitivity of individual rod photoreceptors, plus of course the preservation of this signal throughout the visual projection system. To psychophysics chauvinist DT, the fun of the story is that the psychophysics came first, and yielded a strong prediction about the sensitivity of the individual rods. Direct verification of the exquisite sensitivity of rods was a much sought after goal in retinal physiology for many years. Now that the prediction has been confirmed, this property of photoreceptors changes its status from deduced to observed.

6.5.3 Photoreceptor non-linearities and the detection of gratings

Finally, are there psychophysically detectable effects produced by the saturating nonlinearities seen in photoreceptors? In Chapter 5, we introduced the experiment on the neural CSF executed by He and MacLeod (1996). Remember that He and MacLeod analysed the interoferometric CSF into three regions: a veridically perceived region below about 60 cy/deg, attributed to the ordinary mechanisms of pattern vision; a high spatial frequency region above 60 cy/deg, attributed to alias patterns; and an intermediate region in the neighborhood of 60 cy/deg, attributed to a putative compressive non-linearity, with the explanation for the latter postponed to the present chapter. We now return to this topic.

The model proposed by He and MacLeod is shown in Figure 6.13. Figure 6.13A, at the bottom, shows a diagram of luminance across space for two stimuli: a uniform field and a sinusoidal grating with the same space-average luminance. Figure 6.13B shows a compressive nonlinearity, with the output growing more and more slowly as the input increases. The thin dashed lines projected upward from Figure 6.13A represent the minimum and maximum luminances of the grating, and the luminance of the uniform field. Note that on the Input axis of Figure 6.13B, the minimum and maximum differ equally from the average. Thus, shifting between the uniform field and the grating should produce no change in the space average neural signal.

The thin dashed lines projected rightward from Figure 6.13B show the levels of output for the minimum, maximum and average luminances after passing through the compressive nonlinearity. Because of the nonlinearity, the signal strength for the maximum is closer to the average signal strength than is the signal strength for the minimum. Figure 6.13C shows the result: the space average signal from the grating (shown by the heavy dashed line) is now smaller than the signal from the uniform field (shown by the heavy solid line). In sum, He and MacLeod argue that a compressive nonlinearity in the photoreceptors can produce a spatially uniform change in the magnitude of the neural signal arising from the stimulus field. This change could underlie the detection of non-resolvable gratings in the range of spatial frequencies around 60 cy/deg. In sum, this argument provides us with a theoretical account of the heavy solid line that bridges the Nyquist frequency in Figure 5.14.

This causal story has a couple of minor glitches. First, the nonlinearity that has been seen physiologically in cones is saturating rather than compressive (Figure 5.1B). But either one will work in the He-MacLeod model, provided that the luminance values are well chosen. Second, it turns out that there are later stages of retinal processing at which the signals from individual cones remain separated from each other (as we will see in Ch. xx), and these stages could in principle provide alternative loci for the nonlinearity. But in the meantime, the He-MacLeod model, modified to use a saturating nonlinearity, provides a convincing account of interoferometric detection thresholds in the vicinity of the Nyquist limit.

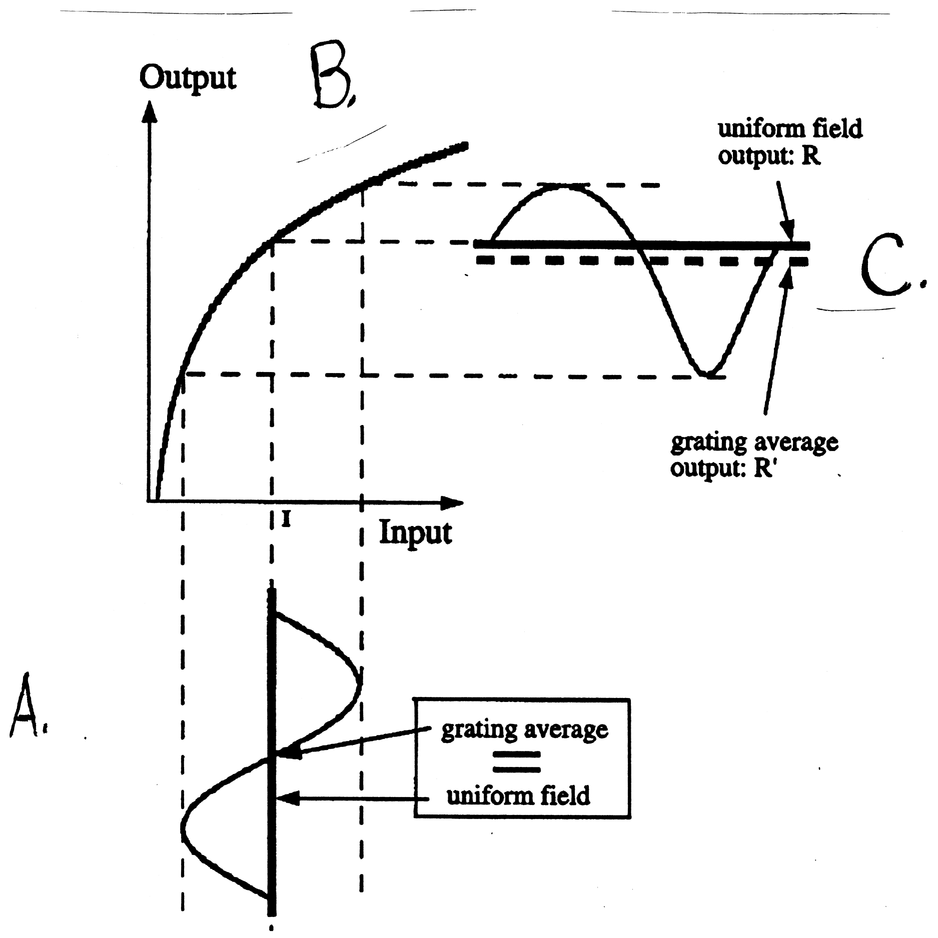


Figure 6.13: The effect of a compressive nonlinearity. A (bottom): A sinusoidal grating, with symmetrical variations of intensity above and below the mean value. The input grating has the same average intensity as a uniform field. B. The grating and the uniform field both pass through a compressive non-linearity. C. The average output for the grating (R') is less than that of the uniform field (R). When the uniform test field is replaced by the grating, a compressive nonlinearity can yield the perception of a still-uniform test field, with a change in intensity at the moment of transition. [From He and MacLeod (1996), Fig. 2, p. 1140.]

6.6 Linking Propositions

As far as DT can see, there are no new linking propositions involved in these three causal stories. All three psychophysical experiments depend only on either threshold or matching judgments, and all three causal stories therefore rely only on Identity propositions. The first story, on spectral sensitivity and equivalence classes, also includes an Analogy between the psychophysical spectral sensitivity curve and the rhodopsin spectrum (Figure 6.12).

6.6.1 The Nothing Mucks it Up proviso

However, we now want to introduce another important linking proposition, having to do with boundary conditions. Philosophers call these the *ceteris paribus conditions* – other (unspecified) things being equal, the argument holds. DT (Teller, 1980) calls it the *Nothing Mucks It Up proviso.* It is the implicit assumption that nothing else in the visual system interferes with the control of the identified physiological processes over the psychophysical phenomenon under study.

For example, in the case of scotopic detection thresholds and equivalence classes, the Nothing Mucks It Up proviso would include the assumption that the rods are the only photoreceptors that mediate vision across the whole range of conditions tested. That is, no neural elements exist that are more sensitive than the rods at any wavelength. Moreover, no code transformations occur that interfere with the Analogy between the psychophysical and physiological spectral sensitivity curves.

In fact, we know that the human retina contains cones as well as rods. And in fact, they do "muck up" the causal story at higher light levels, as we will see in Chapter 7.

6.7 Summary: Photoreceptors and system properties

In this chapter, we introduced the anatomical structure of rods and cones. We learned that rods contain the photopigment rhodopsin, which has a narrow absorption spectrum with a maximum at about 500 nm. We described the molecular structure of rhodopsin, and the cis-trans isomerization of the rhodopsin molecule. It is this small change in the shape of the molecule that accomplishes the transduction from light to physiological signals.

We then reviewed the properties of typical neurons, in order to stand them in contrast to the properties of photoreceptors. Technical stories were developed at the intuitive level, concerning the photocurrents that flow around photoreceptors in the dark, and the way they are changed by the action of light. A brief description was also provided for the chemical cascade, the set of molecular processes that carries the neural signal from the rod disk to the outer membrane, and provides the amplification required to produce a measureable signal in the photoreceptor.

We then had a look at actual physiological recordings from individual photoreceptors. In particular, we saw that the responses to individual quantal absorptions can indeed be recorded from photoreceptors. The changes in photoreceptor response with light levels and the saturating nonlinearities of photoreceptors were also shown in direct physiological recordings.

Finally, we presented three Causal Stories about how rod photoreceptors leave their marks on our visual perception. The first proposes that the transduction process in the rods accounts for the two system properties of scotopic vision developed in Chapter 2 – spectral sensitivity and equivalence classes. The second concerns signal transmission – how the capacity of a rod to signal the absorption of a single quantum determines the value of the absolute threshold of human vision.

And the third concerns how a compressive or saturating non-linearity in the cones can provide a model of the detection of interference fringes in the vicinity of 60 cy/deg, covering the notch between veridical detection and the detection of alias patterns. All three causal stories seem highly credible. We are on a roll, and the question becomes, how much longer can we go on before our causal stories begin to leave more room for doubt?

In Chapter 7, we turn to another of the major ways in which the cone photoreceptors leave their marks. In particular, we examine the consequences that the presence of three cone types has for the processing of wavelength information and for color vision.

Chapter 7

The Trichromacy of Color Vision

In Chapters 2 and 3 we introduced some psychophysical facts concerning how our visual perceptions change (or don't change) with changes in the wavelength and intensity of light. At low light levels (scotopic vision), all wavelengths of light look the same whitish color. Some patches of light will look brighter than others, but given only variations in physical intensity, lights of all wavelengths can be made to look identical. That is, in scotopic vision metamer sets include all wavelengths of light, and all wavelength information is lost.

In Chapter 5 we developed a model to explain this fact. The model assumes that only a single photoreceptor type, the rods, is functional at scotopic light levels. A single photoreceptor type cannot preserve wavelength information because wavelength information is discarded in the transduction process. Each rod sums quantal catches linearly, and any two lights that lead to the same total quantal catch in the rods will be metameric – they will look identical to a human subject.

When stimuli are at photopic light levels, however, subjects experience color variations, as discussed in Chapter 3. Thus, they readily discriminate among lights of different wavelengths because lights of different wavelengths differ in perceived color, and the colors are sustained (at least approximately) across variations of intensity. Using a bumblebees can fly argument, we therefore know immediately that photopic vision cannot be based on a single univariant photoreceptor class like the rods. The physiological model we adopted for scotopic vision must be rejected for photopic vision because it fails to account for the system property – the preservation of wavelength information.

You may be surprised to learn, however, that even in photopic vision there are metamer sets – sets of lights of different wavelength compositions and intensities that are indiscriminable from each other. The nature of these psychophysically defined sets of lights is described by a psychophysical law, the law of trichromacy, which will be presented in detail below. And the question is: why do these metamer sets occur?

To address this question we depart from our practice of avoiding mathematical formulations, and present a mathematical model of trichromacy. We do this because the model is a particularly simple use of algebra (three simultaneous linear equations in three unknowns), yet it is elegant and sufficient to the modeling task. Moreover, the historical interplay between the psychophysical law and the mathematical model, leading on to the discovery of the physiological and genetic entities that instantiate the model, provides one of the loveliest examples of progressive explanation in vision science (see Mollon, 2003 for a full historical account).

In addition, trichromatic matches provide a new and interesting example of the use of matching techniques, and thereby of the Identity family of linking propositions. The Identity family was first introduced in Chapter 2 in our account of thresholds and scotopic matching. Watch for the use of Identity propositions as we go along.

7.1 System properties: The trichromacy of color vision

7.1.1 Three facts about wavelength discrimination

Let us begin by introducing three psychophysical facts about wavelength discrimination. First, as already discussed, we can discriminate among lights of different wavelengths because different wavelengths look different colors. Second, some mixtures of physical wavelengths can be discriminated from other mixtures and from any single wavelength selected from the spectrum. Whites, purples, and desaturated colors (pink, baby blue, light green, etc.) are examples of colors that arise only from wavelength mixtures. They are called *non-spectral* (or *extra-spectral*) colors, meaning that we cannot match them to any individual wavelength.

7.1.2 Metamer sets in photopic vision

But third – and this you may find surprising – metamers occur in photopic as well as in scotopic vision. That is, even at photopic light levels there are sets of lights of very different wavelength compositions and intensities that look identical. The membership in these metamer sets is initially counterintuitive and very odd.

For example, Figure 7.1 shows a plot of the *complementary wavelengths* of light. In vision science, the term complementary wavelengths is used to describe pairs of wavelengths that *look white* when mixed together in proper proportions. This diagram tells us that many different mixtures of wavelengths all look white[1]. Moreover, by varying the intensities of the different mixtures you could match them all in perceived brightness, with the result that they would all look *identical*, even though they are *very* different physically. And there is an infinite number of other combinations of three or more wavelengths that all look white. Similarly, we can make a set of stimuli of many different wavelength compositions that all look identical and a particular shade of yellow; another set that all look identical and a particular shade of light blue; and so forth. Each of these sets of lights is a metamer set.

In ordinary experience we usually don't notice the existence of metamer sets, because metamers are such perfect perceptual facsimiles of each other that the fact that they are physically different passes unnoticed. But here's an example. Outside DT's old office, there was a light fixture that consisted of two light bulbs inside a translucent globe. Around Christmas time one year, someone took out the two ordinary light bulbs and replaced them with a "red" bulb (i.e. a bulb that emitted a band of long wavelengths, say, above 620 nm) and a "greenish-yellow" bulb (i.e. a bulb that emitted a band of middle wavelengths, say, between 530 and 560 nm). When the globe was

[1] Unfortunately the term "white" is used in both physics and psychophysics, and this causes confusion as usual. In physics the term *white light* is often taken to mean an equal energy mixture of all wavelengths, like the last mixture in Figure 7.1A. This definition is troublesome in vision science, because many different physical stimuli (wavelength mixtures) actually appear white, as the rest of Figure 7.1A shows. In fact, a whitish appearance tells us remarkably little about the wavelength composition of a light. To DT's knowledge there is no word in either physics or psychophysics for "physical stimuli that are members of the perceptual white metamer set". The closest phrase is "metameric to an equal energy light".

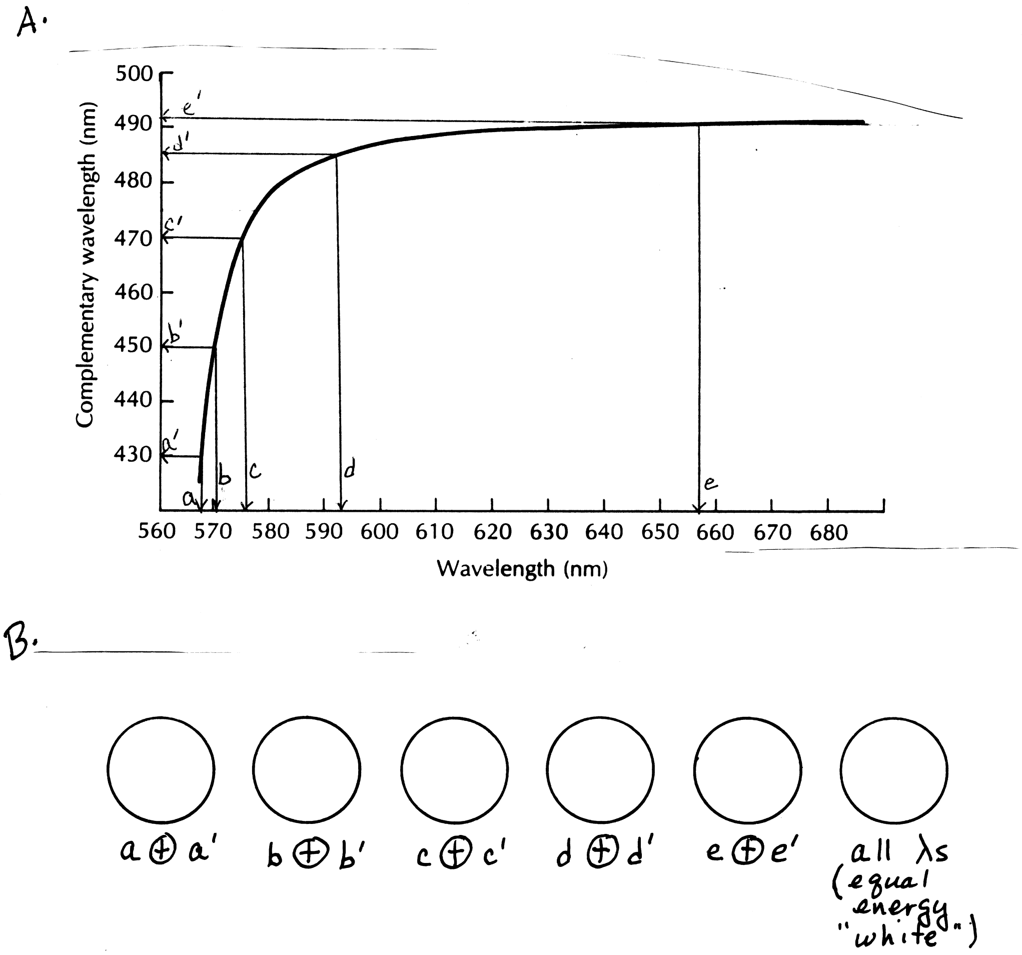


Figure 7.1: Complementary wavelengths: mixtures that look white. A: For each wavelength from about 570 nm to 680 nm, it is possible to achieve a white-appearing spot by combining that wavelength in the proper proportion with the properly chosen complementary wavelength, which will fall somewhere between about 430 and 500 nm. B: For properly chosen intensities, the row of lights shown here would all look white, and would be perceptually indiscriminable. The last spot on the right is a mixture of equal energies of all wavelengths. This is usually the wavelength composition we assume a white-appearing spot to have, but obviously this assumption can be wrong. In addition to those shown, many other mixtures would also look white – for example, any of many mixtures of three wavelengths, four wavelengths, and so on. (The plus sign in the circle is the symbol for superposition: we are superimposing one light on another.) [Modified from Cornsweet (1970), Fig. 10-5, p. 232; After Sinden (1923)].

replaced, one half of the globe looked red and the other half greenish-yellow. But between the two, there appeared a band of very distinct and saturated yellow.

Why was the yellow band there? Not because the band was illuminated by a light of an isolated wavelength that looks yellow (say, 575 nm), but because it contained just the right mixture of light from the "red" and "green" bulbs. The mixture of wavelengths coming from the band belonged to the yellow-appearing metamer set. But most people who walked by would not have even wondered why the yellow band was there. Of those who wondered, most would probably have assumed that the globe must have contained a source of 575 nm light. Only a few would have guessed that nothing but broadband "red" and "green" bulbs were hidden inside the globe.

To emphasize again the oddity of color mixture, notice that the appearances of combinations of wavelengths of light is very different from the sounds we hear when we combine sound waves. When we create a series of vibrations of different temporal frequencies in the air, we hear a series of tones of different pitches. When we play these tones together, we hear *chords* that still perceptually contain the original tones; we don't hear an intermediate tone, much less a completely novel tone or no tone at all. Why is the mixing of lights so different?

7.1.3 The psychophysical law of trichromacy

As it turns out, metamer sets are not as arbitrary as they originally seem. They follow a particular rule, called the *law of trichromacy* (tri = three).

Figure 7.2 shows a laboratory set-up for demonstrating trichromacy. We assemble a set of three slide projectors, fitted out with devices for allowing continuous variation of their intensities. We put a narrow-band color filter in front of each, and overlap the three beams on a projection screen to make a patch of light, A. In other words, patch A is a mixture of three lights of different wavelengths, λ_1, λ_2, and λ_3 [2]. A useful set of choices (cf. Figure 7.4) is to let $\lambda_1 = 460$ nm (which looks predominantly blue), $\lambda_2 = 530$ nm (which looks predominantly green), and $\lambda_3 = 650$ nm (which looks predominantly red). We also set up a fourth projector to make a second patch of light, B. We then use a variety of color and neutral filters in turn, to make patch B appear any color and brightness that we choose.

The law of trichromacy, informally stated, is that by varying only the intensities of the three wavelengths in patch A, we can make a perfect perceptual match to (almost) any other light in patch B; that is, we can make A $\equiv$ B (A is metameric to B). Different colors and brightnesses in patch B will require different intensities of the three wavelengths in patch A, but an exact perceptual match will (almost) always be possible.

In fact, we can exactly match any light in patch B if we are allowed to move one of the three wavelengths from patch A to patch B. This additional provision leads to the formal statement of the law of trichromacy: Given any four lights, we can arrange them with three in patch A and one in patch B, or two in patch A and two in patch B; vary the intensity of any three of them, and end up with a perfect perceptual match between the two patches. Figure 7.3 shows a simulation of the basic trichromacy demonstration.

In 1928, W.D. Wright carried out a classic study of trichromatic color matching. Wright tested 10 subjects with mixtures of 460, 530, and 650 nm in patch A. He set up patch B with each of

[2] The three lights we use as the mixture set are sometimes called *primaries*. However, this term is used in different ways in other contexts. Suffice it to say that the "primary"' colors you learned about in kindergarten make use of a different meaning of the term.

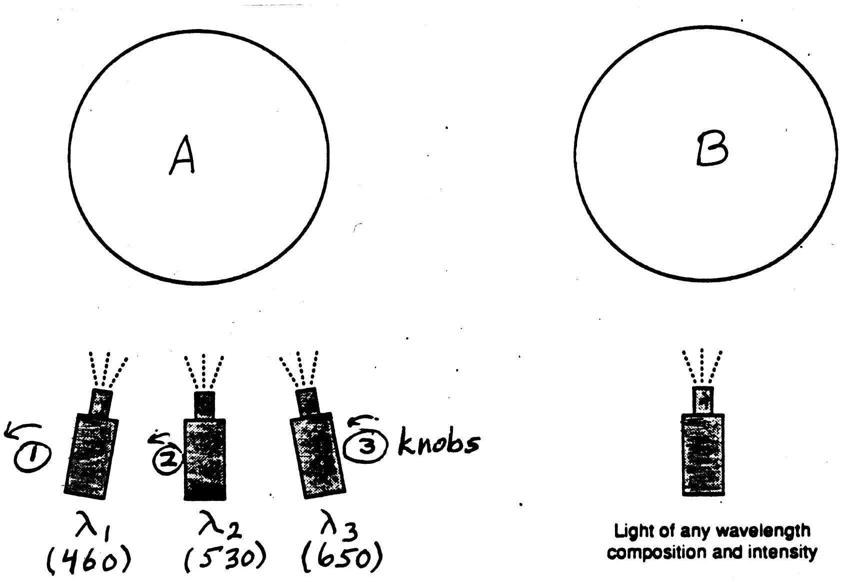


Figure 7.2: Set-up for a demonstration of trichromacy. Light from the three projectors on the left is superimposed to make patch A. The projectors are fitted with narrow-band filters to provide wavelengths of (say) 460, 530, and 650 nm. Each projector also has a knob that controls its intensity. Patch B comes from a fourth projector that can be set to provide any wavelength or combination of wavelengths at any intensity.

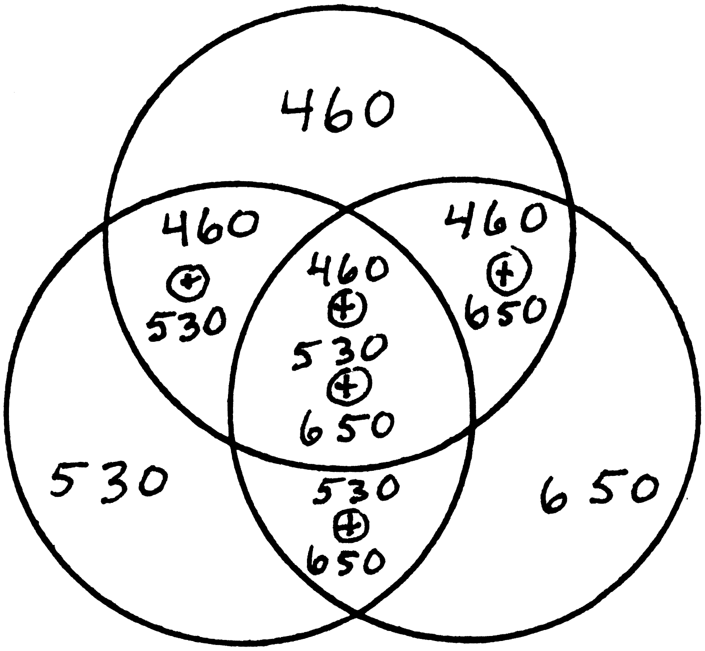


Figure 7.3: COLOR PLATE. A simulation of color mixture. In this figure, the three superimposed beams of Figure 7.2 have been partially separated in space. The outer crescents simulate the colors of each of the three original wavelengths. The outer triangles simulate the colors of combinations of two wavelengths, and the central triangle simulates the color resulting from the combination of all three wavelengths. (Figure 7.3 is only a simulation and not a true demonstration, because the colors of the various segments will be simulated with the broadband pigments used in printing, rather than being made from narrow wavelength bands.)

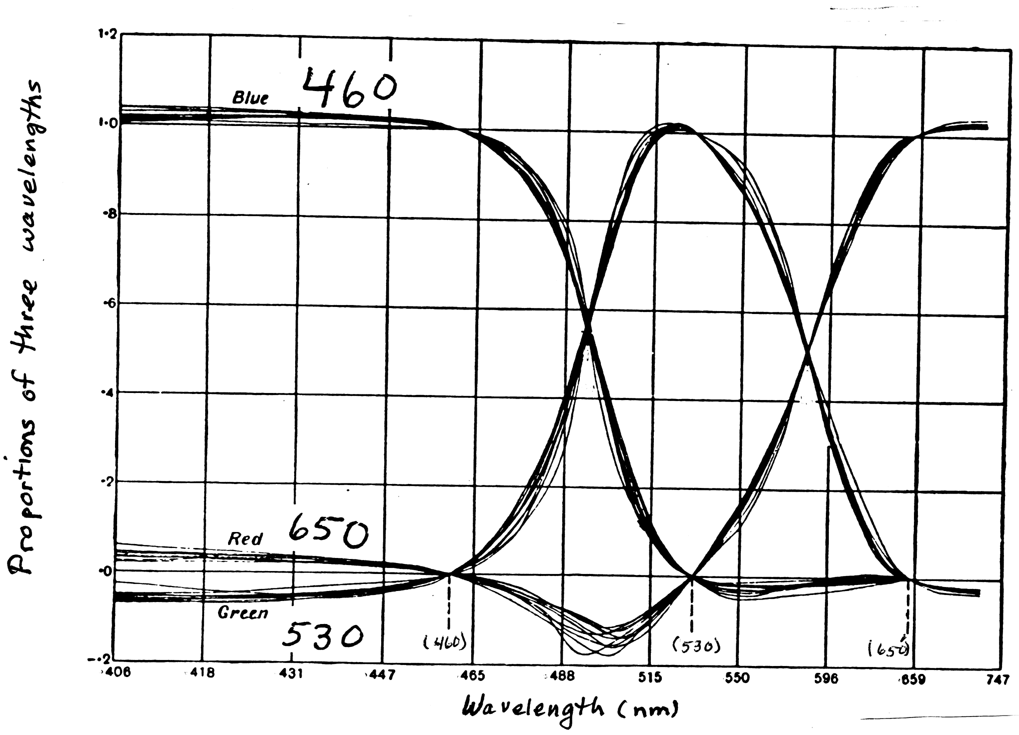


Figure 7.4: Data from a trichromatic color mixture experiment. The abscissa shows the wavelength of light in patch B. The ordinate shows the proportions of 460, 530, and 650 nm lights required to make the two patches metameric. The lines show fits to the data from 10 subjects. Note the "subtraction" of the different primaries in different wavelength regions. [Modified from Boynton (1979). Fig. 5.18, p. 149; data from Wright (1928).]

the different individual wavelengths of light in turn. For each wavelength in patch B, the subjects adjusted the intensities of the three lights in patch A to make metameric matches between the two patches.

Wright's data are shown in Figure 7.4. In the short wavelength range, below about 460 nm, color changes very little with wavelength. These lights all look predominantly violet, and each can be nearly matched with the 460 nm ("blue") primary alone. However, a small amount of the 650 nm ("red") primary must be added, and a small amount of the 530 nm ("green") primary must be subtracted – moved to the other side and mixed with the light in patch B – in order to make the matches. At 460 nm, of course, the 460 nm primary provides an exact match. Between 460 and 650 nm, perceived hue changes more rapidly with wavelength, as do the proportions of the different primaries required for matches. Notably, the 650 nm ("red") primary must be subtracted for all wavelengths between 460 and 530 nm, as must the 460 ("blue") primary for all wavelengths between 530 and 650 nm.

Trichromacy is a remarkable and puzzling system property of photopic vision. Why do wavelength mixtures behave the way they do? Why are the metamer sets as they are? Why are three lights enough? The answer lies in our visual systems.

7.1.4 Reprise on the Converse Identity proposition

Let us do the exercise of ferreting out a linking proposition. The Identity family of linking propositions was introduced in Chapter 2 (Figure 2.8B), in the context of measurements of matching and thresholds. Trichromatic metamers are sets of stimuli that are very different physically but appear identical perceptually. That is, they are another case in which subjects are carrying out a matching task. The data are perceptual, so to explain them physiologically we will be trying to reason from perception to physiology. Thus the two available Identity propositions are the Contrapositive (2) and the Converse (3).

Moreover, the basic perceptual observation is that patches A and B match, so the relevant linking proposition is the Converse: perceptual identity implies physiological identity. Thus, if the theorist assumes the truth of the Converse Identity proposition, metameric matches imply that the signals that arise from a set of metameric stimuli are rendered physiologically identical somewhere within the visual system.

The causal story for trichromacy then becomes a locus and coding question. Where within the visual system do the signals originating from the physically different stimuli become identical, and by means of what physiological processes and computations?

7.2 A mathematical model of trichromacy

7.2.1 Assume three Fundamentals

The mathematical model of trichromacy starts with the set of mathematical or physiological assumptions shown schematically in Figure 7.5. The model assumes that photopic vision is served by three Fundamentals – three mathematical/physiological entities with different spectral sensitivity curves. The peak sensitivities of the three Fundamentals are assumed to differ, but the ranges are assumed to overlap substantially (for simplicity, they overlap entirely in Figure 7.5A). The model also assumes that each Fundamental forms a linear summation of the signals resulting from different wavelengths of light.

For concreteness, in the following pages we identify the Fundamentals with three types of cones. But it's interesting to notice that the mathematical model of trichromacy preceded any direct evidence of the numbers of cone types or their spectral sensitivity curves. We will return to this point below.

As was the case for rods in Chapter 2, we can represent each cone type with a funnel that counts the quanta it catches, without keeping track of their wavelengths (Figure 2.10, in which the marbles can now be identifieid as quanta). In the case of photopic vision there are three funnels, each with its own counter. The three-funnel analogy is shown in Figure 7.5B.

Now we need to develop some symbols. To identify each Fundamental (cone type) with the wavelength range of its maximum sensitivity, the three Fundamentals will be called L, M, and S[3]. The letters *L, M,* and *S* in italics will denote the quantum catch rates generated in the L, M, and

[3] The different cone types have historically been called ”red cones”, ”green cones” and ”blue cones”. Vision scientists avoid this terminology, in order to maintain the clear separation of perceptual and physiological terms. If the terminology is sloppy, it is easy to think that color vision is simple – we see red because we have ”red cones”! We use color names to refer to perceived colors, and a different set of names – *S*, or *short-wavelength-sensitive, M,* or *mid-wavelength-sensitive*, and *L*, or *long-wavelength sensitive* – to refer to cones.(L, M, and S cones) to refer to cone types.

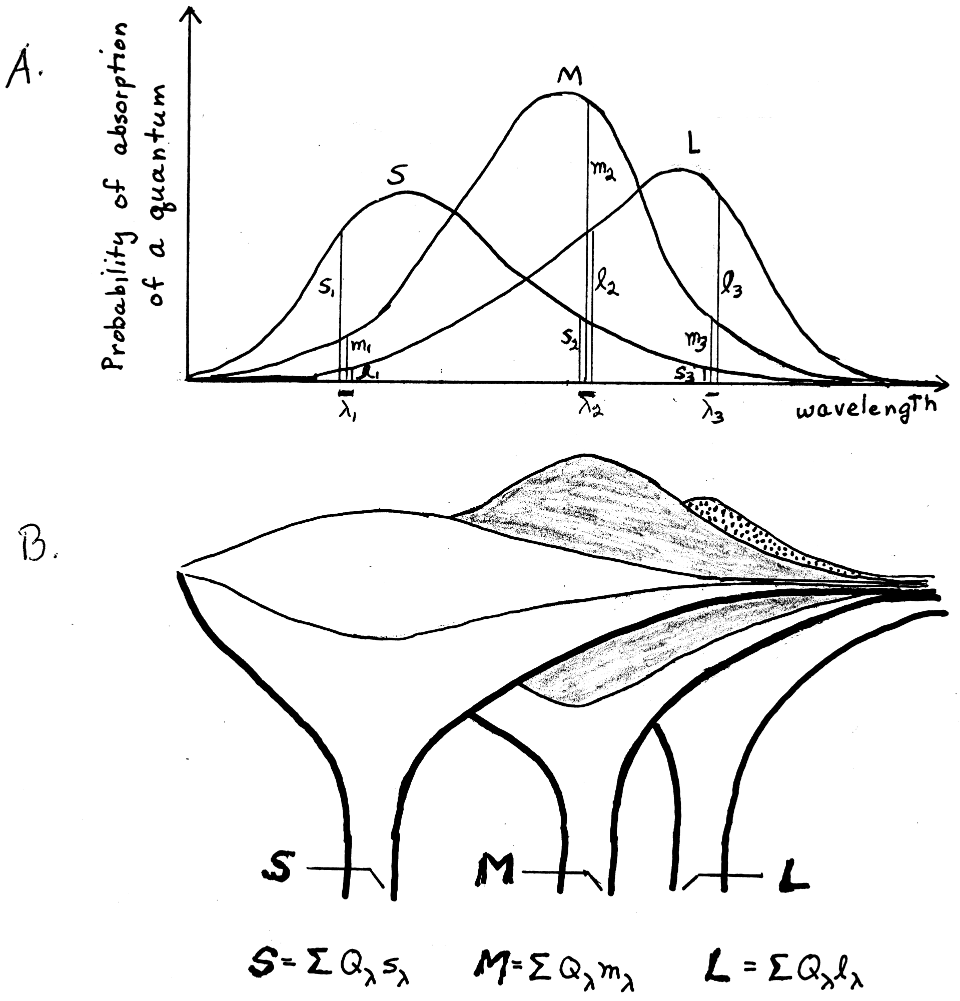

A.
Probability of absorption of a quantum
S
M
L
s_1
m_1
l_1
m_2
l_2
s_2
l_3
m_3
s_3
$\bar{\lambda}_1$
$\bar{\lambda}_2$
$\bar{\lambda}_3$
wavelength
B.
S
M
L
$S = \Sigma Q_\lambda s_\lambda$
$M = \Sigma Q_\lambda m_\lambda$
$L = \Sigma Q_\lambda l_\lambda$

Figure 7.5: A mathematical model of trichromacy. A: Completely overlapping spectral sensitivity curves assumed for the three Fundamentals for purposes of illustration. B: Extension of the funnel analogy to the case of trichromacy. Each funnel captures a broad range of wavelengths with probabilities determined by its width at each wavelength, and counts its total quantum catch without regard to wavelength. The result is a set of three variables, *S, M,* and *L*, whose values are determined by summing the quantal catches within each cone type.

S cones respectively. Because we are dealing with two patches of light, A and B, there will be two sets of cones (two sets of funnels in the analogy), one for patch A and one for patch B. Let the quantum catches resulting from patch A be L_A, M_A, and S_A, and from patch B be L_B, M_B, and S_B respectively.

7.2.2 The condition for metamerism

By hypothesis, metamers occur when lights of different wavelength composition yield identical quantum catches in *each* of the three hypothetical photoreceptor types. That is,

If $L_A = L_B$ and $M_A = M_B$ and $S_A = S_B$, then A ≡ B.

This statement can be called the *condition for metamerism.* But under what circumstances is the condition for metamerism satisfied? Is it a fool's dream, or a realistic basis for a model?

7.2.3 The color equations

What does light of a given wavelength, λ, do when it encounters a three pigment system like that shown in Figure 7.5? It makes a triplet (a set of three) quantum catch rates, one in each of the three cone types. For any given wavelength such as λ_1, the heights of the three curves at that wavelength tell us the probabilities of absorption of a quantum by each cone type.

Let

l_1 = the height of the L curve at λ_1

m_1 = the height of the M curve at λ_1

s_1 = the height of the S curve at λ_1

Let l_2, m_2, and s_2, and l_3, m_3, and s_3 be similarly defined. Since we assumed the shapes of the three pigment curves in Figure 7.5, all of these values of curve heights are constants in the equations we will develop below.

Now we need symbols for the *intensity* of the lights from each of the three projectors; that is, for the rate of arrival of quanta of each wavelength λ_1, λ_2, and λ_3. These intensity values are just the intensities of the three projectors in Figure 7.2. The intensities of projectors 1, 2, and 3 will be called Q_1, Q_2, and Q_3 respectively. Since the subject varies these intensities to make the metameric matches, the Q's will turn out to be the variables in the equations we will develop below.

We are now in a position to write expressions for the rate of quantal absorptions from each of the three wavelengths in each of the three cone types. The rate of quantal absorptions of the wavelength λ_1

by the L cones is: $\mathrm{Q}_1\mathrm{l}_1$

by the M cones is: $\mathrm{Q}_1\mathrm{m}_1$;

and by the S cones is: $\mathrm{Q}_1\mathrm{s}_1$;

and similarly for λ_2 and λ_3. Moreover, to calculate the total quantum catches in each cone type in response to any wavelength mixture, we just add up the quantum catches from all of the available wavelengths for each photoreceptor:

L cone quantum catch : $L = \sum \mathrm{Q}_\lambda \mathrm{l}_\lambda$

M cone quantum catch: $M = \sum \mathrm{Q}_\lambda \mathrm{m}_\lambda$

S cone quantum catch: $S = \sum \mathrm{Q}_\lambda \mathrm{s}_\lambda$

Now let's return to our two patches, A and B. Patch A is composed of three wavelengths, λ_1, λ_2, and λ_3, with variable intensities Q_1, Q_2 and Q_3. We can now write three equations to describe the three cone quantum catch rates produced by patch A.

$$L_A = Q_1 l_1 + Q_2 l_2 + Q_3 l_3$$

$$M_A = Q_1 m_1 + Q_2 m_2 + Q_3 m_3$$

$$S_A = Q_1 s_1 + Q_2 s_2 + Q_3 s_3$$

What about patch B? Patch B is a light of any chosen wavelength composition and intensity. For any specific choice of wavelength composition and intensity, patch B generates a specific triplet of cone signals, L_B, M_B, and S_B. Once the wavelength composition is chosen, these entities are constants.

Now, the fundamental question is, by varying only the intensities Q$_l$, Q$_2$, and Q$_3$, can the triplet of values L_A, S_A, and S_A be made identical to the triplet L_B, M_B, and S_B? That's the condition for metamerism.

Assume for the moment that the two lights *are* metamers. Then by the condition for metamerism we can substitute L_B, M_B, and S_B for L_A, S_A, and S_A to produce:

$$L_B = Q_1 l_1 + Q_2 l_2 + Q_3 l_3$$

$$M_B = Q_1 m_1 + Q_2 m_2 + Q_3 m_3$$

$$S_B = Q_1 s_1 + Q_2 s_2 + Q_3 s_3$$

Notice that this set of three simultaneous equations contain three unknowns – the three intensities Q$_l$, Q$_2$, and Q$_3$. (The little ls, ms, and ss are known because they are specified by the heights of the spectral sensitivity curves L, M and S, and L_B, M_B, and S_B are known because they are specified by the choice of the light B.) Each of the three cone types contributes an equation, and each of the three wavelengths contributes a variable.

But remember that it is an elementary property of linear algebra that three simultaneous linear equations in three unknowns are guaranteed to have a solution, so we know that for any specified values of L_B, M_B, and S_B we can solve for the values of Q$_l$, Q$_2$, and Q$_3$. It follows that for *any* specified set of values of L_B, M_B, and S_B – that is, for any light in patch B – these equations can be solved. Thus logically we know that by varying only the radiances of the three wavelengths in patch A, patch A can be made metameric to light of any wavelength composition in patch B. But that's the informal statement of the law of trichromacy! So in sum, the equations provide a sufficient mathematical model of the law of trichromacy.

But there's one possible flaw in the argument. Remember that, although we are guaranteed a solution to three simultaneous equations in three unknowns, there is no guarantee that all of the values for Q$_l$, Q$_2$, and Q$_3$ will be positive. One or more of them might be negative. But real lights cannot have negative intensities, so how do we interpret the negative values? The answer is, in algebra, we move the negative term to the other side of the equation and make its value positive. In the matching experiment, we move the corresponding light, λ_1, λ_2, or λ_3, from patch A to patch B. From this convention results the formal statement of the law of trichromacy: Given any four lights, we can arrange them in two patches to make A $\equiv$ B.

We now return to a more historically accurate picture. For the sake of concreteness we initially identified the three Fundamentals with three cone types, and the linear summation property with the loss of wavelength information in an individual photoreceptor caused by the properties of the transduction process. But historically, both the psychophysical fact of trichromacy and the mathematical model of trichromacy were established by about 1860, before we had any other evidence of the number of cone types, or the univariance of photoreceptors, or even any modern notion of quantum theory. It was a deep mathematical insight to see that a set of three simultaneous linear equations in three unknowns would provide a sufficient model for the perceptual fact of trichromacy, and to posit three Fundamentals with overlapping spectral sensitivity curves to provide the three variable system of simultaneous linear equations.

Finally, let's return to the properties of wavelength discrimination with which we started this chapter, and review explicitly why they occur. First, we can discriminate among wavelengths of light different wavelengths keep their distinctive colors well across variations in intensity. As shown in Figure 7.5, each different wavelength sets up a different set of *relative* quantum catches, l vs. m vs. s, in the L vs. M vs. S cones. Putting it another way: wavelength information is lost in each individual cone type, but it is preserved in the *ensemble* of three cone types by the *relative* quantal catches among them. This ensemble code is the form in which wavelength information passes through the photoreceptor level of processing. Similarly, information about the *intensity* of the light is preserved in the *absolute* quantum catches in the three kinds of cones.

Second, some mixtures of wavelengths can be discriminated from other mixtures and from any single spectral wavelength. Why? Because not all of the possible ratios of cone signals are created by individual wavelengths of light, and some mixtures of wavelengths create these novel ratios. For example, by inspection of Figure 7.5, there are no individual wavelengths that create L/M/S ratios of 1:1:1, or 2:1:2, and so on; but you can find mixtures of wavelengths that will do so. When mixtures of wavelengths create these ratios, non-spectral colors appear.

The third property, metamerism, blends into the law of trichromacy, and has already been explained in detail.

7.2.4 Reprise on the Initial Identity proposition

Here's a second exercise on linking propositions. Whereas the inference from behavioral trichromacy to three Fundamentals rests on a Converse Identity proposition, the mathematical model from fundamentals to behavioral trichromacy rests on the Initial Identity proposition. The mathematical model begins by assuming (seemingly arbitrarily) the existence of three physiological entities – the three Fundamentals. We argued mathematically that three such entities, acting together, would process physical stimuli in just such a way as to create the metamer sets observed psychophysically in human subjects, and summarized by the law of trichromacy.

The Initial Identity proposition – that identical physiological states imply identical perceptual states – enters the argument because we are trying to reason from (assumed) physiological states to perceptual states. Assuming Initial Identity allows us to use physiological identity to infer perceptual identity, and thus use the model to provide an account of the psychophysical data.

7.2.5 Exact physiological implications of trichromacy

Now let's step back a little. Clearly the system property of trichromacy, together with its mathematical model, places major constraints on physiological models of the visual system. Up until this point, for the sake of specificity and simplicity, we have identified the three mathematical Fundamentals with three cone types. But the true constraints are actually somewhat more general.

What the mathematical model of trichromacy actually suggests is that information available for discriminating among wavelengths and intensities of light passes through a serious bottleneck – or rather, three bottlenecks – somewhere on its way through the visual system. That is, there is a stage at which this information is limited to three variables. (A statistician would say the visual signal has only three degrees of freedom, and an engineer would say the system has only three information channels).

In the specific model we introduced (and in reality in foveal vision) this three-channel stage is instantiated by three kinds of cones with three different photopigments. But logically, behavioral trichromacy could as well come about from having several kinds of photoreceptors, but with the information reduced to three variables by passing through a three-channel stage somewhere later in the visual system. And as it happens, in peripheral vision we do have rods as well as the three cone types, and the reduction to three channels does come later, as well see in Chapter xx.

A final thing to note is that even if the photoreceptors are the bottlenecks, the argument we gave above does not depend on assuming any particular set of shapes or wavelengths of maximum sensitivity for the spectral sensitivity curves of the three photopigments. We assumed three Fundamentals, L, M, and S, and the heights of these three curves at particular wavelengths were constants that entered into the color equations. But logically we could have assumed any of many shapes for the spectral sensitivity curves, as long as they are consistent with color mixture data like that shown in Figure 7.4.

7.3 The search for the Fundamentals of color vision

The psychophysical fact and the three-Fundamental model of trichromacy were both well established by about 1860 (Mollon, 2003). But the situation left vision scientists in a state of acute frustration. We knew that there are (probably) three cone types, but neither the psychophysics nor the mathematical model reveals their spectral sensitivity curves. Determining the spectral sensitivity curves of the three Fundamentals has thus been a fundamental challenge (pun intended) for vision scientists for 150 years, and scientists from several disciplines have set out to determine the shapes of these curves. We will review three historical approaches.

First, some of the earliest relatively accurate estimates of the Fundamentals were derived from psychophysical measurements. Of these the most successful approach was based on the assumption that certain "color-blind" individuals (see below) have lost one of the Fundamentals, but that their two remaining Fundamentals are identical to the Fundamentals of normal subjects. Without the interference of the third pigment, the available pigments in "color-blind" human subjects could be estimated psychophysically, and by hypothesis used to estimate the spectra of the normal pigments.

Excellent psychophysically-based estimates of the sensitivity maxima and the curve shapes of the cone Fundamentals emerged in the 1970s. They are shown with data from other, later techniques in Figure 7.6B, and it can be seen that the data correspond closely. However, the challenge persisted of verifying these estimates with more direct measurements that were free of the assumption that normal pigments occur in color-deficient subjects.

In the next historical iteration, the technique of *microspectrophotometry* was developed. In microspectrophotometry, one dissociates the cells of an excised retina, until individual photoreceptors can be seen floating free under the microscope. One can then shine a tiny beam of light through an individual photoreceptor and onto a photocell, and the percent of light absorbed can be measured for each wavelength in turn.

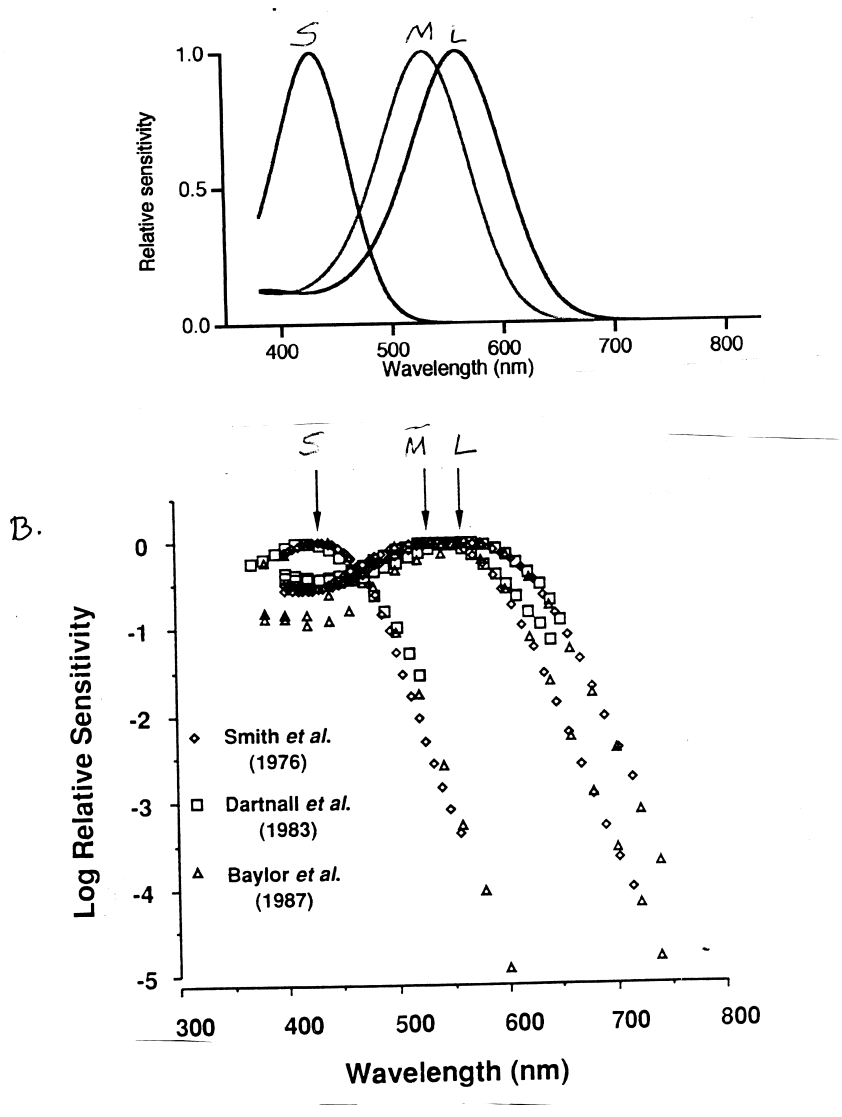


Figure 7.6: The spectral sensitivities of the three cone types in color-normal subjects. A. L, M and S cone spectra from macaque retina, recorded with suction electrodes, shown on a linear ordinate. These empirical curves replace the hypothetical curves shown in Figure 7.5A. B: The three cone spectra estimated with three different techniques, shown on a log ordinate. The results of psychophysical (diamonds), microspectrophotometric (squares), and suction electrode (triangles) techniques are shown. The three sets of measurements agree remarkably well. [A replotted from Bayler, Nunn and Schnapf 19xx, courtesy of J. Schnapf; B modified from Lennie and D'Zmura, 1988, Fig. xx, p. xx.]

Cone spectral sensitivity curves measured with microspectrophotometry are shown in Figure 7.6B. Microspectrophotometry confirms that the spectral sensitivity curves of individual cones are smooth and U-shaped, and these particular data suggest absorption maxima at about 420, 530, and 560 for the three cone types. Microspectrophotometry, however, is beset with signal/noise problems that limit the measurements to a relatively narrow range of wavelengths around the spectral maximum (notice the limited wavelength range of the squares in Figure 7.6B).

In the most recent assault, in the mid-1980's, the problem of cone spectral sensitivities was attacked with suction electrodes. This technique has the advantage that the physiological response of the photoreceptor itself is used for the actual measurements. Since very small amounts of light are sufficient to produce measurable changes in photocurrents, the suction electrode produced an increase in sensitivity over earlier techniques, with a corresponding increase in the wavelength range over which meaningful measurements could be made. Data from an early suction electrode study are compared with those of earlier techniques in Figure 7.6B. The suction electrode data, plotted on a linear sensitivity axis, are shown in Figure 7.6A. These data replace the conceptual Fundamentals introduced in Figure 7.5A.

All of these techniques have evolved over DT's scientific lifetime, and each decade has brought a surer answer. Nowadays there is excellent agreement from all of these very different kinds of measurements, and the cone spectral sensitivity curves are a solved problem. The most recent estimates suggest that the spectral maxima of the S, M, and L cones are very close to 430, 530, and 560 nm. Moreover, as needed to sustain the three channel mathematical model developed above, the spectral curves are broadly overlapping, at least between 400 and 550 nm, so that in this wavelength range, a single wavelength can set up a quantum catch rate in each of the three kinds of photoreceptors, and set up the ensemble code.

Notice, however, that the S cones have negligible sensitivity above about 550 nm. As a result, color-normal subjects actually have only two functional cone types in the mid to long wavelength spectral range. And in fact, color-normal subjects can match any wavelength above about 550 nm with a mixture of only two primaries, such as a 530 nm "green" and a 650 nm "red".

7.4 Trichromacy as a causal story: As good as it gets!

Let us consider where we've been. First, psychophysical measurements (color matching) on human subjects defined and quantified the puzzling system properties of metamerism and trichromacy. Second, a mathematical model was developed to explain these system properties. The model – three Fundamentals with broadly overlapping spectra and linear summation of effects across wavelengths, expressed mathematically in three simultaneous linear equations with three unknowns – provides a sufficient account of the psychophysical data.

But the model also provided specific predictions about neural elements we should find within the visual system: three kinds of cones with different spectral maxima but overlapping spectral sensitivity curves. We then went looking for independent evidence for the existence and spectral properties of three kinds of cone photoreceptors, and found it. Moreover, the linear summation across wavelengths assumed to occur within each Fundamental finds its explanation in the loss of wavelength information in the transduction process. The causal story is complete, and tightly tied into the network of surrounding sciences.

The story of trichromacy also illustrates the difference between mathematical and physiological models. Mathematical modeling is challenging, and it's a major achievement to invent a model that just exactly accounts for a set of psychophysical findings. But given a satisfying mathematical model, the question arises: will the model be instantiated in the real visual system? Vision scientists become keenly interested in finding physiological entities that embody the mathematical entities or parameters assumed by the model.

For DT, it's thrilling to understand that three simultaneous equations in three unknowns provide an account of trichromacy. But it's even more thrilling to learn that the hypothesized discarding of wavelength information within a Fundamental corresponds to the actual discarding of wavelength information by the transduction process, and that the mathematical property of adding up the terms in the linear equations across wavelength corresponds to the physiological property of combining the effects of individual cis-trans isomerizations across wavelength.

In sum, for DT, the causal story of trichromacy is vision science at its very best. Trichromacy and its explanation provide an important case example to which to aspire as one tries to invent causal stories to explain other system properties of vision. In her more thoughtful moments, however, DT ventures the guess that there may never be another equally satisfying causal story in vision science. Part of the implicit basis of the model is that there are three and only three cone types, and that all of the cones of a given type have the same spectral sensitivity curve. [If this were not true, what would happen to the color equations?] This assumption has a chance to be correct because (as we will see) the spectra of photoreceptors are closely controlled by single genes. Such homogeneity cannot be expected in neurons whose critical properties are influenced by a broader range of causes. Moreover, convenient mathematical properties like linear summation will be harder to come by the more deeply we venture into the visual system.

7.5 A Design question: Why three and only three cone types?

A design question: Why did humans evolve to have exactly three cone types rather than two or four? It has been argued that ancestral primates had only two types – an S cone and a prototypical LM cone with a spectral maximum in the mid to long wavelength region. But since the S cones have negligible sensitivity above 550 nm, the ancestral primate would have had only a single pigment available above 550 nm, and would not have been able to discriminate among middle and long wavelengths, nor among surfaces that reflect different combinations of middle and long wavelengths. That is, the ancestral primate would not have been able to discriminate among objects or surfaces that we perceive as yellow-greens, yellows, oranges and reds.

The splitting of the LM prototype into two separate classes – L and M – probably occurred only 30-40 million years ago, and only in old-world primates. It has been proposed that such a trichromatic system allows us to discriminate red, orange, and yellow fruit from green trees, and to tell ripe from unripe fruit. Thus, the third cone type allowed ancestral primates to exploit important new food sources efficiently, and probably carried a major selective advantage in evolutionary terms. This idea is illustrated in Figure 7.7.

Why not keep on evolving, and have more than three cone types? After all, the larger the number of cone types, the smaller the metamer sets, and the more wavelength information is preserved. The speculation here is that, although it is easy to create trichromatic metamers in the lab, they rarely occur in nature. In nature most surfaces reflect broad bands of wavelengths, and most of the objects that produce lights in any one metamer set probably have very similar spectral characteristics. So additional photopigments might not allow us any more genuinely useful color discriminations than we can already make with three.

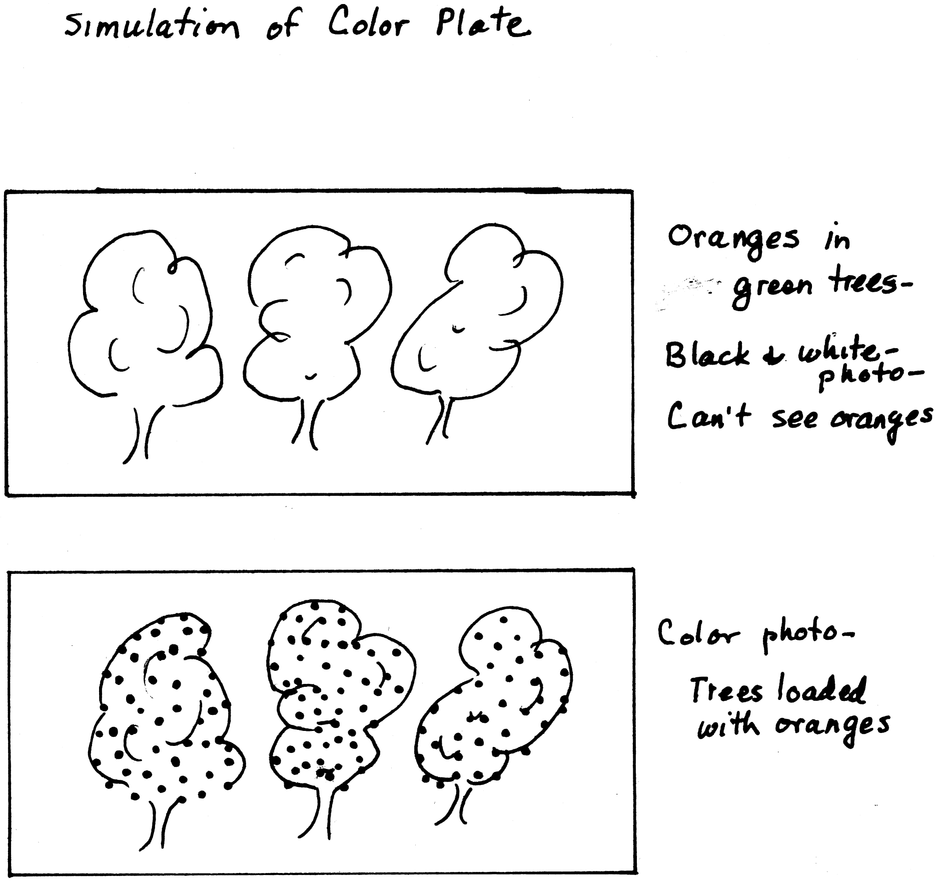


Figure 7.7: COLOR PLATE. Oranges in green trees. This color plate illustrates one theory of why a third cone type conveyed a selective advantage on ancestral primates. Separation of the ancestral long/middle wavelength pigment into the L and M pigments allows discrimination among yellow-greens, yellows, oranges, and reds, allowing trichromatic primates to detect fruit among green leaves, and perceive the ripeness of the fruit. [Source?? Xx]

7.6 Color vision deficiencies: Dichromacies and anomalies

People whose retinas contain the three standard pigments are said to have normal color vision, or to be *color-normal trichromats.* But not everyone is color-normal. Some people are missing one of the three kinds of cones. Others still have three cone types, but one or more of the pigments is shifted in spectral sensitivity. Such changes make predictable changes in color discrimination capacities. Look back at Figure 7.5 as you read the next few paragraphs.

First, what would happen if you were missing one of the three cone photopigments? Suppose you were missing the L photopigment. In that case, you would have two rather than three funnels, and two rather than three cone output signals. You would have only two rather than three equations in your set of color equations, because the equation for quantum catches in the L cones would not be needed. Since there would be only two equations, you would need only two wavelengths in patch A to match any wavelength in patch B (two equations need only two unknowns to be guaranteed a solution). All of the metamer sets of a trichromatic individual would also be metamer sets for you; but your metamer sets would be larger than those of the trichromat – you would confuse patches of light that would be readily discriminable for your trichromatic friend, and you would probably be worse than they are at finding yellow, orange and red fruit in green trees.

This form of color vision deficiency is well known, and occurs quite frequently in human beings. Since the color vision system is reduced from three to two variables, such individuals are called *dichromats* (di = two). The two most common types are *protanopia*, in which the person is missing functional L cones (proto = first; protanopia = the first kind of color deficiency); and *deuteranopia*, in which the person is missing functional M cones (deutero = second; deuteranopia = the second kind). Each of these types of color vision deficiency is sex-linked, and occurs in about 1% of the Caucasian male population. The third kind of dichromacy, *tritanopia* (tri = three; tritanope = the third kind), in which the person is missing functional S cones, is much rarer, and occurs with equal frequency in both males and females.

Second, what would happen if the spectral sensitivity of one of your photopigments were shifted along the wavelength axis? Suppose your L cone pigment were shifted toward your M cone pigment. How would your color equations change? Since the height of the L curve would have changed a little at each wavelength, all of the little l's in the equation for the quantum catch in L cones would change. You would still be trichromatic, because your color vision would still be described by three equations in three unknowns. But for each wavelength composition of patch B, the change in values of l's would make a change in the intensities of the three lights in patch A needed to make the match to patch B. That is, your color matches would be different than those of your color-normal friend. Similar changes would occur if the M or the S cone spectral sensitivity curve were shifted.

This form of color vision deficiency is also well known, and people with trichromatic vision but non-normal metamer sets are said to be *color-anomalous.* The color vision of people with a shifted L cone pigment is called *protanomalous*, while that of people with a shifted M cone pigment is called *deuteranomalous.* Protanomaly and deuteranomaly occur in about 1% and 3% of the Caucasian male population respectively. *Tritanomalous* color vision, which results from a shifted S cone pigment, is much rarer and occurs equally often in males and females.

In sum, in the Caucasian population about 8% (one in 12) of the male population and less than 1% of the female population have a color deficiency caused by losses or spectral shifts of either the L or the M cone photopigment. As a group, these forms of color vision are often called the *red/green color deficiencies.* Another small perceptage (less than 1%) have *tritan* deficiencies –

losses or anomalies of the S cone photopigment.

Color-normal and color-deficient individuals live in different perceptual worlds. To illustrate this point, we here discuss a clinical color mixture test that diagnoses among color-normal, dichromatic and anomalous trichromatic subjects. The test is the *Rayleigh match*, carried out with a device called an *anomaloscope.* The test is illustrated in Figure 7.8. In the anomaloscope field (Figure 7.8A), the subject sees a mixture of 550 and 670 nm lights (which ordinarily look green and red respectively to a color-normal subject) in one half of a circular field, and a 589 nm light (which ordinarily looks a slightly orangish yellow to a color-normal subject) is presented in the other half of the field. The subject is asked to vary the proportion of the 550 vs. 670 nm lights in the one half field, and the intensity of the 589 nm light in the other half field, to try to make a metameric match between the two halves of the field. Figure 7.8, panels B-D show a simulation of the outcomes of Rayleigh matches for color-normal, dichromatic, and anomalous trichromatic subjects, and panel E simulates the appearance of one anomalous trichromat's match to a color-normal observer.

Red/green color deficiencies are so common that in any class of 30 school children, one or more of the boys is likely to have a red/green color deficiency. These children are likely to find it difficult to learn color names, use the "correct" color in drawing with crayons, and so on. An adult color-deficient individual can have trouble choosing two socks that match, and may wear color combinations that seem bizarre to his color-normal friends. If you argue with your friends over what colors things are, you may be a dichromat or an anomalous trichromat and not realize it.

7.7 Genetics of color vision

Each of the red/green color deficiencies described above shows an X-linked pattern of inheritance (the particular deficiency is passed from grandfather to grandson with the mother being a carrier). To geneticists this pattern suggests strongly that the L and M pigment genes are located on the X chromosome. Tritanopia and tritanomaly show an autosomal pattern of inheritance, suggesting that the S pigment gene is located on some other chromosome.

In 1986, a team of geneticists and psychophysicists led by Jeremy Nathans isolated and sequenced the genes that control the production of each of the three human cone photopigments. Genomes were characterized from both color-normal and red-green color-deficient individuals. This research was extremely exciting to vision scientists, since it took the search for the Fundamentals of color vision all the way to the molecular level. Comparisons of the molecular structures of the four normal human photopigments are shown in Figure 7.9.

As expected from the X-linked inheritance patterns of red/green color vision deficiencies, the genes for two photopigments were found next to each other on the X chromosome. Unexpectedly, many individuals had several rather than just one copy of the second gene in the sequence. Protanopes turned out to be missing the first gene of the sequence, which was therefore identified as the L cone pigment gene. Deuteranopes often had only the first gene of the sequence, and the second and later genes were therefore characterized as the M cone pigment genes. More complex genetic patterns – genes made up of pieces from both the L and the M pigment genes – were also found, especially in color-anomalous individuals. Moreover, polymorphisms were found in the normal L and M pigment genes, allowing an explanation of more subtle variations of color vision

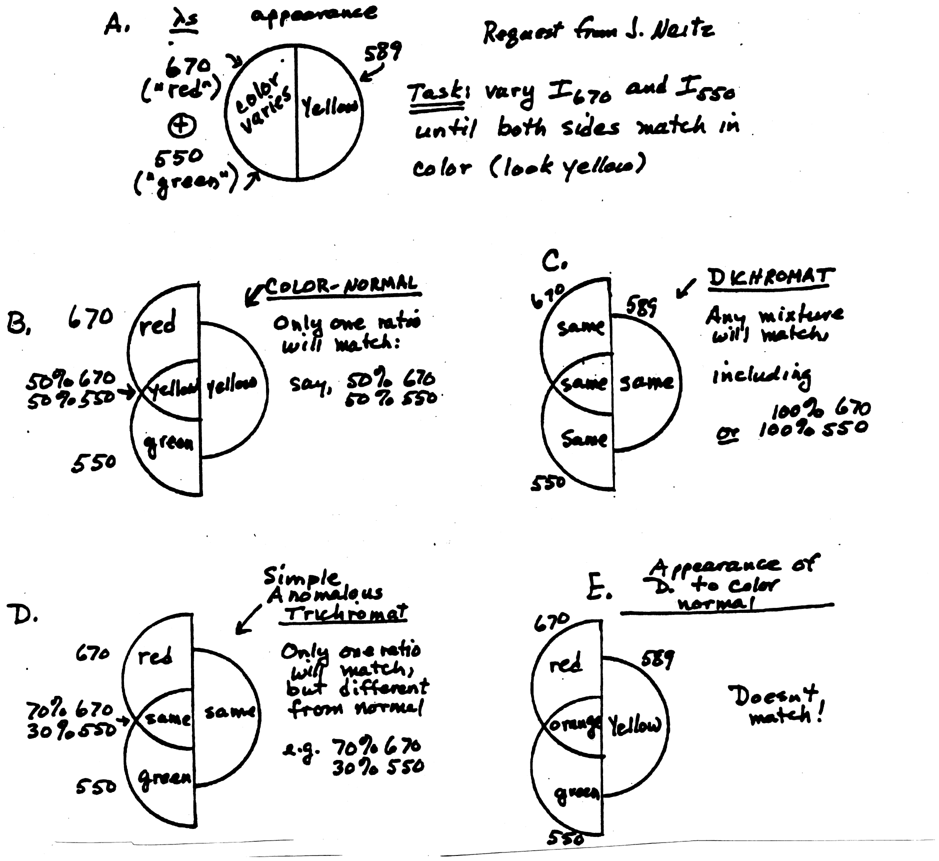


Figure 7.8: COLOR PLATE. Simulated Rayleigh matches. Panel A shows the actual spatial layout of the anomaloscope fields. In panels B-E the two wavelengths in the left half-field have been shifted upward and downward for purposes of illustration. For a color normal subject (panel B), there will be a particular ratio of intensities of the 550 and 670 lights (say, 50:50) that is metameric to the 589 nm light. Both halves of the anomaloscope field will look slightly orangish yellow. For a dichromat (panel C), who cannot discriminate the 550 from the 670 light in the first place, the 589 nm field can be matched by any ratio of the 550 and 670 nm fields, including 100% of either of these wavelengths. In other words, lights that look red, yellow and green to color-normal subjects all look the same – are in the same metamer set – for a dichromat. For an anomalous trichromat (panel D), the normal trichromat's metamers look different colors, but the 550 and 670 lights can be mixed in some *other* proportion (say 70:30) to match the 589 nm light. But (panel E) the half fields that match for the anomalous trichromat look very different to the color normal subject. [Acknowledge J. Neitz, personal communication xx].

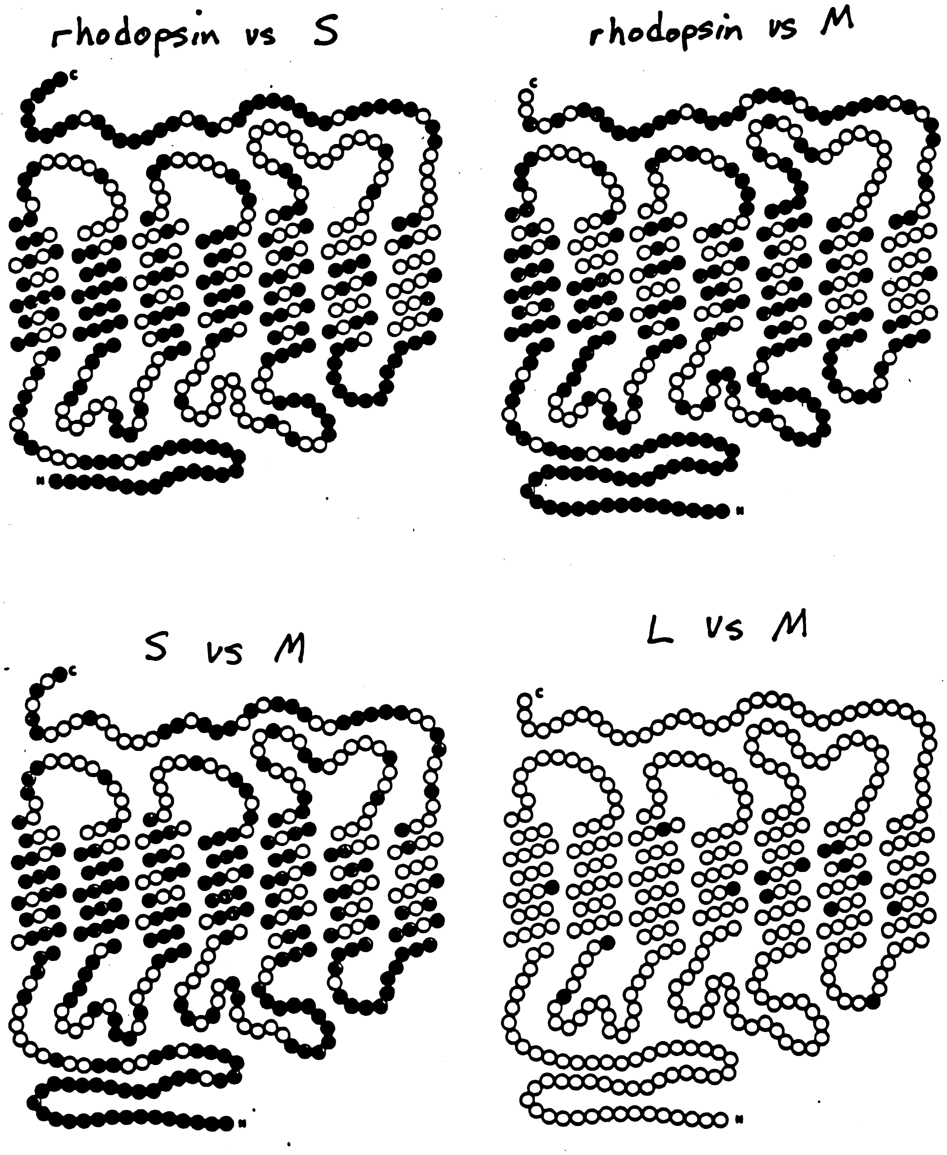


Figure 7.9: Comparisons of the molecular sequences of rhodopsin and the three cone photopigments. Each black dot marks a difference between the two sequences being compared. Note the close similarity of the L vs. M pigments. This similarity is indicative of a recent evolutionary separation. [From Nathans et al (1986), Fig. 11, p. 200.]

among color-normal individuals[4].

7.8 Spatial mosaics for the S, M and L cones

The numbers and distributions of rods and cones across the retina, shown in Figure 6.2, have been known for half a century. But what are the numbers and distributions of each of the three cone types?

These questions have been of major theoretical interest, but the answers proved elusive for many years. The S cone mosaic has been of interest because acuity is poor under conditions that isolate S cones, and it has therefore been speculated that there might be only a small number of S cones. The L and M cone mosaics have been of interest because there are individual differences in photopic spectral sensitivity curves, and it has been suspected that these might be due to individual differences in the proportions of L vs. M cones: the *L/M cone ratio.*

Within the last 20 years or so, the numbers and distribution of S cones has been determined with several different techniques. In the most definitive early study, Christine Curcio and her colleagues (Curcio el al, 1991) used a newly developed stain specialized to reveal the S cone opsin. Use of this stain exposes the whole S-cone matrix.

A sample of Curcio et al's results are shown in Figure 7.10. In their data, S cones provide only about 10% of the cones in the human retina. In fact, although it does not show in the figure, S cones are missing entirely from the central fovea; and they are spaced relatively far apart throughout the rest of the retina.

Why? The suspected design explanation for this sparse representation is that chromatic aberration defocusses the retinal image formed from short wavelength light, so that fine spatial sampling would be wasted in the S cone system. The S cones provide another example of the idea that "poor" optics – in this case chronically defocussed images, due to chromatic aberration – limit our acuity for short wavelength light, and do so through an evolutionary mechanism that matches the spacing of the photoreceptor matrix to the quality of the optical image.

The numbers and distributions of L and M cones are currently being attacked with the techniques of adaptive optics. In our discussion in Chapter 4 we suggested that one of the major uses of adaptive optics will be to *look in* through a person's corrected optics and be able to see the structures in his living retina. Omitting many details, Figure 7.11 shows some of the very first pictures of the distributions of the three different cone types in a living human eye.

Moreover, the three cone mosaics shown in Figure 7.11 can be combined into a single image to reveal the subject's overall retinal mosaic. Figure 7.12 shows pseudocolor images of the retinas of two different subjects. The proportions of S cones were about 5% in both retinas. But the L/M cone ratios differed considerably: 3.8 for JW vs. 1.2 for AN. We will return to these individual differences immediately below.

More recently, adaptive optics have yielded answers to some of the classic questions concerning dichromacy. It has long been speculated that the loss of a photopigment gene could lead to the functional or actual loss of a type of cone: the L cones for protanopes, the M cones for deuteranopes, and the S cones for tritanopes. The images shown in Figure 7.13, taken with adaptive optics in the living eyes of two dichromats, show that this speculation is correct. The retina of the deuteranope

[4] Geneticists (and the rest of us) sometimes slip into talking as though complex human behaviors can be attributed to single genes. A deuteranopic student in one of DT's classes wrote a term paper on the genetics of color vision deficiencies. In parody of the single-gene assumption, his opening sentence was, "I have a gene for mismatched socks".

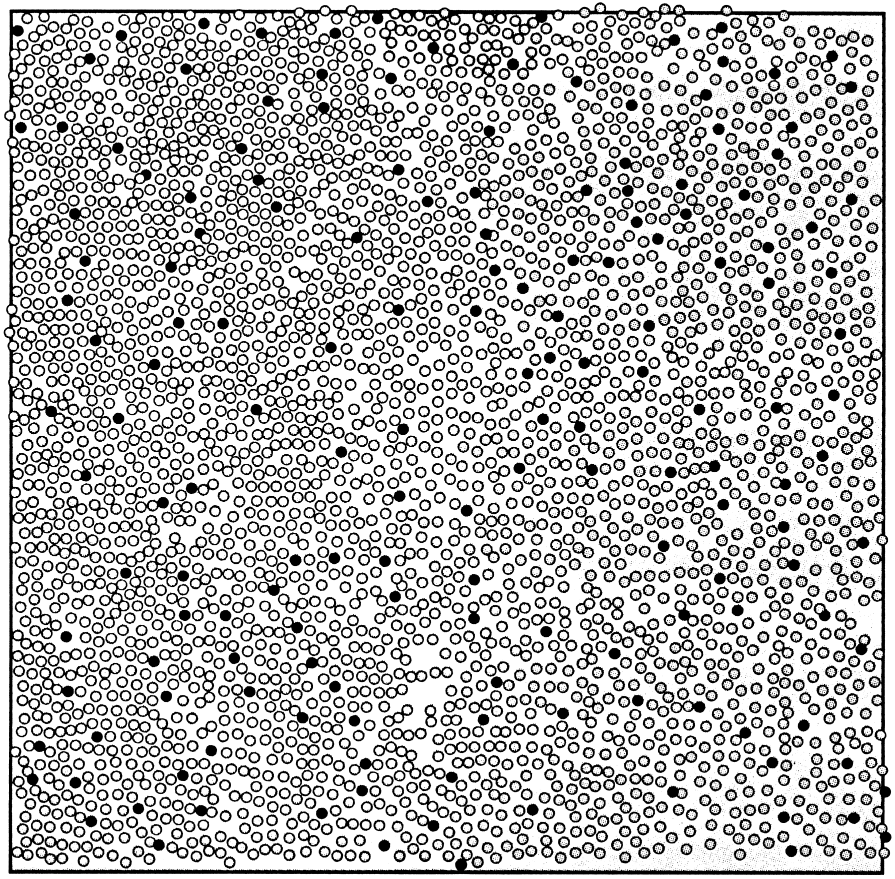


Figure 7.10: The matrix of S cones. The picture shows the distribution of S cones over a small region just off from the center of the fovea. The black dots are S cones; the open dots are L and M cones. [From Rodieck (1998), no fig. number, p. 210; after Curcio et al (1991).

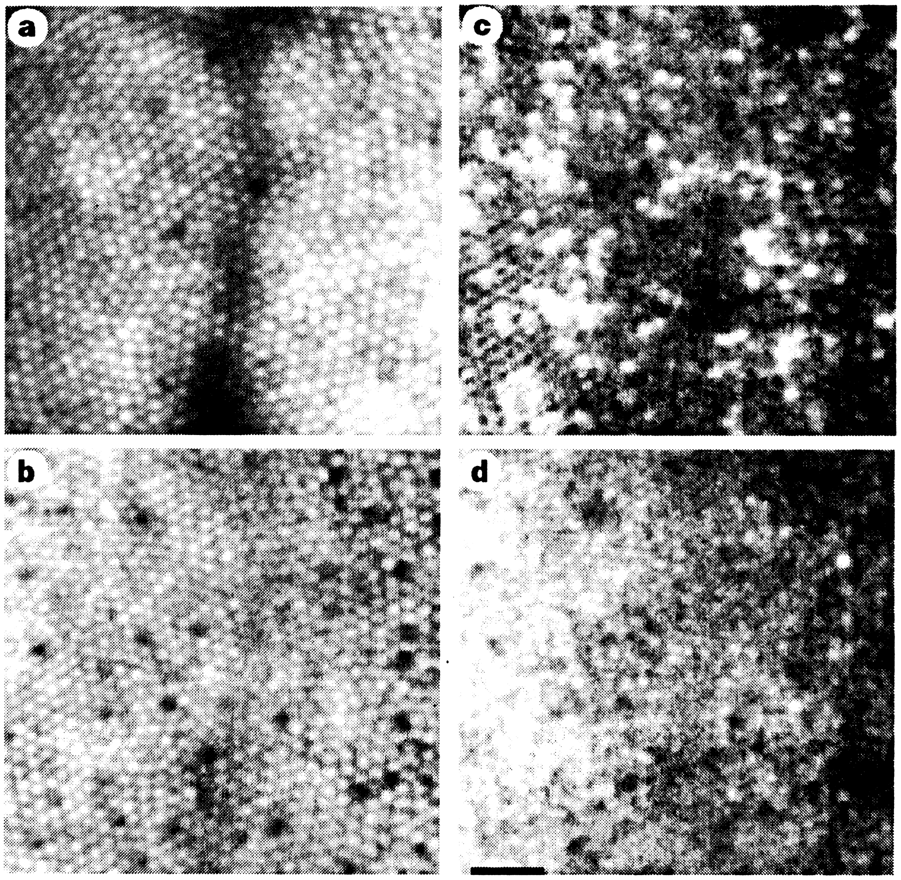


Figure 7.11: Distributions of the three cone types in a color-normal human retina. A: a 1^o patch of the retina of the right eye of subject JW at 1^o eccentricity. The small round dots are individual cone photoreceptors. In Panels B, C, and D the same region of retina was exposed to three different combinations of wavelengths of lights. The different combinations were selected to favor visualization of S cones (the dark spots in B), L cones (the dark spots in C), or, least successfully, M cones (barely visible as the dark spots in D). (In each case the new picture is subtracted from the picture in A. The grey stripe down the middle of A disappears in the subtraction process). [From Roorda and Williams (1999), Fig. 1, p. 520.]

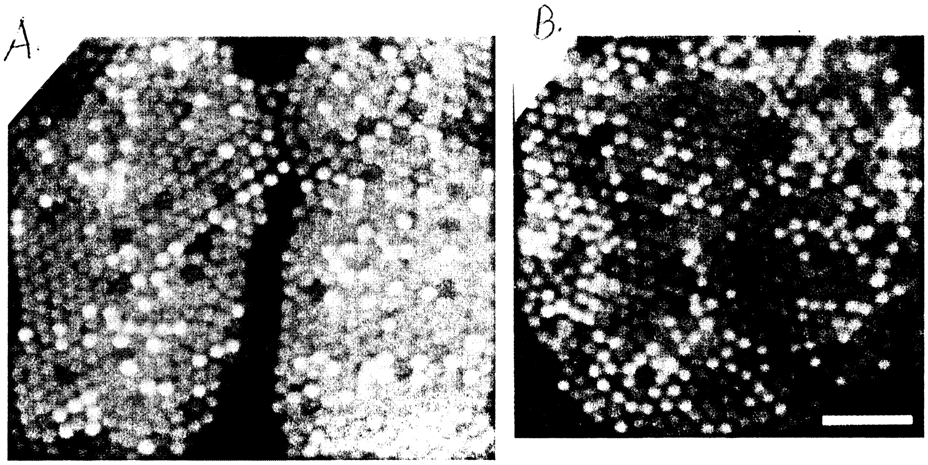


Figure 7.12: [COLOR PLATE]. Cone mosaics in two color-normal subjects. A: the combined matrix for subject JW from Figure 7.11. B: the matrix for a second subject, AN, whose photopic spectral sensitivity curve was elevated in the mid wavelength region, suggesting that he might have a higher than average proportion of M cones. He does. [From Roorda and Williams (1999), Fig. 3, p. 522.]

shown in Figure 7.13A is missing its M cones, and the retina of the protanope shown in Figure 7.13B is missing its L cones.

These images also address a second question. There have been two rival theories for how the loss of a pigment plays itself out at the level of disabling the photoreceptors that would have contained that pigment. *Loss theories* suggest that the cones that would have contained the missing pigment are literally lost from the retina, leaving holes in the mosaic where the missing class of photoreceptors would have been. On the other hand, *replacement theories* suggest that the cones that would have contained the missing pigment are filled with a different pigment – for example, that a protanope would have its potential L and M cones both filled with the M-cone pigment – so that the full complement of cones is retained in the retinal mosaic.

It turns out that both theories are probably right. In studies of the molecular genetics of color vision, two different kinds of genetic changes have been found in different dichromats. A dichromat can be missing the gene for the L or M cone pigment. Alternatively, he can have a mutation that makes the L or M come pigment misshapen and therefore nonfunctional. Perhaps the absence of a gene leads to replacement, and a misshapen pigment leads to loss.

These arguments are supported by the images in Figure 7.13. Two subjects are shown: NC, who is a deuteranope because of a mutant M pigment gene, and MM, who is a protanope because of a missing L pigment gene. The retina of the deuteranope NC is shown in Figure 7.13A. It shows L cones but no M cones, and a reduced overall number of cones, with "holes" between them, as predicted by loss theory. In contrast, the retina of the protanope MM is shown in Figure 7.13B. It shows M cones but no L cones, but a normal number of cones overall. Apparently the L cones have been filled with M cone pigment, and retained in the retinal mosaic, as predicted by replacement theory. Thus ends a controversy that lasted a century – loss of a cone type occurs in some dichromats, and replacement of the missing photopigment by the available one occurs in others.

7.9 Photopic spectral sensitivity

As it turns out, knowing the actual spectral sensitivities of the three Fundamentals provides us with several theoretical bonuses. First, in Chapter 3 we introduced the photopic spectral sensitivity curve, V_λ. Now that the spectral sensitivity curves of the actual L and M cones are available for use, it turns out that V_λ can be readily modeled by a weighted linear sum of L and M cone signals. This idea is illustrated in Figure 7.14, and the physiological model fits the psychophysical data well.

Second, we also mentioned in Chapter 3 that although vision scientists have adopted a standard curve for photopic spectral sensitivity, there are actually small but consistent individual differences in the empirical photopic spectral sensitivity curves for different individual subjects. It has long been speculated that these individual differences could come about from variations in L/M cone ratios among subjects. Inspection of Figure 7.14 reveals that independent sliding of the L and M spectral sensitivity curves up and down will allow interesting changes in the overall photopic curve, and this model provides reasonable fits to the known individual differences.

Moreover, we have had a chance to examine the cone mosaics of two color-normal subjects, JW and AN, in Figure 7.12. The L/M cone ratios differed markedly between these two retinas. The photopic spectral sensitivity curves of these two subjects were also measured, and the differences are in the right direction to be modeled by the differences in L/M cone ratio.

And third, in Chapter 3 we also described the fact that V_λ emerges from many different kinds of psychophysical experiments – flicker photometry, motion photometry and minimally distinct border judgments, among others. We argued that when a characteristic psychophysical ”signature” emerges frequently from the data, it’s a good guess that that characteristic has a physiological instantiation. That is, vision scientists would be drawn to speculate that individual neurons that sum inputs from L and M cones, with spectral sensitivity curves corresponding to V_λ, will be formed within the visual system. However, neurons that embody this prediction are simply not present at the level of the photoreceptors. We are left to speculate that they will emerge at a later level of processing.

7.10 Summary: Photoreceptors and the second transformation

At the end of Chapter 4 we summarized the effects of the First Transformation – from the physical world to the retinal image. We argued that the first transformation rendered the retinal image *two-dimensional* and *low pass filtered.*

In addition to the optics, the incoming visual signal also encounters a stage of discrete sampling by the photoreceptors that rendered it (poetically) *pointillistic.* The discrete sampling stage could be considered either part of optical processing (since the signal is still carried by light), or part of processing by the photoreceptors (since they are the spatially discrete elements). As such, it could be considered part of either the first or the second transformation. We summarized it along with the first transformation at the end of Chapter 4.

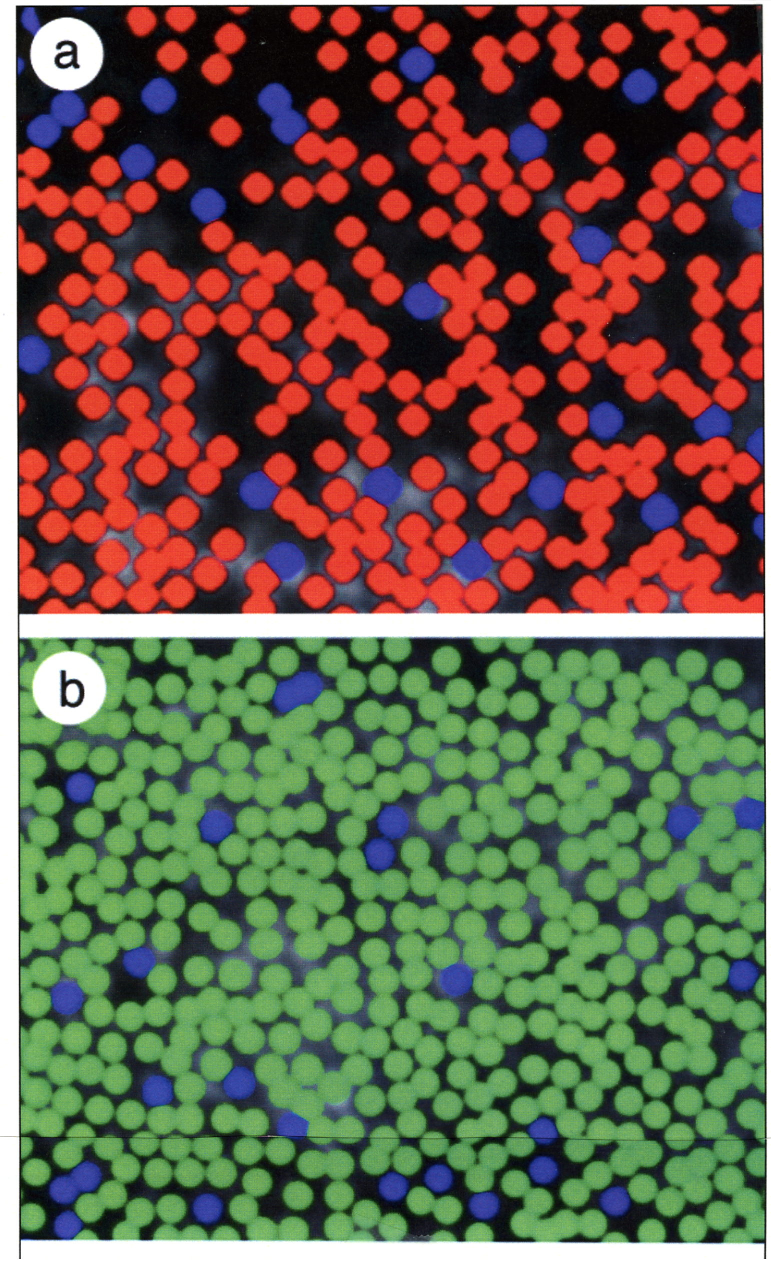


Figure 7.13: [COLOR PLATE]. Cone mosaics in two dichromats. A: Deuteranope NC, missing M cones; B: Protanope MM, missing L cones. NC has a "patchy" retina, with some cones apparently missing, whereas MM appears to have the full complement of cones. [Carroll, Neitz, Hofer, Neitz, and Williams, 2004, Fig. 4, page ??]

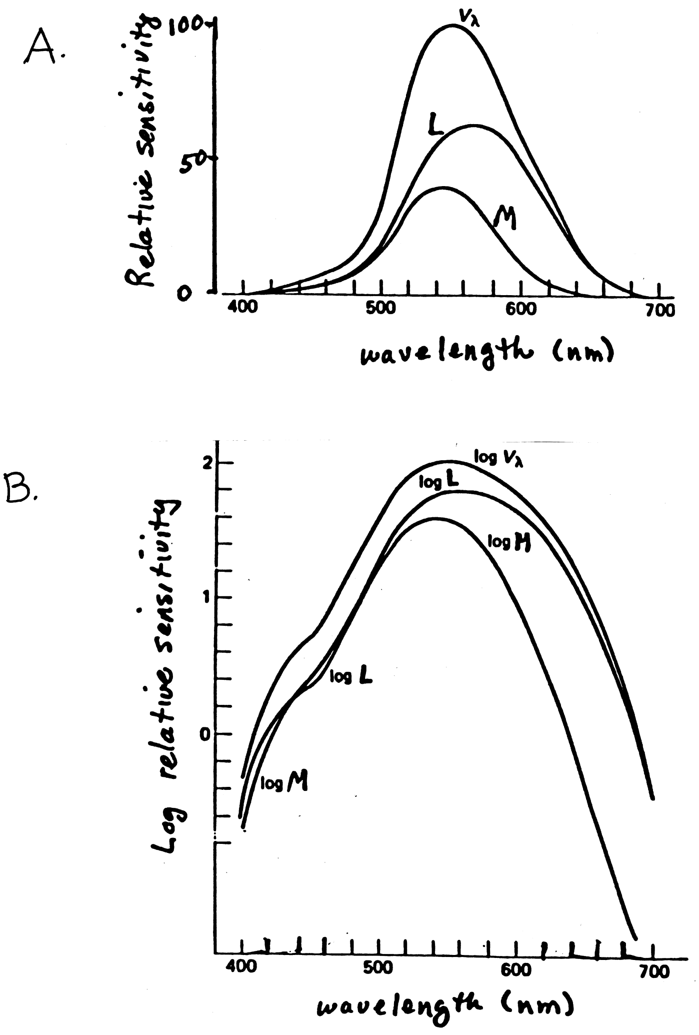


Figure 7.14: A model of V_λ based on a weighted sum of L and M cone inputs. A shows the calculation on a linear ordinate, and B shows it on a logarithmic ordinate. The curves labeled L (or log L) and M (or log M) show the spectral sensitivities of the L and M cones; the curves labeled V_λ (or log V_λ) show the synthesis of V_λ from the sum of the L and M curves. For the average color-normal subject, the weighting needed to fit V_λ is about 2:1 for the L vs. M cone signals; that is, $V_\lambda = 2L + M$. The differential weighting is incorporated into the diagram by increasing the height of the L curve with respect to the M curve. [Modified from Boynton (1979), Fig. 9.3, p. 307.]

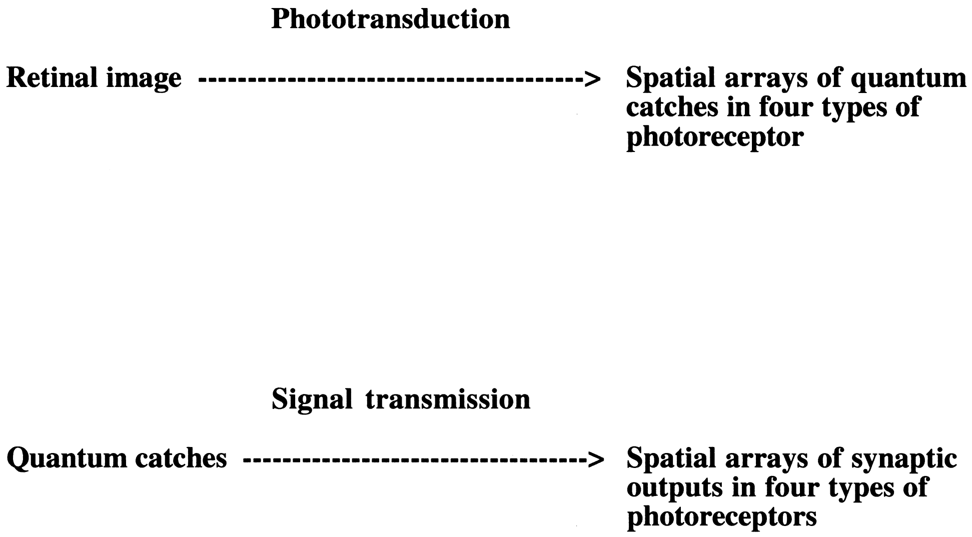

Phototransduction
Retinal image
Spatial arrays of quantum catches in four types of photoreceptor
Signal transmission
Quantum catches
Spatial arrays of synaptic outputs in four types of photoreceptors

Figure 7.15: Photoreceptors: the second transformation.

We are now ready to summarize the (rest of the) second transformation: from the retinal image to the quantum catches in photoreceptors via the phototransduction process, and from the quantum catches to the photoreceptor outputs via a complex set of chemical and electrical information transmission processes. These two stages of processing are summarized in Figure 7.15. In combination these two stages create a spatial array of quantal catches in the rods and L, M and S cones, and transform it into a spatial array of synaptic outputs in the same four kinds of neurons.

The transduction process and the combination of four kinds of photoreceptors provide us with causal stories for some of the system properties of scotopic and photopic vision introduced in Chapters 2 and 3, as well as for the trichromacy of color vision discussed in the present chapter. The scotopic spectral sensitivity curve is maximal at about 500 nm, and wavelength information is lost in scotopic vision, because scotopic vision is served by rods and rods alone. Wavelength information is preserved in photopic vision, up to the limits described by trichromacy, because photopic vision is served by three and only three kinds of cones.

The information transmission process also leaves its mark. For example, at low light levels, each rod is so exquisitely sensitive that it produces a detectable signal in response to the absorption of a single quantum. This sensitivity enables human subjects to detect the absorption of only 5-10 quanta in an extended test field, and therefore of a single quantum in an individual rod. In addition, a saturating non-linearity, probably within the cone photoreceptors, provides a signal that allows the detection of interference fringes in the vicinity of 60 cy/deg. Of course, the properties of photoreceptors influence all aspects of vision, but additional examples are beyond our scope.

Information processing by the three cone types together also illustrates the concept of *ensemble codes* (or *pattern codes*). Because of the nature of transduction, each photoreceptor individually loses wavelength information. Yet, working together as an ensemble, the three types of photoreceptors preserve at least some information about the wavelength composition of each region of the retinal image. By comparing the signals from L, M and S cones, later levels of the system have access to a fair bit of wavelength information. The concept of pattern codes will recur frequently throughout the remainder of this book. Later we will see neurons that compare signals from the L, M and S cones, to create a new wavelength/color code.

Chapter 8

Retinal Processing: Spatial and Temporal Variables

We now return to the question of codes and code transformations. So far we have discussed two (somewhat arbitrarily divided) code transformations that take place early in retinal processing, The first transformation takes place between the physical world and the retinal image, and involves the formation of the *optical image* (or *retinal image*) on the retina. The second takes place between the optical image and the photoreceptor outputs, and involves the formation of what can be called the *photoreceptor image*: the spatially ordered set of outputs from the ensemble of photoreceptors.

In Chapter 8 we arrive at what we will call the *third code transformation*: the recoding of photoreceptor signals within the retina. This recoding takes place between the photoreceptor outputs and the ganglion cell outputs, by means of computations carried out by the retinal interneurons. It produces what we will call the *ganglion cell image* – the spatially ordered set of outputs from the ensemble of retinal ganglion cells. As was the case with the optical image and the photoreceptor image, the ganglion cell image must carry all of the information from the physical world that eventually reaches the visual cortex. The questions of this chapter are, What is the ganglion cell image like? And how is it created within the retina?

The questions may be simple, but of course retinal processing is highly complex. DT's pedigogical strategy is to teach it twice, using history as a tool. First, Chapter 8 provides a relatively simple introduction, by describing three classic studies of retinal processing from the 1950s and 1960s, centering the description around spatial and temporal variables. From these studies you will become acquainted with the major classes of retinal interneurons – the horizontal, bipolar, and amacrine cells – and develop a simplified picture of retinal spatial processing and the origins of the ganglion cell image. Then, in Chapter 9 we pursue some causal stories, trying to account for some of the system properties of human spatial and temporal vision on the basis of properties of the ganglion cell image – the retinal output code.

As it turns out, the early studies of mammalian retinal neurons were not carried out in primates. In fact, the studies of ganglion cells were done in cats, and the studies of interneurons, in salamanders. Thus, for many years, vision scientists' thinking about retinal processing were based on what DT calls the *cat/mudpuppy model*: speculative combinations of information about retinal processing from the data on cat and mudpuppy retinas. At the end of the present chapter we use the cat/mudpuppy model to make some predictions about the responses of retinal ganglion cells to disks and annuli of light.

Recording from mammalian interneurons, including those of primates, became possible in the early 1990s, and since that time we have learned a great deal about primate retinal processing. Thus, in Chapter 10 we retrace our steps, and tell the story of retinal anatomy and neurophysiology over again, but in much greater detail. This time we describe the properties of primate interneurons, emphasizing the major specializations that occur in the primate retina, with particular attention to the question of wavelength processing. Finally, in Chapter 11 we again turn to system properties and causal stories, and ask how the retinal output code leaves its marks on our perceptions of color.

8.1 Anatomical structure of the vertebrate retina

A light micrograph of the cat retina is shown in Figure 8.1. The photoreceptors are at the top, the interneurons in the middle, and the ganglion cells at the bottom. As noted previously, light impinges on the retina from the bottom of the figure, and traverses all of the retinal layers before being absorbed by the photoreceptors at the top.

The first and most immediately striking feature of the cat retina, and vertebrate retinas in general, is its pattern of distinctive layers. Top to bottom, these layers are called the *photoreceptor, outer nuclear, outer plexiform* (or *synaptic*), *inner nuclear, inner plexiform* (or *synaptic*), and *ganglion cell* layers. These names refer to the contents and functions of the layers. The nuclear layers and the ganglion cell layer contain the cell bodies (nucleii) of the photoreceptors, interneurons, and ganglion cells respectively. The plexiform layers contain cell processes and synapses – the outer plexiform layer (OPL) for the synapses among photoreceptors, horizontal cells, and bipolar cells, and the inner plexiform layer (IPL) for the synapses among bipolar cells, amacrine cells, and ganglion cells. The plexiform layers are where the recoding action is: they contain massive concentrations of synapses, and thereby presumably accomplish massive computations and code changes.

A second striking feature of retinal anatomy is that there are far fewer ganglion cells than photoreceptors. That is, the anatomy tells us that signals encoded pointillistically in many neighboring photoreceptors must be combined across space and carried in some other code in the signals of single ganglion cells. The fundamental question of this chapter is: What changes in the spatial code occur within the retina? What is the spatial code by the time we reach the ganglion cells?

8.2 Single unit recording: Eavesdropping on retinal ganglion cells

How can one study the coding properties of individual retinal neurons? The first step is to figure out how to access its ganglion cells while keeping them alive. In cold-blooded vertebrates such as the mudpuppy, it is possible to excise the retina and keep it alive in an artificial medium while studying its neurons. But historically, in warm-blooded animals such as the cat, it was not possible to keep the isolated retina alive, and instead recordings were made in intact, anesthetized animals. Figure 8.2 shows a cartoon of a primitive single unit recording set-up in use on a cat.

The second step is to produce a range of carefully controlled visual stimuli. Most simply a slide projector can be used, with a projection screen (often called a *tangent screen*) placed in front of the animal's eye. The pattern of light can be changed by changing the slide in the slide projector. Alternatively, more elegant optical systems can be used to project patterns of light directly onto the retina. Today video monitors are most commonly used, and the sizes, locations, and sequences of stimuli are controlled by computer.

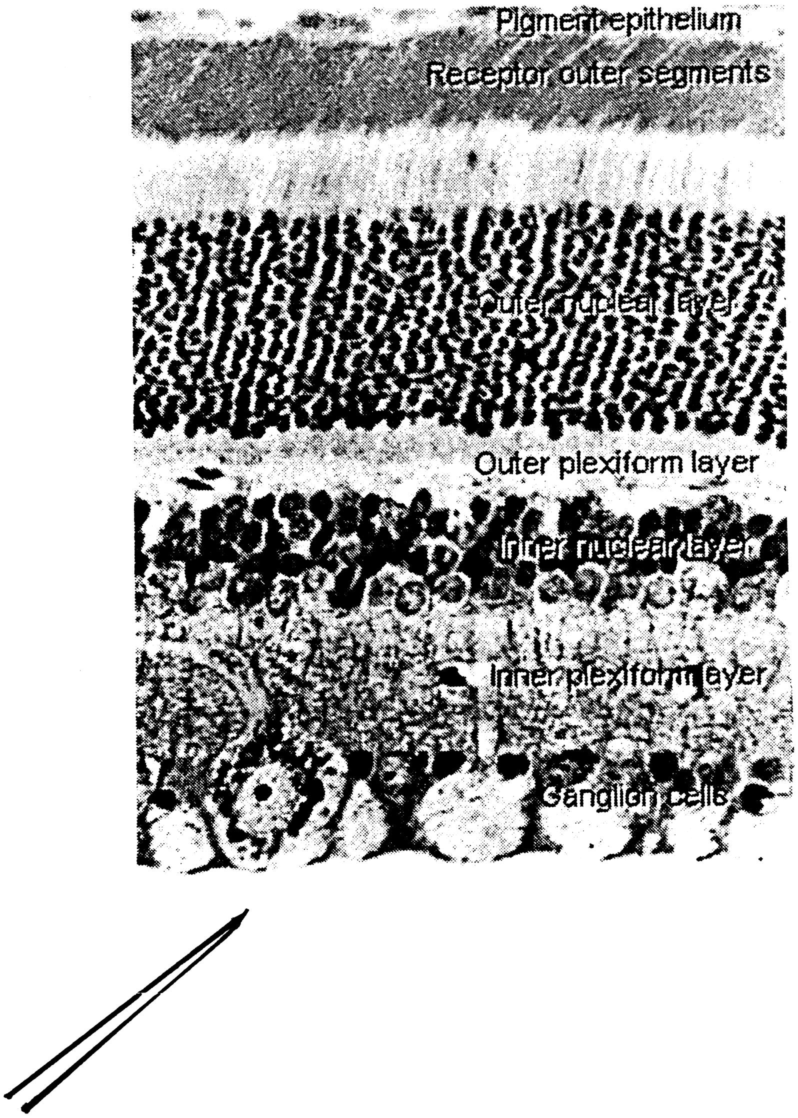


Figure 8.1: The layered structure of the vertebrate retina. The figure shows the retina of a cat. The photoreceptors are at the top and the ganglion cells are at the bottom. The names of the retinal layers are shown. The nuclear layers contain cell nucleii, whereas the plexiform (synaptic) layers contain cell processes and synapses. Light enters from the bottom. [From Levine (2000), p. 50. Permission to be requested. Electrode added by DT.]

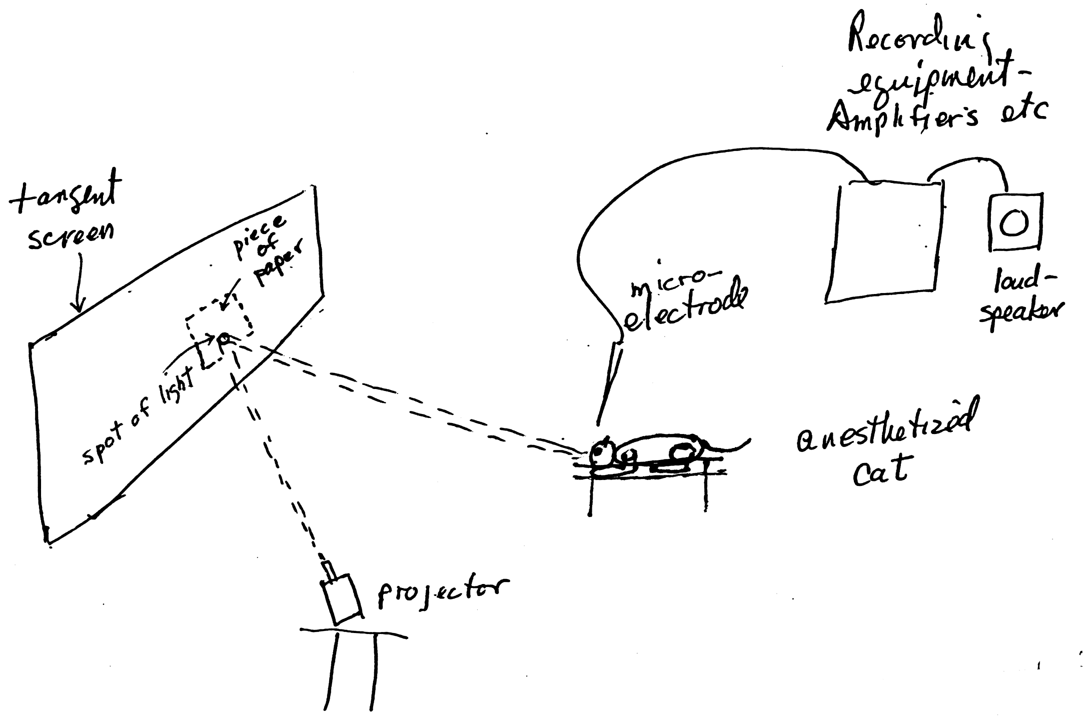


Figure 8.2: Single unit recording from retinal ganglion cells. A spot of light from the projector is shined on the tangent screen and reflected toward the eye of the anesthetized cat. The microelectrode is placed either inside the eye or in the optic tract, and advanced toward the retina or optic nerve. The changes in voltage encountered by the microelectrode are routed to the amplifiers, loudspeaker, and recording equipment. A small spot of light is shined on the retina, and the location of the spot is varied to find the locations and response properties of neurons that respond to it.

The third step is to assemble the apparatus for recording the cells' responses. Unlike photoreceptors, ganglion cells are typical neurons, and fire action potentials (spikes). A tiny wire or hollow glass pipette called a *microelectrode* is placed inside the eyeball or in the optic tract, and used to record the extracellular voltage changes indicative of spike activity. The voltage changes are sent to powerful amplifiers, to allow the experimenter to represent the spikes on a computer screen, and to record them as data. The voltage changes are usually also routed to an audio speaker. When the microelectrode is close to a spiking cell, the experimenter will hear a series of electrical pops, one for each spike generated by the cell.[1]

When the apparatus is ready, the cat is anesthetized and its eye immobilized. The microelectrode is introduced into the inside of the eyeball and moved slowly toward the inner surface of the retina, where the ganglion cells are located. An interesting property of ganglion cells is that they have *spontaneous* (or *maintained*) *activity*; that is, they generate spikes even in the absence of light falling on the retina. Thus, when the microelectrode comes close to the cell body or axon of a ganglion cell, the pops over the loudspeaker announce that a spiking neuron has been encountered.

The final step is to turn on a visual stimulus, such as a small spot of light, and vary its location. We expect, of course, that different ganglion cells will serve different retinal regions, so that if the spot of light is too far from the cell being recorded, the cell will not respond. But if all goes according to plan, a region of the screen will be found within which the spot of light affects the firing rate of the cell. That is, when the spot is in a particular location, the frequency of the series of electrical pops will change – either increase or decrease – as the spot of light is turned on and off.

8.2.1 Receptive fields

We now come to a major new concept: the *receptive field* of a visual neuron. A visual neuron's receptive field is defined as *that region of the retina within which light affects the firing rate of the neuron.* In functional terms, the locations of the photoreceptors that provide input to a visual neuron determine its receptive field.

To plot the receptive field by hand, we present the spot of light at a range of different locations on the screen, as shown in Figure 8.3A. We find a region within which the activity of the cell is influenced by the spot of light, and attach a sheet of paper to the screen in this general location. Now we explore the receptive field more carefully. We put a (+) on the paper in each location at which the spot of light increases the firing rate of the cell, and a (-) in each location at which the spot of light decreases the firing rate. The combination of these two areas defines the receptive field of the cell.

8.3 Three major recodings in the cat retina

The earliest extensive recordings of the responses of cat ganglion cells were published by Stephen Kuffler in 1953. Kuffler was interested in exploring the influence of spatial and temporal parameters on the firing patterns of retinal ganglion cells. He was particularly interested in the interactions of

[1] Sometimes real world signals intrude on the pursuit of science. One cold winter night Mo Powers and Dan Green were recording from single neurons in a rat retina. All of a sudden, over the amplifiers and loudspeaker came a voice: "Rose, Dusty Rose, come in Rose; come in Dusty Rose....". The voice continued for several hours, eventually forcing Mo and Dan to give up recording for the night.

excitatory and inhibitory processes within the receptive fields of these neurons. Kuffler began by recording the firing patterns elicited by small spots of light placed in various locations within the receptive field.

8.3.1 Spatial processing: Center-surround antagonism

Kuffler immediately found two characteristics of ganglion cell responses that are not seen in photoreceptors. First, the ganglion cell's receptive field covers an extended region of the retina. Typically, hundreds of neighboring photoreceptors influence the firing rate of a single cat ganglion cell. And second, the receptive field is not uniform: light on some areas of the screen *increases* the firing rate of the cell, but light on other areas of the screen *decreases* it! These two properties establish immediately that a major recoding is carried out in the retina, and we are forced to bid a fond fairwell to any lingering thoughts of a pointillistic representation of the visual world at the level of the retinal ganglion cells.

The property of center/surround antagonism is shown in more detail in Figure 8.3. Figure 8.3A and B show a cartoon of raw data, whereas Figure 8.3C and D show stylized representations of receptive fields. Kuffler showed that the receptive fields of most cat ganglion cells are made up of two distinct regions: a central more or less circular area, the *center*, and a surrounding annular region, the *surround.* In some cells, light in the center makes the cell fire faster, whereas light in the surround makes it fire more slowly. Other cells have the opposite pattern: light in the center makes the cell fire more slowly, whereas light in the surround makes it fire faster. In all cases, whichever the center is, the surround is the opposite; center and surround act in opposition to each other. This cancellation property is called *center-surround antagonism, spatial antagonism*, or sometimes *spatial opponency*, and it is a critical feature of the processing of spatial information in the retina.

Because of their complementary polarity, the two receptive field areas tend to cancel each others' effects: light on both areas together returns the ganglion cell toward its maintained firing rate. Often the input from the center is stronger than the input from the surround, so that for a homogeneous field of light, the response of the center continues at a diminished level. But if the center and surround are exactly in balance, a homogeneous field of light can yield no change at all in the response of the cell.

Parallel processing: ON-center vs. OFF-center cells

A second, rather puzzling property of retinal ganglion cells was also revealed by Kuffler's recordings: the separation of ON-center and OFF-center types. Both ON-center and OFF-center cells have center-surround antagonism. They differ in that ON-center cells *increase* their firing rate to light falling in the receptive field center, and *decrease* their firing rate to light falling in the surround. OFF-center cells have the opposite pattern: they *decrease* their firing rate to light falling in the center, and *increase* their firing rate to light falling in the surround.

Thus, the onset of a spot of light at a particular retinal location, falling in the centers of the receptive fields of two neurons with overlapping receptive fields, will yield two different but apparently redundant signals, These are an ON response from the local ON-center cells, and an OFF response from the local OFF-center cells. Similarly, the offset of the spot will yield an OFF response from the local ON-center cells, and an ON response from the local OFF-center cells. [Work out the signals that would result from an annulus of light that fell on the surrounds of these two neurons.]

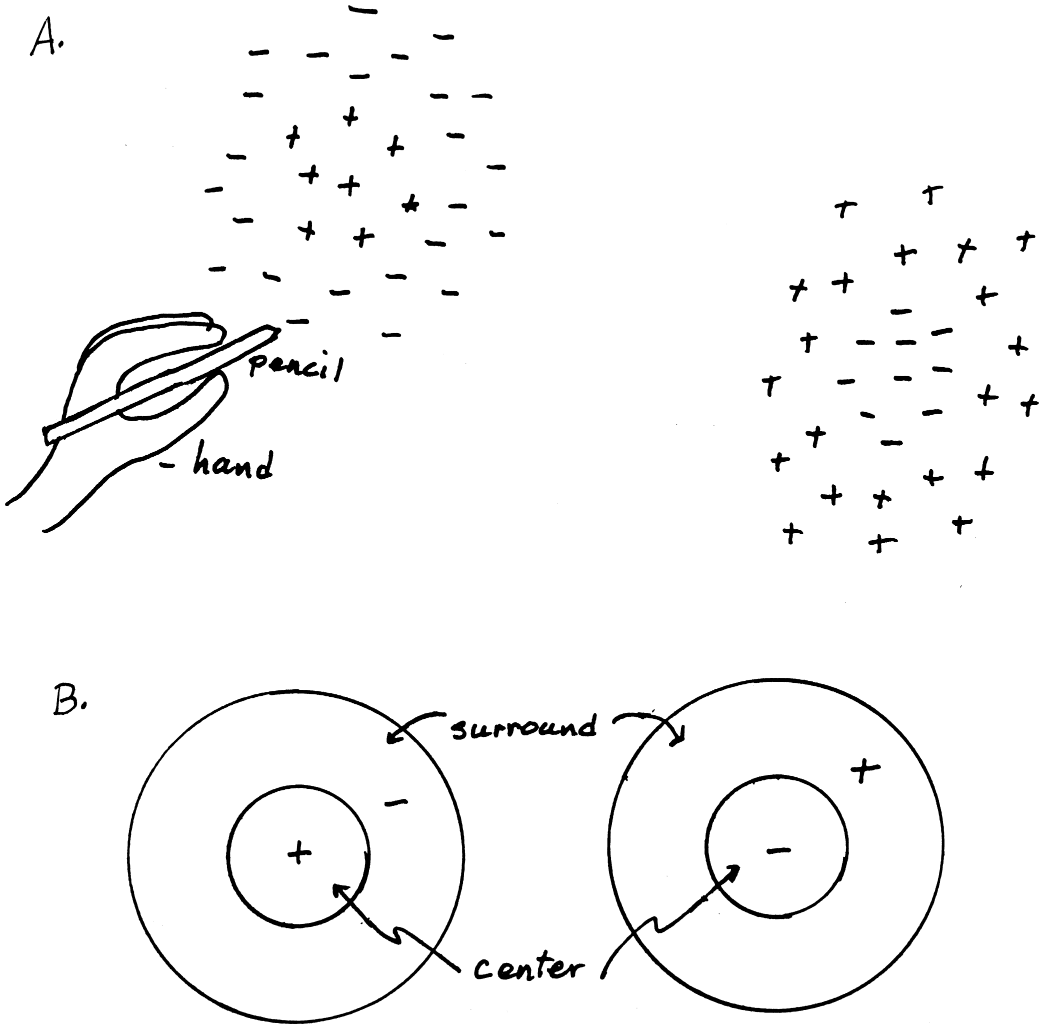


Figure 8.3: Receptive fields of cat ganglion cells. A. Plotting a receptive field by hand. The experimenter records the responses of a single ganglion cell to spots of light falling in different locations on the tangent screen. A (+) signals an increase in the firing rate, and a (-) signals a decrease. The cell on the left has an ON center and an OFF surround, whereas the cell on the right has an OFF center and ON surround. B. Stylized representations of receptive fields of ganglion cells with center-surround antagonism.

8.3.2 Temporal processing: Sustained vs. transient responses

A third set of properties of the cat ganglion cell code concerns the temporal patterns of the responses. As had others before him, Kuffler distinguished among three different temporal response patterns: ON, OFF, and ON-OFF. Samples of these patterns, all recorded from the same neuron, are shown in Figure 8.4A. All three response patterns are defined in terms of increases of firing rate. The *ON response* is an increase in firing rate in response to an increase in stimulus luminance; the *OFF response* is an increase of firing rate in response to a decrease in stimulus luminance; and the *ON-OFF response* is an increase of firing rate both in response to the onset and in response to the offset of a stimulus.

Kuffler also found that ON and OFF responses could be either *maintained* (or *sustained*) throughout the change of luminance, or more *transient*, with the increased firing rate occurring only briefly after the change of luminance (as does the ON-OFF response). The ON and OFF responses shown in Figure 8.4A are actually rather transient, lasting less than the full duration of the stimulus. Figure 8.4B shows an example of a more sustained ON response, along with more transient ON responses recorded from the same cell.

Kuffler showed that a single ganglion cell can produce all three response patterns, and both sustained and transient responses, depending on the location of the stimulus within the receptive field. Thus, he suggested that the different patterns arose from subtle variations in the combination of inputs from the center and surround regions, with the ON-OFF responses, for example, originating near the center-surround border.

In sum, Kuffler's study revealed many novel properties of mammalian ganglion cell activity. We will emphasize three of them throughout this chapter and the next: center-surround receptive fields; differentiation of ON and OFF pathways; and sustained vs. transient responses.

8.4 Neurons do computations

It should be obvious by now that ganglion cells with center-surround receptive fields perform a kind of spatial arithmetic. As such, they illustrate the fundamental point that neurons do not just pass on their inputs unchanged. Neurons do computations, and they signal the results of those computations to the next set of neurons up the line.

We will pause to expand on this important principle with several examples, which will also serve to develop your intuitions about the computations done by neurons with center/surround receptive fields. For simplicity, we will use ON-center cells for these examples, and assume for the moment that they behave more or less linearly in combining signals within their receptive fields.

8.4.1 Addition and subtraction

Instead of confining our stimuli to single small spots of light, what happens if we present two small spots at the same time, as shown in Figure 8.5A? If the two spots are both within the center of the receptive field, Kuffler showed that they will *summate* their effects – the first spot will increase the firing rate, and the second spot will increase it again. Two spots within the surround will also summate their (negative) effects – the first will decrease the firing rate, and the second will decrease it again. And if we put one spot in the center and the other in the surround of the receptive field, the effects of the two spots will *subtract*, and tend to cancel each other out.

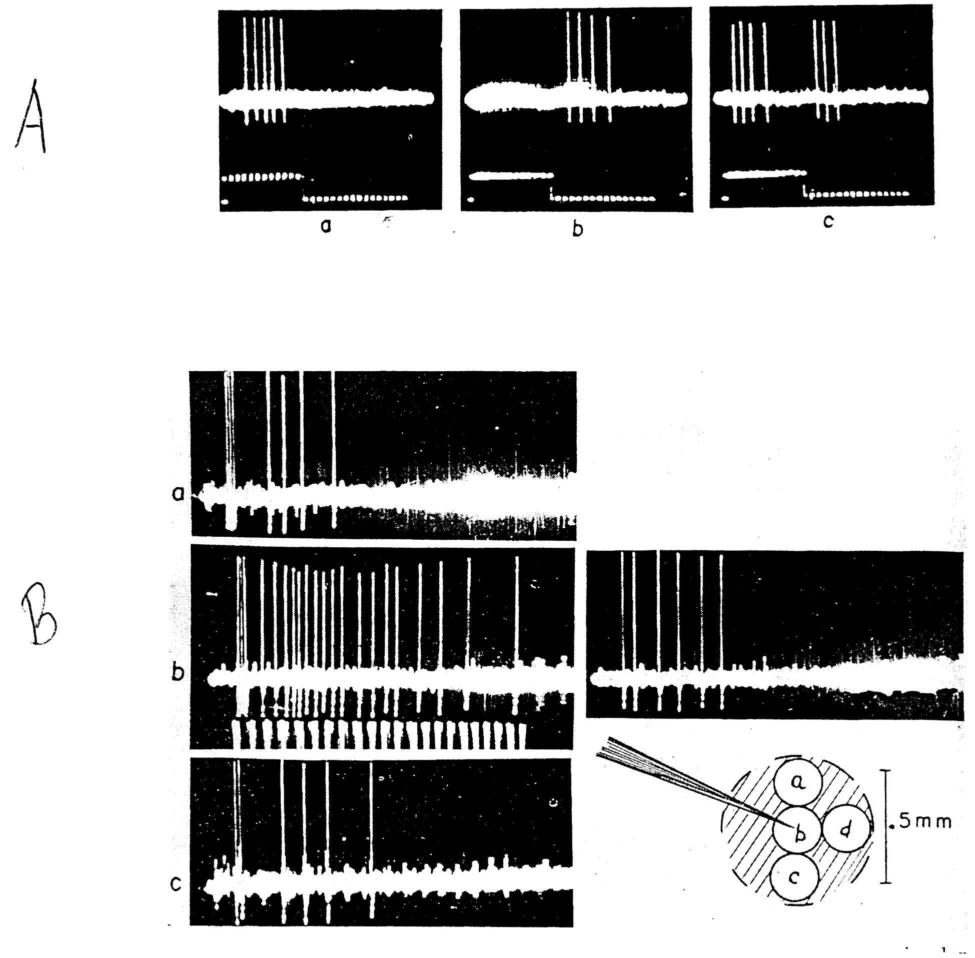


Figure 8.4: Responses of cat ganglion cells to small spots of light. A: ON, OFF, and ON-OFF responses. The lower trace in each panel shows a change (an increase or decrease) in the luminance of the spot. The duration of the increase or decrease was about 500 msec. The upper traces show the resulting patterns of spikes. The left, middle, and right panels show ON, OFF, and ON-OFF responses respectively. All three responses were recorded from the same ganglion cell. B: Sustained vs. transient responses. Responses were recorded from four locations within the receptive field center of a single ganglion cell. The most central location, b, gives the most sustained response. [From Kuffler (1953). A: Fig. 4, p. 47; B. Fig. 5, p. 48.]

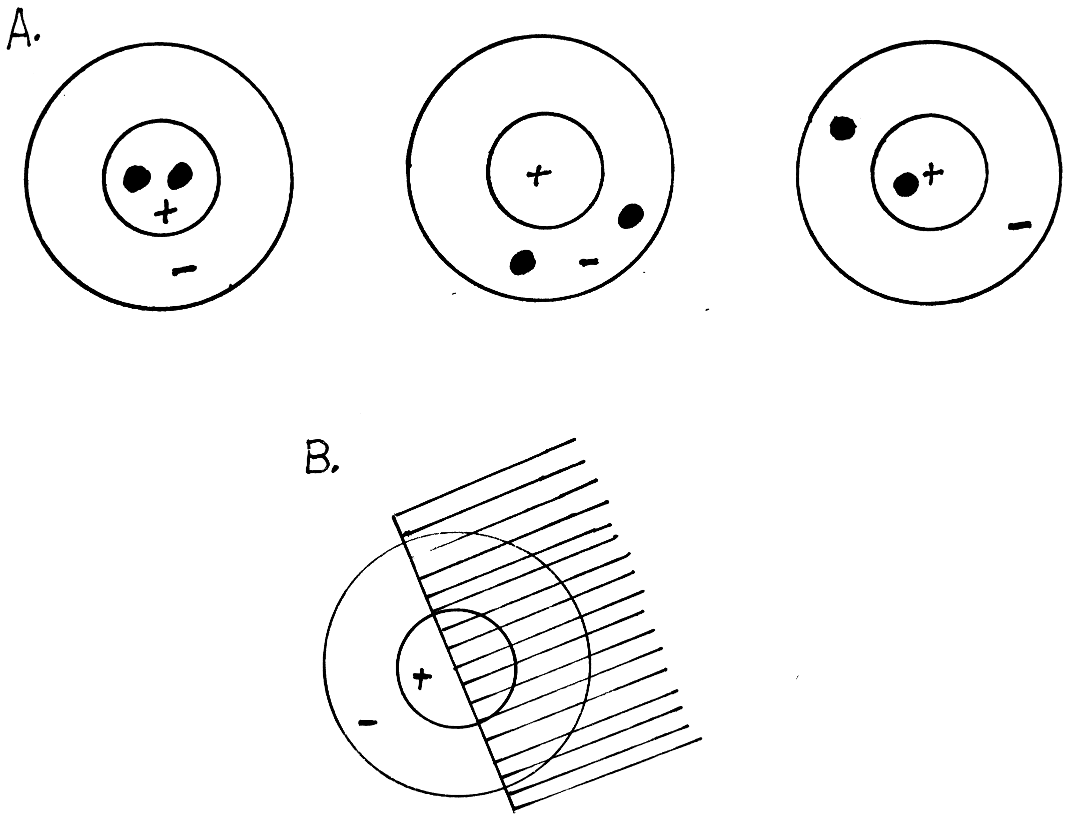


Figure 8.5: Neurons do computations. Spatial configurations of stimuli used to reveal the addition and subtraction properties of center/surround receptive fields. A. Pairs of spots. Two spots within the center add their effects, as do two spots within the surround. Spots presented to the center and the surround simultaneously cancel each other's effects. B. If the cell's arithmetic is linear, and center and surround are in balance, replacing a uniform field with an edge that passes through the center point of the receptive field leaves the firing rate unchanged.

As a second case, imagine a light/dark edge that passes down the middle of the receptive field, as shown in Figure 8.5B. If everything is simple, the effects of the light on half of the center and half of the surround should cancel, so this cell should not respond to the edge. Moreover, the same argument should hold for edges of all different orientations – vertical, horizontal, or any diagonal – as long as they are centered on the receptive field. [Thought question: If the edge is moved away from the center of the receptive field, will the cell respond? Work out the responses of the cell to edges placed at a series of different spatial locations across its receptive field.]

And as a third case, consider the cell's *area-response function*, as shown in Figure 8.6. Suppose we were to start with a tiny disk of light centered on the receptive field of a prototypical ganglion cell, and gradually increase its size. As the disk gets bigger, the effects of light within the receptive field center combine additively. The cell's firing rate will increase with the size of the disk, until the edge of the spot reaches the boundary of the receptive field center. Then, as the light from the disk begins to invade the receptive field surround, subtraction kicks in. The greater the size of the disk, the greater the subtractive effect. If center and surround processes are exactly in balance, the cell will return to its original maintained firing rate as light from the disk reaches the outer edge of the surround. Further changes in disk size only serve to put light outside the receptive field, so they have no effect on the activity of the cell. In short, the maximum response of the ganglion cell occurs for a disk of light that just exactly covers the center of its receptive field, and the idealized ganglion cell does not respond at all to a uniform field.

8.4.2 Equivalence classes and ensemble codes

Finally we return to the problem of equivalence classes, or null sets. Remember that ganglion cells, like photoreceptors, are univariant – a ganglion cell's only output is a spike rate, and many different stimuli will produce a given spike rate. Therefore, like photoreceptors, ganglion cells must also have equivalence classes.

For example, consider the patterns of dots in Figure 8.5. There are many combinations of sets of dots in the center of the receptive field (Figure 8.5A) that will yield the same firing rate as will a disk of light of a particular intensity that just covers the center of the receptive field. Variations of disk size and intensity will also yield new members of the equivalence class, as will many combinations of spots (or disks and annuli) falling on both center and surround. In fact, for an ON-center cell, responses to the onset of light on the center and the offset of light on the surround can be in the same equivalence class. [Work out some more examples of equivalence classes for ganglion cells with center-surround antagonism.]

It is important to think through the implications of the fact that equivalence classes exist for individual neurons. Logically, like the photoreceptor with matched quantum catches from light of two different wavelengths, the individual ganglion cell (or any neuron) confounds all of the stimuli within a given equivalence class. Clearly, if we are to discriminate among the stimuli that fall in an equivalence class for one neuron, we must do so on the basis of the responses of other neurons in the ensemble; that is, we must rely on an ensemble code. And as we change the properties of individual neurons from one level to the next, it is worth remembering that all we can do is exchange neurons with one set of equivalence classes for neurons with another.

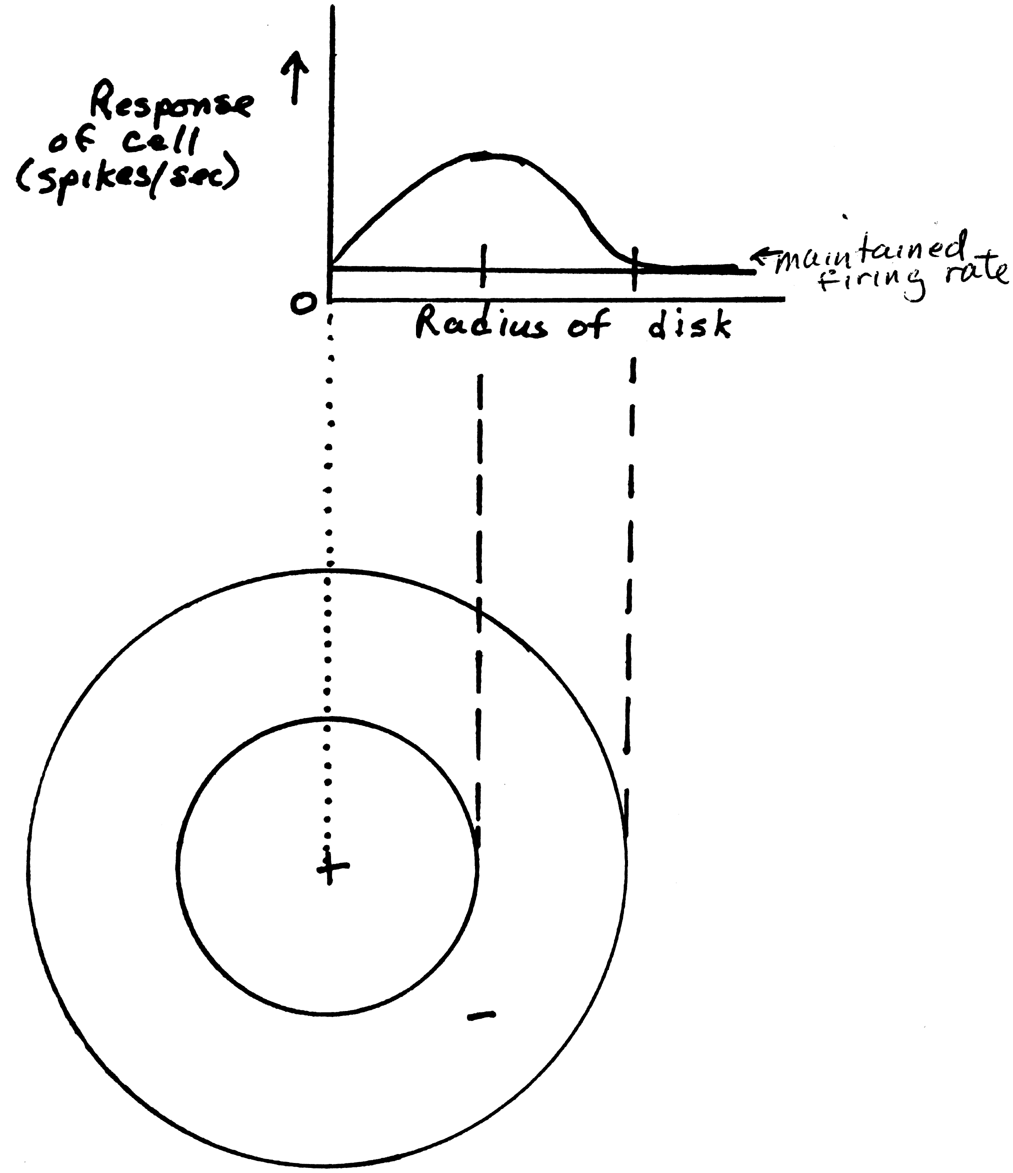


Figure 8.6: An area-response function. The lower part of this figure shows the receptive field of a well-balanced ON-center ganglion cell. The upper part shows the response of the cell to disks of light of different sizes. As the size of the disk increases, the response of the cell first increases, and then decreases back to its original maintained firing rate.

8.5 Spatial weighting functions: DOGs in cats?

The *spatial weighting function* of a cell provides a quantitative description of the manner in which the cell weights inputs from the different parts of its receptive field. For example, it could be that all of the receptors included within the receptive field center are equally weighted and excitatory; all of the receptors within the surround are equally weighted and inhibitory; and for a homogeneous field of light, the summed signals from center and surround exactly balance and cancel each other out. A squared-off, perfectly balanced spatial weighting function of this kind is shown in Figure 8.7A.

In 1965, Robert Rodieck and Jonathan Stone proposed a more sophisticated model. They suggested that the center and surround processes might each be Gaussian curves (also called normal or bell-shaped curves), with the surround broader than, but concentric with, the center. This model is shown in Figure 8.7B. In this model, for both the center and the ”surround” process, the light is most effective at the center point of the receptive field. The ”surround” process actually continues through the center region of the receptive field (setting a terminological trap for the unwary). When the two Gaussians are subtracted from each other, the result is a *difference of Gaussians*, or *DOG* function, as shown in Figure 8.7C.

As it happens, DOG functions provide a reasonably accurate description of the spatial weighting functions of some real ganglion cells, and are also very mathematically tractable. For this reason, in mathematical models of retinal processing, DOG functions are frequently used (even in cats) to represent the spatial weighting functions of ganglion cells with center-surround antagonism. Similarly, an array of DOG functions is often used to represent the array of ganglion cell receptive fields across the retina.

8.6 More about the cat retina

8.6.1 X cells and Y cells

The use of sinusoidal gratings as stimuli came into vision science in the early 1960s. In a classic study in 1966, Christina Enroth-Cugell and John Robson used sinusoidal gratings to characterize the contrast sensitivity functions of individual ganglion cells, and to study their spatiotemporal properties. Notice that sinusoidal gratings are very different from the stimuli used by Kuffler, in that sinusoidal gratings cover a broad region of the retina that includes the whole receptive field of a ganglion cell, rather than concentrating light in a small spot, or on the center or the surround.

One of the major stimulus sets used by Enroth-Cugell and Robson is illustrated in Figure 8.8. In each case the stimulus sequence started with a uniform field, changed to a grating, and then changed back to a uniform field. As shown by the diagrams in the center of Figure 8.8, the grating was positioned at various spatial locations with respect to the receptive field of the cell. Thus, on each stimulus presentation some portions of the receptive field were subjected to an increase in light level, whereas others were subjected to a decrease.

Two of the positions of the grating – positions 1 and 3 – lead to changes in the average illumination on the center and on the surround of the neuron, so the neuron was predicted to respond to the onset and/or the offset of the grating. But importantly, for the other two positions – 2 and 4 – a light/dark transition in the grating straddles the center of the receptive field, and its onset or offset leads to no change in the average light falling on the center or on the surround. Thus, if

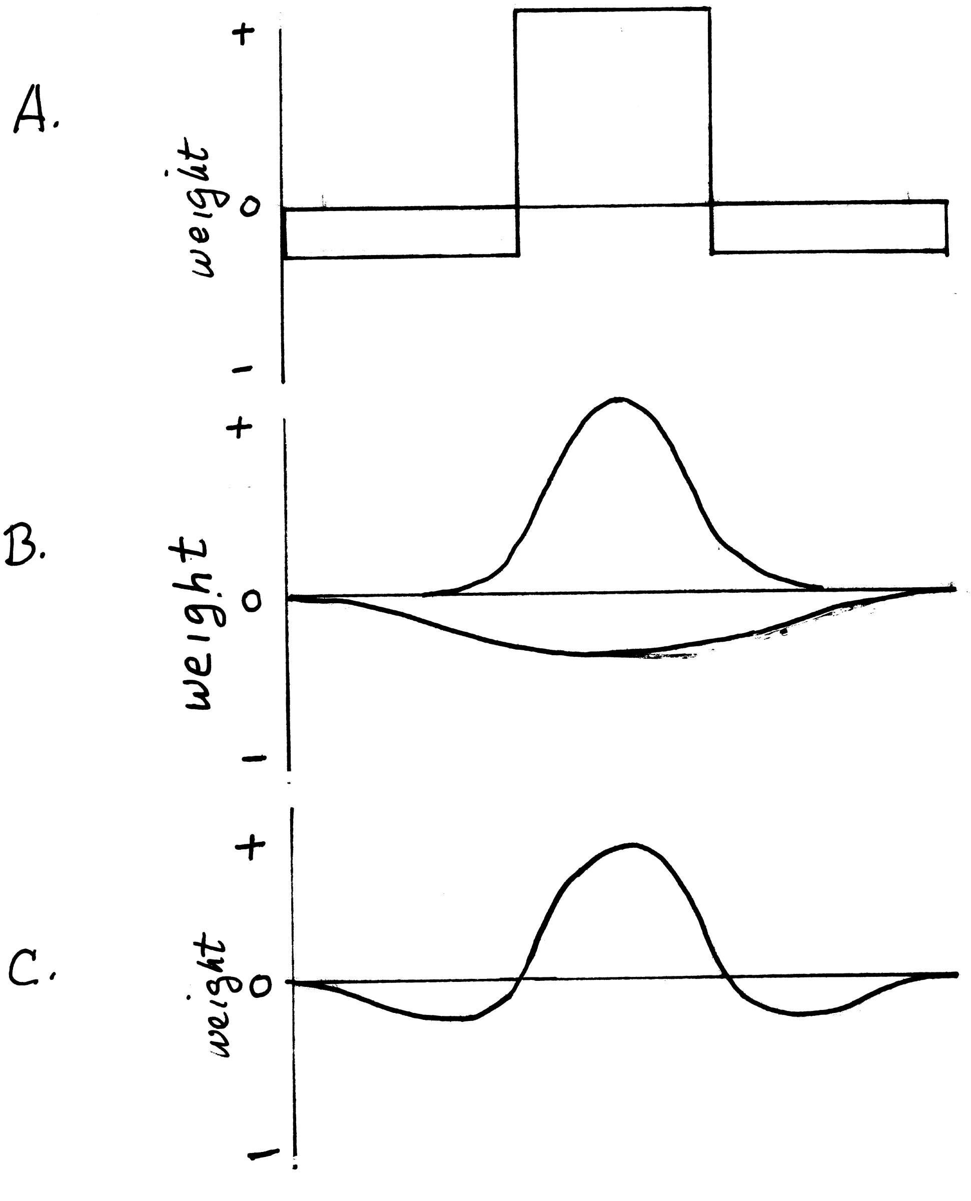


Figure 8.7: Spatial weighting functions. A spatial weighting function describes the weights that are assigned to light at each location within the receptive field. A. A squared off weighting function, like those implicit in our previous drawings of center/surround receptive fields, with abrupt transitions between center and surround. The weights assigned to the surround are smaller because the surround covers more area on the retina. B. A weighting function made up of two gaussian (bell-shaped) curves, a narrower one for the center process and a broader one for the "surround" process. C. A difference-of-gaussians (DOG) function resulting from subtracting the center and "surround" processes in B.

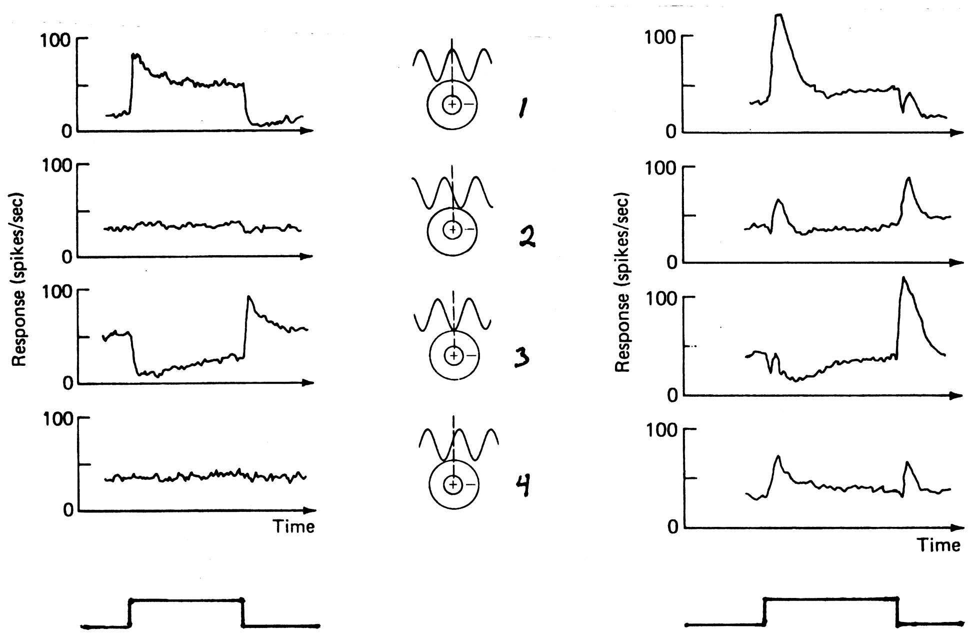


Figure 8.8: Responses of X and Y cells to sinusoidal gratings. The central column of this figure shows an ON-center receptive field, with a sinusoidal grating presented at four different locations. The traces at the bottom of the left and right columns show the onset and offset of the grating; that is, the exchange of the grating for a uniform field of matched space-average luminance. In locations 1 and 3 a single bar of the grating is centered on the receptive field, and at locations 2 and 4 a light-dark boundary straddles the center . Cells with linear summation properties will not respond in cases 2 and 4. The cell on the left, an X cell, behaved linearly; the cell on the right, a Y cell, did not. [After Enroth-Cugell and Robson (1966), Fig.1, p. 523.]

the neuron is linear, these two positions are potentially null positions for the grating.

The responses of an ON-center ganglion cell are shown on the left side of Figure 8.8. The cell responded to the onset of a light or a dark bar centered on its receptive field, but indeed showed null positions – failed to respond – whenever a light/dark transition straddled the center. Enroth-Cugell and Robson argued that these neurons could be modelled by combining photoreceptor signals linearly (see below). They named these neurons *X cells.*

Other neurons, however, were not so well-behaved. Responses from a second ON-center cell are shown on the right side of Figure 8.8. This neuron produced a large ON transient when the bright bar was centered in its receptive field, and a large OFF transient when the dark bar was centered. When the light/dark transition straddled the center, however, no null position was found; instead, the cell produced similar transient responses at both onset and offset. Enroth-Cugell and Robson argued that these neurons are importantly non-linear. They named these neurons *Y cells.*

Enroth-Cugell and Robson also found many ganglion cells in the cat retina that didn't fit into either the X or the Y category. These cells had a wide variety of characteristics, and were much less systematically studied at the time. They were lumped together and called *W cells.*[2] More recent studies reveal many distinct types of W cells, with many different destinations in the brain.

Inspired by Enroth-Cugell and Robson, many other vision scientists studied X and Y cells, and many consistent differences between the two types of cells were found. X cells have relatively small receptive fields and relatively sustained responses, and are most common in the central part of the cat's retina. They constitute about 55% of cat retinal ganglion cells. Y cells, on the other hand, have relatively larger receptive fields and relatively transient responses, and occur throughout the cat's retina. They constitute only about 5% of cat retinal ganglion cells. The remainder, about 40% of cells, are W cells.

Enroth-Cugell and Robson also recorded the responses of X and Y cells to sinusoidal gratings of many different spatial frequencies, and plotted contrast sensitivity functions (CSFs) for individual cat retinal ganglion cells. Some of these contrast sensitivity functions are shown in Figure 8.9. Figure 8.9A shows the CSF from the X cell with the highest high-frequency cut-off. Figure 8.9B shows two CSFs from a Y cell, scored by two different response criteria.

Based on data like these, Enroth-Cugell and Robson developed an influential model of cat retinal ganglion cell receptive fields. They assumed, as did Rodieck and Stone, that the spatial weighting functions for both center and surround processes are gaussian functions. Each model neuron sums photoreceptor signals across the receptive field center, and separately across the receptive field surround. The two signals converge on the ganglion cell, and produce a DOG function in the cell's response.

Aside from its intrinsic interest, why was the discovery of X and Y cells so important? The major reason is that they provided a successful solution to the difficult problem of cell classification. It is easy to observe a population of neurons that varies more or less continuously in some characteristic from one end of a continuum to the other. It is always tempting to divide the continuum in some more or less arbitrary place and name the two resulting clumps of cells with different names, and many examples of this practice can be found in the visin literature. But it is another thing entirely to show that two kinds of cells have characteristics so non-overlapping that they deserve to be

[2] When Enroth-Cugell and Robson first submitted the paper describing their results, they called the X cells D, for dull and the Y cells I for "Interesting". An anonymous reviewer, uncharmed by whimsey, suggested the more prosaic terms X and Y respectively. W cells are widely known as "Wastebasket" cells, because in the early days, if you were studying X and Y cells and you encountered a W cell, you just put it in the wastebasket and moved on.

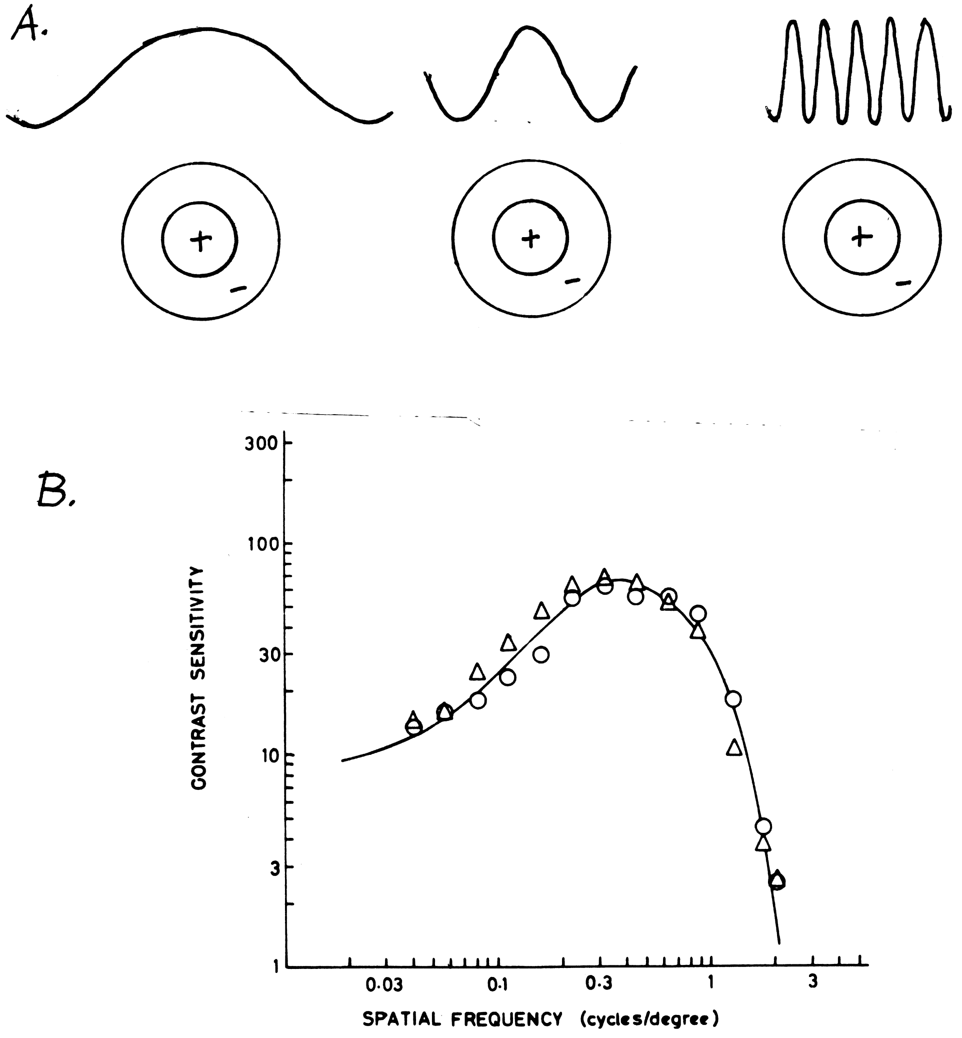


Figure 8.9: Contrast sensitivity functions of cat retinal ganglion cells. A: An ON-center X cell; B: An OFF-center Y cell scored in two different ways. Note the inverted U shapes of the functions. [After Enroth-Cugell and Robson (1966). A: Fig. 9, p. 533; B: Fig. 17, p. 544.]

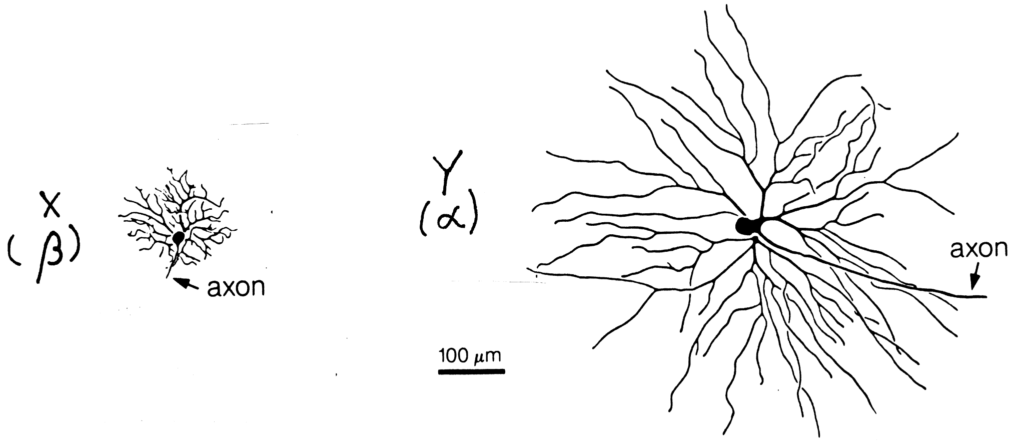


Figure 8.10: Anatomical identification of X and Y cells. The physiologist's X cell is the anatomist's β (beta) cell; the physiologist's Y cell is the anatomist's α (alpha) cell.

regarded as truly distinct classes of cells.

Enroth-Cugell and Robson's study was the first to provide a justifiable classification scheme for categorizing retinal neurons. Their choice of stimuli and experimental paradigm, and the careful quantitative descriptions they derived for individual neurons, made it possible to derive a categorization scheme that has stood the test of time. Their observations have been replicated in many laboratories. Moreover, as noted above, their classification scheme has been extended to include other physiological and anatomical characteristics of these neurons. X cells and Y cells are a true dichotomy rather than just the ends of a continuum.

The defining characteristics of X and Y cells are physiological. However, it turns out that these two classes of cells are also anatomically distinct. To show this correspondence, a technique informally called "stick and stain" is used. In this kind of work, hollow glass micropipettes are used as electrodes. First, one records extracellularly from the neuron to determine its physiological characteristics. When recording is finished, the electrode is thrust forward to "stick" the neuron, and a stain is infused into it through the electrode. At the end of the experiment, the stained retina is examined. By correlating the positions of the different types of neurons as recorded physiologically with the positions of the stained neurons in the retina, one can demonstrate the correspondence between physiological and anatomical cell types.

As it turns out, X cells correspond to a set of relatively small neurons which anatomists call *beta cells*, and Y cells correspond to a set of relatively larger cells called *alpha cells*. Examples of alpha and beta cells are shown in Figure 8.10.

8.6.2 The importance of discrete cell classes

When such a distinctive difference in properties truly exists, it provides a justification for arguing that the two cell types may serve different *functions*. That is, X cells and Y cells appear to perform different recodings of the incoming visual information. Populations of X and Y cells send forward two different analyses of the spatiotemporal pattern of the light that falls within their receptive fields, and the various types of W cells contribute many more. Moreover, as we will see, the different cell types send their axons to different places in the brain, and contribute to different aspects of the animal's responses to visual stimuli.

The concept that different cell types within one anatomical level of the visual system do different calculations, and thereby perform different functions, has come to be called *parallel processing.* Like Kuffler's discovery of ON-center and OFF-center pathways, Enroth-Cugell and Robson's characterization of X and Y cells set the stage for continuing discoveries about the widespread occurrence of parallel processing within the visual system.

8.7 Retinal interneurons: Horizontal, bipolar, and amacrine cells

Historically, a major problem in discovering the properties of retinal interneurons – horizontal cells, bipolar cells, and amacrine cells – is that, like the photoreceptors, all of these types of neurons are non-spiking cells. Thus, their responses could not be observed with the usual extracellular recording techniques. But in the 1960s, important innovations in microelectrode techniques allowed individual retinal interneurons to be penetrated with a microelectrode without killing them, with the result that intracellular recordings could be made. The first major report of recordings from retinal interneurons was that of Frank Werblin and John Dowling (1969).

Three features of Werblin and Dowling's study are of particular note. First, Werblin and Dowling speculated that, since the major anatomical features of the retina are preserved throughout all vertebrates, much that is of relevance to primate neural processing could be discovered by studying the retina of any convenient vertebrate species. As it turns out, the mudpuppy, *Necturus* – a salamander – has a relatively simple retina with very large neurons, making it a good choice for an initial attempt at intracellular recording. Second, Werblin and Dowling were able to use ”stick and stain” techniques to mark the individual neurons from which they recorded signals. In this way they could trace the recorded signals back to individual types of interneurons. And third, Werblin and Dowling chose a set of stimuli that clearly revealed the spatial and temporal processing properties of each neuron.

8.7.1 Light microscopy

Figure 8.11 shows a schematic view of the mudpuppy retina, highlighting the five major types of neurons. As we already know, the photoreceptors and the ganglion cells provide the retinal input and output respectively. The gap between them is bridged by the horizontal, bipolar, and amacrine cells. (We will treat the characteristics of these neurons in much more depth in Chapter 10xx.)

A little study of Figure 8.11 reveals that the horizontal cells spread their processes widely across the OPL, and the amacrine cells do the same in the IPL. Bipolar cells, on the other hand, seem to confine their processes to narrow retinal neighborhoods. The anatomy thus suggests that the photoreceptors, bipolar cells, and ganglion cells might form a straight through, or serial processing pathway, whereas the horizontal and amacrine cells might transfer the incoming information horizontally, comparing and combining the incoming information across the retina[3].

[3] Figure 8.11 also shows a sixth kind of neuron, the *interplexiform cell.* The dendrites of the interplexiform cell lie in the inner plexiform layer (IPL), and the cell sends its axon back to the outer plexiform layer (OPL). The

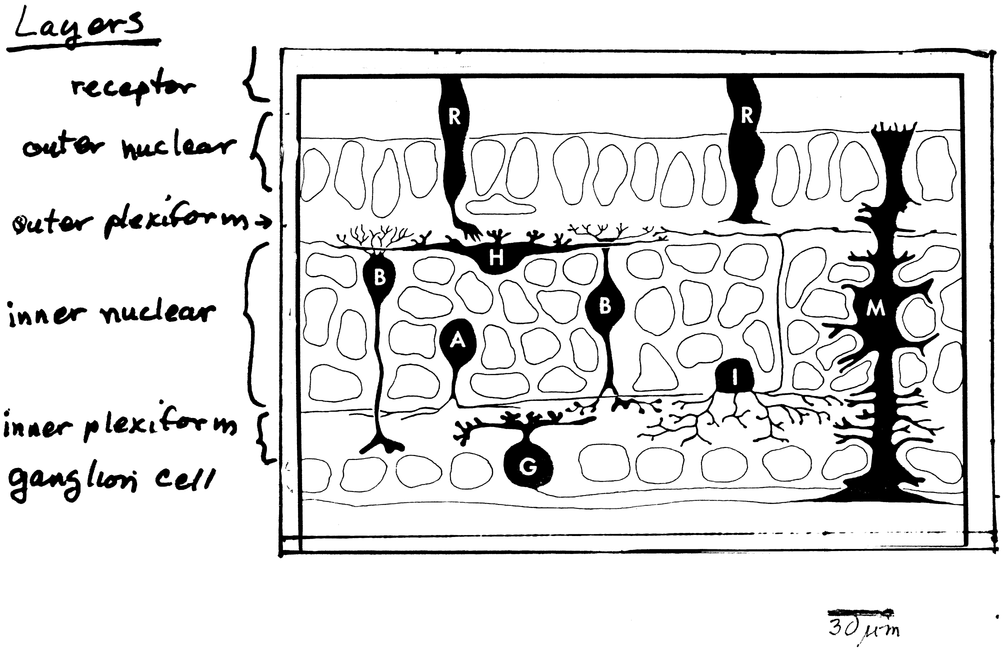


Figure 8.11: Anatomy of the mudpuppy retina. This highly schematic diagram portrays individual neurons much as they would be revealed by a Golgi stain. Note the large sizes of the neurons, which made early intracellular recording possible. R is a photoreceptor; H, a horizontal cell; B, a bipolar cell, A, an amacrine cell; and G, a ganglion cell. Other cells shown are I, an interplexiform cell; and M, a Mueller cell. (From Dowling (1987), Fig. 2.2, p.15)

8.7.2 Physiological recordings

Werblin and Dowling managed to penetrate and record from all five of the major types of neurons in the mudpuppy retina, and to stain examples of all of them. The question then becomes, what are the properties of the receptive fields of the different kinds of interneurons? In particular, where will the three most novel properties of the ganglion cell code – center/surround antagonism, ON-center vs. OFF-center pathways, and sustained vs. transient signals, arise?

Werblin and Dowling used three main stimuli: a small spot of light, a small annulus, and a large annulus, each centered on the receptive field of the recorded neuron. The annuli was designed to test the influence of light from distant parts of the retina, and seek out center/surround antagonism if it occurred.

Figure 8.12 shows Werblin and Dowlings recordings. Recordings from a photoreceptor are shown in Figure 8.12A. As we should expect from Chapter 6, the photoreceptor responded to the small spot of light with a hyperpolarization. Interestingly, the annulus – light at a distance from the photoreceptor – also produced small hyperpolarizations, probably due to the combination of scattered light and input from other photoreceptors through horizontal cells or gap junctions. These and more recent results suggest that the receptive field of a photoreceptor may be larger than the photoreceptor itself. But no spatial antagonism is seen.

Recordings from a horizontal cell are shown in Figure 8.12B. Like photoreceptors, horizontal cells gave a hyperpolarizing response to both disks and annuli of light. In this case the magnitude of the response increased with the area covered by the stimulus, as if the horizontal cell were indeed summing signals across a large retinal area. As might be guessed from the anatomy, horizontal cells have large receptive fields; but note that they, like photoreceptors, showed no center-surround antagonism[4].

Recordings from a bipolar cell are shown in Figure 8.12C. In contrast to photoreceptors and horizontal cells, bipolar cells change their response dramatically with the spatial configuration of the stimulus. The bipolar cell, like the photoreceptor and the horizontal cell, hyperpolarizes to the small spot. But it also *depolarizes* to the annulus, revealing center-surround antagonism! Thus, in Werblin and Dowling's hands, the first novel property of ganglion cells – center-surround antagonism – has its onset at the level of the bipolar cells.

Moreover, the bipolar cell in Figure 8.12C hyperpolarizes to the small spot and depolarizes to the annulus. But other bipolar cells depolarize to the small spot and hyperpolarize to the annulus (not shown). That is, like center-surround antagonism, the second major novel property of ganglion cells – the presence of both ON-center and OFF-center cells – originates at the bipolar cells.

Finally, the responses of an amacrine cell are shown in Figure 8.12D. When a stimulus (whether a small spot or an annulus) comes on, this particular amacrine cell depolarizes briefly; and when

interplexiform cell complicates the retinal picture by allowing the possibility that processing in the IPL could feed back to influence retinal processing in the OPL. Interplexiform cells are prominent in bird retinas, but not in cats or primates, and they will be ignored here for simplicity. The schematic also shows a Mueller cell. Mueller cells are not neurons; they are *glial cells* – structural cells that support and maintain the neurons.

[4] In general, horizontal cells hyperpolarize to light anywhere in their receptive fields. A major exception occurs in some non-mammalian retinas, such as those of fish. In fish retinas, there are horizontal cells that hyperpolarize to some wavelengths of light and depolarize to others. As we will see in Chapter 12, this kind of subtractive arrangement is critical to wavelength processing. For lack of a better alternative, the same form of processing was presumed for many years to hold for primate retina as well – one might call our implicit model the cat/mudpuppy/ goldfish model of primate retinal processing. But as we will see in Chapter 13, primate horizontal cells are now known to hyperpolarize to all wavelengths of light, and the chromatically opponent interactions are postponed to later processing levels.

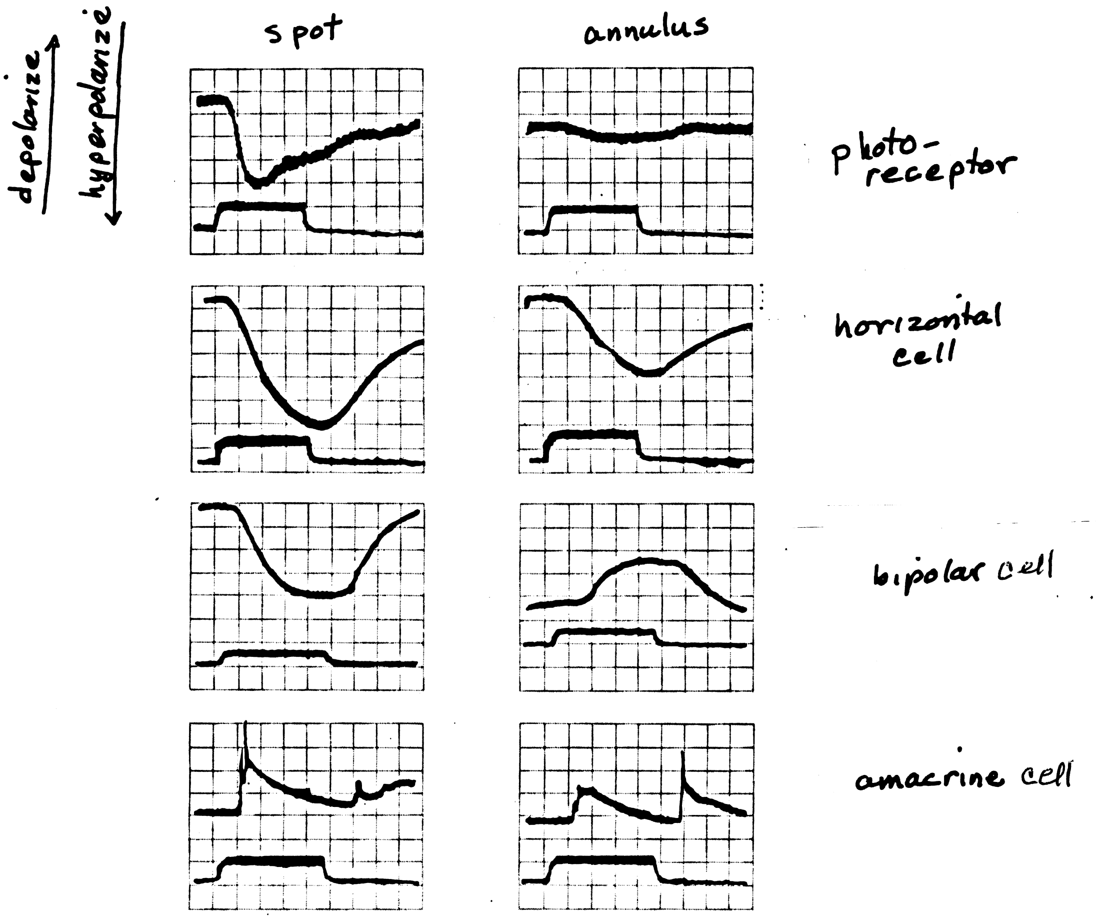


Figure 8.12: Physiological responses of mudpuppy retinal neurons. The left column shows responses to a small spot of light placed at the center of the receptive field. The middle column shows responses to a small annulus, and the right column to a larger annulus. The data are Werblin and Dowling's (1969) classic recordings. Rows A, B, C and D show the responses of a photoreceptor, a horizontal cell, a bipolar cell, and an amacrine cell respectively. [After Werblin and Dowling (1969), Fig. 3, p. 344.]

the stimulus goes off, the cell again depolarizes briefly. In between, the cell tends to return to its maintained voltage. Here for the first time in the retina, we see a cell type that responds transiently to changes in illumination, and this type of neuron is now called a *transient amacrine cell.* Since no earlier neurons in the retina produce transient responses, the transient amacrine cells appear to be responsible for creating the third novel property – temporally transient as opposed to solely sustained signals.

Another important feature of mudpuppy retinal neurons is that they usually respond to *contrast* rather than to absolute luminance. That is, the higher the contrast, the higher the response rate. Moreover, a fixed response is maintained by a fixed ratio of light falling on the center vs. the surround of the receptive field, rather than by a fixed luminance. We will return to the importance of this coding scheme below.

We close with another reminder that the purpose of this chapter is to introduce a simplified picture of retinal processing, upon which we can subsequently build a truer but more complicated story. In fact, there are many different subtypes of horizontal, bipolar, and (especially) amacrine cells, with much of the complexity intertwined with issues of chromatic processing. The subtypes and their synaptic connections will be covered in much more detail in Chapter 10.

8.8 Recipes for ganglion cells

Our next task is to use the data from mudpuppy retina to make a speculative model of the novel aspects of retinal recoding in the cat retina. How and at what level might each of the novel properties of the ganglion cell code – center-surround antagonism, ON-center vs. OFF-center cells, and sustained vs. transient cells – be created within the retinal processing network?

Figure 8.13 presents a speculative recipe for producing ON-center ganglion cells, of both sustained and transient types. From top to bottom, the figure shows two photoreceptors, one horizontal cell, two bipolars, one amacrine cell, and three ganglion cells. The retinal image of the stimulus light is shown by the solid bar at the top left. The stimulus falls on the center of the receptive field of the ganglion cell on the left, and on the surround of the ganglion cell on the right. (Also notice the ganglion cell in the middle; we will come back to it later.)

We need one more modeling tool to make our neural circuit: the distinction between *sign conserving* and *sign inverting synapses.* When one neuron synapses with another, the sign of the electrical potential that carries the signal can be conserved – for example, a hyperpolarization in a photoreceptor can produce a hyperpolarization in a horizontal cell. Alternatively, depending on the neurotransmitter and/or the properties of the post-synaptic membrane, the electrical potential can be reversed – for example, a hyperpolarization in the photoreceptor can yield a depolarization of the bipolar cell. In Figure 8.13, the symbols at each synapse specify the nature of the synapse, with open and closed circles showing sign conserving and sign inverting synapses respectively.

First examine the pathway on the left – the straight-through pathway. The bipolar cell in this pathway is ON-center; it depolarizes to light on its receptive field center. But (working backwards), because photoreceptors hyperpolarize to light and this bipolar cell depolarizes to light on its receptive field center, a sign inverting synapse must occur between the receptor and the center process of this bipolar cell.

Following along the straight through pathway, a sign conserving synapse then produces the ON-center ganglion cell from the ON-center bipolar cell. The ON-center ganglion cell receives a depolarizing input to light falling on its receptive field center, and depolarization in a spiking cell

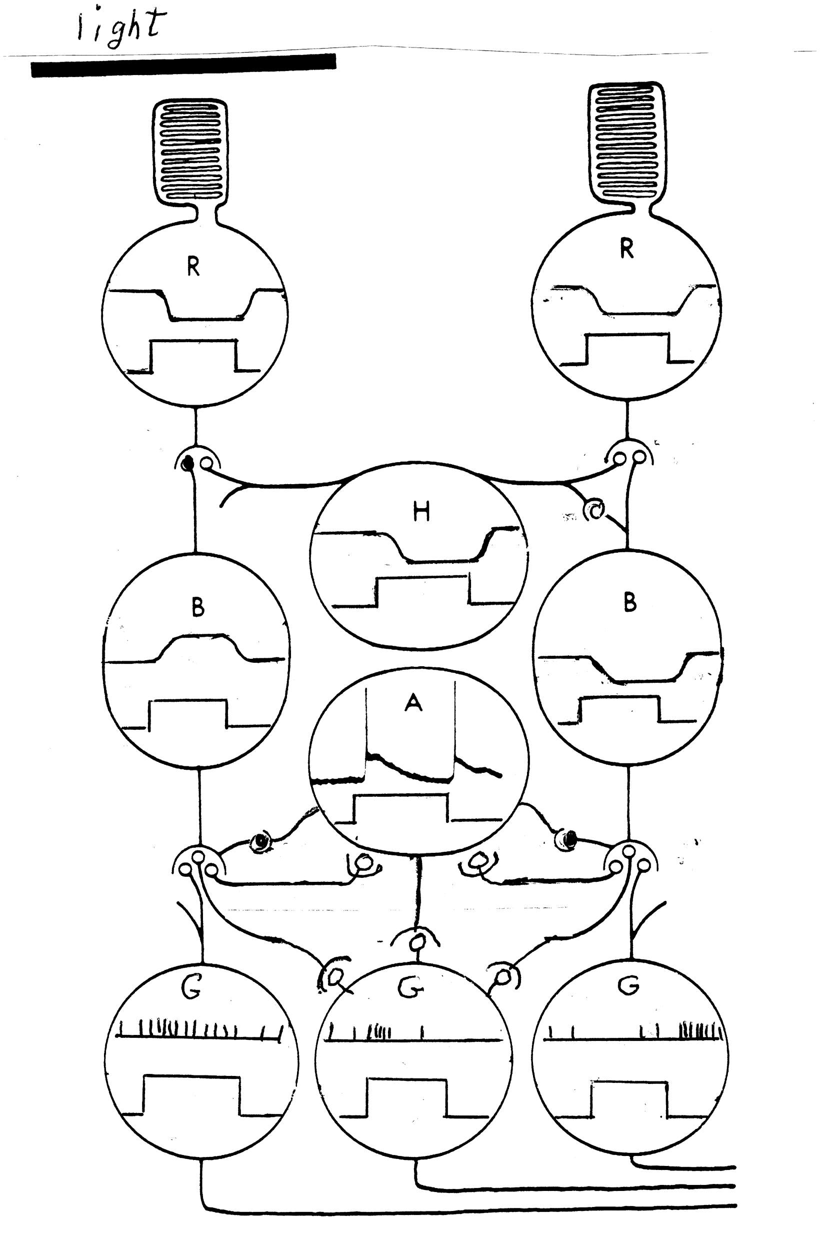


Figure 8.13: Recipes for ON-center ganglion cells. The black line at the upper left shows the location of a small spot of light. Open and closed dots at the synapses represent sign-conserving and sign-inverting synapses respectively. Inside each schematic neuron, the square trace shows the onset and offset of a stimulus light; the voltage trace or spiking pattern shows the response of the neuron. Three ON-center ganglion cells are shown in the bottom row of the figure. Those on left and right have relatively sustained ON and OFF responses respectively; the one in the middle is more transient. [After Dowling (1987), p. 109.]

leads to an increase in firing rate. In other words, the left-hand path provides the center process of a sustained ON-center ganglion cell.

Now examine the right-hand pathway – the indirect pathway via the horizontal cell. The photoreceptors hyperpolarize to light onset, as always. The horizontal cell hyperpolarizes as well, via sign-conserving synapses with the photoreceptors. The horizontal cell then feeds into the bipolar cell on the right, again with a sign-conserving synapse. When the horizontal cell indicates that light has fallen in its receptive field surround, the ON-center bipolar cell hyperpolarizes. This response feeds through to the ganglion cell on the right, and this ganglion cell decreases its firing rate in response to light falling on its surround. Simply put, it is the horizontal cell that creates the surround of the bipolar cell's receptive field.

To frost the cake, combine a center made from the pathway like the one on the left with a surround made from pathways like the one on the right. You have just created an ON-center ganglion cell.

In the middle of the row of three ganglion cells at the bottom of Figure 8.13, we have represented an ON-center ganglion cell with transient properties. Here, for both center and surround, the transient amacrine cell is involved in the circuit, to produce a transient change in the firing rate of the ganglion cell at both onset and offset of the stimulus. The details of how the transient is created are left unspecified, but note that some form of negative feedback from the amacrine cell back to the bipolar cell (hinted at in the synaptic arrangement shown in Figure 9.7) might just do the trick.

Figure 8.13 gave you the recipes for ON-center ganglion cells. After you have worked through Figure 8.13, test yourself by figuring out the recipes for OFF-center ganglion cells, using the framework laid out in Figure 8.14. (Hint: You only need to swap the locations of one sign-inverting and one sign-conserving synapse.)

8.9 The third code transformation: Creating the retinal output code

8.9.1 The ganglion cell image of an edge

Suppose you are looking at an edge – say, a dark/light transition in the physical world. What will the neural image of the edge be like at each of the three coding levels we have considered? That is, what will the retinal (optical) image, the photoreceptor image, and the ganglion cell image, be like? And how does the ensemble of ganglion cells carry information about the various features – the location, orientation, contrast polarity, and contrast – of the edge?

The ganglion cell image of a dark/light edge is derived schematically in Figure 8.15. Figure 8.15A shows the retinal image – a lower luminance field on the left, and a higher luminance field on the right. In the photoreceptor image, we would expect to find a similar pattern – two broad regions of retinal photoreceptors, with two different quantum catch levels (not shown).

The ganglion cell image, however, is very different. The optical image of the edge will fall upon the receptive fields of many ganglion cells of various types. For simplicity, lets consider just the ensemble of sustained ON-center cells, as shown in Figure 8.15B. What will the neural image of an edge be like in this ensemble of cells?

Four of the ganglion cells in Figure 8.15B are labeled E, F, G and H. Due to the presence of center-surround antagonism in their receptive fields, cells E and H, lying under the homogeneous

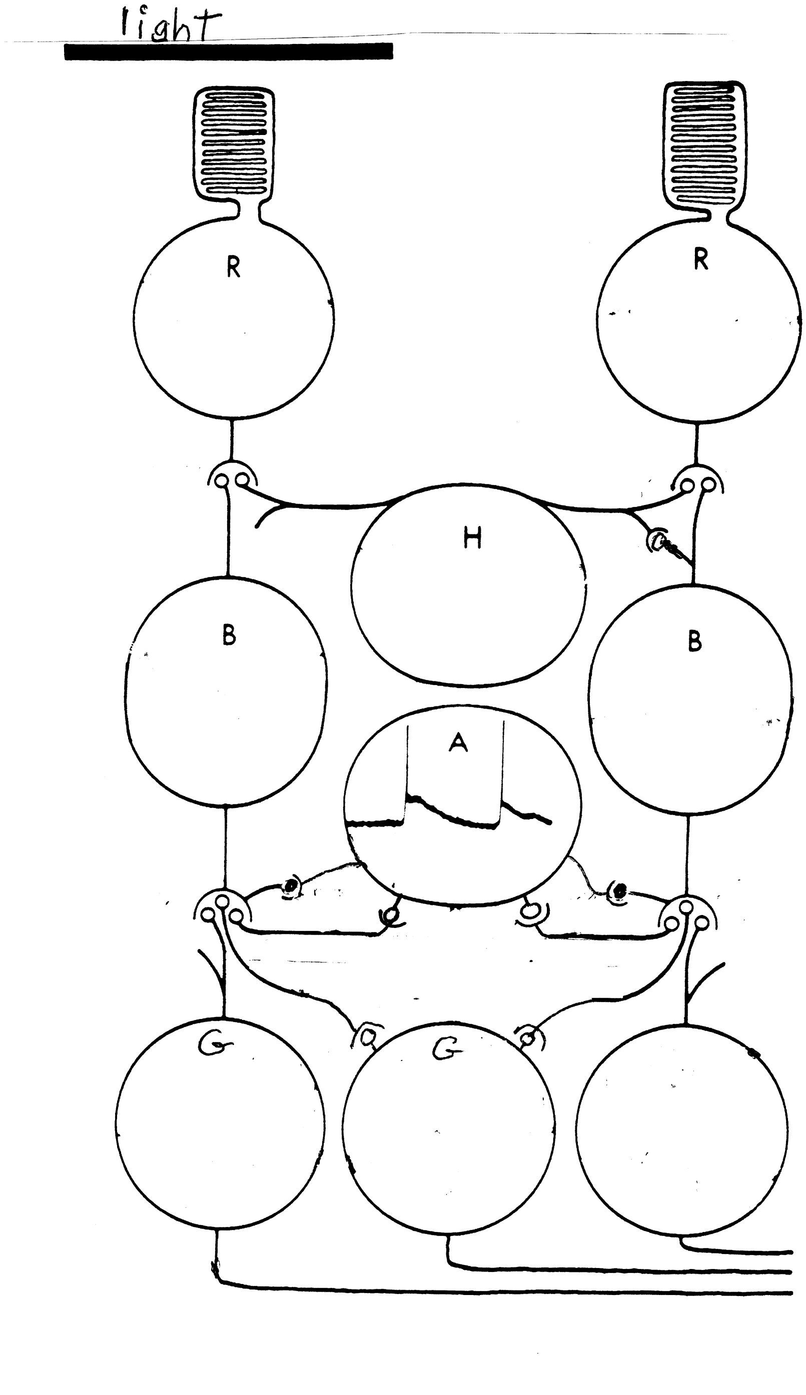


Figure 8.14: Recipes for OFF-center ganglion cells. A do-it-yourself kit. [After Dowling (1987), p. 109.]

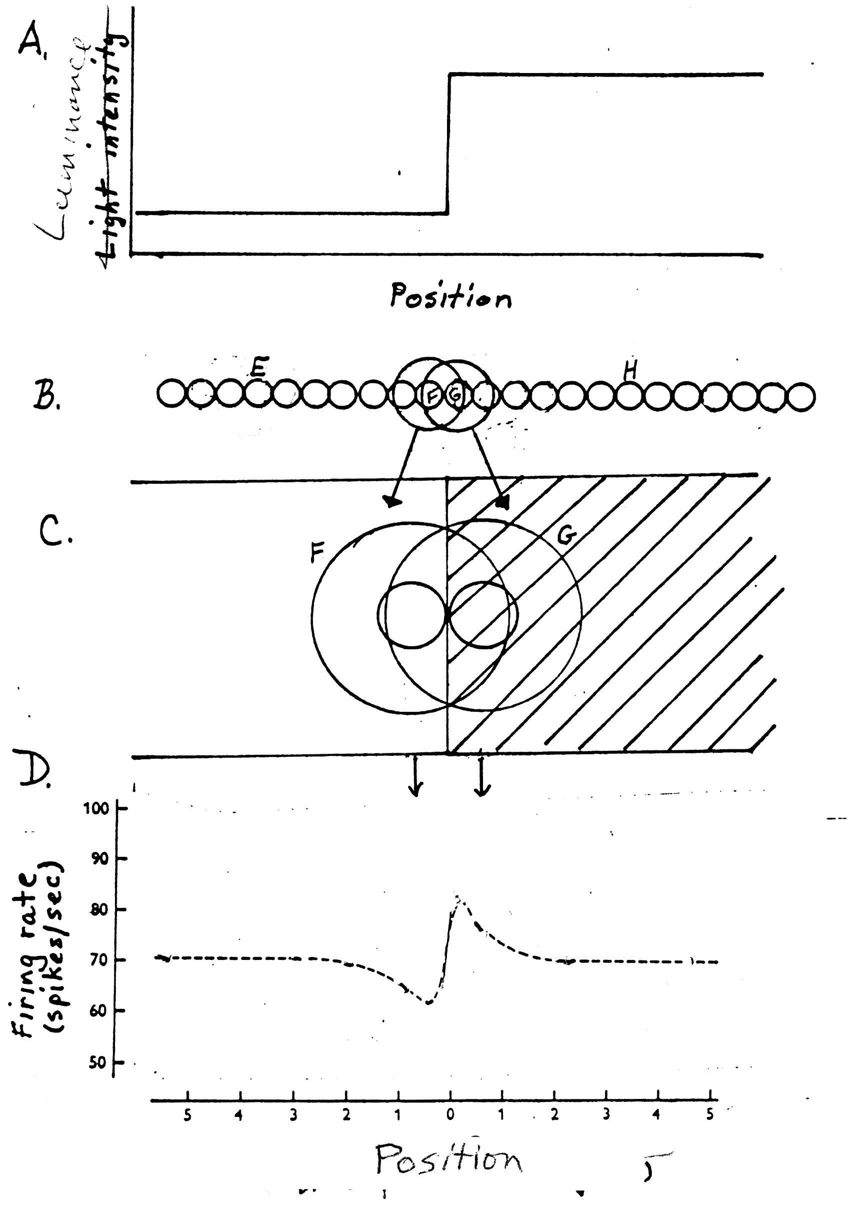


Figure 8.15: The ganglion cell image of an edge. A: The distribution of light across the edge, both in the physical world and in the retinal image. B: A row of ON-center ganglion cells. Except for neurons F and G, the two neurons closest to the edge, only the centers of the receptive fields are shown. E and H are neurons that lie far from the edge. C. Magnified view of the receptive fields of neurons F and G, with the edge lying across them. For cell F, light covers only a portion of the surround, and none of the center. For cell G, light covers all of the center and much of the surround. D. The predicted pattern of activity across the ensemble of ganglion cells, with the location of the edge coded by an ordered pair of "dog-ears".

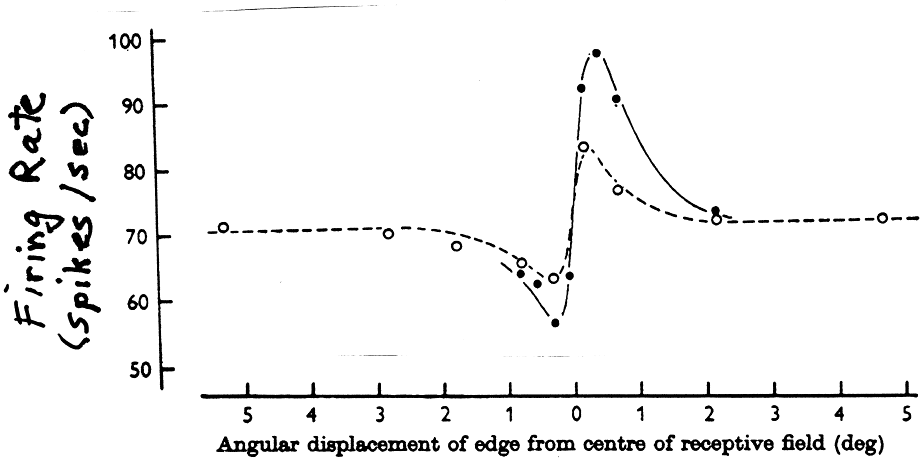


Figure 8.16: Response of a real ON-center cat ganglion cell to an edge. The edge was placed at a series of different distances from the center of its receptive field. The open and closed symbols show responses to 20% and 40% contrast respectively. [Enroth-Cugell and Robson, 1966, Fig. 14, p. 541.]

lower and higher luminance portions of the pattern respectively, will remain near their maintained firing rates. Similarly, since the stimulus is symmetrical, a cell that is directly under the border (not shown) should also remain near its maintained firing rate.

But the responses of cells F and G will be different. Cell F, just to the left of the edge under the lower luminance portion of the pattern, will have lots of light on some of its surround and less light on its center, so it will decrease its firing rate. Cell G, just to the right of the edge under the higher luminance portion of the pattern, will have lots of light on all of its center and only part of its surround, so it will increase its firing rate. The consequence of center-surround antagonism, then, is to create a set of "dog-ears" in the ganglion cell neural image, as shown in Figure 8.15D.

Figure 8.16 shows the results of such a physiological experiment on a cat ganglion cell (Enroth-Cugell and Robson, 1966). In a real experiment, recording in turn from each of the ganglion cells in the neural image is actually not feasible. Instead, these data were recorded from a single cat ganglion cell, by placing the edge at various locations with respect to the center of the cell's receptive field. The ganglion cell's response follows the predicted pattern. And conversely, a spatial pattern of activity like this one, in a particular local region of the ganglion cell image, would signal that a dark/light edge is occupying a particular position in the retinal (optical) image.

Finally, Figure 8.17 shows an edge in the original physical stimulus, together with a summary of the three images of the edge – the retinal (optical) image, the photoreceptor image, and the ganglion cell image – in visual processing. The retinal image closely resembles the physical stimulus, and the photoreceptor image differs from it only a little. But the ganglion cell image is very different, with its dog-ears coding the location of the edge.

In sum, in an ensemble of ON-center ganglion cells, a dark/light edge does not create a group of ganglion cells firing uniformly slowly under the dark side of the edge, and another group firing uniformly rapidly under the bright side of the edge. Instead, the information is recoded. The center/surround receptive fields of the ganglion cells create a pair of dog-ears – a downward blip to the left of an upward blip – in the ganglion cell image. This is the kind of code in which the location of an edge in the retinal image is carried by one ensemble of ganglion cells. And similar patterns would be created in other ensembles. [Work out the ganglion cell image of the same edge for an ensemble of OFF-center ganglion cells.]

8.10 The design question: Why this code at the retinal output?

The studies reviewed in this chapter center attention on three major features of the neural code created in retinal processing: center/surround antagonism, ON vs. OFF pathways, and sustained vs. transient signals. From a design perspective, why did these particular coding features evolve? What are the advantages of using this particular code at the retinal output? Put on your speculator's cap!

Let's begin with *center-surround antagonism.* What is the functional significance of this code feature? Remember that the goal of visual processing is not to keep track of the matrix of quantum catches *per se* – the loss of the pointillistic photoreceptor image is entirely to be expected. Instead, we presume that the goal is to produce codes whose parameters will eventually correspond more closely to the properties of objects, on the assumption that such coding would provide an optimal basis for mapping the incoming sensory information to the conscious perception of objects and our motor responses to them.

Let us take the perception of *lightness* as an example. In vision science, the lightness of an object is defined as its perceived shade along the white-grey-black continuum. The corresponding physical variable turns out to be the surface reflectance of the object – the fraction of the incident light that it reflects. In natural scenes the surface reflectance of an object is typically perceived quite accurately. The higher the physical reflectance of the object the whiter it appears perceptually, and the lower, the blacker. Moreover, subjects often show good lightness constancy: the perceived lightness of an object remains remarkably constant over changes in illumination on the scene.

But how can our perceptual systems determine the lightness of an object? It's a hard problem, because the retinal illuminance of the object is determined by two independent factors: the object's reflectance and the level of illumination on the scene. Moreover, the reflectance of the object varies only between (say) 10% and 80%, a factor of 8, whereas the illumination may vary by a factor of 100,000 or more, with the result that variations in retinal iluminance introduced by variations in reflectance will be swamped by variations due to variation in illumination.

On the other hand, an object with a fixed reflectance, viewed against a background with a fixed reflectance, will yield a fixed *contrast* (object reflectance/background reflectance) in the retinal image, across variations of the level of illumination. If the visual system were to code local regions of the incoming scene in terms of contrast, then a particular rate of firing in the ganglion cell image could be taken to indicate a particular surface reflectance.

As Werblin and Dowling showed, most ganglion cells respond to contrast rather than to absolute luminance. That is, due to center-surround antagonism, the magnitude of the neural signal will reflect the contrast in the visual scene. Thus, the sorting out of surface reflectance from luminance

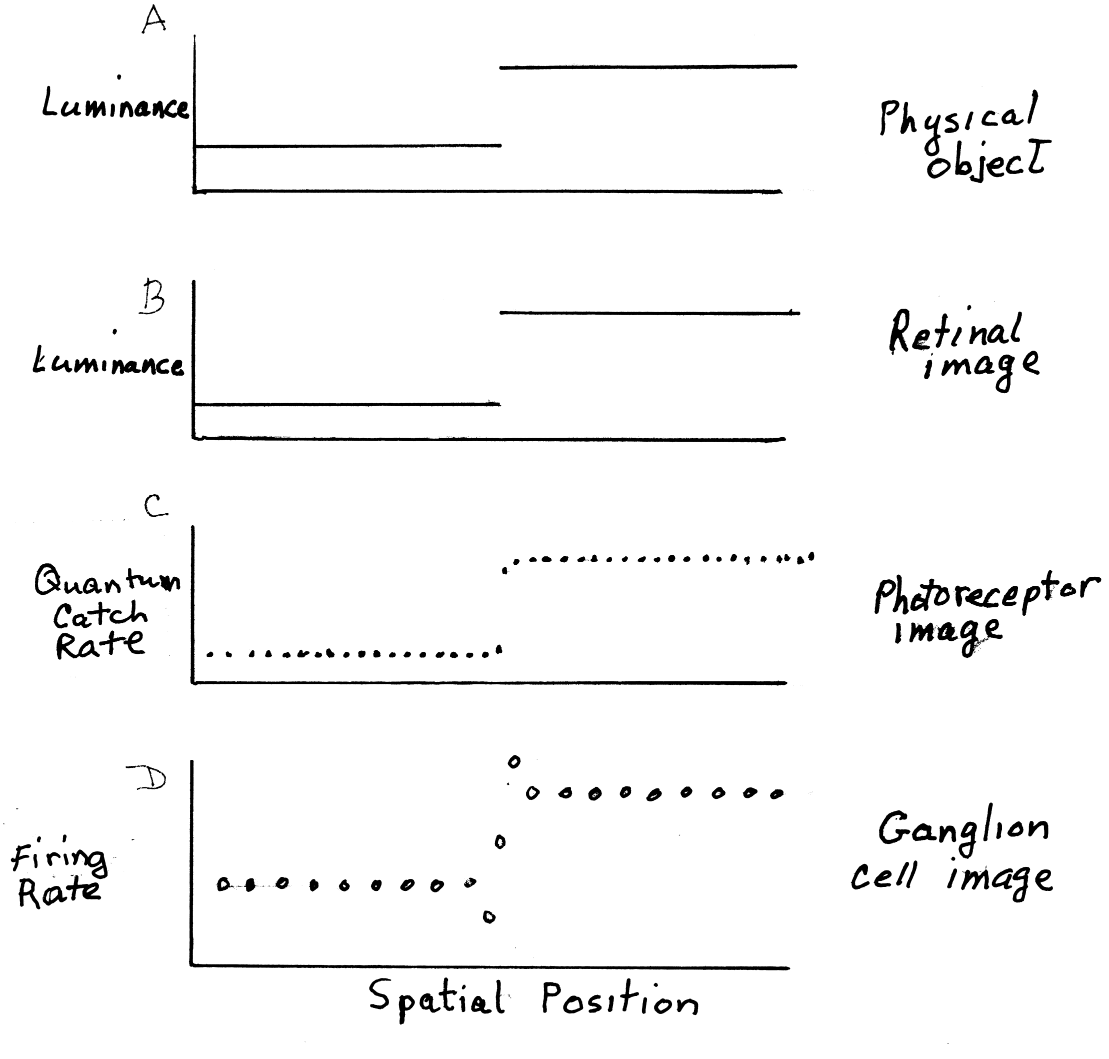


Figure 8.17: Summary of the sequence of codes for the location of an edge in retinal processing. A: The physical object – the edge – shows low-luminance and high-luminance regions. B: The retinal image retains these attributes. C: The photoreceptor image rounds the corners slightly. D: The response is amplified in the ganglion cell image by the introduction of "dog-ears".

has already begun at the level of the ganglion cell image. We return to the discussion of lightness coding in Chapter xx.

What about the *ON and OFF pathways*? The usual design argument about ON and OFF pathways begins with the fact that the maintained firing rates of cat retinal ganglion cells is only about 20-30 spikes per second. Since the maximum firing rate is 500 spikes/sec or more, there is a much greater range of increases than decreases available for use in coding. That is, decreases of firing can provide only a coarse indication of the magnitude of change in the stimulus, whereas increases of firing can provide a much finer grained code. The argument is that the ON pathway codes increases of contrast with increases of firing rate, whereas the OFF pathway codes decreases of contrast with increases of firing rate, thus allowing a fine-grained coding of both.

What about *sustained vs. transient signals*? The usual design argument here is that it is much more efficient – fewer spikes are used – to signal the temporal changes in the stimulus than it is to use a temporally continuous signal to indicate that the stimulus is remaining constant. Since spikes use energy, fewer is better. Notice, too, that the spatial coding of edges with dog-ears constitutes a similar strategy in the spatial domain.

8.10.1 Sparse vs compact codes

The above discussion of the retinal output code brings us to the distinction between *sparse* and *compact* neural coding schemes. A *compact* code is one in which each individual neuron changes its firing rate to varying degrees in response to each of many different stimuli. In the extreme, each different firing rate of the neuron is used equally often in the code, so that a maximum amount of information can be squeezed into a minimum number of active neurons. A *sparse* code is one in which, in response to most stimuli, an individual neuron remains at its resting level; it changes its firing rate only in response to a few very specific stimuli. In consequence, activity in this neuron is likely to signal the presence of some particular stimulus or object in the world. In most respects, the retinal output code is closer to the compact than to the sparse end of this continuum.

Why would the CDC choose a relatively compact code at the retinal output stage? The most common argument comes from anatomy. That is, the retinal ganglion cell axons make up the optic nerve. But the optic nerve, traversing out of the eyeball and up toward the brain, must be of limited size. In human beings, the optic nerve contains only about one million axons, one for each retinal ganglion cell. Thus, the incoming information flow from each eye must be packed into the ongoing activity of only one million transmission lines, and a compact code is needed.

As it turns out, there are many more cortical neurons than optic nerve axons devoted to vision. In consequence, different types of cortical neurons will have the luxury of sampling different subsets of ganglion cell outputs, and combining them in many different specialized ways. These recoding tactics will allow individual cortical neurons to specialize their responses to represent much narrower classes of visual stimuli than the retinal ganglion cells can afford. In sum, as we will see, cortical neurons will have the luxury of using a sparser code, and a change of activity in a cortical neuron will have a much more specific meaning than would a change of activity in a retinal ganglion cell.

8.11 Summary: Spatial and temporal recoding in the retina

In this chapter we used three classic studies to introduce a simplified description of mammalian retinal processing. After an introduction to retinal anatomy, we described Kuffler's (1953) early report of center-surround antagonism, the existence of both ON-center and OFF-center ganglion cells, and the presence of temporal variations such as sustained vs. transient responses, in the responses of individual ganglion cells.

We then recounted Enroth-Cugell and Robson's (1966) use of linear systems analysis to document the presence of two distinct types of ganglion cells: X and Y. The permutation of ON-center vs. OFF-center with X vs. Y dichotomies produces four basic types of ganglion cells: ON-center X, ON-center Y, OFF-center X and OFF-center Y. These four types of neurons make up our simplified retina. Finally we used Werblin and Dowling's (1969) study to introduce some of the simplest properties of three types of retinal interneurons – horizontal, bipolar, and amacrine cells. These properties were used to make qualitative models of the four types of ganglion cells.

Werblin and Dowling's data suggest that both center-surround antagonism and the ON-OFF dichotomy are first seen in the bipolar cells in the OPL. In contrast, the creation of transient signals arises from the properties of transient amacrine cells in the IPL. These interneurons will receive much more detailed treatment in Chapter xx.

In summarizing the form of the neural code at the level of the photoreceptors, we argued that the neural code is initially *pointillistic.* What is explicit in the activity of individual photoreceptors is the local quantum catches – the quantum catch by each photoreceptor at its particular location on the retina.

How can we capture the essence of the retinal output code with equal brevity? The sustained firing of, say, an ON-center X cell signals that more light is falling on the center than on the surround of its receptive field – there is local spatial contrast. Similarly, a temporal transient signals a change in the light level falling on part or all of a ganglion cell's receptive field – there is local temporal contrast. In sum, in the briefest terms, what is explicit in the firing patterns of individual retinal ganglion cells is *local contrast in space and time.*

Chapter 9

Spatiotemporal Processing and Perception

As discussed in Chapter 1, a major goal of this book is to examine some of the more esoteric forms of argumentation that arise in vision science. In particular, we wish to examine the arguments that vision scientists use when they try to explain perceptual states on the basis of neural states. In this book such attempted explanations will be called *causal stories*.

In Chapters 2-7, we explored the optical and transduction stages of the visual system. We addressed some causal stories, by asking how the properties of these early processing stages leave their marks on our visual perception. For example, we ascribed the high-frequency fall-off of the CSF to the optical quality of the eye (Chapter 5), the"zebra stripes" seen with laser interferometry to aliasing at the photoreceptors (Chapter 5), the scotopic spectral sensitivity curve and scotopic equivalence classes to transduction by rhodopsin (Chapter 6), and trichromacy to transduction by three and only three types of cones (Chapter 7). All of these examples are cases of elegant and well established causal stories.

In Chapter 8, we finally ventured beyond the photoreceptors. Setting aside psychophysics, we described the anatomy and physiology of the retina, and the spatio-temporal recodings of stimulus information that take place within it. We examined the novel properties of the retinal output code, as embodied in the responses of retinal ganglion cells to patterns of light in the retinal image. Importantly, we identified three novel properties of the retinal output code: center/surround antagonism; separate ON vs. OFF pathways; and parallel processing by several distinctive ensembles of ganglion cells.

In the present chapter we reintroduce psychophysics into the mix, and examine the question, how do the spatio-temporal recodings that take place in the retina leave their marks on our visual perception? We attack this goal by introducing and analyzing several classical causal stories that depend upon analogies between particular perceptual phenomena and particular aspects of the retinal output code. Some of these causal stories depend on center-surround antagonism, and others depend on the separation of ON and OFF pathways. (Causal stories that depend upon the third novel property, parallel processing, will be postponed to Chapter xx.)

The retinal output is a particularly advantageous vantage point from which to examine causal stories. This is so because, as Chapter 8 makes clear, the information flow from the retina to the brain is all in one direction. There is no feedback from the brain back to the retina. Thus, the retina can be seen as an early, separate module of visual processing. The one-way information flow has encouraged classical vision scientists to speculate upon the influence of retinal processing on perception. Part of the goal of the chapter is to examine the arguments by which they do so.

In sum, part of the art of being a good vision scientist involves acquiring skill in the evaluation of proposed causal stories. Whenever you encounter a causal story, you need to be in the habit of asking: What is proposed to be causing what? What are the hidden assumptions? What linking propositions are involved? Is the assumption structure reasonable? How compelling is the causal story? And especially, How could it be tested further? The goal of this chapter is to raise these questions, and to use them to critique some commonly accepted causal stories, in order to give you some practice at such analysis. In the meantime, start thinking about the following question: what does the statement "ψ is retinal" mean to you?

9.1 Causal stories

9.1.1 Causal stories and linking propositions

We begin by reviewing some conceptual themes originally introduced in earlier chapters. As discussed above, DT claims that vision scientists relate neural states to perceptual states by means of arguments that she calls *causal stories.* Moreover, she claims that causal stories always include at least one of a set of premises that she calls *linking propositions* – premises intended to specify the mapping rules that obtain between perceptual and neural states.

In Chapter 1, several kinds of linking propositions were identified. In particular, the *universal linking proposition* states that all perceptual states arise from neural states. And the *relational propositions* state that the same relationships that hold among a set of perceptual events will also hold between the explanatory set of neural events. Three relational families have been introduced: identity, similarity and mutual exclusiveness.

9.1.2 The roles of analogies in causal stories

Beyond the universal linking proposition and the families of relational propositions, another common element of causal stories in vision science is that of *analogy.* In ordinary English, an analogy can be defined as a partial similarity between like features of two things. In causal stories, the analogy usually concerns curve shape – a similarity between the graphical shapes of data sets derived from neural and perceptual domains. The use of an analogy as a causal story sometimes simply rests on the premise that if the two graphs are plotted on similar axes, and *look alike*, then the neural events provide an explanation of the perceptual events. Notice that the argument is purposefully loose, because there is no set definition of *looking like* across the vision literature.

In DT's view, analogies play at least two major roles in vision science. First, they encourage the development of neural models of perceptual phenomena. In particular, an analogy can be seen as a promissory note, to the effect that starting from the analogy between data sets, it would be relatively simple to fill in the remaining arguments and premises, and make a formal model of the perceptual effect based on the neural effect (Hood, 1998). For example, in Chapter 5, the analogy – the similarity of curve shape – between the scotopic spectral sensitivity curve and the absorption spectrum of rhodopsin leads to the expectation that a particularly simple causal story could be filled in. However, the causal story is incomplete as stated here, and it would be a useful exercise to fill it in explicitly. [Try it.]

A second role of analogies is that they provide communication across the subdomains of vision science (Westheimer, 19xx). In early vision science, we knew much more about psychophysical phenomena (like trichromacy) than we did about the details of anatomy and physiology. In consequence, theory and speculation flowed from psychophysics to the inferred properties of visual system physiology. But since the 1950s, we have had increasing knowledge of both kinds available. Consequently, arguments and speculations flow in both directions: from psychophysics to physiology and vice versa. Through the use of analogies – that is, not necessarily through rigorous argument – we exchange informal, intuitive hints between the subdomains, with novel data in each subdomain suggesting new experiments in the other.

A final observation about analogies is that they can enter into different causal stories in different ways. In some cases, the analogy is drawn between the properties of the psychophysical data and the properties introduced by retinal processing into the responses of *individual* ganglion cells. In other cases, it is drawn to the properties introduced by retinal processing into the properties of the *ensemble* of ganglion cells. We will see both kinds of cases as we examine the causal stories below.

9.2 Locus questions: What does "ψ is retinal" mean?

9.2.1 Criteria for assigning a retinal locus

Another concept introduced earlier in the book is that of a *locus question.* A locus question is a question about the anatomical location at which a perceptual phenomenon occurs. For example, the question, "at what anatomical location does trichromacy arise?", is a locus question. Similarly, "trichromacy arises at the photoreceptors" is a locus statement.

Locus statements are paradoxical. That is, we all agree that information originating in the physical domain is carried toward the brain through a complex neural information processing system, and is coded and recoded as it passes from one processing level to the next. But when we make a locus statement we talk as though we attribute a given perceptual phenomenon to one and only one link of the causal chain. The question is, what do we mean by such statements?

In DT's view, the answer is closely related to the use of analogies. To invent an analogy, we are looking for a similarity of curve shape between data sets from physiological and perceptual realms – the location at which the physiological signal first starts *looking like* the perceptual one. We can ask, at what anatomical stage, or locus, does this similarity of curve shapes first arise? When we say, "perceptual phenomenon ? occurs at anatomical locus A", we probably mean that the two data sets start looking alike at A. Specification of the exact criteria for "looking like" are left to the theorist in any individual case.

Locus questions, of course, can be posed at any stage of processing. However, by far the most common use of a locus question is to ask whether a particular perceptual phenomenon is or is not *retinal* (Hood, 1998). That is, does a sufficiently compelling analogy first arise within the retina, at or prior to the ganglion cell outputs? As we said at the beginning of the chapter, the retinal output is a particularly clean and convenient stopping place for sorting out causal stories, because there is no feedback from the brain to the retina. Hence, the question – Is ψ retinal? – has historically been the most commonly asked locus question.

In 1998, Donald Hood formulated a generic model that allowed him to propose a more explicit definition of locus questions, particularly with respect to the retina. His formulation is shown in Figure 9.1. It consists of two processing modules, retinal and cortical, with the cortical module

Figure 9.1: Hood's generic model of visual processing. The model is intended to provide a vocabulary for distinguishing "early" (retinal) from "late" (cortical) processing. It consists of three modules, the retinal, cortical preprocessing, and decision operations. The model is described further in the text. (Hood, 1998, Fig. 1, p. 505.)

being further separated into CP (cortical preprocessing) and DM (decision mechanism).

In Hood's formulation, E(λ, x, t) is the number of quanta of each wavelength falling on the retina at point x, as a function of time t. R(t) is the output of the retina: the responses of the ensemble of ganglion cells to E(λ, x, t). E(λ, x, t) is mapped to R(t) by the retinal processing module. R(t) is then processed by the cortical preprocessing module and then by the detection, or decision, module to yield the final output, ψ.

Given this formulation, Hood's proposal is that the locus statement "ψ is retinal" means that ψ can be predicted from R(t) very readily, with no more than "a very simple set of CP (cortical processing) assumptions followed by a traditional DM (decision mechanism)." (Hood, 1998, p. 507, words in parentheses added). We will see illustrations of this principle below. [Meanwhile, try to decide as you go along what "ψ is retinal" means to you.]

9.2.2 Evaluating causal stories

Finally, we raise the question: What are the criteria on which causal stories are judged? The answer is, the same criteria on which any scientific theory is judged. There are three basic criteria.

The first criterion is the domain of facts that the theorist intends the theory to unite; the broader, the better. In his classic chapter, Brindley argued that "Arguments of this kind depend on first securely establishing a psycho-physical linking hypothesis by *correlating very many features of a sensory phenomenon with corresponding features of an objectively determined one.*" (p. 149). For example, if center/surround antagonism really causes and explains the Westheimer effect, as argued below, then additional properties of center/surround antagonism should predict additional properties of the Westheimer effect, and vice versa. Ideally, vision scientists would carry out a program of research exploring these parallels. The more interpredictability between the two domains, the greater the support for the causal story. We will examine the Westheimer effect from this perspective at length below.

The second criterion of excellence for any theory (or causal story) is the degree to which the theory gives a convincing account of the facts within its chosen domain. And the third criterion is related: the absence of other theories that provide equally well or better for the facts. In other words, many philosophers of science believe that a theory is not displaced by contradictory facts, but only by a better theory.

9.3 Manifestations of center-surround antagonism

We now present and analyse three causal stories based on the presence of center-surround antagonism in ganglion cell receptive fields. The first example – the perceptual phenomenon known as the Westheimer effect – was discovered only after the physiological fact of center-surround antagonism had been discovered. But the second example – Mach bands – is much older. Remarkably, the perceptual phenomenon of Mach bands was used to argue for the presence of center-surround antagonism, 100 years before center-surround antagonism was found to occur in the cat retina. The third example – the low-frequency fall-off of the CSF – is included because it provides an account of a perceptual phenomenon that has been waiting patiently for several chapters for an explanation (see p. xx).

The first and second cases are interesting together because they depend on two different uses of analogy propositions. The proposed explanation of the Westheimer effect depends on an analogy between a perceptual phenomenon and the properties of individual ganglion cells. In contrast, the proposed explanation of Mach bands depends on an analogy between a perceptual phenomenon and the joint responses of an ensemble of ganglion cells. In addition, notice that the Westheimer effect involves thresholds, and the causal story is an attempt to account for a Class A phenomenon, whereas Mach bands are supra-threshold perceptual judgments, and the causal story is an attempt to account for a Class B phenomenon.

9.3.1 The Westheimer effect

As we now know, a ganglion cell with center-surround antagonism shows an inverted U-shaped area-response curve (Figure 8.6). In 1965, Gerald Westheimer pursued an analogy between psychophysics and physiology, and predicted that certain novel psychophysical experiments should also yield U-shaped functions.

The basic stimulus configuration used by Westheimer is shown in Figure 9.2A. In designing his stimuli, Westheimer began with a tiny *test spot*, only 6' in diameter. He did this on the conjecture that such a tiny spot might be detected by only one ganglion cell (or a very few at most), and thus that variations in the threshold for the test spot might reveal variations in the threshold of that individual ganglion cell. He then added a second stimulus component – a concentric *background disk* of variable diameter. The goal of the experiment was to measure variations in the threshold for the tiny test spot, as a function of the size of the background disk.

Typical results from an experiment with this paradigm are shown schematically in Figure 9.2B. As the size of the background disk increases, the threshold for the tiny test spot first rises, indicating a decrease of sensitivity. But as the background disk passes a certain critical size, the threshold for

Figure 9.2: The Westheimer effect. A: The stimuli. The subject's detection threshold is measured for a tiny test spot centered on a disk of variable diameter, as a function of the disk diameter. B: The results. As the disk size increases, the threshold for the tiny test spot first rises, and then falls again. Light in an annular region at a distance from the test spot lowers its threshold.

the test spot begins to fall again, indicating an *increase* of sensitivity. The surprising result is that light at a distance from the test spot can increase the subject's sensitivity to the test spot. This *spatial sensitization* (or *Westheimer*) *effect* can be large – the threshold can be reduced by nearly a log unit under optimal conditions.

An intuitively appealing model of the Westheimer effect is also shown in Figure 9.2. We start by noticing the analogy – the similarity of like elements – between the two different U-shaped curves: the physiologically derived area-response curve of an ON-center ganglion cell (Figure 9.2B) and the psychophysically derived Westheimer function (Figure 9.2D). Both curves are inverted U's; both show variations in a response (a firing rate or a psychophysical threshold) with the diameter of a background disk; and the maxima of the two curves can reasonably be set to similar disk sizes. These similarities make the analogy a tempting promissory note for a causal story. From it stems a model that has been called the *center-surround*, or *net excitation* model of the Westheimer effect.

Notice that, importantly, the center/surround model requires a major nothing mucks it up proviso – that over the whole range of conditions used in the experiment, this particular ganglion cell is in control of the subject's thresholds for the tiny test spot. For this to be true, two interesting sets of ancillary assumptions are required. First, the ganglion cell must have the right intrinsic properties to generate the U-shaped psychophysical data; and second, it must have the right properties to dominate all of the other neurons in the visual system in controlling the subject's threshold.

In regard to the first set of ancillary assumptions: Notice that the data in Figure 9.2B are neurophysiological, showing the firing rates of a neuron. The data of Figure 9.2D are perceptual, showing detection thresholds for the tiny test spot. To explain the latter by the former, we are implicitly making an assumption that links the two ordinates. That assumption is that this ganglion cell's detection threshold will vary with its firing rate – the higher the firing rate, the higher the detection threshold for the test spot. This assumption is made explicit in Figure 9.2C.

Is this assumption reasonable? So far we have not had any reason to consider it one way or the other. However, it does seem to be a simple empirical prediction, and so eminently testable. We will return to this issue below.

In regard to the second set of ancillary assumptions, we need to turn our attention to the whole ensemble of ganglion cells and the central visual neurons that receive inputs from them. The assumption we are searching for is that detection of the test spot by the subject is controlled by the response of this single retinal ganglion cell. It isn't controlled by some other ganglion cell; and it isn't changed by visual processing at any more central level. In other words we must be assuming that *over the whole range of sizes of the background disk, this particular ganglion cell is the weakest link in the most sensitive neural channel available to the subject for detection of the test spot.* We are assuming that nothing else – no other ganglion cells, and no other links of the causal chain – interferes with the control of this ganglion cell over the subject's detection threshold.

Is the second set of assumptions reasonable? Maybe or maybe not. Look back at the stimulus configuration used in Figure 9.2B – the area-response curve of the ganglion cell. Notice that the ganglion cell providing the area-response curve is a cell on which the stimulus configuration is *centered.* If a non-centered cell were used, the area-response curve would not be such a neat U-shaped function, and the analogy would break down [try it].

Is it reasonable to assume (as we implicitly do) that the cell on which the test spot is centered is the most sensitive cell in the ensemble for that test spot? Maybe. Remember that the signal from the ganglion cell has to play many roles in the retinal code, including signaling the *location* of the test spot. Perhaps the neuron centered at a given location location has evolved to be the most sensitive cell in responding to the test spot at that location, and the identity of the cell in which activity occurs also signals the location of the stimulus.

Of course, the psychophysical curve in Figure 9.2D can also be taken as a prediction of the outcome of a physiological experiment. It suggests that we go back and look again at area-response functions in ON-center ganglion cells. We could seek out the cells that are the most sensitive to tiny test spots, and see whether the thresholds of these cells for detecting tiny test spots vary in concordance with their area-response functions. In other words – and note this as a general experimental strategy – we could press the analogy by carrying out a neural experiment using the exact stimulus configuration used in the psychophysical experiment.

In fact, some such experiments have been done, but the issue is not yet fully resolved. U-shaped functions have been seen in some individual retinal neurons – both bipolars and ganglion cells – in some species, but not in other neurons in other species. Few studies have been done in primates, and none that address the question in individual neurons that are selected to be the most sensitive to the tiny test spot. More physiological experiments need to be carried out, and the more exactly parallel they are to the psychophysical ones, the greater their power. In short, the ganglion cell-based center-surround model of the Westheimer effect still awaits full confirmation or rejection.

9.3.2 Parametric studies of the Westheimer effect

To gather additional evidence for the center-surround model, we can look for the parallel influence of stimulus parameters and other experimental factors on the psychophysical and physiological data. Detailed studies of the Westheimer effect on the one hand and center-surround antagonism

Figure 9.3: Westheimer functions in individual human and monkey subjects. A: a monkey; B: a human subject. The parameter on the curves is the retinal eccentricity of the stimuli. For both species, Westheimer effects are seen at all eccentricities, with the spatial parameters increasing as eccentricity increases. The similarity across species is striking. (Spillmann et al 1987, Fig. 15, p. 59).

on the other have indeed indicated many similarities of features. For example, stimuli intended to isolate or to light adapt rods vs. cones have predictable effects on rod-mediated vs. cone-mediated Westheimer effects.

Variations of the Westhemer effect with retinal eccentricity are also relevant. As discussed earlier, the sizes of receptive fields vary with eccentricity, becoming regularly larger as one moves from the fovea to the retinal periphery. Do the spatial parameters of the Westheimer effect vary in concordance?

An exemplary set of experiments in this domain was performed by Lothar Spillmann and his colleagues (Spillmann, Ransom-Hogg and Oehler, 1987). These authors carried out systematic psychophysical experiments on variations of the Westheimer function with eccentricity in human subjects. Moreover, importantly, monkey subjects were also tested, and human subjects were tested in the monkey apparatus with identical techniques and stimuli. As discussed earlier, psychophysical data from monkey subjects is particularly valuable because quantitative, within-species comparisons can be made to monkey neuroanatomy and neurophysiology.

The psychophysical results from both species are shown in Figure 9.3. The data on the left are from a monkey subject, and those on the right from a human subject. Two points are obvious. First, Westheimer functions are remarkably similar across the two species. And second, as eccentricity increases, both the peak and the plateau of the Westheimer function shift systematically toward larger background disk diameters.

Spillmann and his colleagues also made systematic comparisons of their psychophysical data from monkey subjects to physiological data on receptive fields in the same species. They predicted that if the centers and surrounds of ganglion cell receptive fields create the desensitization and

Figure 9.4: Comparison of the maxima of human Westheimer functions (x's) to the sizes of monkey ganglion cell receptive field centers (dots) . The close similarity strengthens the argument that the properties of ganglion cell receptive fields control the properties of Westheimer functions. (Spillmann et al, 1987, Fig 17, p. 59.)

sensitization arms of the Westheimer effect, the peak of the Westheimer function would be correlated with the sizes of receptive field centers, and the diameter of the plateau would be correlated with the size of the receptive field surround, across variations in eccentricity. This proved to be the case, as shown in Figure 9.4.

Moreover, the peaks of the Westheimer functions were also tightly correlated with the sizes of ganglion cell dendritic fields. Spillmann et al attribute the Westheimer effect to one particular class of ganglion cells (the P alpha cells) in the monkey retina. In general, both the receptive field center sizes of P alpha cells and the peak sizes of the Westheimer functions were about a factor of two larger than the dendritic fields of these cells. The discrepancy is attributed to the spatial spread of signals by other retinal neurons prior to the ganglion cells. Spillmann et al argue that the tight correlations among retinal anatomy, retinal physiology, and visual function, across eccentricity, provide strong evidence for the center/surround model.

However, the issue is not yet finally settled, for two reasons. First, recent neurophysiological work has challenged the correlation between the variations in activity levels of cortical neurons and their detection thresholds for superimposed test spots (Kunken, Sun, and Lee, 2005), one of the major premises adopted above. And second, viable alternative theories have been proposed (Westheimer, 19xx).

In summary, the Westheimer effect provides an instructive example of a causal story. At its best, the analogy between perceptual and neural effects is tight, and is bolstered by additional parallels between neural and perceptual responses across stimulus parameters such as eccentricity and light level. Further parametric studies at the physiological level, if successful, would further strengthen the analogy.

Figure 9.5: Mach bands. Dashed line: A ramp pattern in the physical stimulus. Solid line: The common perception of the ramp pattern. Most subjects see illusory bright and dark bands at the transitions between the ramp and the flanking fields. (Ratliff, 1965, Fig. 2.3, p. 41.)

9.3.3 Mach Bands

Our second example of a causal story – the Mach band phenomenon – is illustrated in Figure 9.5. Suppose that we create a stimulus composed of a high luminance field, a low luminance field, and a gradual shading off – a ”ramp” – of luminance in between. Most subjects report seeing an extra light band at the high luminance edge of the ramp, and an extra dark band at the low luminance edge. These extra light and dark bands are illusory, in that they occur in perception but not in the stimulus. They are called Mach bands after the physicist Ernst Mach, who first called scientific attention to them in 1865.

Mach bands are easy to demonstrate with a projector and a piece of cardboard. Hold the cardboard in the projector beam about halfway between the projector and the screen, varying the distance for maximal effect. The cardboard creates a shadow and its penumbra creates the ramp. Prominent illusory bright and dark bands should be seen. In fact, in the ordinary physical world you can learn to notice Mach bands on any shadow that has a prominent penumbra.

Now, why do Mach bands occur? In Chapter 8, we discussed the fact that at the level of the ganglion cells, the code for a sharp edge is a pair of dog-ears in the neural image of the edge. Like the sharp edge in Figure 8.15, a ramp pattern will yield maxima and minima – dog ears – in the pattern of activity across the ensemble of ganglion cells, roughly in the locations where the bright and dark Mach bands are seen. Yielding to the analogy, one can imagine that these maxima and minima in the neural image of the ramp pattern cause the perception of Mach bands.

But let's think more critically about this causal story. For DT there are two main issues that need to be addressed. The first is the question of intended domain. Notice the implicit assumption that the perceived brightness at each point in the visual field is determined by the level of activity in the ganglion cell corresponding to that point in the neural image – the greater the firing rate of the ganglion cell, the greater the perceived brightness. But over what domain of stimuli and perceptual phenomena is this mapping rule assumed to hold? Does perceived brightness at a point always correspond to the firing rate of a ganglion cell corresponding to that point? If so, why don't we see Mach bands at sharp edges, which also set up dog ears in the neural image? To make this causal story work, one must develop an argument as to when dog-ears lead to the veridical perception of a sharp edge, and when they lead to the perception of illusory bright and dark bands.

And second, there's an enormous "nothing mucks it up" proviso involved – we must be assuming that nothing in the recodings between the ganglion cells and whatever underlies our conscious perception, destroys the correspondence between ganglion cell firing rates and perceived lightness. It is certainly legitimate to argue that the coding that causes a Class B phenomenon such as Mach bands is introduced as early as the retina. But such an "early" causal story imposes a powerful constraint on the kinds of recodings that can be allowed to happen in the intermediate levels of the visual system. These transformations must preserve the retinally established lightness code, or else Mach bands wouldn't occur.

In sum, this explanation of Mach bands is appealing. But it would be strengthened if it were expanded to provide an account of the conditions under which Mach bands are and are not perceived at edges. An account would also be welcome regarding how the appropriate retinal code is preserved up to the neural level needed for perception, and not lost in the intervening post-retinal processing.

9.3.4 The spatial CSF

In Chapter 4, we introduced the spatial contrast sensitivity function (CSF) as a way of characterizing the system properties of spatial vision. As noted in Chapter 8, inasmuch as the psychophysical CSF is band-pass, there is no ready explanation for it at the photoreceptoral level. But as Enroth-Cugell and Robson showed (Chapter 8), the CSFs of many ganglion cells are band-pass, with a prominent low-frequency fall-off. As such, neural CSFs provide a tempting analogy for explaining the low frequency fall-off of perceptual CSFs.

Let's dig into this analogy further. As was the case for the Westheimer effect, the psychophysical and neural CSFs are from different experimental realms, with different dependent variables, and a variety of ancillary assumptions will be needed to relate them to each other. The ordinate for the psychophysical data concerns psychophysical detection thresholds, whereas the ordinate for the neural data concerns neural detection thresholds. To complete the analogy we need an assumption that links the two ordinates. The necessary assumption is that the observer's threshold will be controlled by the ganglion cell's threshold. This assumption is probably reasonable, and if it is accepted, the neural CSF provides an analogy – and thereby a causal story – for the perceptual CSF.

The convincingness of the explanation can be strengthened by showing an interpredictability of the influence of experimental parameters between the two realms. In particular, as shown in Figure 9.6, psychophysical and neural CSFs change in similar ways with such variables as light levels. And as Brindley argued, the similarity of details – the fact that both perceptual and neural CSFs become less band-pass and more low-pass at low light levels – strengthens the causal story.

Figure 9.6: Contrast sensitivity functions (CSFs). A: A cat ganglion cell. B: A human subject. Both sets of CSFs shift leftward and flatten out as luminance decreases. The low frequency falloff of the CSF is usually attributed to center-surround antagonism. (A: Enroth-Cugell and Robson, 1966, Fig. 15, p. 542. B: After van Ness and ?ouman, 1967, via Olzak and Thomas, 1986, Fig. 7, p. 7.)

9.4 Manifestations of separate ON and OFF pathways

We now turn to causal stories that arise from the existence of the separate ON and OFF pathways. As discussed in Chapter 8, this separation begins in the outer plexiform layer and is preserved throughout the retina and the lgn, all the way up to the first stage of cortical processing. Not until area V1 do signals from ON and OFF pathways finally converge on individual cortical neurons. Can we design psychophysical experiments that reveal the early separation of processing between the ON and OFF systems, and their later convergence in the cortex?

9.4.1 Detection of increments vs. decrements

In pursuit of this goal, it has often been suggested that incremental stimuli – increases of luminance – are detected and processed via the ON pathway, whereas decremental stimuli – decreases of luminance – are detected and processed via the OFF pathway. If this were true, the use of incremental vs. decremental stimuli would obviously become available as a key experimental strategy for exploring the properties of the ON and OFF systems.

In 1986, Peter Schiller and his colleagues (Schiller, Sandell, and Maunsell, 1986) undertook a set of experiments designed to probe this question. The work is exemplary because it combines the use of psychophysics, physiological recordings, and selective neural lesions, all in the same animals.

The approach was to train a monkey to carry out psychophysical tasks involving the detection of increments and decrements. Schiller and his colleagues then produced localized, specific, temporary lesions within the monkey's retina, and tested the monkey a second time on the same psychophysical tasks. Finally, they allowed time for the monkey to recover from the temporary lesion, and tested her a third time after her visual functions had returned.

How was this possible? In the early 1980's a chemical substance was discovered that disables the retinal ON system while leaving the OFF system intact. The poison, 2-amino-4-phosphonobutyric acid, or APB, disables the ON bipolar cells by hyperpolarizing them. The effects of APB last for several hours, but disappear within 24 hours. Thus it is possible to produce an experimental animal that is selectively missing a functional ON system one day, and whole again the next.

Given the possibility of reversible lesions, Schiller and his research group made two important predictions, one for cone- and the other for rod-mediated vision. First, cone vision is mediated by both the ON and the OFF systems. Since APB disables the ON system but leaves the OFF system intact, in cone vision APB should interfere with the detection of increments but have little influence on the detection of decrements. But second, since all of the rod bipolars are ON neurons, rod vision is mediated only by the ON system; and all rod-initiated signals must pass through the ON bipolars before finally entering the OFF system at the OFF cone bipolar (see Chapter 10 for a detailed treatment of retinal connections). Therefore, the second prediction was that for rod-mediated vision, APB should lead to major deficits in the detection of both increments and decrements.

The first experiment was carried out using stimuli designed to isolate cone-mediated vision. The stimulus configuration is shown in Figure 9.7A. The first step was to train a monkey subject to fixate the center of a homogeneous field of light, and record her eye movements as she did so. Then, using light levels designed to isolate cone-mediated vision, these researchers presented a brief incremental or decremental test flash at one of six possible locations around a fixation point. The animals task was to fixate the fixation point, watch for an increment or decrement at any of the six locations, and when she saw one, quickly move her eyes to look at the location where the increment or decrement appeared. A rapid and accurate eye movement to the target position produced a reward (a drop of juice) for the monkey, and indicated that the monkey had detected the target.

The pattern of eye movements produced by a monkey for suprathreshold targets is shown in Figure 9.7A. Both increments and decrements were readily detectable. Moreover, the monkeys' normal eye movement reaction times were relatively short and regular, as shown in the top two panels of Figure 9.5B.

Next, the researchers treated the monkey with APB, disabling the ON-center system while leaving the OFF-center system intact, and retested the monkeys in the increment and decrement tasks. The data are shown in the bottom two panels of Figure 11.5B and in the deviant points in Figure 11.5C. With APB treatment the monkeys showed little change in reaction times to decrements, nor in decrement thresholds. But they had a very hard time with increments! Moreover, their capacity to respond to increments returned with the recovery from APB.

Putting the argument in slightly more formal terms, these data provide strong evidence that in an otherwise normal animal, a functional retinal ON system is a *necessary* condition for high sensitivity to increments of light, but not for high sensitivity to decrements. A functional retinal OFF system is apparently *sufficient* for processing decrements.

A similar experiment was carried out at light levels designed to test rod-mediated vision. In this case, the monkeys lost sensitivity to both increments and decrements of light. Thus, both of the original predictions were correct. These experiments, then, provide strong evidence to support the premise that ON cells do mediate our sensitivity to increments, and OFF cells to decrements. Given these results, can we design psychophysical experiments intended to reveal further characteristics

Figure 9.7: The experiment of Schiller, Sandell, and Maunsell (1986). A: The stimulus set-up. The monkey fixated a central fixation square presented against a mid-intensity background field. One of six surrounding test squares was presented, as either an increase or a decrease in luminance. The monkey’s task was to shift his eyes to fixate the test square as quickly as possible. The figure also shows the pattern of the monkey’s eye movements in response to clearly supra-threshold stimuli. All of the responses were prompt and correct. B: In the experiment proper, the test stimuli were presented at moderately supra-threshold luminances. The upper two panels show the distributions of reaction times in normal animals, to light increments (”normal, light”) and decrements (”normal, dark”) respectively. The lower two panels show the distributions of reaction times in APB-treated animals. C: The abscissae show days of the experiment. The arrow labelled ”APB” indicates the day of APB treatment; the arrow labelled ”Saline” indicates a control treatment with a saline solution. The upper panel shows reaction times, and the lower panel shows the monkey’s per cent correct eye movements (chance is 16.5%). APB clearly devastates visual function for light increments but has little if any effect for light decrements. (After Schiller et al, 1986, Fig. 1, p. 825.)

of these pathways in human subjects?

9.4.2 Effects of contrast polarity

Experiments designed to probe the differences between ON vs. OFF pathways often make use of stimulus patterns that vary in *contrast polarity.* The concept of contrast polarity refers to the direction of difference in luminance between a figure and its background. A white figure on a black background – an increment – has *positive contrast polarity.* A black figure on a white background – a decrement – has *negative contrast polarity*, and the two have *opposite contrast polarity.* Similarly, a stimulus of *uniform contrast polarity* is composed of elements that all have the same contrast polarity – e.g. either all white-on-black or all black-on-white (Figure xxM); whereas a stimulus of *mixed contrast polarity* is made up of both black-on-white and white-on-black elements. We will describe two experiments that make use of variations of contrast polarity to reveal manifestations of the ON vs. OFF pathways.

The first experiment relies on a simple argument. Given that there are separate ON and OFF pathways, and that ON and OFF ganglion cells have similar spatial properties, similar spatial effects should be seen for stimuli of opposite contrast polarities. A good example is the Westheimer effect, discussed above (Figure 9.2). In the Westheimer paradigm, thresholds for a tiny incremental test spot follow a U-shaped function with variations in the size of an illuminated background disk. If ON and OFF pathways have similar spatial properties, then detection thresholds for *decrements* of luminance of the same tiny test spot, on *dark* background disks, should follow a similar U-shaped function.

Experiments of this kind were performed by Sinai, Essock, and McCarley (1999). Sanai et al tested thresholds for incremental test stimuli on incremental background disks, and for decremental test stimuli on decremental background disks. The luminance of the disks was kept the same for both conditions, and the perceived brightness vs. darkness of the disk was varied by lightness contrast, by varying the luminance of a larger field that surrounded the disk (called the surround). Foveal viewing was used.

Sanai et al's results for the two conditions are shown in Figure 9.8. The upper panel shows increment thresholds on incremental disks. These increment-on-increment data replicate the classical U-shaped Westheimer function, with a threshold maximum on a disk of about 6' and a plateau at about 13'. The lower panel shows decrement thresholds on decremental disks. The decrement-on-decrement data reveal a highly similar U-shaped function! As the disk size increases from 3' to 6', thresholds increase – larger decrements are needed for detection. But as the disk size increases from 6' to about 13', thresholds decrease again – smaller decrements are sufficient for detection. Thus, the prediction that a U-shaped function would be found for decremental stimuli was confirmed. In sum, the parallel effects found with increments and decrements can be seen as a behavioral manifestation of the presence of separate ON and OFF pathways, with similar spatial properties, in the human retina.

The second experiment relies on a different line of reasoning. It begins with the premise that neural signals can only interact when they converge upon the same postsynaptic neuron. It follows that, since ON and OFF signals are processed separately throughout the retina, they cannot interact within the retina, but only after they converge on individual neurons in the visual cortex. Thus, if performance is the same with uniform and mixed contrast polarity, cortical processing is implied; whereas if performance decreases with mixed contrast polarity, retinal processing seems more likely.

Figure 9.8: Westheimer effects with incremental vs. decremental stimuli. A: Incremental test stimuli on incremental disks; B: decremental test stimuli on decremental disks. In both cases, thresholds increase as disk size increases from 2 to 6 minutes, and then decrease for larger disk sizes. The two curves are remarkably similar, in concordance with the receptive fields with highly similar spatial properties seen for ON-center vs. OFF-center ganglion cells. (Sanai et al, 1999, Fig 1, p. 1850).

O'Shea and Mitchell (1990) applied this logic to a study of vernier acuity. The stimuli were pairs of vertically oriented lines, as shown in Figure 9.9A. In each case the location of the upper line was held constant, and that of the lower line was varied in small steps. The subject's task was to say whether the lower line was to the left or to the right of the upper one. The pairs could be made up entirely of black bars on a white background (uniform contrast polarity), or of mixed white-on-black and black-on-white bars (mixed contrast polarity).

O'Shea and Mitchell's results for their most favorable subject are shown in Figure 9.9B. For this subject, across a wide range of conditions, vernier thresholds were about twice as high for mixed contrast polarity as for uniform contrast polarity stimuli. Although there were large individual differences, all three subjects did better most of the time with uniform contrast polarity. Thus, under the premises adopted, these data provide a second psychophysical manifestation of separate ON and OFF pathways in the human retina.

9.4.3 Perception of whiteness and blackness

Finally, Westheimer (200xx) has recently called renewed attention to Hering's (1974) original discussion of the perception of whiteness and blackness. Hering argued that like the sensations of redness and greenness, and yellowness and blueness (cf. Chapter xx), the sensations of whiteness and blackness are categorically distinct. He further argued that the perceptual distinctness of whiteness and blackness implies a major physiological dichotomy between the neural signals giving rise to them. In Hering's view, signals of entirely different kinds must underlie our perceptions of

Figure 9.9: Vernier acuity with same-contrast vs. opposite-contrast stimuli. A: The stimuli were pairs of vertical bars (left) or dots (right), of either the same (SC) or opposite (OC) contrast polarity. The upper and lower elements of the stimulus pair varied in their vertical separation. B. With both bars and dots, vernier acuity is better for SC than for OC stimuli, except at the largest separations. (O'Shea and Mitchell, 1990, Fig. 2, p. 210.)

whiteness and blackness. Westheimer renews and updates Hering's argument, adopting the premise that it must be the activity in the ON and OFF pathways that maps to the perception of whiteness and blackness respectively.

9.5 Manifestations of parallel processing

We have argued that the first two novel features of the retinal output code are center-surround antagonism and the separation of ON vs. OFF pathways, and we have explored possible psychophysical manifestations of these properties. The third novel feature is that of parallel processing. In the cat retina, parallel processing is manifested in the separation of X and Y cells. The human retinal shows similar separations of functions among classes of cells called M, P and K cells. However, since the definitions of these cell types are tightly entwined with their distinctive inputs from different cone classes, we will postpone further discussion of parallel processing in the primate retina to Chapters 10 and 11.

9.6 Some things are not retinal

In summary, we have explored a variety of examples of perceptual phenomena whose origins and/or properties are commonly attributed to retinal processing. Many similar causal stories can be found in the vision literature.

On the other hand, of course, there are many other visual effects to which no analogy is found in the retinal output signal. For example, in many cases the orientation fo a line or pattern influences its perception (see Chapter xx). Since in general the responses of retinal ganglion cells do not vary with the orientation of the stimulus, it is difficult to make an account of any perceptual phenomenon that includes differential processing of stimuli of different orientations. As we will see, cortical neurons are tuned for orientation, and vision scientists usually rely on them for accounts of orientation effects. This topic is discussed at length in Chapter xx. And all higher-level perceptual and cognitive process, such as object recognition and reading, presumably rely largely on processing at cortical levels.

9.7 Summary: Retinal recodings leave their marks

In the present chapter, we examined several different causal stories in which vision scientists attempt to explain perceptual phenomena on the basis of the functioning of retinal ganglion cells. In dong so, we emphasized the use of analogies and the central role they play in causal stories. In the realm of detection thresholds we explored the analogy between center/surround antagonism and the Westheimer phenomenon, and between ON-center and OFF-center ganglion cells and the detection of light increments and decrements. And considering ensembles of ganglion cells and taking a more perceptual tack, we also examined the argument that activity in an ensemble of ganglion cells might provide the explanation for the perceptual illusion known as Mach bands. These examples were intended to give you some practice at working through the kinds of causal stories that are typical of visual science.

Finally, a fundamental question. Is this kind of argumentation useful? Some vision scientists have argued that causal storieis relating particular perceptual phenomena to particular coding stages at early levels of the visual system is not a useful enterprise; just describe the anatomy, physiology and the psychophysical effects separately, they say, and leave it at that. Each student is encouraged to decide for himself whether causal stories of the kinds we have discussed are interesting and useful, or not. Of course to DT the answer is an unqualified YES.but the stories must be told very carefully, with all their assumptions and limitations made explicit.

Chapter 10

Light and Dark Adaptation

The terms *light adaptation* and *dark adaptation* refer to the changes in visual function that occur with changes in light level, and to the changes in physiological processing that bring about these functional changes. Adjustments to increases of light level are called light adaptation, and adjustments to decreases are called dark adaptation.

Human beings need to function in a wide range of visual environments. From finding our way through the woods in starlight to skiing on a snowfield, we encounter changes of (say)10^{10} – 10 orders of magnitude – in the average environmental light level. To be most useful to us, our visual systems must function well over this entire range. In particular, the EDC faced two difficult design requirements. First, at the low end of the range, our visual systems need to function under conditions of extreme quantum scarcity, and make the most of every available quantum of light. But second, small relative variations in light levels – low *contrasts* – are useful in defining the textures, shapes, and surfaces of objects (see Fig. xx on depth from shading), and are important to object recognition. So to be most useful to us, the visual system also has to be able to detect low contrasts, and do so over as wide a range of environmental light levels as possible.

The overall design specification, then, would be something like: *make a system that can detect individual quanta at very low light levels, and respond to low contrasts over a wide range of higher light levels.* This was a staggering request, and we are told that some members of the EDC chose to resign rather than accept this mission.

We begin our treatment of light and dark adaptation with a dramatic demonstration of the adjustments the visual system makes to different light levels. We then review some of the classical psychophysical data – light and dark adaptation curves, and changes in spatial contrast sensitivity functions. Next we explore a pair of theoretical approaches used to explain the major adaptation effects. We then return to a second set of psychophysical experiments, designed to place further constraints on the properties of the adaptation process.

Finally, we review what is known about the physiology of light and dark adaptation. We warn you at the outset, however, that the physiological mechanisms of light and dark adaptation are both complicated and not yet fully known. Thus, there is as yet no complete or universally accepted causal story relating the psychophysical data to their neural underpinnings. Causal stories of light and dark adaptation are still complicated works in progress, and when they are finished they will still be complicated. In fact, they serve as our first example of an incomplete and multifaceted causal story.

As you will see, most of the changes in visual processing that control light and dark adaptation occur within the retina. Thus, another reason for taking up light and dark adaptation at this point is to exercise your new knowledge of retinal neurons and retinal circuitry.

10.1 A demonstration and a theoretical overview

Light and dark adaptation bring about enormous changes in our vision. Yet these changes occur so automatically that we usually take them for granted, and most people are only dimly aware of them (no pun intended). To help bring the everyday manifestations of light and dark adaptation to your attention, here are three questions. First, where are the stars in the daytime?[1] Second, when you first turn out the lights on your way to bed in a strange hotel room, why do you have to grope your way across the room? And when you wake up later, why is it so much easier to move about? And third, why did your acuity change with light level in the demonstration of Figure 10.1? We will return to these questions as the chapter proceeds.

10.1.1 A lesson from the closet

One reason that light and dark adaptation usually pass unnoticed is that usually both of your eyes are in the same adaptational state. A spectacular demonstration of the differences in processing between light and dark adapted states can be produced by creating different levels of adaptation in your two eyes, and then looking out alternately at the world through one eye and then the other.

There are two ways to do this demonstration. The first is to close one eye and cover it with a black eye patch (or a piece of black construction paper held on tightly with tape). Put yor hand over it to keep out as much light as possible, and wait 15 minutes or more. The patched eye will become your *dark adapted eye.* Then, go into a brightly lit room (or better, outdoors on a bright day) with the other eye left open, for a minute or so. It will become your *light adapted eye.* Now quickly go to a very dark room (say, a closet or a windowless bathroom with just the tiniest bit of light coming in under the door). Quickly observe your surroundings through your light adapted eye. Then compare what you see through light and dark adapted eyes by alternately opening one eye and then the other – use about two seconds for each eye.

The second way to do this demonstration is to wake up with your wits about you in the middle of the night (this option is not available to DT), when both eyes are fully dark adapted. Close one eye and put an eye patch and your hand tightly over it, leaving the other eye open. Now go into that bathroom and turn the lights on for at least a minute, so that your open eye is light adapted. Turn the lights off again, and quickly compare what you see through the two eyes.

With your light adapted eye, you will probably see a lot of sparkly "noise", but discern nothing at all about the visual environment. Many people report a weird feeling of pressure, as though something were wrong with their light adapted eye. But when you switch to your dark adapted eye, the room and the objects within it should jump into view. Then over the course of several minutes, as the light adapted eye becomes dark adapted, your visual impressions through the two eyes should come to be more and more the same.

[1] DT once put this question on a quiz. The class wag answered: Each star has a little hole in the sky that it lives in, like the holes on a golf course. When the sun starts to come up, the star jumps into the hole and closes the door, leaving nothing but sky behind. (He got full credit, of course.)

10.1.2 Changes in the processing state of the retina

Here's the puzzle: How is it that you perceive the world so differently with your two eyes, when both are receiving essentially the same incoming stimulus? The presumed answer is that changes in the physiological processing states of the retinas must underlie the perceptual changes, and that these changes go on independently in the two retinas. We will call the presumed physiological processing changes that control light and dark adaptation the *adaptation process(es).* But what kinds of changes take place? And in which retinal neurons?

10.2 Psychophysics: Classical system properties

The first step of our quest is to quantify the psychophysics of light and dark adaptation. Precisely what are the changes in the system properties of vision with light and dark adaptation? Are there some landmarks with which to anchor our thinking? And what constraints do the system properties place on models of the adaptation process?

10.2.1 Light adaptation curves

Where are the stars in the daytime? Stars are maximally visible when the sky is darkest. The dimmest visible stars begin to disappear as dawn is imminent, and stars of higher and higher intensities disappear in sequence as scattered light from the sun increasingly illuminates the sky. But do the stars leave the sky? Of course not. What happens is that you require higher and higher intensity stars in order to detect the *increase* or *increment* in illumination provided by the star against the background of the sky. In full daylight, none of the stars are visible – none provide a sufficient increment to exceed your detection threshold against the daytime sky. With light adaptation, your *detection threshold* (or *increment threshold*) goes up – you lose your sensitivity to dim lights.

To quantify these elevations of detection threshold in the laboratory, we substitute a *test spot* (a relatively small field of light) for the star, and an *adapting* (or *background*) *field* (a second, larger field of light) for the sky, as shown in the inset in Figure 10.1A. We ask the subject to fixate a fixation target, in order to present the test spot in the retinal periphery. The subject's task is to determine a detection threshold for the test spot – an absolute threshold against zero background (minus infinity on the log axis in Figure 10.1), and an increment threshold for each of a series of intensities of the adapting field.

In 1937, M.L.J. Crawford carried out a set of classic experiments in light adaptation. One of Crawford's light adaptation functions, measured at 14^o peripheral, is shown in Figure 10.1A. On the abscissa is the intensity, I, of the adapting field, and on the ordinate is ΔI, the subject's detection threshold for the incremental test spot. Both I and ΔI are plotted on log axes. Psychophysical data of this kind are variously called *light adaptation, field adaptation*, or *threshold-vs-intensity (tvi) functions*, and they display the absolute sensitivity of the subject to the test flash.

As we learned in Chapter 2, at absolute threshold a subject can detect the absorption of one or a very few quanta of light – the EDC fulfilled the first design requirement. But as shown in Figure 10.1, as the adapting field is turned on and increased in intensity, the detection threshold rises – the subject *loses absolute sensitivity* to the test spot. In fact, as the adapting intensity changes by a factor of 10^{10}, the detection threshold changes by about 10^5 overall in this particular experiment.

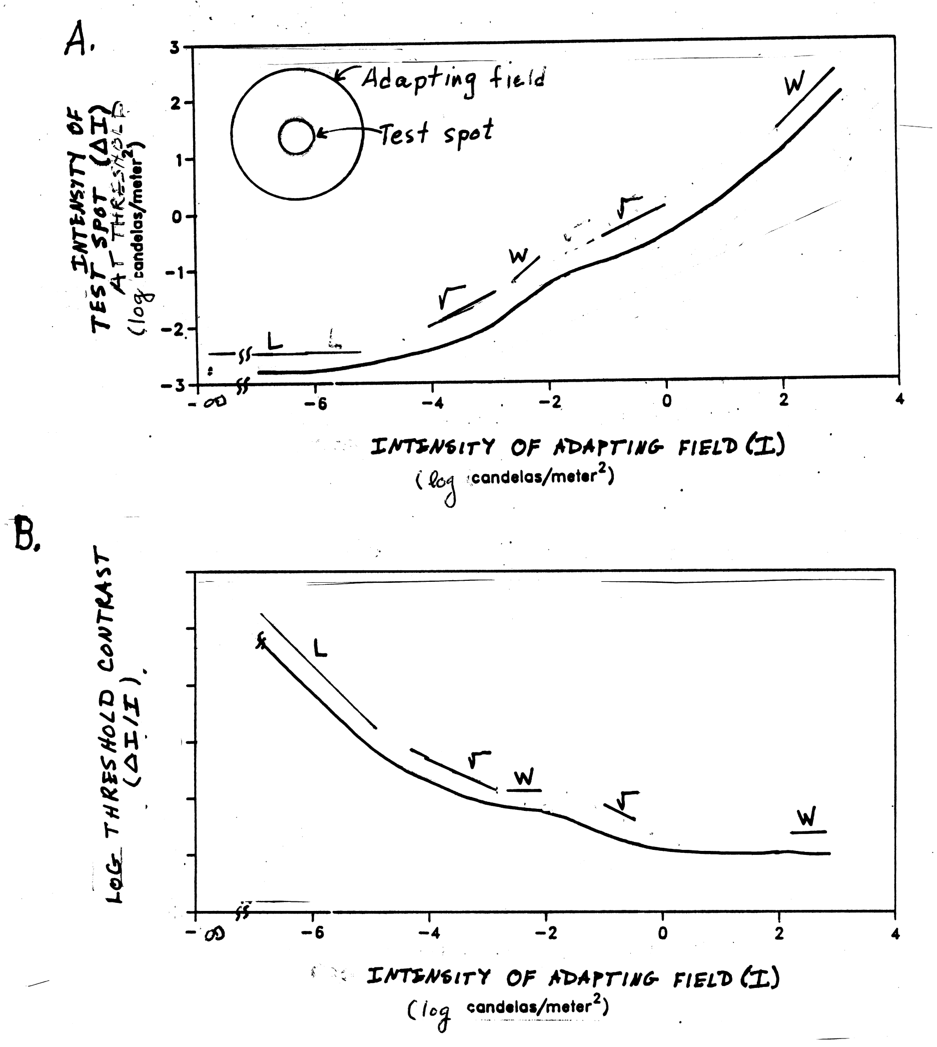


Figure 10.1: Light adaptation functions. A. Light adaptation with white lights at 14^o in the peripheral retina. The abscissa shows the log intensity of the adapting field, ΔI, and the ordinate shows the log intensity of the test spot required for detection threshold, ΔI. As the intensity of the adapting field increases over a range of about 10^{10}, the threshold for the incremental test spot increases over a range of about 10^5. The light adaptation curve shows a marked kink at about -2 log cd/m^2. Line segments marked with W have a slope of 1 (Weber's Law). Line segments marked with the square root sign have a slope of 1/2 (the square root law). And line segments marked with an L describe the ”linear” portions of the curve. This plot emphasizes the large losses of *absolute sensitivity* involved in light adaptation. B. The same data plotted in terms of contrasts – the ratio of the increment threshold to the adapting field intensity. As the level of light adaptation increases, the threshold contrast decreases – the subject becomes increasingly sensitive to stimulus contrast. This plot emphasizes the excellent *contrast sensitivity* achieved at intermediate and high levels of light adaptation. [A. after Crawford, 1937, via Hood and Finkelstein, 1986, Fig. 5.39, p. 5-32.] B. Invented by DT; to be replaced by calculations.]

In other words, compared to the threshold measured under dark adapted conditions, high levels of light adaptation involve *enormous losses of absolute sensitivity.* Nonetheless, even at the highest adaptation levels the subject can still detect the test spot if its intensity is high enough.

But now let's look at the same data in relative rather than in absolute terms. To do this we convert the detection thresholds in Figure 10.1A to *contrast thresholds*, as shown in Figure 10.1B. In the context of light adaptation paradigms, the contrast threshold is defined as ΔI/I – the detection threshold for the test spot, ΔI, divided by the intensity of the adapting field, I.[2] Figure 10.1B reveals that contrast thresholds are high in the dark adapted eye, but decrease rapidly, leveling off briefly at perhaps 10%xx just below an adapting field intensity of -2 log cd/m^2. As the adapting intensity increases further, contrast thresholds decrease again, followed by a second leveling off at about 1%xx at about 0 log cd/m^2. Under Crawford's testing conditions, above an adapting field luminance of about -2 log cd/m^2, the EDC achieved its second design requirement – detection of contrasts of 1% or less.

Finally, let's return to the original light adaptation function in Figure 10.1A, and look at some of its additional properties. First, because of their theoretical interest (see later), the different portions of the light adaptation curves with different *slopes* have acquired specialized names. These regions are indicated with short line segments in Figure 10.1A and B. A region with a slope of zero is sometimes called a *linear* region of the curve – the increment threshold is constant with variations in adapting field intensity (ΔI = a, where a is a constant). A region with a slope of 0.5 is said to follow the *square root law*, because the threshold increases with the square root of the adapting field intensity (ΔI = c sqrt I, or log ΔI = 0.5 log I + c', where c is a constant and c' is the log of c). Finally, a region with a slope of 1 is said to follow *Weber's law* (ΔI = kI, or log ΔI = log I + k', where k is a constant and k' is the log of k). Depending upon testing parameters and conditions, each branch of the light adaptation curve can begin with a linear portion; transition gradually through the square root law; and achieve a final slope that approaches Weber's Law. In the contrast plot of Figure 10.1B, the linear region becomes a region with slope of -1, the square root region has a slope of -0.5, and the Weber region has a slope of 0.

Second, notice that the light adaptation curve seems to have two branches, which intersect at about -2 log cd/m^2 in Crawford's experiment. The two-branched function suggests that two different retinal processes might be controlling the light adaptation function in its lower vs. upper branches. Using variations of retinal location and wavelength, we can dissect the curve and manipulate its two major branches selectively, as shown in Figure 10.2.

If we continue to test in the retinal periphery, and vary the *wavelength* of the test spot, the two branches of the curve shift vertically, quite independently of each other. For example, if we change the wavelength of the test spot from 580 to 500 nm, the threshold for the test spot goes down in the lower branch, and up in the upper branch, as shown in Figure 10.2A. In fact, the lower branch shifts by amounts predictable from the scotopic spectral sensitivity curve, and the upper branch shifts by amounts predictable from the photopic spectral sensitivity curve (Chapters 2, 3). These spectral characteristics suggest that the lower branch is mediated by rod-initiated signals, and the upper branch by cone-initiated signals. The idea that cones mediate the upper branch is further supported by the fact that if we make the test spot very small and confine it to the fovea, where there are no rods, we see only the upper branch of the curve.

[2] Compare this definition to the definition of contrast for a sinusoidal grating in Chapter 5. The two definitions differ in detail. But in both cases the concept of contrast refers to the change of light level created by the test stimulus, compared to the space average luminance (or adaptation) level.

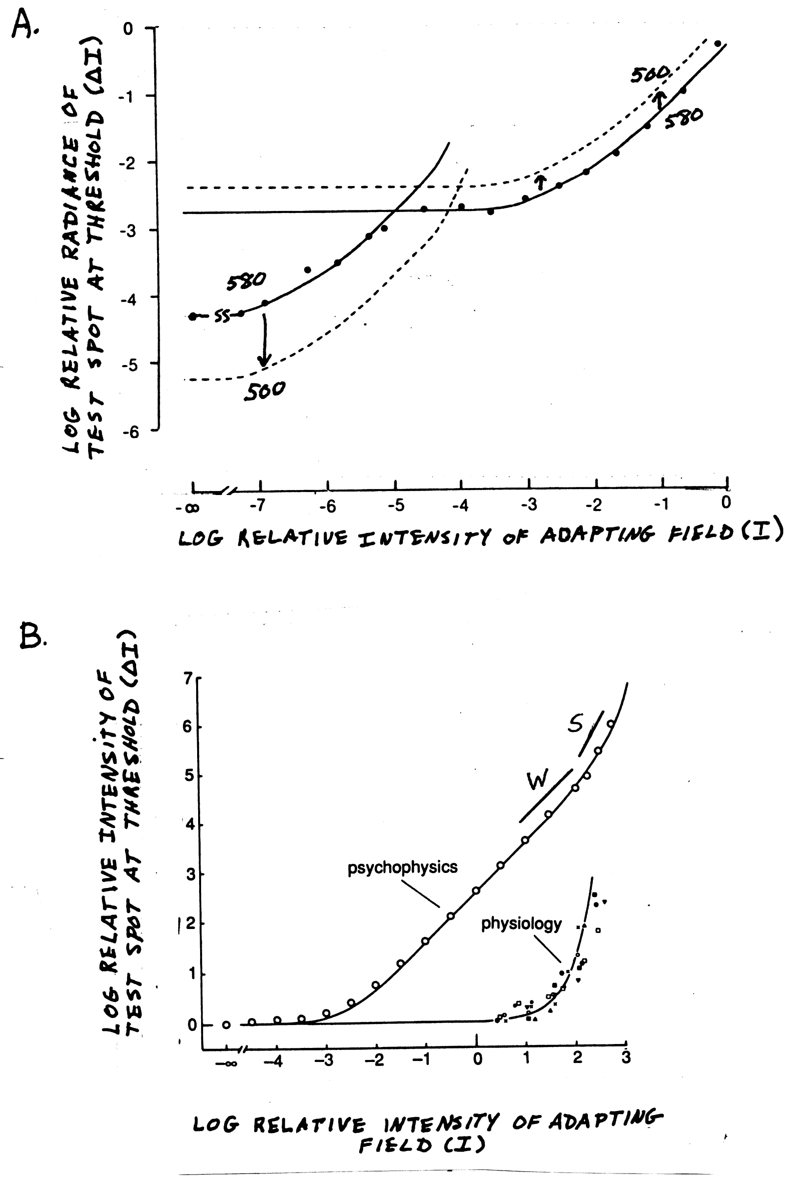


Figure 10.2: Rod and cone branches of the light adaptation curve. A. As the wavelength of the test spot is varied, the rod and cone branches move vertically in accord with the spectral sensitivities of rod and cone vision respectively. For example, a switch from a 580 to a 500 nm test light moves the lower portion of the curve downward (rod vision is more sensitive at 500 than at 580 nm), whereas it moves the cone portion upward (cone vision is more sensitive at 580 than at 500 nm). B. When the rod-mediated portion of the curve is isolated by an optimal choice of stimulus parameters, it shows a marked increase of slope, above that predicted from Weber's Law, at its high end. This phenomenon is called rod saturation. The points at the lower right show physiological recordings from rods (to be discussed later). [A after Hood and Finkelstein, 1986, Fig. 5.42, p. 5-35. B after Aguilar and Stiles, 1954, via Walraven et al, 1990, Fig. 21, p. 83.]

And fourth, if we select conditions that shift the cone branch upward and the rod branch downward as much as possible, we can see the rod branch over the broadest possible range, as shown in Figure 10.2B. Under these conditions, the rod-mediated curve shows Weber's Law over a range of several log units, but then steepens abruptly *to a slope greater than 1* at the high end of its range. It is as though the rod system has reached the top of its functional range, and can only barely continue to signal further increases in light level. This psychophysically defined phenomenon is called *rod saturation*. (We will return to the curve labeled "physiology" later in the chapter).

So in summary, psychophysical light adaptation curves reveal several novel system properties: elevations of detection thresholds, reductions of contrast thresholds, rod- and cone-mediated branches, and a remarkable variety of slopes. Our eventual question will be, what properties of retinal physiology cause these changes in detection thresholds?

10.2.2 Dark adaptation curves

Now, when you first walk into the movie theater on a sunny day, why are the seats so hard to see, and why do they become more visible after a minute or two? And what about that hotel bedroom? In fact, for some odd reason the EDC gave you a physiological system that takes nearly an hour to recover fully from high levels of light adaptation.

In a dark adaptation experiment in the laboratory, the subject is first light adapted to a large, high intensity adapting field. The adapting field is left on for perhaps a minute, and then abruptly turned off. As soon as the adapting field goes off, the subject fixates a fixation target, placing the test light in her peripheral retina. But this time the test spot is presented against a completely dark background. The subject's task is to adjust the intensity of the test spot until it is just visible, and to repeat this measurement at a series of times after the offset of the adapting field.

In 1937, Selig Hecht and his colleagues (Hecht, Haig, and Chase, 1937) studied the time course of dark adaptation. Typical results of their experiment are shown in Figure 10.3A. The subjects detection threshold for the test spot, in logarithmic units, is plotted on the ordinate. On the abscissa is time in the dark. The adapting field intensity is zero throughout the experiment. But the detection threshold decreases dramatically – sensitivity increases by perhaps six orders of magnitude – as a function of time in the dark. That is, the adaptation process must be readjusting retinal processing, so that detection of dimmer and dimmer test spots becomes possible over time.

As is the case in light adaptation, there are two clear branches to the typical dark adaptation curve. The upper branch descends relatively rapidly. For very intense adapting fields, the asymptotic value of the upper branch is reached in about 5 minutes. This asymptotic threshold value – also called the *cone plateau* – is then maintained up until about 12 minutes after the offset of the adapting field. At this point the threshold begins to descend again, this time more slowly. The second branch of the curve reaches its asymptotic value – we are back at absolute threshold – after about 45 minutes to an hour in the dark.

As we did in the case of light adaptation, we can use standard manipulations of the stimulus to sort out the receptors that mediate the two branches of the dark adaptation curve. If we test dark adaptation in the fovea with small, long-wavelength test fields, we see only the upper branch of the curve, as shown in Figure 10.3B. Thus, we attribute the upper branch to cone-initiated signals. If we test with different wavelengths of light, we can predict the spectral characteristics of the lower branch from the scotopic spectral sensitivity curve; thus, we attribute the lower branch to

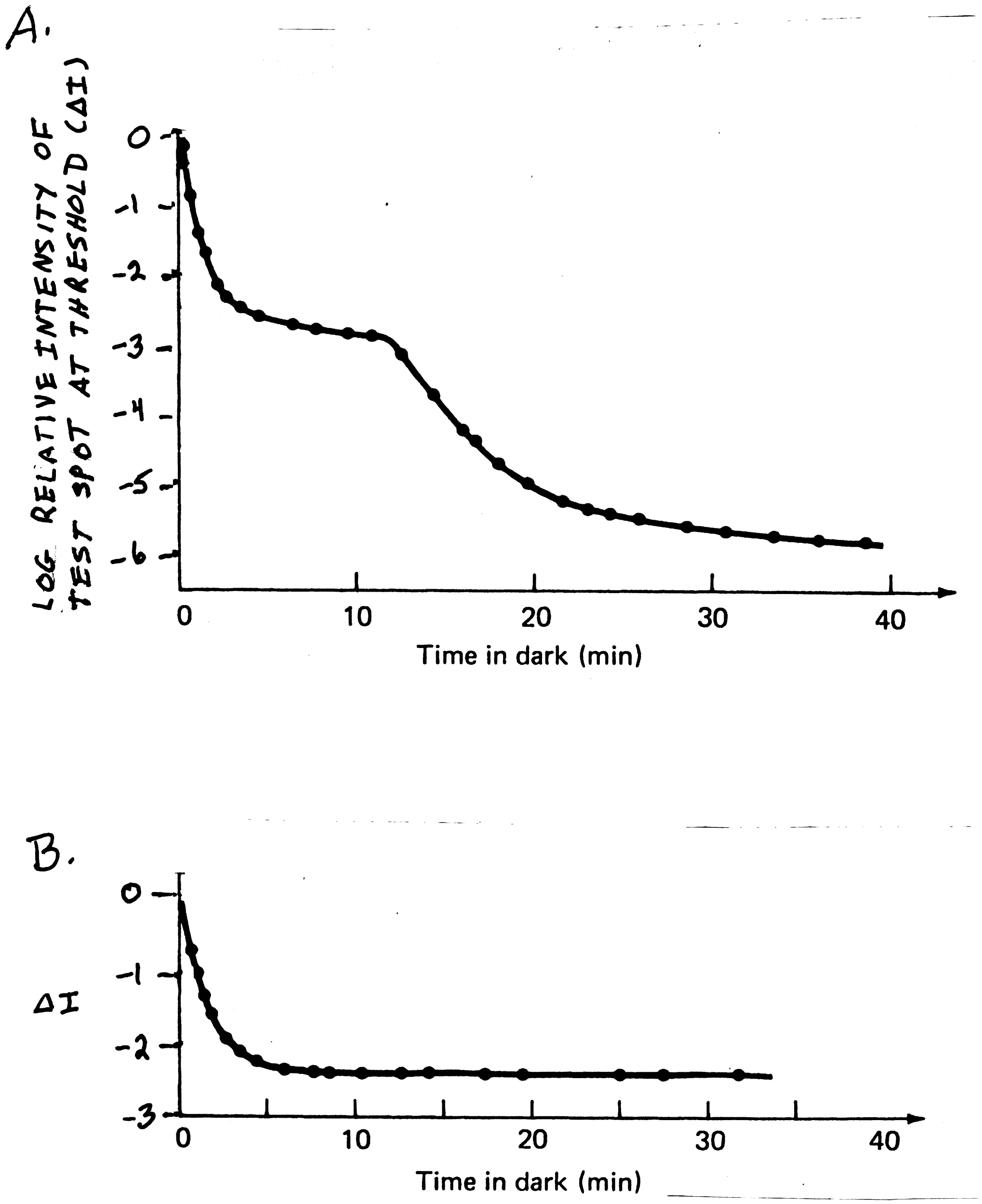


Figure 10.3: Dark adaptation. A. A dark adaptation curve measured in the peripheral retina after a very intense preadaptation. The adapting field was extinguished at time zero. Notice the two branches of the dark adaptation curve, with the cone plateau extending from about five to about ten minutes after the adapting light was extinguished. The upper branch is attributed to cone-initiated signals, and the lower branch to rod-initiated signals. B. A dark adaptation curve measured in the fovea. Only the upper (cone) branch is present. [From Hecht, Haig, and Chase, 1937; via Levine, 2000, Figs. 6.3 and 6.4, p. 99.]

rod-initiated signals[3].

Why did the EDC allow dark adaptation to proceed so slowly? One speculative answer is that photopigments regenerate slowly, and may exert some indirect control over the state of adaptation. Another level of speculative answer is that under natural circumstances light levels usually change gradually, and that somehow there is little evolutionary pressure for dark adaptation to be rapid. Perhaps the slow dark adaptation curve was just the easiest way to fit the overall design together. (The EDC didn't anticipate light switches).

10.2.3 Contrast sensitivity functions and acuity

Now, how do the spatial properties of vision change with light adaptation? This question can be addressed by using sinusoidal grating stimuli, and measuring *contrast sensitivity functions* (CSF; see Chapter 5) at a range of different levels of adaptation. To measure a CSF, we use a large homogeneous field of fixed space-average luminance (which becomes the adapting luminance). We then test the contrast threshold for a small patch of a sinusoidal grating embedded in the adapting field, for a series of different spatial frequencies. The experiment is repeated at several space average luminance levels, up to the maximum available on the video monitor needed to generate the gratings.

The results of such an experiment are shown in Figure 10.4A. At the lowest level tested – xx log cd/m^2 – the CSF appears to be low pass, and the contrast threshold is below 10% (a contrast sensitivity of 10) at all spatial frequencies. As the space average luminance increases the CSF shifts upward – contrast sensitivity increases for each spatial frequency – and the curve becomes bandpass at about xx log cd/m^2. As we shift from rod-mediated to cone-mediated vision, the peak of the CSF shifts to higher spatial frequencies. In consequence, at low spatial frequencies contrast sensitivity reaches a maximum value at about 1 Td xx, but at high spatial frequencies it keeps on increasing for another two log units, up to perhaps 100 Td xx.

[Try choosing a single spatial frequency, drawing a vertical line at that frequency, and plotting the *contrast threshold* (1/contrast sensitivity) as a function of the space average luminance. You should get a graph similar to that shown in Figure 10.4B. Both data sets are telling us that contrast thresholds decrease – contrast sensitivity increases – with increasing light levels.]

What about grating acuity? The upward shift of the CSF shown in Figure 10.4 produces a rightward shift of the (extrapolated) high-frequency cut-off, at least up to a luminance of about xx. Since grating acuity is closely related to the high frequency cut-off of the CSF (see Chapter 5), it too should increase with increasing luminance. These ideas integrate the changes in grating acuity we saw in Figure 1.2 into the broader context of light adaptation.

Many other characteristics of vision also change with light and dark adaptation. Most strikingly, of course, color vision becomes available at photopic levels. [Explore other differences in perception with light level in your own eyes as the occasion arises.]

[3] The theory of independent rod and cone mediation of light and dark adaptation is classically called *duplicity theory* (another odd name – always sounds a bit sneaky to DT). Notice the assumptions that the rod and cone systems light and dark adapt independently, and that the most sensitive system available determines the detection threshold. That is, the overall curve follows the lower envelope of the rod- and cone-mediated branches.

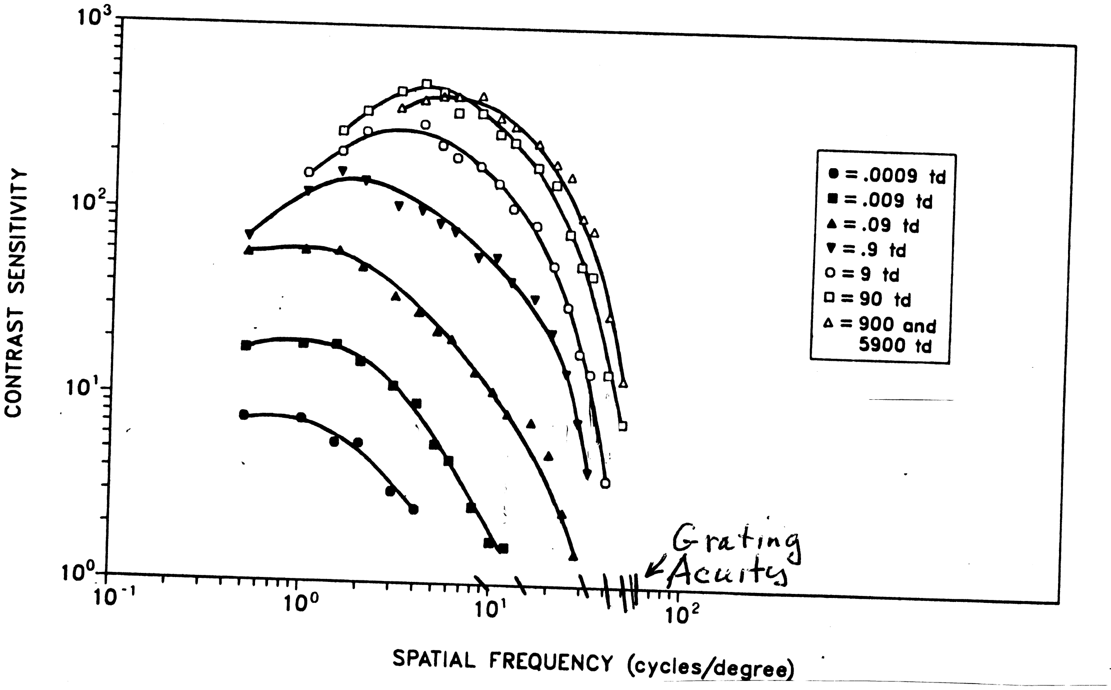


Figure 10.4: The effects of light adaptation level on the CSF. A: CSFs measured at different space average luminances (light adaptation levels), from xx to xx cd/m^2. The solid symbols indicate rod mediation; the open symbols, cone mediation. As the space average luminance increases, contrast sensitivity increases at each spatial frequency. The CSF is lowpass at low light levels, and begins to show a peak at about xx log cd/m^2. The peak shifts from about two to about seven cy/deg across the higher range of space average luminances. Extrapolating each CSF downward to the abscissa at high frequencies yields estimates of grating acuities, which also improve as the space average luminance increases. [A from van Ness and Bouman, 1967, via Olzak and Thomas, 1986, Fig. 7, p. 7-18. B – best option not yet located.]

10.3 Modeling tools

The threshold changes involved in light and dark adaptation are so large and striking, and have been described in so much detail for so long, that they h ave attracted the attention of many scientists who enjoy modelling. Two major approaches have often been used in models of light adaptation. We will call them *noise processes* and *multiplicative processes*. (A third category, called *subtractive processes*, are ignored for the sake of simplicity.)

10.3.1 Noise processes

As we saw in Chapter 4 and 6, near absolute threshold the quantal nature of light has profound effects on detection thresholds. At absolute threshold, a human subject can detect the absorption of a single quantum of light (or at most a very few), and most of the variability represented in the subject's responses is thought to be due to quantal fluctuations in the stimulus. That is, at absolute threshold we have exquisite *absolute* sensitivity – sensitivity that approaches the absolute limit of responding to a single absorbed quantum. Notice that at this level, however, *contrast* sensitivity must be very limited, again because of the quantal nature of light. After one quantum, the next available amount of light is two quanta – a 100% change! So the EDC's design specification of responding to a 1% change is theoretically impossible at such low light levels.

Moreover, as we turn on a very dim adapting field, we find that detection thresholds are at first unaffected – the slope of the light adaptation function at very low light levels is zero (the linear portion of the light adaptation function). That is, when quanta from the adapting field arrive too infrequently, they rarely coincide with quanta from the test spot. Thus they provide no mechanism for raising the detection threshold.

As the adapting level is increased further, detection thresholds eventually rise above the absolute threshold, and the slope of the light adaptation function begins to increase. Under some combinations of stimulus parameters (particularly with very small test spots) a region with a slope of 0.5 can extend over several log units of adapting field intensity. This region has played a particularly large role in guiding theories of rod-mediated light and dark adaptation.

Theoretical accounts of the lower part of the rod-mediated light adaptation function have usually relied heavily on signal/noise considerations. As we saw in Chapter 4, the quantal nature of light leads to instantaneous fluctuations in the actual quantal catch from a test spot of a nominally fixed intensity, as well as from an adapting field of a fixed nominal intensity.[4] In the context of some simple theoretical elaborations, these considerations lead to predictions of the square root law (a slope of 0.5). In addition to quantal noise, the visual system also adds intrinsic noise – noise generated within the visual neurons by spontaneous isomerizations of photopigment molecules and other factors. A detailed account of signal/noise theories of light adaptation is beyond the scope of this book, but we will be on the lookout for noise processes as we go along.

[4] The distribution of numbers of quanta in the test spot varies as a Poisson distribution, in which the the standard deviation is equal to the square root of the mean. The light from the background field is also Poisson distributed, and constitutes noise against which the signal from the test spot must be detected. Since the value of (sqrtN)/N diminishes with N, contrast thresholds can decrease with increasing light adaptation.

10.3.2 Multiplicative processes: Dark glasses and automatic gain controls

Multiplicative processes have been the centerpiece of many models of light adaptation. The underlying concept is simple – it's just the idea that the retina produces light adaptation by multiplying the signals from all incoming lights by a common factor less than one.[5] Over the years, multiplicative processes have been described with three different metaphors, each with a different name: *dark glasses effects, automatic gain control*, and (at a more physiological level) *shifts in dynamic range.* Let's walk through these metaphors to see how multiplicative processes work.

The first metaphor is that of *dark glasses.* Suppose we wanted to build a robot with eyes that could process visual inputs equally well over a large range of levels of ambient illumination. One fanciful but effective way to do this is shown in Figure 10.5. We could outfit the robot with a little external gizmo consisting of a pair of photocells and a pair of dark glasses made from neutral density filters of variable density. The photocell could absorb quanta, create a signal that corresponds to the ambient light level, and feed back this signal to control the density of the filter over each eye. In the dark glasses metaphor, the higher the light level – the higher the adaptation state – the denser the glasses.

For example, suppose the illumination in the world increases by a factor of 10 (or 100, or 1000...). The luminance of every physical surface will increase by a factor of 10 (or 100, or 1000...). The feedback gizmo could increase the quantal absorption rate of the filter by a factor of 10 (or 100, or 1000...), and exactly restore the original retinal illuminance at every point in the retinal image. We will call this version of the dark glases model the 1:1 version – an increase of illumination of a given factor produces an increase in absorption in the dark glasses by the same factor, exactly restoring the retinal image to its initial state.

The beauty of the dark glasses model is the simplification to which it gives rise. Since nothing in the retinal image would change with ambient light level, there would be no need to change anything whatsoever in retinal processing. The visual system could operate exactly the same way over the whole range of light levels for which the dark glasses gizmo is assumed to work.

Moreover, notice that the 1:1 dark glasses model predicts Weber's Law. This is true because as the ambient illumination is increased by a factor of 10 (or 100, or 1000....), the gizmo scales the retinal illuminance back to its original value; and it automatically scales back the intensity of the test spot by the same factor. [Work this out on the axes used in Figure 10.1A.)

In addition, the dark glasses approach can also be used to model other slopes of light adaptation curves. Suppose the gizmo divides the incoming signal by a factor less than the change of adapting intensity. For example, if the gizmo allows the retinal illuminance to increase by 0.5 log unit for each log unit increase of ambient illumination, the result would be a light adaptation function with a slope of 0.5 – the square root law. In a multiplicative model, the multiplicative factor – the factor by which one divides the signal – becomes a parameter, and any desired slope can be generated.

The second metaphor for a multiplicative process is that of *gain control.* Why don't we have neutral density gizmos built onto our noses? The EDC doubtless thought it would be inelegant. Variable neutral density eyelids were a possibility, but the materials were not available. Better, they thought, to build a multiplicative feedback mechanism *inside* the visual system.

[5] DT was confused by light adaptation theories for a long time, because she supposed that a ”multiplicative” process would *increase* the strength of a signal. Nothing about models of light adaptation makes sense when one starts out with this error. In fact the multiplication is always by a factor less than one. This booby trap could be avoided by calling the process *divisive* (division) rather than multiplicative (but then, maybe no one but DT was ever confused on this point ; >).

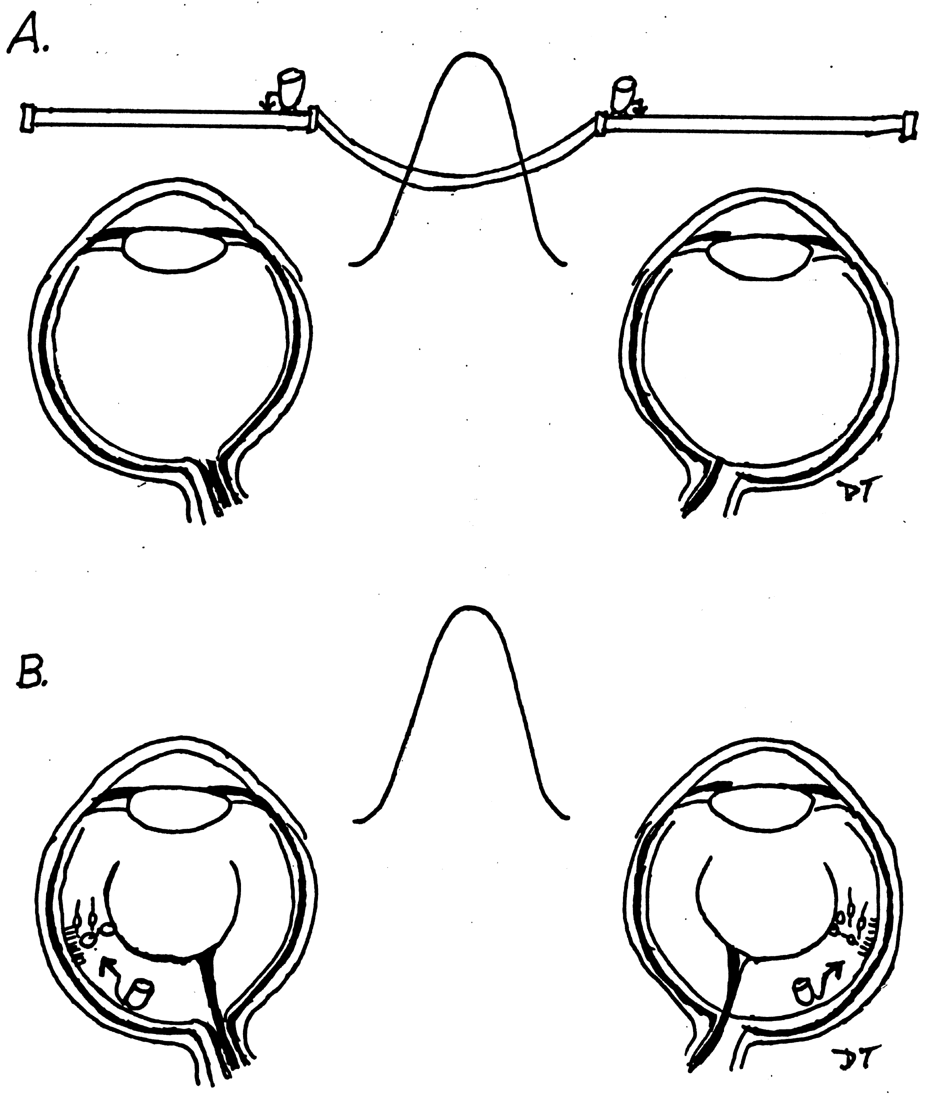


Figure 10.5: Metaphors of light adaptation. A. Dark glasses theory. The photocells sense the average light level in the scene, and increase the density of the glasses in proportion to the average light level. B. Automatic gain control. The retinas are shown expanded in thickness within the eyes. In the simplest case, the automatic gain control has the same effect as the dark glasses, but sensors and their effects are inside the visual system.

A metaphoric gain control mechanism is shown in Figure 10.5B. As discussed in Chapter 5, the *gain* of a system is defined as the output/input ratio – the magnitude of the output with respect to the magnitude of the input. The lower the gain, the smaller the fraction by which the input is multiplied to produce the output. So in a theory of light adaptation, the higher the light level – the higher the state of adaptation – the lower the metaphorical gain.

In the simplest case, dark glasses and gain controls have identical effects, in the sense that a factor of 10 increase in absorption by the dark glasses is equivalent to a factor of 10 decrease of gain. And just as in the case of the dark glasses, the gain changes can be 1:1 with the increases in illumination, producing Weber's Law, or less than 1:1, producing light adaptation curves with shallower slopes.

The internal gain control mechanism has, however, a distinct advantage. Variations of the state of adaptation across the retina were not feasible with the dark glasses gizmo. Yet it would be desirable to vary the adaptation state across the retina, using a lower gain for regions of the visual field illuminated by direct sunlight, and a higher gain for regions that lie in shade. In principle, an internal gain control mechanism can be made local, and this provides another reason for putting the gain control mechanism within the retina rather than on the nose.

10.3.3 Shifts of dynamic range

A third and more physiologically realistic version of a multiplicative light adaptation process is called a *shift of dynamic range*. The concepts we need are developed in Figure 10.6.

We begin with the concept of *dynamic range*. As shown schematically in Figure 10.6A, the dynamic range of a neuron is defined as the range of *inputs* over which the neuron's *output* changes. The level of adaptation is fixed, and the retina is exposed to incremental test flashes of varying intensity. The dynamic range of this neuron under these conditions is from about -7 to about -5 units on the log abscissa of Figure 10.6A – about a factor of 100, or a two log unit range. This is a typical dynamic range for many kinds of visual neurons.

Below the low end of the dynamic range, all test stimuli are below the threshold for the neuron, and the neuron does not respond. Then, once we enter the dynamic range, we find a range of intensities over which progressively higher inputs produce progressively larger responses – the neuron responds differentially to different light levels. At the top of the dynamic range, the neuron *saturates*: it reaches its maximum response level, and further increases in input make no further change in output. So the lower end of the dynamic range is marked by the detection threshold, and the upper end by saturation.

In short, a neuron with a limited dynamic range is useful in coding variations in luminance only within its dynamic range. But what if the sun comes out? If all of the light levels reaching the eye were to shift upwards by, say, a factor of 1000, they would all saturate the cell, and the cell could no longer code luminance variations in the range provided by the environment.

A solution to the problem of limited dynamic range is to *shift the dynamic range* of the neuron along the abscissa in concert with the ambient illumination, as shown in Figure 10.6B. Such a shift in dynamic range will allow the cell to respond differentially to stimuli in a new range of luminances. If the ambient illumination and the dynamic range shift by exactly the same factor, as shown in Figure 10.6B, the neuron will be re-tuned to respond differentially to the new range of light levels provided by the environment. True, it will fail to detect or differentiate among lights at levels below its current dynamic range, or above it. But it will hang in there for the observer, responding differentially to lights of different intensities in just exactly the range that is likely to be most important at any particular time.

Like dark glasses and gain controls, shifts of dynamic range are multiplicative models of light adaptation. Assuming that a constant response from the neuron is required for detection threshold, when the dynamic range shifts the neuron's detection threshold will shift by exactly the same factor (see the black dots in Figure 10.5B). And as in the dark glasses and gain control examples, if the shift of dynamic range is 1:1 with the illumination, Weber's Law will prevail. Smaller shifts of dynamic range will produce flatter light adaptation functions. For example, shifts of 0.5 log unit for every log unit change in illumination will produce the square root law; and intermediate shifts will produce intermediate slopes. [Sneak a peak at Figure 10.13.]

10.3.4 Adaptation as rescaling: Discounting the illumination

Finally, let's turn our arguments around and think of multiplicative processes as proceses of adaptational *rescaling*. In the dark glasses case, the dark glasses produced a complete rescaling of the input signal, restoring the illuminances in the retinal image to their original values at every point. The consequence was that changes in ambient illumination required no adjustments in retinal processing. In other words, a dark glasses process can be seen as a mechanism for *factoring out*, or *discounting the illumination*. In the dark glasses case, the discounting is literal and exact, since variations due to the ambient light level are factored out of the stimulus before the stimulus ever reaches the retina.

Similarly, shifts of dynamic range (or equivalently, the use of a gain control mechanism) can be viewed as a rescaling process, just applied slightly later in the processing sequence. In Figure 10.6B we illustrated a shift in a neuron's dynamic range by plotting its responses to test flashes for a series of adapting fields of different luminances. But think of it the other way around. If neurons early in the retinal circuit were to shift their dynamic ranges in a 1:1 fashion with the incoming light level, they could send a rescaled input to the next neuron down the line, completely factoring out the effects of changes in the ambient illumination. This perspective is shown in Figure 10.6C, in which the abscissa shows the rescaled input to the next neuron. If the neurons at the next level always get the same range of inputs regardless of the environmental illumination, then these neurons (and the whole rest of the visual system) need make no adjustments to deal with changes in ambient illumination. What a simplification!

This point also makes a difference when we try to define the *locus* of a light adaptation process. If we were to record from retinal ganglion cells, we might well see shifts of dynamic range, and we might be tempted to give the ganglion cells the credit for light adaptation. Yet ganglion cells might be contributing nothing to the adaptational process, because all of the shifting of dynamic range might have been introduced by earlier stages of processing. In other words, just because a neuron shows a shift in dynamic range, it isn't necessarily part of the adaptation process. To deal with the locus question we need to seek out the *earliest* retinal neurons that exhibit the shifts of dynamic range, and inquire into the mechanisms that produce them.

Finally, if we buy into a dark glasses/gain control/multiplicative model, several questions immediately arise. Can we use psychophysics to reveal more hints about these multiplicative processing changes? Over how broad a region of the retina do they average the adaptation state? How fast do they work? And, in physiological studies, do the appropriate shifts of dynamic range occur within retinal neurons? Which retinal neurons – how early in retinal processing? Let's look at a couple

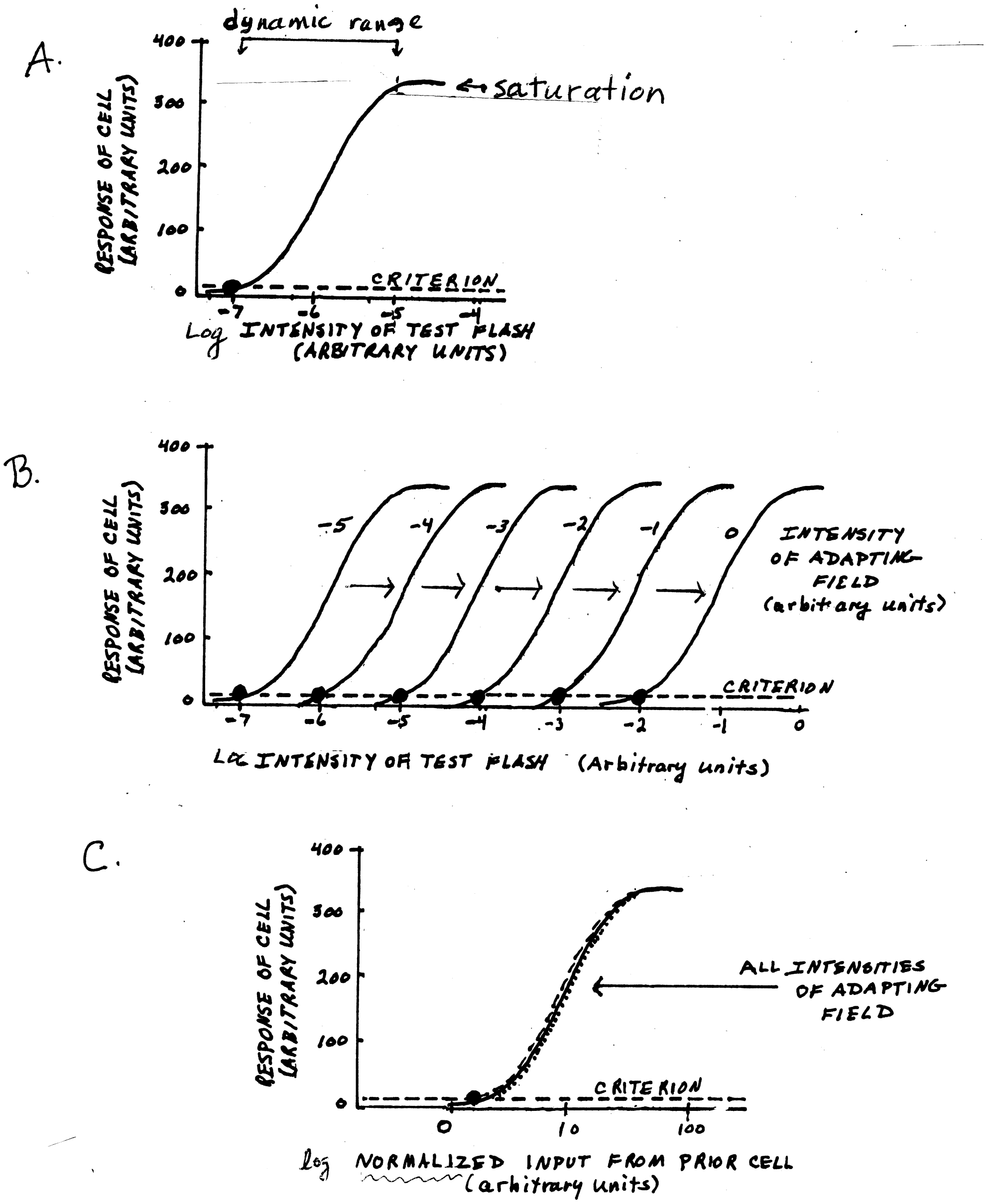


Figure 10.6: Shifts of dynamic range. A: Dynamic range. The dynamic range of a neuron is that range of inputs over which the neuron's output changes. Here, the dynamic range is about -7 to -5 in arbitrary (log) units. B: Shifts of dynamic range. As the intensity of an adapting field increases, the neuron's dynamic range is shown shifting along the abscissa. A one log unit increase in adapting field intensity is shown causing a 1 log unit shift of dynamic range. C: Rescaling. The neuron in B need not actually be changing any of its properties. If its inputs have been rescaled by earlier parts of the neural circuit, the same intensity of test flash, delivered on all intensities of the adapting field, can yield the same input to the neuron in question. In this case, even though the neuron exhibits a shift in dynamic range, it takes no part in the adaptation process.

more psychophysical experiments before we move on to retinal physiology.

10.4 More psychophysics: On the trail of the adaptation process

Prior to the availability of techniques for recording from single neurons within the retina, logic, psychophysics and psychophysically-based models provided several strong clues concerning the physiological locus of light and dark adaptation. These data and arguments are interesting examples of constraints that system properties place on models of the visual system.

10.4.1 What about the pupil?

As we saw in Chapter 4, the pupil of the eye changes size with ambient illumination, becoming larger in the dark and smaller in the light. Since the pupil acts in the direction of rescaling the amount of light reaching the retina across variations in environmental illumination, one is tempted to propose that it accounts for changes in sensitivity with light and dark adaptation. But in fact, the smallest pupil diameter is about 2 mm, and the largest about 8 mm. Pupil area thus varies by only a factor of about 16, or about 1.2 log units. Thus, changes in pupil size with light level still leave about 9 log units of variation in the retinal image for the rest of the visual system to deal with.

10.4.2 Is the adaptation process within the retina?

At the beginning of this chapter, we demonstrated the remarkable independence of light and dark adaptation processes with the two eyes, and registered the strong speculation that adaptation processes lie within the retinal circuit. In the 1930s, Kenneth Craik did an ingenius experiment that also supported the retinal locus of adaptation processes. He used a technique called *pressure blindness.* If a subject presses firmly on her eye from the side for about 15 seconds, vision in that eye ceases temporarily, probably because of an interruption of the blood supply to the retinal ganglion cells. (Don't try this experiment, as it can conceivably damage your eye.)

Craik pressure blinded one of his eyes, and then exposed that eye to a field of light of high intensity (which of course he could not see because of the pressure blindness). He then released the pressure, traced a dark adaptation curve, and found it to be the same as the dark adaptation curve traced without the use of presssure blindness. Craik argued that because the signal from the adapting light did not reach his brain, it could not have left its long-term effects there. Thus, he argued, the physiological mechanisms of light and dark adaptation must lie within the retina[6].

10.4.3 Is it within the individual photoreceptor?

A general paradigm designed to probe for the locus of light adaptation is to use a briefly flashed, spatially patterned adapting field (such as a set of stripes). One can then ask whether the effects of

[6] The young DT, fresh from graduate school, had an overwhelming insight one day – she thought up Craik's experiment. She raced into the lab, spent two months building the equipment and getting the experiment right, and risked her own eye with pressure blindness. She found that pressure blindness did not change the dark adaptation curve, wrote up the experiment and submitted a paper. The next day she went to the library and found that Craik had done the experiment before she was born! As a wise mentor is reputed to have said, it's amazing how a few months in the laboratory can save you an hour or two in the library.

the adapting pattern remain after the adapting field exposure is terminated. The argument is that if the spatial pattern of the adapting field is visible later, the adaptation process must be confined to small local retinal regions – perhaps even to the individual photoreceptors. But if the spatial pattern of adaptation makes no detectble difference, the adaptation process must be averaged, or combined, or spatially *pooled*, across the photoreceptors. Moreover, the largest spatial grain of adaptation that has no effect on our subsequent vision provides an estimate of the maximum size of the so-called *adaptation pool*.

The most elegant use of this approach, called the *difference frequency* paradigm, was developed in the 1980s by Donald MacLeod and his associates. This paradigm, which employs sinusoidal grating stimuli, is illustrated in Figure 10.7. First, as shown in the top panel, an adapting grating of a particular spatial frequency, F1, is flashed briefly, leaving a striped pattern of isomerized photopigment molecules across the ensemble of photoreceptors. Second, as shown in the middle panel, the subject views a test grating of a slightly different spatial frequency, F2. The question is, will the subject see a difference-frequency pattern – a spatial *beat* – with a spatial frequency F1 - F2, like that shown in the bottom panel of Figure 10.7? If the subject's answer is no, I see no pattern, then most or all of the adaptation process must be pooled across retinal regions at least as big as the pattern elements. But if the subject's answer is yes, I see bars at the beat frequency, then at least part of the adaptation process must be confined to local regions at least as small as the pattern elements.

Interestingly, the results of these experiments differed depending on whether the experiment was done at low light levels, in the domain of rod-mediated vision, or at higher light levels, in the domain of cone-mediated vision. At low light levels, MacLeod, Chen, and Crognale (1989) found no beat patterns for adapting gratings of frequencies higher than about 7 cy/deg (8-9 minutes of arc per cycle). That is, for rod-mediated vision, the adaptation process must be pooled across several minutes of arc – much too far to be confined to individual rods.

For cone-mediated vision, however, MacLeod, Williams and Makous (1992) found the opposite result. In this case, very high frequency adapting and test gratings were created by means of laser interferometry (Chapters 6 and 7). Beats were seen for adapting gratings of spatial frequencies as high as 130 cy/deg (more than 1 cycle per photoreceptor!). This fine grain of adaptation suggests that each individual cone contains a private adaptation mechanism. That is, at least part of adaptation process in cone-mediated vision must lie within the individual cone[7].

In short, these psychophysical experiments lead to some rather strong physiological predictions about the loci of adaptation processes. For the rod system, the prediction is that the adaptation process lies somewhere beyond the rods, in the retinal circuitry that pools and processes rod-initiated signals. But for the cone system, the prediction is that at least part of the adaptation process occurs within the individual cone. Later in the chapter we will see whether or not these predictions are borne out in single unit recording.

10.4.4 How fast does it work?

In Figure 10.3 we saw that dark adaptation involves a long, slow process. What about light adaptation? This question can be addressed by turning on an adapting field abruptly, and measuring thresholds for a small, brief superimposed test spot at each of a series of times after the onset of

[7] If there were a postreceptoral pathway devoted to each individual cone, the adaptation could take place in that pathway rather than withinin the individual cone (see Chapter 13xx).

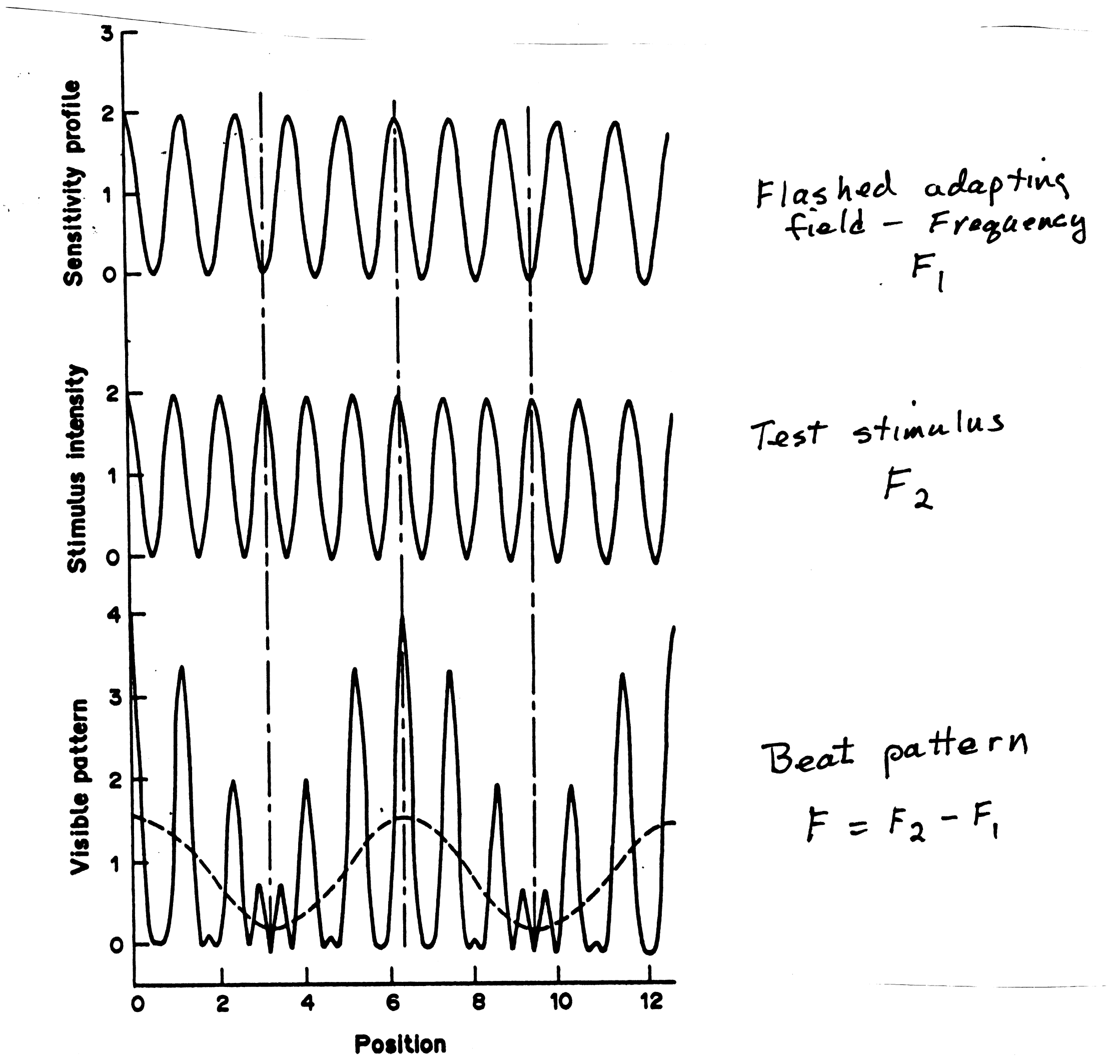


Figure 10.7: The difference frequency paradigm developed by Donald MacLeod and his associates. The subject's retina is exposed to a flashed adapting field (top panel) of a spatial frequency F_1, which leaves a spatially varying pattern of pigment isomerizations across his retina. He then views a grating of a slightly higher spatial frequency, F_2. If the signal left by the adapting flash is confined to the individual photoreceptors, the subject should see a pattern at the beat frequency F_2 - F_1. But if the adapting signal is pooled over many photoreceptors, an F_1 signal would no longer exist, and no beat pattern should be seen. MacLeod and his associates saw beat patterns in cone-mediated but not in rod-mediated vision. [Modified from MacLeod et al, 1989, Fig. 4, p. 970.]

the adapting field. The first experiment of this kind was carried out by Crawford in 1947. The results of Crawford's experiment are shown in Figure 10.8.

At the onset of the adapting field, the threshold for the test spot changes very rapidly indeed. It passes through a maximum when the test spot is approximately coincident with the onset of the adapting field, changes rapidly for the first 100 msec or so, and then more slowly. Similar fast and slow phases occur at adapting field offset. This and more recent experiments suggest that the major changes in retinal processing that underlie light adaptation take place very rapidly, but that slower changes also occur. There is also an early rapid component to dark adaptation, followed by the beginning of the long, slow dark adaptation curve we saw in Figure 10.3.

10.5 Physiology of light adaptation: Who gets the credit?

Finally, we come to the question: what is known about the loci of adaptation processes within the retina? Before launching our attack on this question, we need to state a couple of caveats. First, to date, most of the data come from photoreceptors and ganglion cells. Very little is known about the functioning of retinal interneurons in light and dark adaptation, especially in primates. And second, important (and confusing) species differences are known to exist. Although we will use data from turtles and cats as well as primates, we must take our species hopping with a grain of salt. Similarly, the credibility of comparisons between psychophysics and physiology will suffer until the full story is known for primate retina.

10.5.1 An overview: Rod- and cone-mediated ranges of vision

Let's think again about that 10^{10} dynamic range of primate vision, but this time from a quantal perspective. In Chapter 4 (Figure 4.xx) we introduced a specification system for the luminances of light coming from environmental surfaces, in units of candela/meter squared (cd/m^2). We also introduced the conecept of retinal illuminance, specified in Trolands (Td). We noted that units of luminance and retinal illuminance do not relate 1:1 to each other because as luminance increases the size of the pupil decreases, causing retinal illuminance to increase less rapidly than luminance over the range of intermediate light levels over which the area of the pupil changes most dramatically.

In the present chapter, we will base our discussion on units of retinal illuminance. Figure 10.9A shows retinal illuminance values in Td, now on a linearized scale, along with the environmental light level landmarks used in Chapter 4. The classical designations of scotopic (rod-mediated), mesopic (mediated by both rods and cones) and photopic (cone-mediated) light levels are also shown.

Now, from knowing the retinal illuminances and the sizes of the photoreceptors, vision scientists can estimate the numbers of quanta absorbed per rod or per cone, given the luminances of adapting lights. Estimates of rod and cone quantal catch rates are shown in Figure 10.9B.[8]

Figure 10.9B allows us to emphasize again the Design problem posed by the wide range of light levels presented to our eyes by our natural visual environments. At the lower extreme, near absolute

[8] Figure 10.9B incorporates some simplifications, mainly caused by suppressing the difference in spectral sensitivity between scotopic and photopic vision. However, notice the horizontal arrow drawn between the upper and lower scales (rod vs. cone quantum catches) in Figure 10.9B. It serves as a reminder that were we to vary the wavelength of adapting and test fields, these two scales would slide horizontally with respect to each other and with respect to the Td scale, and the slides could be large. [Figure out the relationship between this statement and Figure 10.2A.] The simplifications we have used are very small with respect to these factors.

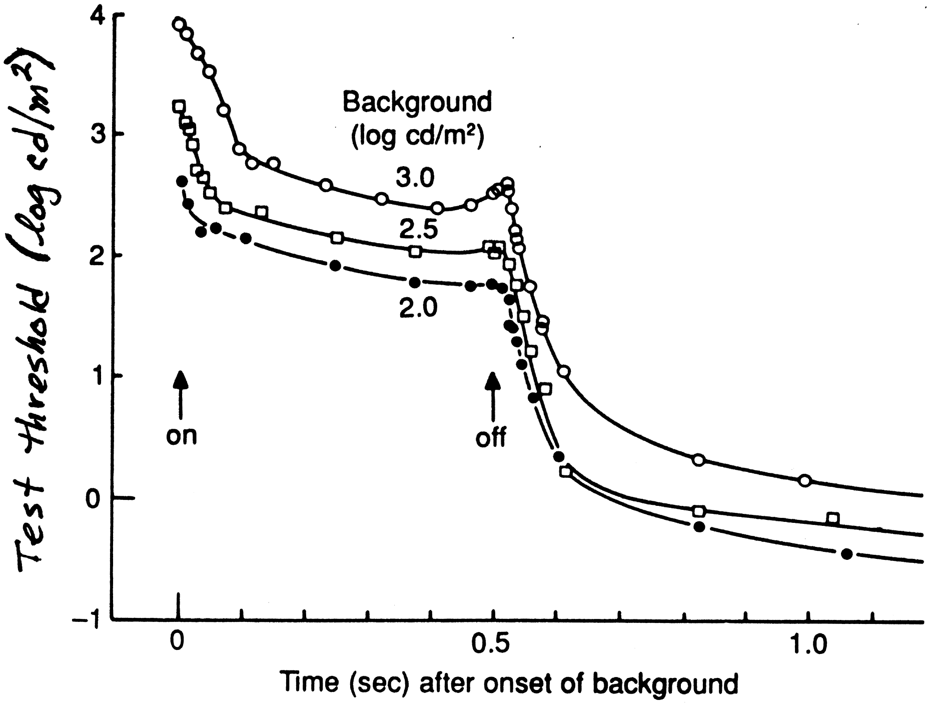


Figure 10.8: Crawford transients. Crawford's data describe changes in the threshold for a 30' test spot, superimposed upon a 12^o adapting field. The adapting field was turned on at time 0 and off about 1/2 second later. The large changes and abrupt threshold maxima at adapting field onset and offset are taken to indicate that a major part of the light and dark adaptation process takes place very quickly – within 100 msec or so of the change in adapting field intensity. The continuing, slower decreases in thresholds are taken to indicate the presence of additional, slower adaptational processes. [From Crawford, 1947, via Walraven et al, 1990, Fig. 14, p. 75.]

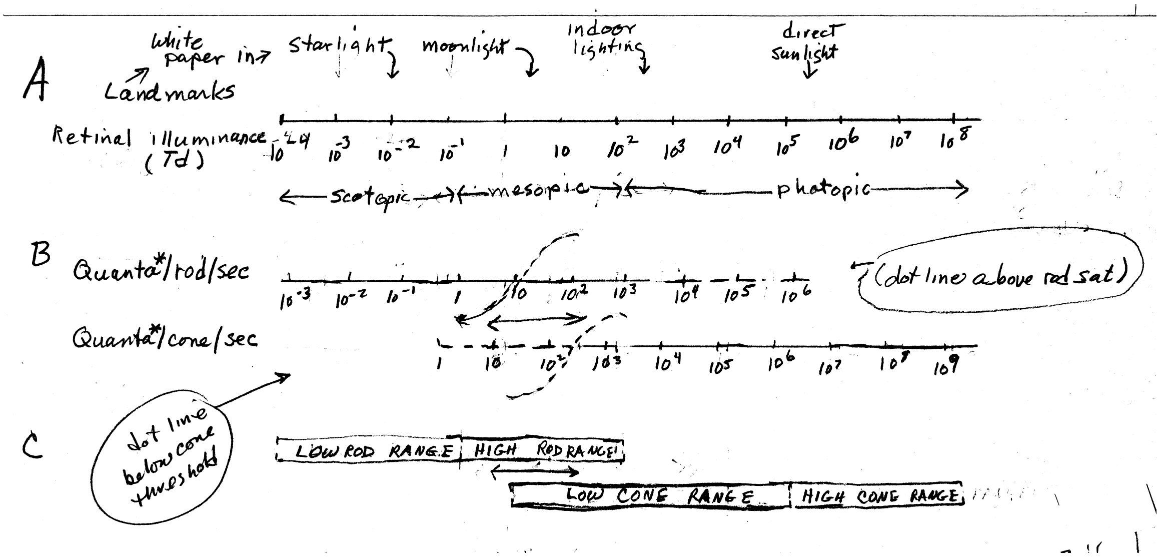


Figure 10.9: Quantum catches in rods and cones. A: The full dynamic range of human vision, in units of Td, on an even, logarithmically spaced scale. The landmarks above the Td scale are repeated from Figure 4.xx. The scotopic, mesopic, and photopic ranges are those usually given in introductory treatments. B: Conversion to approximate quantal catches for rods and for cones. The dotted curves show the approximate dynamic ranges of dark adapted rods and cones respectively. The arrow between the two scales serves as a reminder that these two scales would shift horizontally with variations in wavelength. C: The low rod, high rod, low cone, and high cone ranges, as discussed in the present chapter. We divide the total dynamic range into these four subparts because different mechanisms underlie the light adaptation process in the four different ranges.

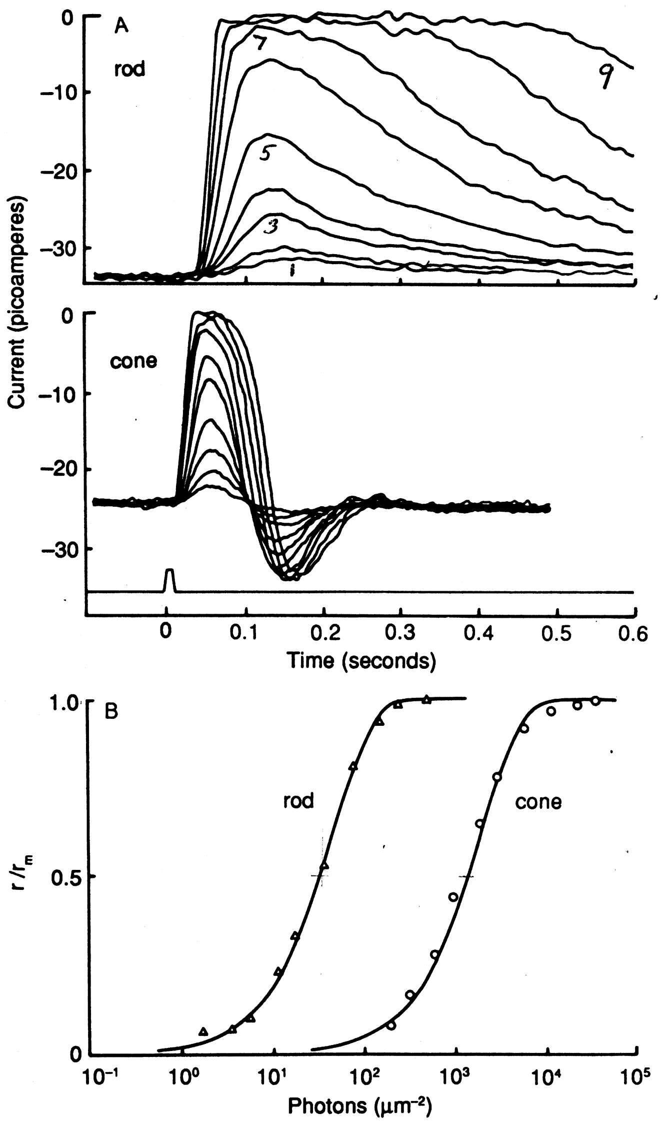
threshold, the problem is quantum scarcity: we make use of light levels so low that only about one rod in 10,000 catches a quantum. At the upper extreme, the problem is quantum overload: we continue to respond to small contrasts when each cone is bombarded with millions of quanta per second.

10.5.2 Responses of photoreceptors to flashes of light

To begin our coverage of the physiology of light and dark adaptation, we will look at suction electrode recordings from dark adapted primate photoreceptors. Figure 10.10 shows the responses of a dark adapted rod and a dark adapted cone. Each individual photoreceptor was exposed to flashes of light of varying intensity. The results are shown in Figure 10.10A. Both the rod and the cone show increasing peak responses with increasing intensity of the test flash, and both eventually saturate when the flash intensity gets too high.

Figure 10.10B shows the peak amplitudes of these responses. Both photoreceptors show dynamic ranges that cover about two log units. The rod's dynamic range covers the stimulus range from about 1xx to about 200xx quanta absorbed per flash, whereas the cone's dynamic range covers the stimulus range from about 200xx to 20,000xx.

Notice that the dynamic range of the dark adapted rod lies several log units above the absolute



Number of isomerizations for half-maximal response ↑ 30 (rod) ↑ 2000 (cone)

Figure 10.10: Dynamic ranges of rods and cones. A, top panel: Superimposed responses to flashes of light of increasing intensity, recorded from a monkey rod with a suction electrode. The average number of quanta absorbed per rod per flash increases by about a factor of two from each trace to the next. For traces 1-7, the higher the intensity the higher the peak membrane current. For traces 7-9 the peak response shows little if any additional increase; these traces reveal physiological saturation in the rod. A, bottom panel: Saturation in a monkey cone. B: Peak responses of a rod and a cone to flashes of light.

threshold. We will therefore divide the rod-mediated range into two parts: the*low rod range*, over which the average quantal catch per rod is less than one quantum per flash; and the *high rod range*, over which it is above one quantum per flash. Similarly, we will divide the cone-mediated range into two parts: the *low cone range*, which covers a broad range that includes the dynamic range of a dark-adapted cone; and the *high cone range*, in which individual cones absorb many quanta.

In the next few sections, we discuss what is known and/or speculated about the different adaptation processes that occur in each of these ranges. A Big Picture summary is presented at the end of the chapter (Figure 10.14). You may wish to locate the processes discussed for each of the four ranges in the Big Picture as we go along.

10.5.3 Rods and rod circuits in the OPL

We define the low rod range as the range of light adaptation over which each individual rod catches, on average, less than one quantum per test flash. Obviously, below this landmark, questions of the dynamic range of an individual rod are irrelevant. Yet a check back at Figure 10.1 shows that the detection threshold is elevated by a factor of at least 100 over this range. To explain these threshold elevations we will need to rely on some process that pools signals spatially across rods. This process must be in the rod circuit, and not in the individual rod. This is the range in which quantum scarcity and noise models are prominent.

The rod circuit in the OPL consists of the horizontal cells that receive rod inputs, and the rod bipolar cells. As we have seen, each horizontal cell contacts many rods. In consequence, the horizontal cell is a good candidate to provide the spatial pooling process demonstrated psychophysically. When very dim adapting fields are turned on, the horizontal cell gets sporadic inputs from the individual rods, but its summed input increases with the total number of quantal absorptions in the whole pool of rods that it serves.

A problem arises in that the horizontal cell and the rod bipolar cell receive input from many rods, with the consequence that their inputs varies enormously over the range of rod vision. Yet the dynamic ranges of these neurons, like those of other visual neurons, are limited to about two log units. What to do?

Rodieck (1998) speculates as follows: "Each (horizontal cell) contacts a pool of a few hundred rods.... If only a few rods are receiving photons, the rod signals pass to the rod bipolar cells unaltered. But as the photon catch rises, the (horizontal) cell becomes activated, and acts to attenuate the effect of each rod on the rod bipolar cell...." The exact mechanisms are unknown at the present time, but the bottom line is that horizontal cells are believed to modulate the input from the photoreceptors to the rod bipolars at the rod/rod bipolar synapse, and rod bipolars are believed to shift their dynamic ranges in concert with the level of environmental illumination.

What do the incoming signals look like from the perspective of the rod bipolars? There are a variety of options. At one extreme, if the horizontal cell were to attenuate the effect of each rod on the rod bipolar in accord with its summed signal, it could in principle rescale perfectly the inputs to the rod bipolar. If so, the rod bipolars need contribute nothing to the adaptation process, and we could attribute all of the adaptation process in the low rod range to the modulatory influence of the horizontal cells. At the other extreme, the bipolars could create part or all of their own shift of dynamic range, and be a major player in the adaptation process in the low rod range.

Next we discuss the high rod range: the range from one to a few hundred quantal absorptions per rod per test flash. We have already seen in Figure 10.10 that this range constitutes the dynamic

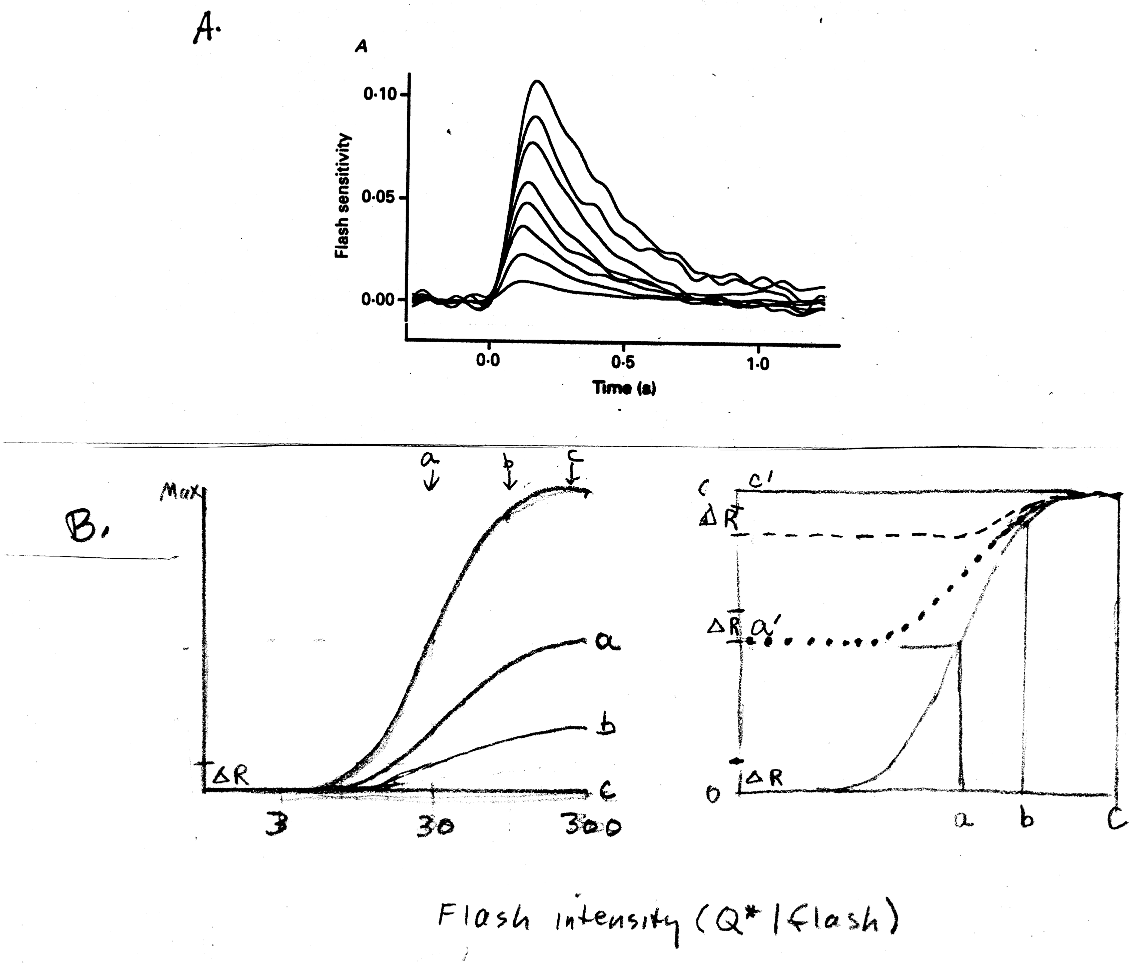


Figure 10.11: The effect of adapting fields on the responses of human rods. A: Response of the rod to a test flash. The top trace shows the dark adapted rod; lower traces show estimates of responses to the same test flash in the presence of adapting fields of increasing intensity. At the highest adapting intensity, the adapting field nearly saturates the cell, and the cell barely responds to the test flash. B: Mechanism of light adaptation in the rod. Adapting fields a, b, and c, of increasing intensity, diminish the responive range of the rod. The effects of test flashes are diminished and eventually eliminated. [xx Expand, depending on which versionis used.] [A. After Kraft, Schneeweis and Sshnapf, 1993, Fig. 9, p. 760; B: DT]

range of the dark adapted rod, and that the signals of dark adapted rods saturate at the top if this range.

Now, remember that our primary interest is in light adaptation – changes in processing with changes in the ambient illumination. From this perspective the next question is, how does the response of the rod change in the presence of *adapting fields* of various intensities? This question is addressed in Figure 10.11. Figure 10.11A shows a calculation of the responses from a human rod to a test flash of a fixed intensity, when the flash is presented in darkness or superimposed on steadily presented adapting fields of various intensities. As the intensity of the adapting field increases, the rod's response to the test flash decreases, nearly disappears for the highest adapting fields shown, and would disappear completely on even higher intensity adapting fields.

Why does this loss of signal occur? The answer is shown schematically in Figure 10.11B. The higher the adapting intensity, the closer the cell's response *to the adapting field* comes to saturating the rod all by itself. Thus, less and less of the rod's dynamic range is left to signal the presence of the test flash. When the adapting field itself saturates the rod, the rod can no longer provide any differential response to superimposed test flashes. Moreover, notice that the rod's dynamic range sems to be fixed. In short, the human rod seems to be a neuron that shifts its dynamic range little if at all with changes in adapting field intensity, and saturates if the adapting field takes the rod to the limit of its response range.

What implications does physiological rod saturation have for psychophysical rod saturation? First, the data at the bottom of Figure 10.2B show the physiological responses of rods, saturating over just about the same range of adapting field intensities over which psychophysical saturation occurs. When the stimulus conditions are chosen to isolate rod vision, and stimuli above physiological rod saturation are used, the rods cannot signal the presence of the test flash, information about the presence of the test flash is lost, and the subject cannot see it. Thus, the physiological data have the right properties to explain the psychophysical ones, and give us a good account of psychophysical rod saturation – slopes greater than 1 at the very top of the rod-mediated light adaptation curve.

Second, one can speculate that there is a Design-related reason that variations among rod-initiated signals dwindle and disappear at the top of the rod-mediated range. Perhaps the transition to cone-mediated vision is most readily accomplished if the rod-initiated signals max out and vanish, and let the cone-initiated signals dominate the retinal input.

In sum, the best bet on a causal story at present is to believe that light adaptation in the low rod range is mediated by shifts of dynamic range in the rod bipolars, caused by modulation of the incoming signal by the horizontal cells; and in the high rod range by saturation within the rod photoreceptors themselves.

10.5.4 Cones and cone circuits in the OPL

The upper half of the dynamic range of human vision is mediated by cones and cone-initiated signals. Like rods, cones have a limited dynamic range, and as shown in Figure 10.10, they eventually saturate to flashes of light, just as rods do. But unlike the case of rods, when steady adapting fields are used, cones escape the consequences of saturation by shifting their dynamic range. the result is that instead of gong silent, they are able to respond differentially to a new, higher range of test flashes. The causes of the shifts of dynamic range are different in the low vs. the high cone range.

No extensive recordings of adaptation processes in individual primate cones are yet available, so

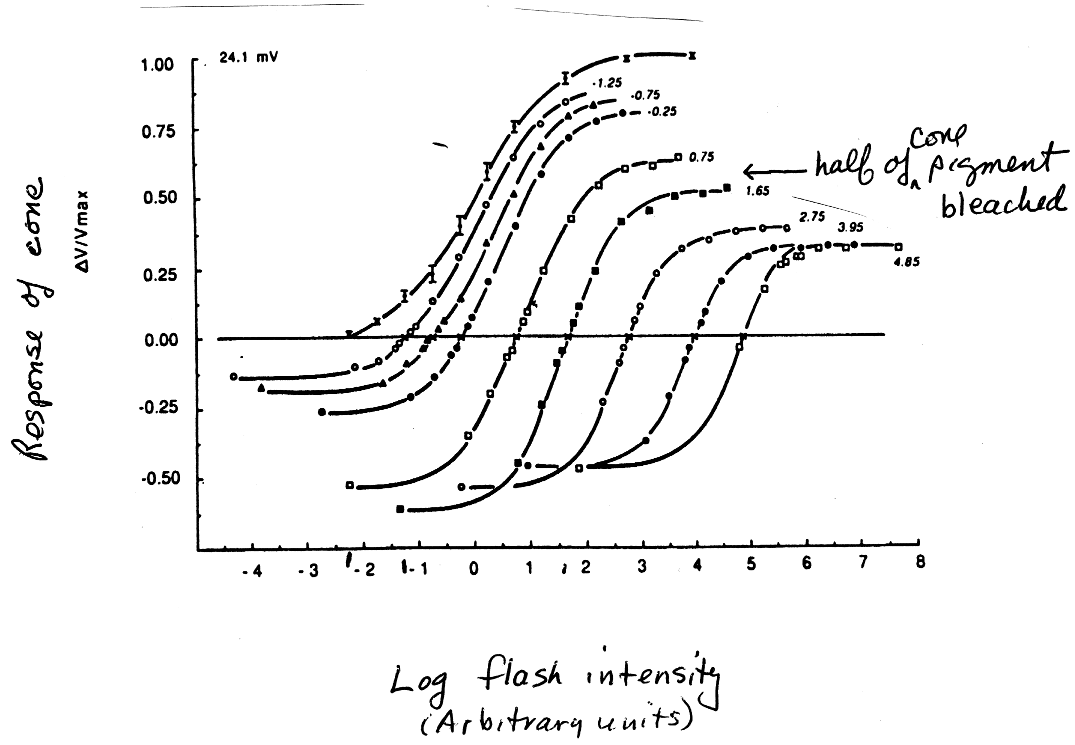


Figure 10.12: Shifts of dynamic range in an individual turtle cone. The abscissa shows the log test flash intensity in arbitrary (turtle-eye-based) units. The ordinate shows the response of the cone. The parameter on the curves is the intensity of the adapting field, again in arbitrary units. The leftmost curve shows the dark adapted cone; curves displaced to the right show responses in the presence of adapting fields of increasing intensity. Except for the upright triangles, the change from one adapting field intensity to the next was approximately 1 log unit, and the rightward shifts are also close to 1 log unit, giving 1:1 shifts of dynamic range. At an intensity of about 1.65, marked by the arrow, the adapting field depleted the pigment concentration to about 1/2 of its original value. Displacements to the right of this value can be attributed to pigment depletion.

we will need take the risk of changing species in midstream. In 1994 Dwight Burkhardt published an elegant study of the responses of individual turtle cones to flashes of light, presented against adapting fields of a series of intensities. Burkhardt used a laser to achieve very high levels of light adaptation. He also made careful measurements of the amount of bleached pigment in the retina as a whole at each level of light adaptation; and he tested the cones' responses to *decrements* as well as increments of light.

The data from Burkhardt's best behaved turtle cone are shown in Figure 10.12. The leftmost curve shows the cone's responses to test stimuli under full dark adaptation (of course, all of the stimuli must be increments in this case). The curves displaced to the right show the responses to increments and decrements on steady adapting fields of increasing intensities. The data are plotted such that the response of the cell to each steadily presented adapting field falls on the horizontal line. The points above the horizontal line are responses to increments, whereas those below the horizontal line are responses to decrements.

There is some vertical shifting from one curve to the next, but the striking effect is a set of curves of remarkably constant shape, displaced rightward with increasing intensities of the adapting field – that is, shifts of dynamic range within the individual cone. Moreover, from the second curve rightward, the spacing between the curves shifts by close to a log unit for each log unit increase in adapting field intensity – a 1:1 shift of dynamic range with adapting field intensity, which is the signature for Weber's Law.

As it turns out, the closed squares in Figure 10.12 show responses measured on an adapting background that isomerized about half of the cone photopigment (we call adapting fields above this level the high cone range, and it is discussed further below). But substantial shifts of dynamic range also occur among the leftmost six curves – the low cone range.

At least two mechanisms are thought to contribute to the shifts of dynamic range seen in the low cone range. First, it turns out that the effect of a quantal absorption on the synaptic output of the cone varies with the cone's level of hyperpolarization. The greater the hyperpolarization already brought about by earlier quantal catches, the less the effect of a new quantal absorption on the cone's output. This mechanism is tightly tied into the chemical cascade (Chapter 6), and its details are beyond the scope of this book. Second, there may also be negative feedback onto cones from cone-related horizontal cells.

In primate cones, we do not yet know the extent of shifts of dynamic range in the low cone range. Some experts believe that these shifts are quite small, whereas others believe they could be as large as the three log unit shifts shown in the turtle cone. This question remains open at the present time.

In the high cone range, a final light adaptation mechanism – cone pigment depletion – comes into play. At these high light levels, so many of the photopigment molecules hav ealready been isomerized by the adapting field that the rate of regeneration cannot keep up with the rate of isomerization. In consequence, the number of available cone photopigment molecules is depleted, and therefore so is the number of quanta caught from each test flash. In the turtle, the large, beautiful shifts of dynamic range among the final four curves at the right of Figure 10.12 are largely or entirely due to the photopigment depletion mechanism.

In primates, a pigment depletion mechanism limits the maximum quantal absorption rate to about 1,000,000 quanta per cone per sec, and accounts for light adaptation above about four log Td. Pigment depletion reduces the numbers of quantal absorptions in proportion to the intensity of a steady adapting field. [Could we call this a dark glasses mechanism?]

Now, what about the role of horizontal cells in the cone circuit? Here we come to another of the limits of current knowledge. As discussed in Chapter 9, cones make sign-conserving synapese onto the dendrites of horizontal cells. It seems likely that horizontal cells feed back onto cones in a sign-inverting manner, reducing the magnitude of the cone output signal, and in this way provide part of the adaptation mechanism in the low cone range.

Another major unknown is the degree to which these putative horizontal cell signals are pooled across space. On the one hand, MacLeod et al's (1992) psychophysical experiment suggests that at least part of cone light adaptation is specific to the individual cone. On the other hand, since each horizontal cell contacts many cones, perhaps it combines its cone inputs across space, and use its feedback to set at least part of the adaptation level of a whole group of cones together. Moreover, it's not known whether the effects of the horizontal cells in creating the center-surround antagonism seen in the cone bipolars should be treated separately from the effects of the horizontal cell's control of light and dark adaptation, or whether both effects should be modeled together.

What about the cone bipolars? In all probability cone bipolars shift their dynamic ranges throughout the whole cone-mediated range of light adaptation (as well as the rod- mediated range – see below). Over the high cone range, due to photopigment depletion, they undoubtedly receive a rescaled input, and thus their shifts of dynamic range are probably secondary to those of earlier neurons. But over the low cone range, in primates, the cone bipolars may well contribute to their own shifts of dynamic range, and thereby function as part of the light adaptation mechanism.

10.5.5 Contributions of the IPL

We complete our account with two comments about the role of the IPL. First, as we have seen, the IPL contains *dyads* – synaptic complexes that include reciprocal synapses between bipolar cell axons and amacrine cells. These reciprocal synapses provide a second potential feedback loop within the retina. Some authors have argued that the dyads in the IPL make major contributions to light adaptation.

And second, of course, the AII amacrine cell receives inputs from rod bipolars and feeds these inputs into cone bipolars. In consequence the cone bipolars provide combined rod and cone inputs to ganglion cells. At low light levels, the input to the ganglion cells would come from rod-initiated signals; at mid levels, from both, and at high levels, from cone-initiated signals.

But do the ganglion cells play an active role in adaptation? At one extreme, the cone bipolars might deliver a fully rescaled input to the ganglion cells, allowing them to function identically regardless of the light level. At the other extreme, intrinsic properties of the ganglion cells might help to control their own shifts of dynamic range. As was the case for bipolar cells, this question remains to be sorted out.

In any case, in 1969 Bert Sakmann and Otto Creutzfeld carried out a classic study of light adaptation in cat ON-center ganglion cells. These investigators varied the intensity of an adapting field that covered the whole receptive field of the ganglion cell. A series of six adaptation levels were used, all within the scotopic to mesopic range for the cat. At each adaptation level, they tested the cell's responses to a series of incremental test spots of different intensities, confined to the center of the receptive field. Their results are shown in Figure 10.13A. For an individual cell, the dynamic range shifted cleanly along the abscissa, maintaining the same shape and slope at each adapting level.

However, notice that in cat ganglion cells under these conditions, the dynamic range shifts by factors less than one log unit per log unit change in the adapting field luminance. That is, a 1 log unit increase in adapting field luminance led to a shift of dynamic range of only about 0.7 log units. Sakman and Creutzfeldt went on to define the detection threshold for the cell as the test spot luminance needed to produce a criterion response of five extra spikes per sec above the maintained discharge rate. The resulting light adaptation curves are shown in Figure 10.13B. The ganglion cells' light adaptation curve rose with a slope of about 0.7, in accord with the shifting dynamic range of the cell.

In sum, by the time we get to the ganglion cells, we find neural phenomena that make it easy to model psychophysical light adaptation curves, including sections with different slopes. It's a good bet that parallel, parametric studies of human or monkey light adaptation curves and monkey ganglion cell responses would reveal even closer resemblances, but this remains to be proved. And the more subtle question remains unanswered: how early in the retinal circuit do the critical light adaptation processes occur? If Burkhardt's turtle cone is any guide to primate cones, much of the answer could lie in the individual cone. But the truer answer is that we still don't know for sure – the search for a complete causal story for light adaptation is still a complicated work in progress.

10.5.6 A role for central processing after all?

Finally, let's look back at our original claim that light and dark adaptation porcesses occur mostly or entirely within the retina. In fact, an argument can be made that cortical processes are also involved. In 1983, Vaijo Virsu and Barry Lee studied light adaptation in cells in the lateral geniculate nucleus (LGN) of primates – the next level after the ganglion cells, as we will see (Chapter xx). They found that for small test fields there was good agreement between absolute thresholds measured behaviorally and physiologically. But for large test fields, covering more than the receptive fields of single LGN cells, the absolute thresholds of primate LGN cells are considerably higher than is the psychophysically defined absolute threshold. Therefore, they argued, a more central processing level must monitor inputs from many ganglion cells, and summate their signals to account for the very low absolute thresholds measured psychophysically for large test fields. [Notice the use of a bumblebees-can-fly argument.]

10.6 Changes in spatial processing

Space limitations preclude a discussion of the spatial aspects of light and dark adaptation – the shift in the peak of the CSF toward higher values with increasing luminance levels, and the increase of acuity with light level. Theorists who wish to account for these phenomena at the retinal level usually point to certain decreases in receptive field size known to occur with light adaptation. Other theorists would probably wish to postpone the account of these effects to higher levels, at which spatial frequency is more explicitly coded than it is in the retina (see Chapter 16).

10.7 Summary: Many processes work together

The classic phenomena of light adaptation document our capacity to detect immensely dim lights, and to perceive small percentage changes in light levels over a range of illumination of 10^{10}. At the physiological level, many factors work together to allow our vision to function across the range of environmental light levels with which we are challenged. These factors are summarized in the

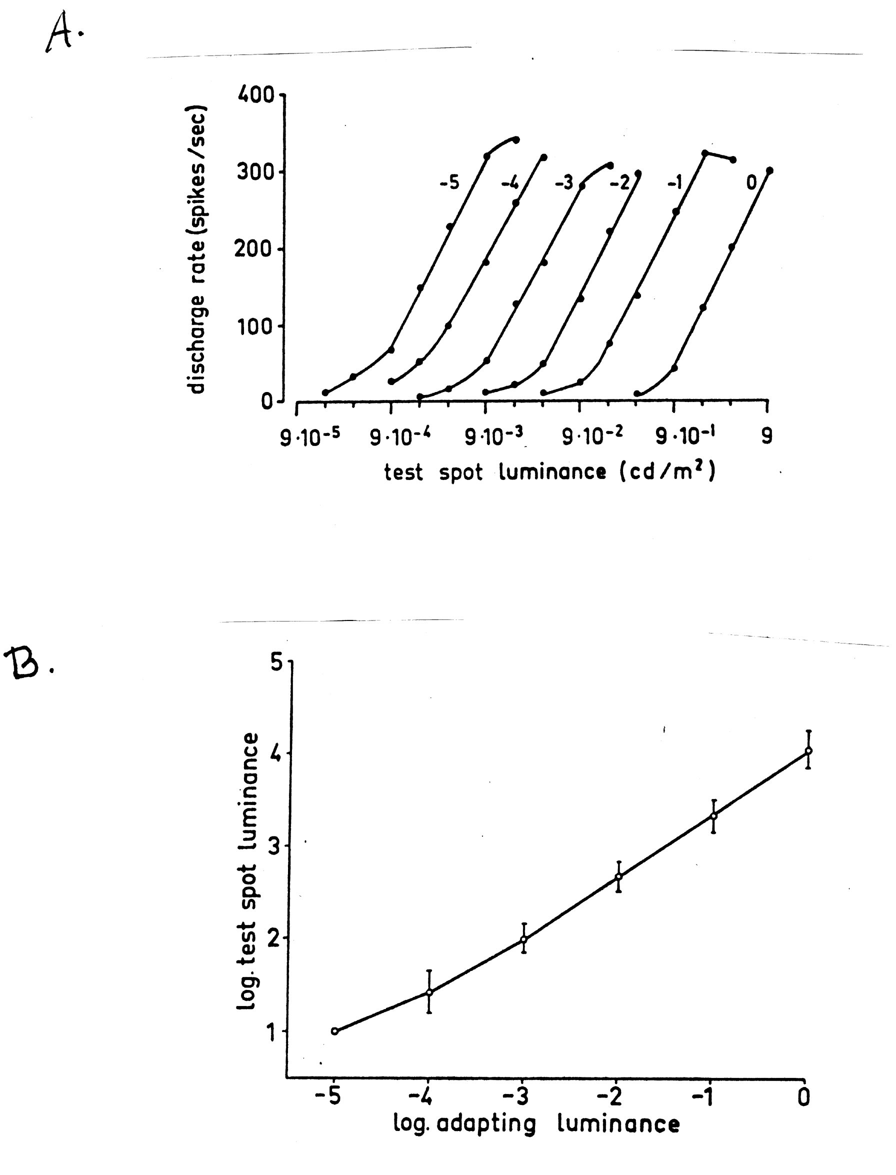


Figure 10.13: Shifts of dynamic range in cat ganglion cells. A: The responses of a single cat ganglion cell to a test spot, as a function of the luminance of the test spot superimposed on an adapting field. The parameter on the curve is the luminance of the adapting field in log cd/m^2. The cell shows a beautiful set of shifts of dynamic range, of about 0.7 log unit per 1 log unit of adapting field intensity. A cell like this one would generate a light adaptation curve like the one in B. B: The test spot intensity needed to elicit a criterion response is plotted as a function of the luminance of an adapting field. The light adaptation curve for the ganglion cell rises with a slope of 0.7, between the square root law and Weber's Law. [From Sakmann and Creutzfeldt, 1969. A, Fig. 8, p. 179; B, Fig 6, P. 177.]

Big Picture in Figure 10.14. The low rod range is controlled largely by summation in horizontal cells and shifts in dynamic range in rod bipolars. The high rod range is controlled largely by saturation mechanisms in the individual rods, and possibly small shifts in rod dynamic range. The low cone range and its extent are the least well understood aspect of light adaptation in primates, but elements of the chemical cascade, and negative feedback from horizontal cells onto cones, are probably both involved. And the high cone range can be attributed with considerable certainty to photopigment depletion.

Finally, once a shift of dynamic range is introduced in neurons at a given level, it should be observable as a secondary shift of dynamic range at all later levels of visual processing. The converse of this statement is that a shift in dynamic range is not a sufficient for claiming that a given cell participates in the light adaptation process. To get credit for participating in light adaptation, a neuron must instigate the adaptive rescaling process, not just passively pass it on.

In summary, light and dark adaptation are classic psychophysical phenomena whose properties have been well described since the 1930s. The basic separation into rod-mediated and cone-mediated branches was deduced on the basis of psychophysical experiments with variations of wavelength and eccentricity. However, the striking thing about light adaptation and its causal story is its complexity. In this book, light and dark adaptation are used to examplify the potential complexity of the answers to simple questions. And by the way, where *do* the stars go in the daytime?

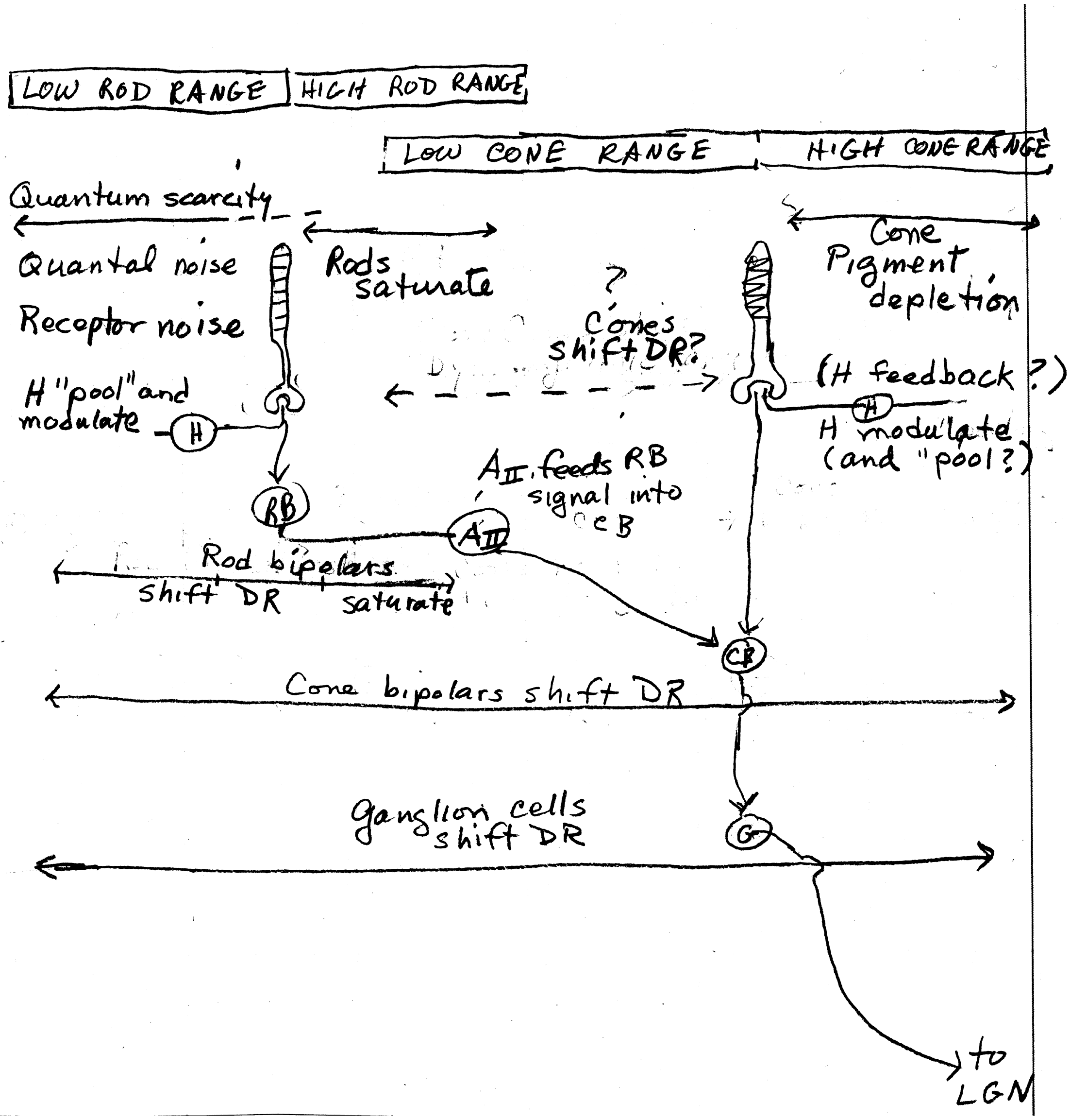


Figure 10.14: Mechanisms of light adaptation: The Big Picture. A speculative summary of the retinal processes that make up the overall adaptation process. For primates, the biggest unknowns are a) the existence and form of feedback from horizontal cells onto cones; b) the extent to which cones shift their dynamic range within in the low cone range; and c) the combination rule for rod- and cone-initiated signals in the AII amacrine cells.

Chapter 11

Retinal Processing and Perception

In the early chapters of this book, we explored the optical and transduction stages of the visual system. We also addressed some Type 3 – causal – questions by asking how the properties of these early elements of the visual system act to shape our visual perception. For example, we ascribed the high-frequency fall-off of the CSF to the optical quality of the eye (Chapter 5), the"zebra stripes" seen with laser interferometry to aliasing at the photoreceptors (Chapter 5), the scotopic spectral sensitivity curve and scotopic equivalence classes to transduction by rhodopsin (Chapter 6), and trichromacy to transduction by three and only three cone types (Chapter 7). All of these examples are cases of elegant and well established causal stories.

In Chapters 8-9, we moved beyond the photoreceptors. We confined ourselves to describing the anatomy and physiology of the retina, and some of the important neural recodings that take place within it. In Chapter 10 we explored both the psychophysics and the physiology of light and dark adaptation, developing tentative causal stories concerning how the various elements of the retinal circuitry may control changes of visual sensitivity.

In the present chapter, we build further on Chapters 8-10, and examine several more causal stories involving the retinal code at the level of the ganglion cells. Our goal is to examine the question, how do retinal recodings leave their marks on our neural signals and our visual perceptions? We develop three causal stories relating to detection thresholds as they are influenced by three properties of retinal neurons: center/surround antagonism, the convergence of rod-initiated and cone-initiated signals, and ON vs. OFF pathways. We then examine the ensemble properties of retinal ganglion cells, and work out how an edge is likely to be coded in the activity of such ensembles. Finally, we develop a fourth causal story, concerning the Class B phenomenon of Mach bands.

As you read this chapter you will notice that the causal stories are usually less elegant than the ones in earlier chapters. At this stage in history, causal stories relating the properties of ganglion cells to the properties of perception are usually more complex, less logically tight and more speculative – they often involve a larger number of more questionable assumptions. At the same time, it is important to take these causal stories seriously and analyse them carefully, because the creation and improvement of causal stories is a major goal of visual science. Perhaps these causal stories are just at an earlier stage of development, and only need more work to bring them to maturity. Or perhaps causal stories based on neural elements beyond the photoreceptors will always require more complex logical structures than did those based on optics and transduction.

The subject matter of this chapter also lends itself to illustrating some important observations about how visual science works. We begin the chapter by examining three metathemes: the role of analogies in the origins of theories; the use of selective lesions – destruction of selected parts of the neural tissue – in conjunction with animal psychophysics; and the use of custom-designed stimuli.

11.1 Three metathemes

11.1.1 Analogies and co-evolution

The first metatheme concerns the use of *analogies* in scientific argument. An analogy can be defined as a partial similarity between like features of two things. As we use the term here, an analogy is a partial similarity between a phenomenon defined at the psychophysical or perceptual level and a phenomenon defined at the anatomical or physiological level – in this case, the properties of ganglion cell activity.

Analogies play a major role in visual science, because a similarity between phenomena at perceptual and physiological levels usually hints at a causal story. For example, in Chapter 4, the similarity of curve shape between the optical MTF and the high-frequency fall-off of the psychophysical CSF suggested that the former *causes* the latter. And in Chapter 5, the similarity of curve shape between the scotopic spectral sensitivity curve and the absorption spectrum of rhodopsin led to similar insights.

Obviously, however, a causal story that arises from an analogy is incomplete, and the implied causal story should not be accepted without further justification. When we dig into such an analogy, it usually turns out to contain interesting implicit assumptions. As we pointed out in Chapter 4, the hidden assumption in the MTF-CSF causal story is that the psychophysical threshold occurs at a fixed physical contrast in the retinal image; and a very general hidden assumption in the rhodopsin story is that nothing mucks it up – no other photoreceptor type contributes to the detection of light at absolute threshold.

We would argue that a big part of the art of being a good vision scientist involves acquiring taste in the evaluation and amplification of analogies. To acquire such taste, whenever an analogy poses as a causal story, each of us needs to be in the habit of asking: What are the hidden assumptions? Are they reasonable? How compelling is the analogy, and how could it be tested further?

A related metatheme is the concept of *co-evolution* between scientific disciplines. In early vision science, we knew much more about psychophysical phenomena (like trichromacy) than we did about the details of anatomy and physiology. In consequence, theory and speculation tended to flow from psychophysics to anatomy/physiology. But since the 1950s, we have had increasing knowledge of both kinds available. Consequently, arguments flow in both directions: from psychophysics to substrate and from substrate to psychophysics. Through the use of analogies – that is, not necessarily through rigorous argument – we *exchange hints* across the subdisciplines, and each subdiscipline evolves in parallel to the other.

In the early part of this chapter we will emphasize three cases in which the discovery of a new property of ganglion cells suggested a possible property of perception. Such hints have led to the discovery of new perceptual phenomena. These examples thus illustrate the co-evolution of scientific disciplines, via the use of analogies and the exchange of hints across disciplines.

11.1.2 Lesions and behavioral studies in the same animal

Second, in this chapter we review some experiments that make use of an important new paradigm: the combined use of psychophysics and physiological manipulations in the same animal. That is, it is possible to train a monkey subject to carry out psychophysical tasks, and establish the presence of some visual function – call it V. Subsequently, one can make a localized lesion intended to deactivate a particular class of neurons – call them the N neurons – and then retest the monkey on the same psychophysical task. Moreover, if the lesion is reversible, the animal can be retested again after the function of the N neurons has returned.

As with all other kinds of studies, the results of lesion studies need to be interpreted with caution. But the intended logic is that if a particular visual function V is present when the N neurons are functional, and impaired or absent when the N neurons are selectively deactivated, then normal N neurons are a *necessary* condition for having V – they form a link of the causal chain that enables V to occur. Moreover, if a different visual function V' remains normal, then the N neurons are *not* a necessary condition for V'. In a sense, the remaining neural types N' are *sufficient* for V', but of course only in the context of normal functioning of the remaining parts of the causal chain.

11.1.3 Custom-designed stimuli

The third metatheme is that of *custom-designed stimuli* (sometimes called *designer stimuli*, like designer clothing). That is, vision scientists now know a great deal about the properties of many individual elements of the visual system – for example, the spectral sensitivities of photoreceptors and the center-surround nature of ganglion cell receptive fields. When we know this much about the stimuli to which individual neural elements will respond, and about their equivalence classes, we can custom-design stimuli with the goal of *maximizing the responses of some types of neurons while minimizing or silencing the responses of others.*

For example, when we know the properties of the area-response curves of ganglion cells, we can design stimuli to just fill the centers of their receptive fields, and thus maximize their firing rates, or fill their whole receptive fields, and thus minimize their firing rates. Or when we know about ON-center and OFF-center cells, we can attempt to limit detection to the former by presenting increments of light, and to the latter by presenting decrements.

And as a second kind of example, we could custom-design a spatial pattern to silence the rods, so that a subject's detection of the spatial pattern must depend only on input from cones. There are two ways we can do this. One way is to make the pattern of a high enough intensity that even the darkest parts of the pattern deliver more than 500 quanta per second to the rods, thus saturating all of the rods and rendering the rod ensemble incapable of signalling variations of intensity (Chapter 10). Alternatively, and more cleverly, we could make the pattern out of stimuli of different wavelengths, and vary their intensities until all parts of the pattern are in an equivalence class for rods (Chapter 2 and 6). In both cases, if the subject can see the pattern, he must be seeing it with cone-initiated signals, and we can go on to explore the properties of vision when only cone-initiated signals are available. [Think about the reverse case. Can you think of two ways to custom-design stimuli such that they would silence the cones and be detected only by the rods?]

In fact, you may notice a general similarity between the use of lesions and the use of custom designed stimuli. In the first case, a class of neurons is rendered non-functional by means of a lesion. In the second, a class of neurons is rendered non-functional by custom designing stimuli to fall within a equivalence class for those neurons. The difference is that lesion studies require invasive techniques, and can only be used in animals, whereas custom-designed stimuli can be used readily in human subjects.

11.2 Retinal processing and detection thresholds

We now consider in detail three cases of candidate causal stories. All three of them arise from analogic reasoning from the properties of ganglion cells to potential psychophysical phenomena. All three use custom-designed stimuli in attempts to isolate particular kinds of retinal neurons, and all three attribute control of the resulting detection thresholds to particular kinds of retinal neurons.

11.2.1 Case 1: Center-surround antagonism and the Westheimer phenomenon

As discussed in Chapter 8 and 9, the receptive fields of retinal ganglion cells exhibit spatial (center-surround) antagonism, and a ganglion cell with spatial antagonism shows an inverted U-shaped area-response curve (Figure 8.8). In 1965, Gerald Westheimer picked up a hint from these physiologically defined U-shaped curves, and perceived an analogy that led him to perform some psychophysical experiments.

As background, we begin by asking: In the psychophysical domain, what would happen if we varied the *size* of a test spot, and measured the detection threshold for the test spot as a function of its size? The result of such a *spatial summation* experiment is shown schematically in Figure 11.1. As the size of the test spot increases, its threshold decreases monotonically, until it levels out at some constant level. Why? Probably because as the test spot gets bigger, it falls into the receptive fields of more and more ganglion cells. Perhaps the odds that a highly sensitive ganglion cell will be encountered increase with test spot size; or perhaps some later stage of the system summates the signals from all of the ganglion cells affected by the test spot. The asymptote would represent the limits of the presumed physiological summation. In any case, the classical spatial summation paradigm does not reveal anything that impresses us as being analogous to center/surround antagonism.

Westheimer (1965) used a different approach – he custom-designed the stimuli with the goal of revealing center/surround antagonism psychophysically. To do this, he began with a tiny test spot, only 6' in diameter. The reasoning is that perhaps a tiny test spot will be detected by only one (or at most a few) ganglion cells – the one (or few) on which the test spot is centered – instead of a constantly varying number as was the case in the spatial summation paradigm. Westheimer then added a second stimulus component – a background field of variable diameter – with the idea that this field would yield systematic variations in the excitation level of the ganglion cell. He then varied the size of the background field, and measured the detection threshold for the tiny test spot for each different background field size[1].

[1] In Westheimer's 1965 experiment, the test spot was made from 500 nm light, and the background from a broad band of wavelengths above 630 nm, in order to insure that the subject detected the test spot only via rod-initiated signals. (The use of wavelength variations in custom-designing stimuli will be developed further in Case 2 below.) In later experiments he also used conditions designed to insure detection by cone-initiated signals, and found similar U-shaped functions. (A second, larger background field was also included in these experiments to reduce the effects of stray light. This field is ignored here for simplicity.)

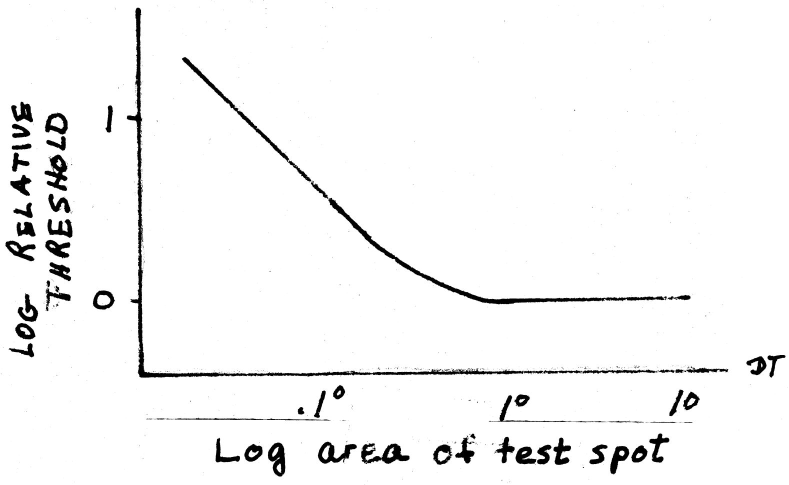


Figure 11.1: A classical spatial summation curve. The size of a test spot is varied, and its threshold is measured as a function of test spot size. As the size of the test spot increases, the intensity of light (per unit area of the stimulus) required for detection decreases. Such a classical spatial summation curve yields no hint of processing by neurons with center/surround antagonism.

Typical results from Westheimers (1965) experiment are shown in the right panel of Figure 11.2A. As the size of the background field increases, the threshold for the tiny test spot first rises, indicating a loss of sensitivity – no surprise here, given what we know about the pooling of adaptational processes across the retina (Chapter 10). But as the background field passes a certain critical diameter (about 40' in this example), the *threshold for the test spot begins to fall again*, indicating an *increase* of sensitivity. The curve finally levels out at a background field size of about 3^o. The surprising result is that light at a relatively large distance from the test spot – in the annular region from 40' to 3^o – can increase the subject's sensitivity to it. The *spatial sensitization effect* (or *Westheimer effect*) can be large – the threshold can be reduced by nearly a log unit under optimal conditions. Moreover, the effect has slightly different dimensions under rod-isolating vs. cone-isolating conditions, and scales up in size as one moves from foveal to peripheral retina.

Now, how do we model the Westheimer effect? An intuitively appealing model is shown in Figure 11.2. We start by noticing the analogy – the similarity – between two different inverted U-shaped curves: the area-response curve of prototypical ON-center retinal ganglion cell (Figure 11.2A) and the psychophysically derived Westheimer function (Figure 11.2B). The similarity between the two curves encourages us to think further about a causal story.

The implicit analogy goes like this. Suppose that the stimuli used in Westheimer's experiment were centered on the receptive field of a prototypical ganglion cell. As the background field increased in size, it would first cover more and more of the receptive field center of the cell, elevating its firing rate. Perhaps the threshold amount of light needed for the cell to detect the tiny test spot would increase as well. But as the size of the background field exceeds the size of the receptive field center, and begins to invade the receptive field surround, the cell begins to return toward its maintained level of firing; perhaps its threshold for the tiny test spot would also decrease again. The analogy further suggests that the psychophysical detection threshold flattens out at large background field sizes because eventually the edge of the background field reaches the edge of the receptive field surround, and adding more light outside the receptive field no longer influences the activity level of the cell. So far, so good.

The next step is to notice, however, that the analogy is not perfect. Close examination of Figures 11.2A and 11.2B reveals that the physiological experiment is not exactly the same as the psychophysical one, and the quantities plotted on the two abscissae, and on the two ordinates, differ. In the physiological experiment, the independent variable is the size of a disk of light, and the dependent variable is the firing rate of a ganglion cell in response to disks of various sizes. But in the psychophysical experiment, the independent variable is the size of a background disk, and the dependent variable is the detection threshold for a different stimulus component – the tiny test spot. To move this analogy toward being a more complete causal story, we need to establish some believable connections between the variables in the two experiments.

There are two major assumptions involved in this analogy. First, we need to assume a connection between the ordinates of the two U-shaped graphs (Figure 11.2A and B). The connection would have to be the assumption that the cell's threshold for detecting the tiny test spot will vary with the activity level of the cell – the higher the cell's firing rate, the higher its threshold for the test spot is assumed to be. This assumption is made explicit in Figure 11.2C. Among other things, a criterion for the cell's threshold would also have to be made explicit – perhaps the intensity of the test spot required to produce a small, fixed number of extra spikes in the cell, above the maintained spike rate.

And second (and easy to overlook), we also need to make a connection between the physiological

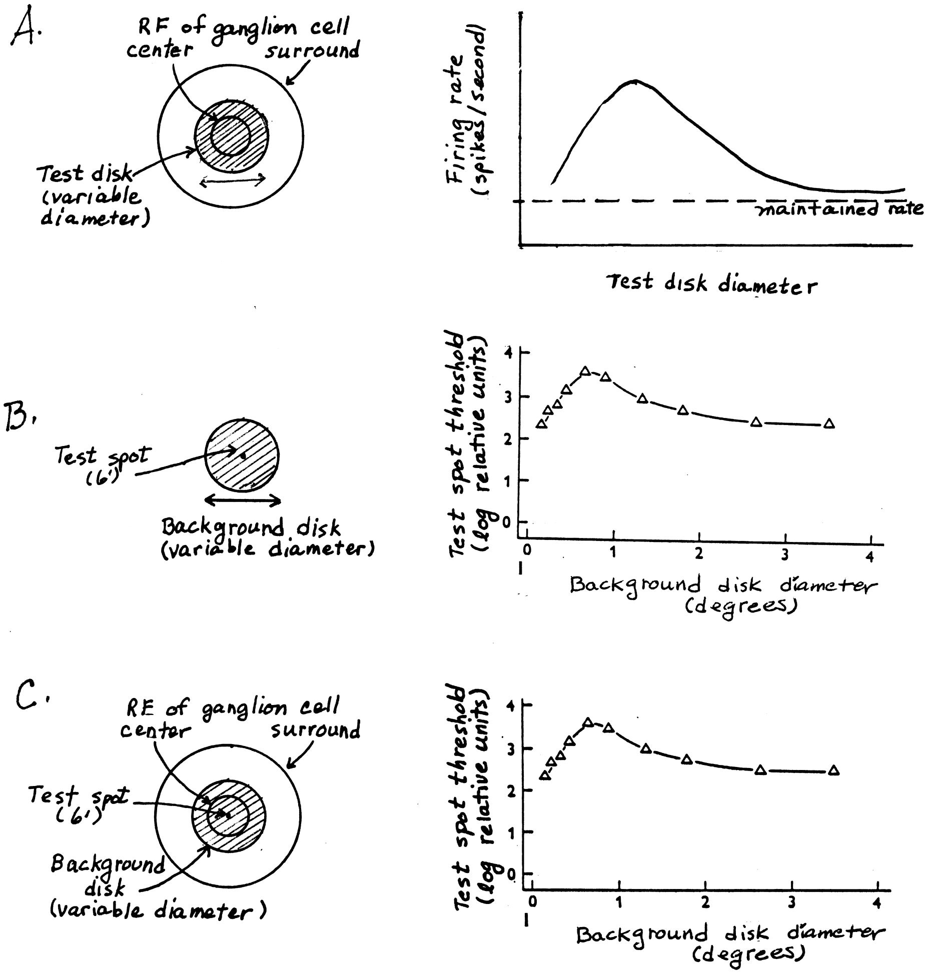


Figure 11.2: The Westheimer effect. A: An area/response curve for a prototypical ON-center ganglion cell with center/surround antagonism (cf. Figure 8.8). A disk of light is centered on the receptive field and varied in size (the test disk shown is just larger than the receptive field center). The neuron's firing rate varies with the size of the disk, showing a maximum when the disk just fills the center of the receptive field. B: The Westheimer effect. The psychophysical threshold is measured for a tiny test spot centered on a background disk, as a function of the size of the background disk. The threshold first rises and then falls again. C: A physiological prediction, based on the common explanation of the Westheimer effect. A tiny test spot is centered on the receptive field of an ON-center ganglion cell, and the size of a background disk is varied. The prediction is that the cell's threshold for the test spot will first rise, reaching a maximum when the disk just fills the center of the receptive field. It should then fall again, reaching an asymptotic value as the disk fills the whole surround of the receptive field. In B and C the zero on the ordinate is the person's, or the cell's absolute threshold. (B. modified from Westheimer, 1970, Fig. 2, p. 112.)

threshold of this cell for a tiny test spot and the psychophysical threshold of the whole human subject for the same tiny test spot. That connection is the assumption that detection of the test spot by the subject would have to be controlled by the response of the single retinal ganglion cell on whose receptive field the stimulus configuration is centered. In other words we must be assuming that this particular cell is the *weakest link in the most sensitive neural channel available to the subject for detection of the test spot, over the whole range of sizes of the background disk.* This assumption is another example of a "nothing mucks it up" proviso, because one is assuming that nothing else – no other cell anywhere in the system – interferes with the control of this ganglion cell over the subject's threshold.

To complete the cycle of co-evolution, notice that the curve in Figure 11.2C can actually be taken as a prediction of the outcome of a physiological experiment. It suggests that we go back and look again at area-response functions in ON-center ganglion cells. We could seek out the cells that are the most sensitive to tiny test spots, and see whether the thresholds of these cells for detecting tiny test spots vary in concordance with their area-response functions. In other words – and note this as a general experimental strategy – we could press the analogy by carrying out a physiological experiment using the exact stimulus configuration used in the psychophysical experiment.

In fact, some such experiments have been done, but the issue is not yet settled, particularly in primates. U-shaped functions have been seen in some individual neurons – both bipolars and ganglion cells – at some times in some species, but not at other times or in other species. And the question has not yet been studied systematically in primates. So the explanation of the Westheimer effect by means of center-surround antagonism still remains a speculation.

What about the question of locus? Can we say that the Westheimer phenomenon is an "early" phenomen, whose causes are located in the retina? According to our criteria for deciding locus questions, we should ask whether the phenomenon can be modelled on the basis of the known properties of ganglion cells, without the need for any complex central processing. We might well answer with a tentative yes, because the analogy between the two U-shaped functions seems quite compelling. However, again, before voting it would be nice to look systematically at physiological Westheimer functions in real retinal neurons, to make sure that the analogy doesn't break down when we actually measure thresholds for tiny test spots in retinal neurons[2].

In summary, the case of the Westheimer phenomenon is an instructive example of the use of custom-designed stimuli and the exchange of hints across disciplines in the derivation of a causal story.

11.2.2 Case 2: Convergence of rod-initiated and cone-initiated signals

Our second case examines a psychophysical experiment done by Walter Makous and Ronald Boothe in 1974. The story starts from the fact that inputs to the retinal circuits come in through the photoreceptors – the rods and cones. In the outer plexiform layer, rods and cones are served by separate bipolar cell types – the rod bipolars and cone bipolars. But as we noted in Chapter 8 and 9, there are several opportunities for the interaction of rod-initiated and cone-initiated signals within the retina. In particular, there are kissing junctions between rods and cones; horizontal cells that receive input from both rods and cones; and probably most importantly, AII amacrine cells that feed rod-initiated signals into cone bipolar cells.

[2] Walter Makous (1997) has recently proposed a very different theoretical approach to theWestheimer effect, and gives it a much more central locus. It remains to be seen which view prevails.

We have previously established that light and dark adaptation curves, like those shown in Chapter 10, are two-branched. The upper branch is attributed to detection via cone-initiated signals, and the lower to detection via rod-initiated signals. Historically the abrupt transition between the two branches has been interpreted to mean that rod-initiated and cone-initiated signals do not interact, and in particular do not influence each others detection thresholds. But this conclusion no longer seems right, given the anatomical possibilities. So the question becomes: how might we custom-design a psychophysical experiment to demonstrate the interaction of rod-initiated and cone-initiated signals?

Makous and Boothe (1974) designed such an experiment. The design of their experiment is shown in Figure 11.3. Figure 11.3A shows the stimuli. Makous and Boothe used an 8 diameter test spot superimposed on a 10^o adapting (background) field. The test spot was centered at 2^o in the peripheral retina, where both rods and cones are numerous. Figure 11.3B and 11.3C show reminders of two sets of facts you've learned before. Figure 11.3B a dark adaptation curve, with its rod (scotopic) and cone (photopic) branches; and Figure 11.3C shows the standard photopic and scotopic spectral sensitivity curves, with their different sensitivity maxima – the scotopic curve at about 500 nm, and the photopic curve at about 555 nm.

Makous and Boothe used two sets of tricks in custom-designing their stimuli. First, they custom designed the test flash to be detected only via rod-initiated signals. To do this, they set the wavelength of the test flash to 491 nm, near the maximum sensitivity of rods (Figure 11.3C). Moreover, in a preliminary experiment (shown as Figure 11.3B) they traced out a dark adaptation curve with the 491 nm test flash, and determined the intensity of the test flash at the cone plateau. In the main experiment they worked only with stimuli whose thresholds fell *below* the cone plateau, so that rods alone would be detecting the test stimulus.

Second, Makous and Boothe created two different adapting fields that were in an equivalence class for rods, but had very different effects on cones. As shown in Figure 11.3C, they started by choosing two adapting fields – 491 nm and broadband ”red” – that differ greatly in their relative effectiveness for rods vs. cones. They measured absolute thresholds for these two adapting fields, and equated the two fields in effectiveness for rods by setting their intensities at equal multiples above their respective absolute (rod) thresholds. The trick is that 491 nm stimuli and broadband ”red” stimuli equated for rods are not equated for cones. In fact when these two stimuli are equated in effectiveness for rods, the ”red” stimulus is about a factor of 10 more effective for cones than is the 491 nm stimulus.

Finally, in the main experiment, Makous and Boothe traced out two light adaptation curves. That is, they measured the subject's threshold for the 491 nm test spot as a function of the intensity of each of the two adapting fields in turn.

As shown schematically in Figure 11.4A, there are two possible outcomes for the experiment. If rod thresholds are elevated only by rod-initiated signals, the two light adaptation curves must be identical, as shown by the two superimposed curves located toward the right of the graph. But if cone-initiated signals also elevate rod thresholds, then the light adaptation curves will differ. Since the ”red” adapting field is about a factor of 10 more effective for the cones than is the 491 nm adapting field when the two are equated for rods, the light adaptation curve on the ”red” background should fall about a log unit to the left of the light adaptation curve on the 491 nm background, as shown by the ”red” curve displaced to the left.

Figure 11.4B shows the results obtained by Makous and Boothe. In fact, the two field adaptation curves do not coincide. The field adaptation curve for the ”red” background is shifted to the left

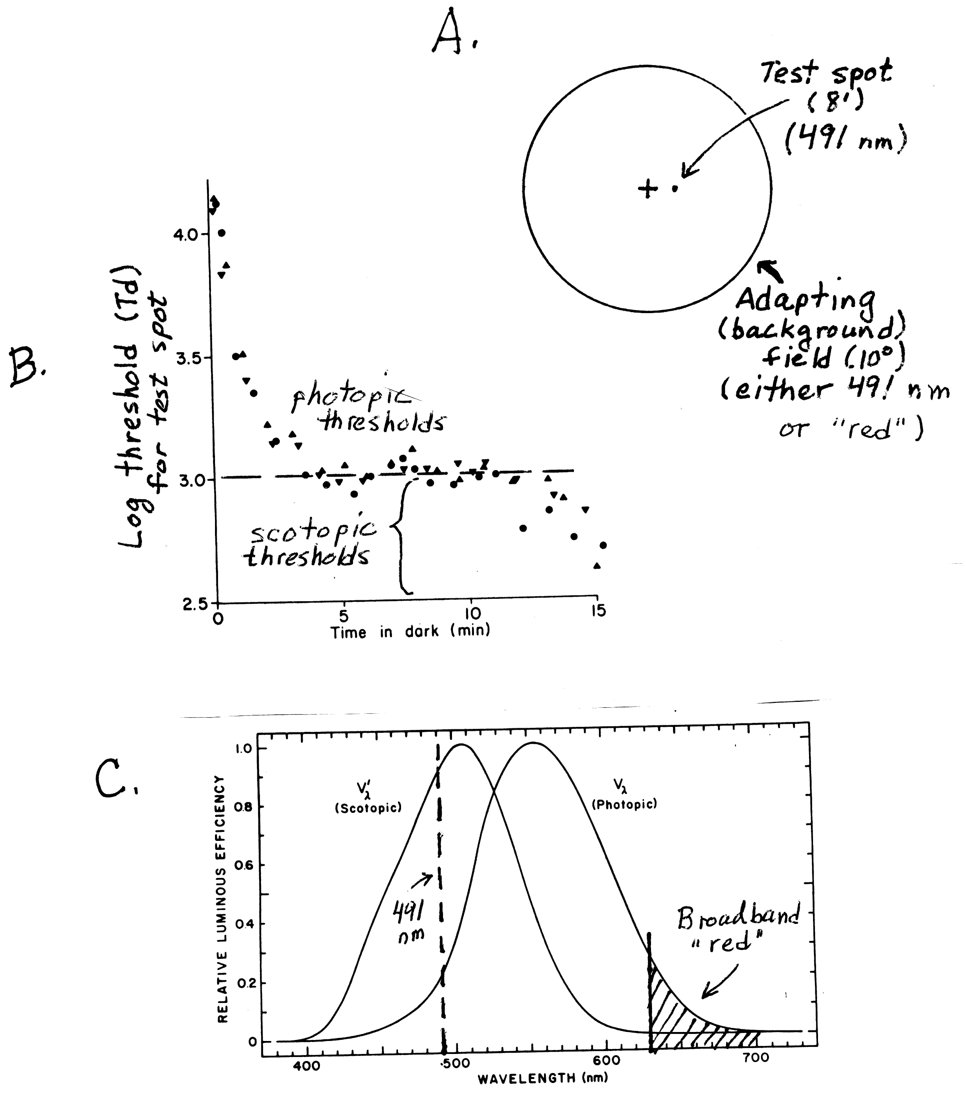


Figure 11.3: The Makous and Boothe experiment: experimental design. A: The stimulus configuration. Light adaptation curves were measured for a tiny 491 nm test spot on each of two different adapting (background) fields: 491 nm, and broadband "red". B: A dark adaptation curve measured with the 491 nm test spot. The cone plateau (marked by the dashed horizontal line) occurred at a test spot intensity of about 3.0 log Td. All thresholds below this value are presumed to be based on rod-initiated signals. C: Scotopic and photopic spectral sensitivity curves. The 491 nm test spot is designed to favor detection by rods. The 491 nm adapting field is much more effective for rods than for cones, whereas the broadband "red" adapting field is much more effective for cones than for rods. When the intensity of the "red" adapting field is increased to make the two adapting field equally effective for rods, the "red" adapting field is about a factor of 10 more effective for cones than for rods. (B modified from Makous and Boothe, 1974, Fig. 2, p. 288; C after Pokorny et al, 1979, Fig. 2.2, p. 28.)

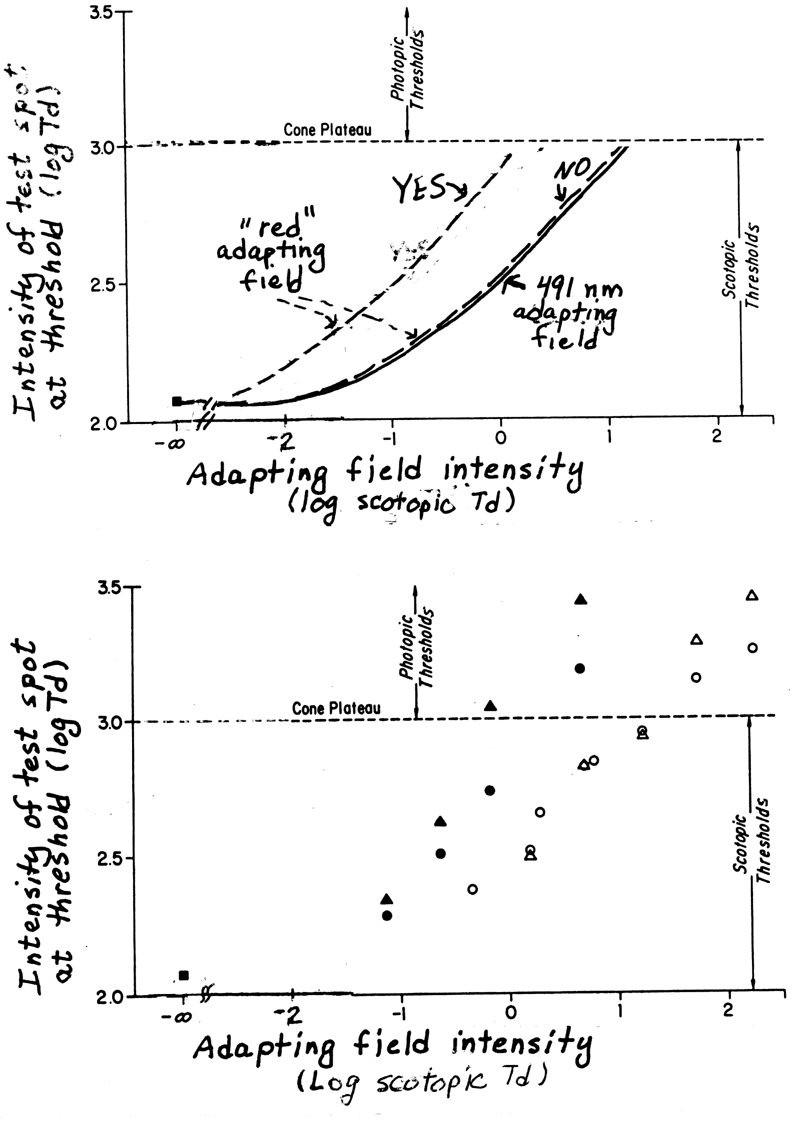


Figure 11.4: The Makous and Boothe experiment: Predictions and results. A: Two possible outcomes. The solid line shows a hypothetical light adaptation function for the 491 nm test spot, measured against a 491 nm adapting field. The dashed lines show two hypothetical light adaptation functions measured against the broadband "red" adapting field, predicted on the hypothesis that cone-initiated signals do (YES) or do not (NO) elevate the threshold for rod-initiated signals from the test spot. The abscissa shows the intensities of the two adapting fields, equated in effectiveness for rods (i.e. in log *scotopic* trolands). B: Results. Open symbols: thresholds on 491 nm background; closed symbols: thresholds on "red" backgrounds. The data on "red" backgrounds are displaced to the left, in accord with the prediction that cone-initiated signals block signals initiated by rods. (The difference between the circles and triangles is beyond the scope of this book, and can be ignored for our purposes.) (B modified from Makous and boothe, 1974, Fig. 5, p. 289.)

by about a factor of 10 – the same factor by which the red adapting field exceeded the 491 nm adapting field in effectiveness for cones. This difference cannot have been caused by signals initiated by rods, because the two adapting fields are in an equivalence class for rods. Therefore, Makous and Boothe concluded (and we agree) that under these conditions, rod-iniated and cone-initiated signals interact, and that cone-initiated signals from the adapting field elevate the threshold for rod-initiated signals from the test field.

Now, let's back away from specifics and ask, what is the most general constraint that this experiment places on models of the visual system, and what class of models does it reject? The most general constraint is that at least under these conditions, rod- and cone-initiated signals interact. Assuming only that neural signals cannot interact unless they pass through the same neuron, rod-initiated and cone-initiated signals must pass through the same neurons. In other words, we can reject all theories that predict that cone-initiated signals have no effect upon detection thresholds for rod-initiated signals, or that the two kinds of signals do not pass through the same neurons.

The question then becomes, which of the anatomical possibilities mediates the effect? The kissing junctions, the horizontal cells, the convergence of signals on cone bipolars, or something else? Makous and Boothe's experiment does not distinguish among these options[3]. Instead, it challenges vision scientists to figure out how to perform physiological experiments at all of these levels, in order to track down the locus of the particular interaction of rod- and cone-initiated signals that causes the elevation of detection thresholds. Such experiments are rapidly becoming feasible in primate retina, as we will see in Chapter 13.

The suggestion of new physiological experiments, of course, feeds into the circle of coevolution, just as it did for the Westheimer phenomenon. In this case, an anatomical fact – the apparent convergence of rod and cone pathways – suggested a psychophysical experiment; and the results of the psychophysical experiment suggested that a new set of physiological experiments would be worthwhile.

11.2.3 Case 3: ON-center vs. OFF-center cells and the detection of increments and decrements

A third and final major aspect of retinal recoding is the creation of two variants of cells with center-surround antagonism: ON-center vs. OFF-center cells. Why did the EDC give us these two mirror image processing circuits?

The most usual Design argument starts from the premise that increases in firing rate are a more reliable neural signal than are decreases in firing rate. That is, suppose that the maintained firing rate of a ganglion cell is low (say, 50 spikes per second), and variable in time; and suppose also that the cell's maximum firing rate is high (say, 500 spikes per second). Then *increases* in firing rate – decreases in interspike intervals – probably provide a readily detectable signal, and many different increases in firing rate are available to code different magnitudes of stimulus change. On the other hand, *decreases* in firing rate – increases in interspike intervals – form a less detectable signal, because long interspike intervals will often occur spontaneously in the maintained spike pattern; and because far fewer differentiable decreases in spike rate are available. A latency argument can also be made, to the effect that an increase in firing rate (a decrease in the interspike interval) can

[3] DT is led to wonder whether there are psychophysical experiments remaining to be custom-designed, that would sort among the anatomical options. Perhaps such experiments will become possible as we learn more about the detailed physiological properties of retinal interneurons.

be detected *faster* than a decrease in firing rate (an increase in the interspike interval). In sum, many vision scientists argue that increases in firing rate provide the better neural signal.

The Design argument continues by noting that important events in the environment can be represented by either increases or decreases in light level in the retinal image (for a decrement, think of a black cat). Moreover, rapid changes in the environment can cause either rapid increases or rapid decreases in light level (think of the shadow of an approaching predator). Therefore, the argument continues, it is functionally important to signal both increases and decreases of the physical intensity of light with increases in firing rate. It follows that the organism would benefit from having two classes of cells: cells that respond to light *increments* with an *increase* of firing rate, and cells that respond to light *decrements* with an *increase* of firing rate. These arguments combine to suggest that ON-center cells serve the detection of increments and OFF-center cells serve the detection of decrements.

The plot thickens. In the early 1980's a chemical substance was discovered that disables the retinal ON-center system while leaving the OFF-center system intact. The poison, 2-amino-4-phosphonobutyric acid, or APB, disables the ON-center bipolar cells by hyperpolarizing them strongly for several hours. Moreover, the effects of APB are *reversible*: they disappear within 24 hours. Thus it is possible to produce an experimental animal that is missing a functional ON-center system on one day and whole again the next.

Combining the functional arguments given above and the reversible lesions provided by APB, Peter Schiller and his research group made two important predictions. APB disables the ON-center system but leaves the OFF-center system intact. Therefore, their first prediction was that for cone-mediated vision, APB should interfere with the detection of increments but have little influence on the detection of decrements. But in the case of the rod system, all of the rod bipolars are ON-center, and all rod-initiated signals must pass through these ON-center bipolars before finally entering the OFF-center system at the OFF-center cone bipolar (Figure 9.12xxx). Therefore, their second prediction was that for rod-mediated vision, APB should lead to major deficits in the detection of both increments and decrements.

Since it is obviously impossible to carry out such invasive experiments in human subjects, Schiller and his group used monkeys as subjects. The first experiment was carried out by Schiller, Sandell and Maunsell (1986), using stimuli custom-designed to isolate cone-mediated vision. Their paradigm and results are shown in Figure 11.5. The first step was to train a monkey subject to fixate the center of a homogeneous field of light, and record her eye movements as she did so. Then, using light levels designed to isolate cone-mediated vision, these researchers presented a brief incremental or decremental test flash at one of six possible locations around a fixation point. The animals task was to fixate the fixation point, watch for an increment or decrement at any of the six locations, and when she saw one, quickly move her eyes to look at the location where the increment or decrement appeared. A rapid and accurate eye movement to the target position produced a reward (a drop of juice) for the monkey, and indicated that the monkey had detected the target.

The pattern of eye movements produced by the monkey for suprathreshold targets is shown in Figure 11.5A. Both increments and decrements were readily detectable, and the monkeys' eye movement reaction times were relatively short and regular, as shown in the top two panels of Figure 11.5B and in Figure 11.5C.

Next, the researchers treated the monkey with APB, disabling the ON-center system while leaving the OFF-center system intact; and retested the monkeys in the increment and decrement tasks. The data are shown in the bottom two panels of Figure 11.5B and in the deviant points

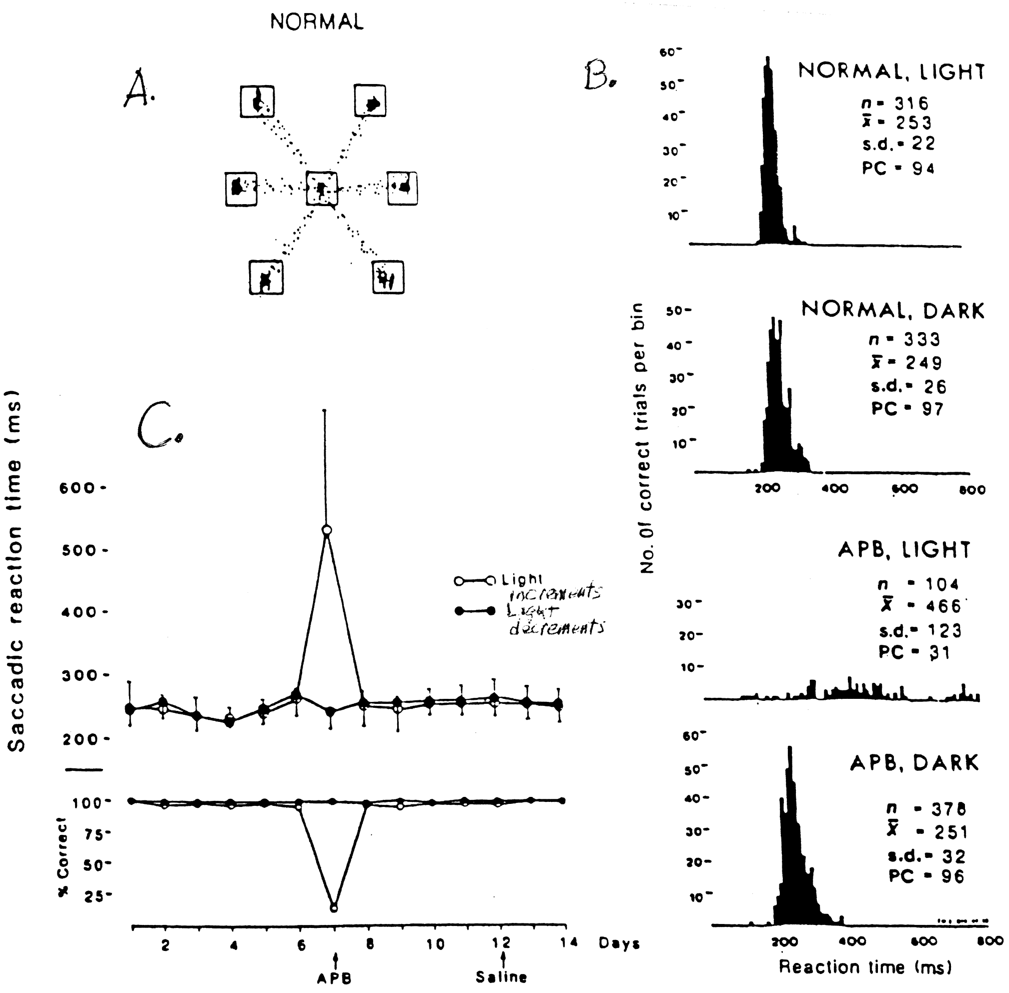


Figure 11.5: The experiment of Schiller, Sandell, and Maunsell (1986). A: The stimulus set-up. The monkey fixated a central fixation square presented against a mid-intensity background field. One of six surrounding test squares was presented, as either an increase or a decrease in intensity. The monkey's task was to shift his eyes to fixate the test square as quickly as possible. The figure also shows the pattern of the monkey's eye movements in response to clearly supra-threshold stimuli. All of the responses were prompt and correct. B: In the experiment proper, the test stimuli were presented at moderately supra-threshold light levels. The upper two panels show the distributions of reaction times in ormal animals, to light increments (”light”) and decrements (”dark”) respectively. The lower two panels show the distributions of reaction times in APB-treated animals. C: The abscissae show days of the experiment. The arrow labelled ”APB” indicates the day of APB treatment; the arrow labelled ”Saline” indicates a control treatment with a saline solution. The upper panel shows reaction times, and the lower panel shows the monkey's per cent correct eye movements (chance is 16.5%). APB clearly devastates visual function for light increments but has little if any effect for light decrements. (Modified from Schiller et al, 1986, Fig. 1, p. 825.)

in Figure 11.5C. With APB treatment the monkeys showed little change in reaction times to decrements, nor in decrement thresholds. But they had a very hard time with increments!

These data provide strong evidence that in an otherwise normal animal, an intact retinal ON-center system is a *necessary condition* for high sensitivity to increments of light, but not for high sensitivity to decrements. Thus, one can argue that at the retinal level the OFF-center system is *sufficient* for processing decrements. (One of the ”nothing mucks it up” provisos involved in this argument is that there are no other cell types available. We are relying on the fact that the ON-center and OFF-center cells we have described are the only two cell types that enter into the processing of increments and decrements at the retinal level.)

A similar experiment was also carried out by Dolan and Schiller (1989) at light levels designed to test rod-mediated vision. In this case, the monkeys lost sensitivity to both increments and decrements of light. Thus, both of the original predictions were correct. These experiments, then, provide evidence to strengthen the constellation of arguments made above: that ON-center cells give us sensitivity to increments, and OFF-center cells to decrements; and that coding an event by increases in firing rate is important[4].

Finally, these experiments tie back to the themes of this chapter. They provide the first example we have seen of the use of lesions, and also of the value of combining animal psychophysics with lesion procedures. Also notice the custom-designing of stimuli – high vs. low adaptation levels to isolate rod-mediated vs. cone-mediated vision, and increments and decrements to isolate ON-center and OFF-center neurons. Moreover, the custom-design is strongly influenced by co-evolution – Schiller and his colleagues had to know a lot of retinal physiology to design their essentially psychophysical experiments.

11.3 Ensembles and neural images

In all of the above examples of psychophysical experiments, vision scientists custom-designed stimuli to influence individual cells, or they at least used models that could be phrased in terms of the activities of single cells. Moreover, all of these experiments have been Class A – experiments centered around detection thresholds. But in fact, most everyday stimuli are above detection threshold; and most stimuli will fall on the receptive fields of *many* cells. As in the case of trichromacy, we now consider the fact that the joint action of an *ensemble* of cells is necessary to preserve and convey most facets of visual information in real-world settings.

The *retinal image* of a stimulus is its two-dimensional optical image on the retina. Similarly, let’s define the *neural image* of a stimulus to be the set of states of the neurons at a given level of the visual system that result from the presence of the stimulus. That is, for a given stimulus, there is a retinal image, a photoreceptor-level neural image, a bipolar cell neural image, a ganglion cell neural image, and so on up the line. (In fact there are several at each level, one for each type of neuron.) To illustrate this idea, we now consider how the retinal ON-center ganglion cells code the presence and location of an edge – a light-dark transition in the visual stimulus.

[4] DT finds it remarkable that mirror image coding is so important, given the many other classes of environmental events that must be competing for representation in the retinal code.

11.3.1 How is an edge represented at the retinal output?

Figure 11.5A shows a sharp *edge* – a lower intensity field on the left, and a higher intensity field on the right – falling on the retina. Such an extended visual pattern will fall upon the receptive fields of many retinal ganglion cells. For simplicity, lets consider just the ensemble of prototypical ON-center ganglion cells, as shown in Figure 11.5B. What will the *neural image* of an edge in the ensemble of ON-center ganglion cells be like?

Consider a row of prototypical ON-center ganglion cells., as sketched in Figure 11.5B. Four of these are marked cells E, F, G and H. Due to the presence of center-surround antagonism in their receptive fields, cells E and H, lying under the homogeneous lower- and higher-intensity portions of the pattern, will remain near their maintained firing rates. But the responses of cells F and G (Figure 11.6C) will be different. Cell F, just to the left of the edge under the lower intensity portion of the pattern, will have lots of light on some of its surround and less light on its center, so it will decrease its firing rate. Cell G, just to the right of the edge under the higher intensity portion of the pattern, will have lots of light on all of its center and only part of its surround, so it will increase its firing rate. And by symmetry, a cell that is right under the border should not change its firing rate. Figure 11.5D shows the results of an actual physiological experiment designed to mimic the one described, and it follows the predicted pattern[5].

In sum, in an ensemble of ON-center ganglion cells, a dark/light edge does not create a group of ganglion cells firing uniformly rapidly under the light side of the edge, and another group firing uniformly slowly under the dark side of the edge. Instead, the information is recoded. The center/surround receptive fields create a set of *dog-ears* – a downward blip to the left of an upward blip – in the neural image. Oppositely, a light/dark edge would create dog-ears with an upward blip to the left of adownward blip. Such a pattern in a neural ensemble has often been called *edge enhancement.* This is the code in which the location of an edge on your retina is carried in your visual system at the ganglion cell level.

Why did the EDC create this recoding? From a design perspective it has been suggested that signalling the locations of edges – rather than the intensity of light at each point – might be an advantageous neural code. If neurons corresponding to large areas of homogeneous activity can remain at their maintained levels, while only the neurons near the edges need change, a coarsely patterned scene can be coded by signals (deviations from resting level) in a relatively small proportion of neurons.

[Now, some thought questions. What would the pattern of activity in the ensemble of ON-center ganglion cells be for a low frequency square-wave grating? Think about mid and high spatial frequencies too. What about sinusoidal gratings of different spatial frequencies? What about a triangle? And what about ensembles of OFF-center cells? Finally, the receptive fields of ganglion cells were created from a spatial weighting of photoreceptor inputs. Would it be possible to combine inputs from a set of ganglion cells, in such a way as to create a single neuron whose activity represents the presence of an edge? How? (As you will see in the visual cortex, the answer is yes).]

In sum, there is an important similarity between an ensemble of three cone types carrying wavelength information through the photoreceptor layer, and an ensemble of many ganglion cells

[5] Recording from each of the ganglion cells in the neural image is actually not feasible. Instead, Enroth-Cugell and Robson (1966) recorded from a single cat ganglion cell, placing the edge at various positions with respect to the center of the cell's receptive field.

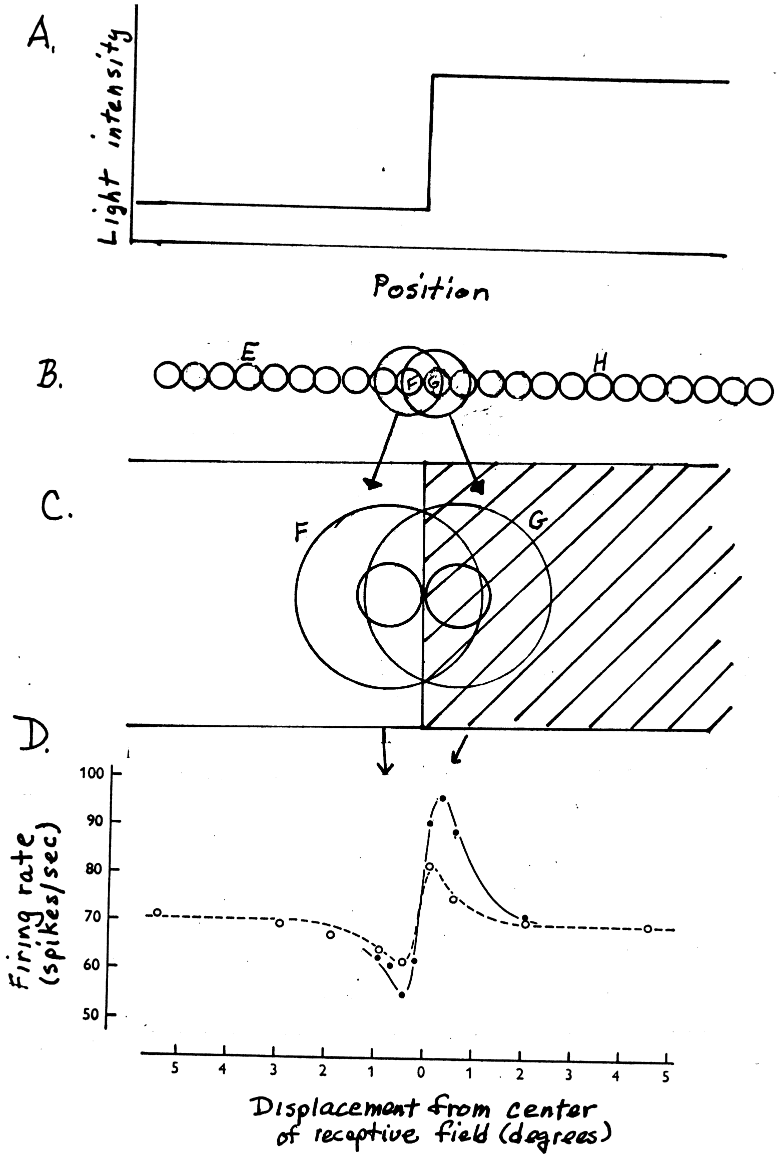


Figure 11.6: The neural image of an edge in an ensemble of ON-center ganglion cells. A: Cross section through the light distribution across the edge. B: A row of prototypical ON-center ganglion cells. Except for neurons F and G, the two neurons closest to the edge, only the centers of the receptive fields are shown. E and H are neurons that lie far from the edge. C. Magnified, face-on view of the receptive fields of neurons F and G, with the edge lying across them. For cell F, light covers only a portion of the surround, and none of the center. For cell G, light covers all of the center and much of the surround. D. Results of an experiment designed to simulate the pattern of activity in such an ensemble of neurons. Closed circles: 20% contrast; open circles: 40% contrast. The neural image shows edge enhancement. (A-C by DT, D modified from Enroth-Cugell and Robson, 1966, Fig. 14, p. 541.)

carrying spatial information through the ganglion cell layer. In both cases, no single neuron tells us what the stimulus pattern is. We have to consider the relative activities in three – or many – cells at once.

11.3.2 Case 4: Mach Bands

All of the cases we examined in detail above had to do with Class A psychophysical experiments – detection thresholds. However, retinal processing has also been used from time to time in attempts to explain Class B phenomena. Arguments that involve Class B phenomena will necessarily be more complicated and dicey than those based on Class A phenomena, because they will involve linking propositions other than Identity, and because to many of us it seems to model fundamentally perceptual (conscious) phenomena on the basis of physiological activity in a region as remote as the retina. Let's examine a case in detail.

The Mach band phenomenon is illustrated schematically in Figure 11.7. Suppose that we create a stimulus composed of a high intensity field, a low intensity field, and a gradual shading off – a "ramp" – of intensity in between. Most subjects report seeing an extra bright band at the bright edge of the ramp, and an extra dark band at the dark edge of the ramp. These extra bright and dark bands are illusory – they occur in perception but not in the stimulus. They are called Mach bands[6].

Why do Mach bands occur? An appealing analogy arises from considering an ensemble of ganglion cells with center-surround antagonism. Like the sharp edge in Figure 11.6, a ramp pattern is likely to yield maxima and minima – dog ears – in the firing patterns across the ensemble of ganglion cells, roughly in the locations where the bright and dark Mach bands are seen. Yielding to the analogy, one can imagine that these maxima and minima in the neural image of the ramp pattern cause the perception of Mach bands.

But now let's think hard about this analogy. First, of course, we would want to test some actual ganglion cells, to wee whether or not they show the right pattern across the ensemble. Second, notice the implicit assumption that the perceived brightness at each point in the visual field is determined by the level of activity in the ganglion cell corresponding to that point in the neural image – the greater the firing rate of the ganglion cell, the greater the perceived brightness.

But examining this assumption only leads to more questions. First, over what domain of stimuli and perceptual phenomena is this rule assumed to hold? Does perceived brightness always correspond to the firing rate of a ganglion cell? If so, why don't we see Mach bands at sharp edges, which also set up dog ears in the neural image? Will all brightness phenomena fit this rule? To make this analogy work, one must develop a principled argument as to when dog-ears lead to the veridical perception of a sharp edge, and when they lead to the perception of illusory bright and dark bands.

And third, there's an enormous "nothing mucks it up" proviso involved – we must be assuming that nothing in the recodings between the ganglion cells and whatever underlies our conscious

[6] The phenomenon is named after Ernst Mach (1838-1916), a physicist whose "skepticism and independence" were noted by Einstein, and who was brave enough to declare that what he saw was not always a veridical representation of the physical stimulus. Mach bands are easy to demonstrate with a projector and a piece of cardboard. Hold the cardboard in the projector beam about halfway between the projector and the screen, varying the distance for maximal effect. The cardboard creates a shadow and its penumbra creates the ramp. Prominent illusory bright and dark bands should be seen. In fact, Mach bands are easy to see directly once you know what to look for. By now, for DT every shadow that has a ramped edge is accompanied by Mach bands.

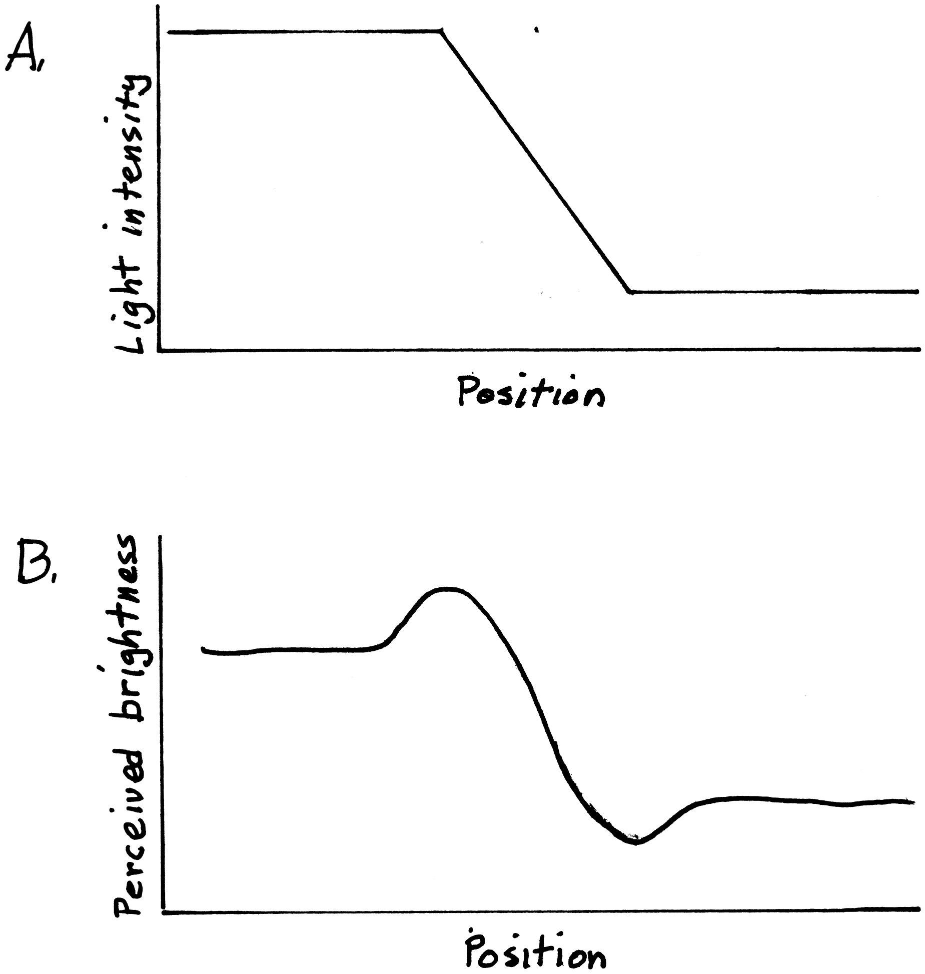


Figure 11.7: Mach bands. A: A ramp pattern in the physical stimulus. B: The common perception of the ramp pattern. Most subjects see illusory bright and dark bands at the ends of the ramp.

perception, mucks up the correlation between ganglion cell firing rates and perceived brightnesses. This assumption must impose a powerful constraint on the kinds of recodings that can happen in the intermediate levels of the visual system. These transformations must, in some vague and poorly understood sense, preserve the retinally established brightness code, or else Mach bands wouldn't occur.

It's not that this explanation of Mach bands is necessarily wrong. It's just that explaining the features of conscious perception by means of patterns of neural activity in early stages of the visual system is a great deal more complicated than might be thought when the analogy first arises. And yet, it is certainly legitimate to orgue that the coding that causes a Class B phenomenon is introduced at a particular physiological level, even as early as the retina. We just have to know what we mean when we say it, and be ready to accept the assumptions involved.

11.4 Summary: Retinal processing and perception

In the present chapter, we examined four different causal stories that attempt to explain perceptual phenomena on the basis of the functioning of retinal ganglion cells. In the realm of detection thresholds, we explored the analogies between center/surround antagonism and the Westheimer phenomenon; between the physiological convergence of rod-initiated and cone-initiated signals and the psychophysical demonstration of rod-cone interactions; and between ON-center and OFF-center ganglion cells and the detection of light increments and decrements. Then, considering ensembles of ganglion cells and taking a more perceptual tack, we also examined the argument that activity in an ensemble of ganglion cells might provide the explanation for the perceptual illusion known as Mach bands.

These examples were intended to give you some practice at working through the kinds of experimentation and argumentation that are typical of visual science. In particular, we emphasized the use of analogies for exchanging hints across sub-disciplines; the joint use of lesions and psychophysics in animal subjects; and the use of custom-designed stimuli to silence particular elements of the visual system and stimulate others, in order to explore the properties of vision when only some of the elements of the visual substrate are available to the perceiver.

Chapter 12

Opponent Color Codes and Color Space

In this chapter we return to the topic of color vision. Let's review briefly the main themes of earlier color chapters. In Chapter 3, we explored some of the classic Class B observations about color appearance. We asserted that for isolated patches of light and color-normal observers, variations in wavelength map regularly to variations in perceived color. We reviewed Hering's observations of unique vs. binary hues (red, green, yellow and blue being the unique hues), and of mutually exclusive hue pairs (red vs. green and yellow vs. blue). We also unearthed Hering's major linking proposition: that mutually exclusive *perceptions* imply mutually exclusive *physiological processes.*

Using these facts and assumptions, we explored a neo-Heringian model of the neural code for color appearance (Figure 3.8). This model suggests that color appearance is coded by two separate *oppponent* physiological processes – processes that can deviate from their resting states in either of two directions. The two putative opponent processes are "redness/greenness" neurons that (say) increase their output to signal redness and decrease it to signal greenness, and "yellowness/blueness" neurons that (say) increase their output to signal yellowness and decrease it to signal blueness.

Then, in Chapter 7 we explored the fact that color vision is behaviorally trichromatic, and explained this behavioral fact with a mathematical model. From this model we predicted that the retina contains three and only three cone types, but were unable to specify their spectral sensitivities. More recently, experiments of several kinds have verified the prediction of three cone types, and converged to establish their spectral maxima at about 430, 540, and 570 nm (Figure 7.6).

Historically, trichromatic theory and opponent process theory were often set up as competing alternatives, and some color scientists construed the successes of trichromatic theory as evidence against opponent process theory. But other color scientists argued that the two coding schemes could be instantiated at different stages or *zones* of visual processing. The three-channel bottleneck could be imposed by the photoreceptors, and the feature of mutual exclusiveness could be introduced in postreceptoral processing.

In this chapter we will play out the theme of opponent processing in color vision. We will introduce the idea of *three-channel linear models* of postreceptoral processing, and show how the geometry of three dimensional spaces can be used to reveal their properties. We will introduce the concepts of *null planes* and *isolation axes*, and from them derive new paradigms for studying color coding at postreceptoral levels.

Although this material is somewhat specialized, we feel it is worth including in an introductory textbook for several reasons. First, although the theoretical framework is mathematical, the major ideas can be conveyed with the use of geometry. Second, this approach continues and sustains the theoretical elegance – the complementary use of psychophysics, mathematics, and physiology – introduced earlier with the trichromatic theory of color vision. And third, some of the most exciting current research in visual science concerns physiological studies of the recoding of color in the primate retina – *our* retina. The paradigms used in that work are based on the material in this chapter.

The question addressed in this and the next two chapters is, what recombinations of the signals L, M and S take place to form the early postreceptoral color codes? How can we model these changes, and how can we search for them with psychophysical and/or physiological experiments? The present chapter explores the modelling and geometry of postreceptoral color processing; Chapter 13, the physiology; and Chapter 14, the psychophysics.

12.1 Opponent coding schemes

The neo-Heringian model, with its two types of putative color-opponent channels – redness/greenness and yellowness/blueness – carries the seeds of an explanation for the perceptual mutual exclusiveness of the mutually exclusive hue pairs. The argument is that since a neuron cannot both increase and decrease its output simultaneously – the two directions of the response are inherently mutually exclusive – the two hues coded by such a process must also be mutually exclusive. Moreover, the distinction between unique and binary hues also has an explanation in the neo-Heringian model. The unique hues red and green occur when the redness/greenness channel is activated in one or the other direction, but the yellowness/blueness channel is at its resting state; and conversely for the unique hues yellow and blue. These two properties – opponent coding that produces mutual exclusiveness, and null points in each channel that correspond to the unique hues coded by the opposite channel – will guide the upcoming modelling of postreceptoral processing.

12.1.1 Subtraction of signals from different cone types

A very simple scheme for creating a color-opponent neuron is shown in Figure 12.1. Let us assume, as shown in Figure 12.1A, that signals initiated in an M cone and an L cone combine their signals on a postreceptoral neuron, but with *opposite signs.* In this example, let's say that the L-cone-initiated signal depolarizes the postreceptoral neuron, and the M-cone-initiated signal hyperpolarizes it. This neuron and its inputs embody the mathematical process of *subtraction*, and we have just made a postreceptoral neuron that computes the signal L - M.

As shown in Figure 12.1C, the response of the postreceptoral neuron will vary systematically with the wavelength of light, depolarizing to a range of long wavelengths and hyperpolarizing to a range of mid-wavelengths. The neuron will also have a *crossover point* or *null wavelength* – a wavelength at which there is *no* deviation from the resting level – at some particular intermediate wavelength. Thus this simple subtraction scheme produces the raw material for a model that embodies both mutual exclusiveness (increases and decreases of the signal in a single neuron to different regions of the spectrum) and unique hues (crossover points).

Moreover, as shown in Figure 12.1A and C vs. B and D, the *weightings* given to the L and M cone inputs will influence the crossover point for the postreceptoral neuron. Let us say that in

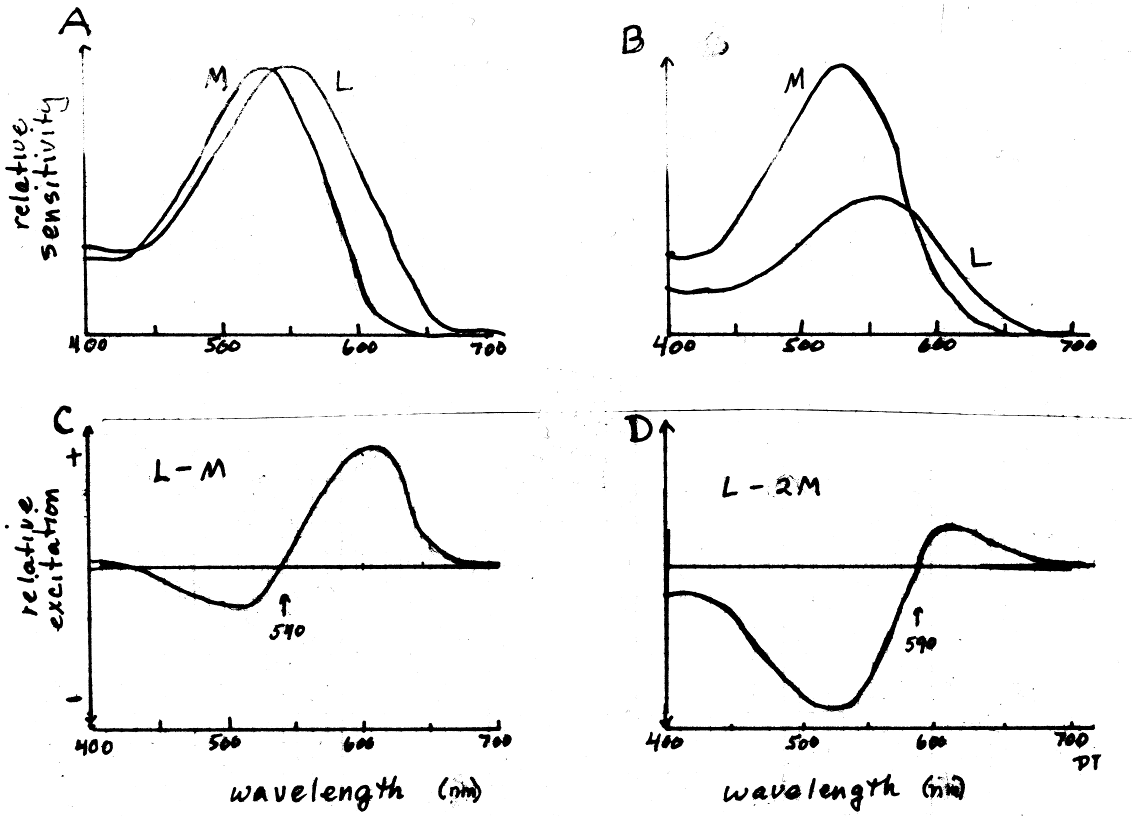


Figure 12.1: Subtraction of signals from two different cone types. A: The spectral sensitivity curves for the L and M cones, both set to the same maximum sensitivity. B: The same two spectral sensitivity curves, but with the height of the L cone curve cut in half to decrease its weighting. C: The difference between the L and M cone signals in A. The postreceptoral neuron increases its excitation for long wavelengths, and decreases it for short wavelengths, with a crossover point (null point) at about 540 nm. D: The difference between the L and M cone signals in B. Notice the shift in the crossover point to about 590 nm.

Figure 12.1A the weightings of the L vs. the M cone inputs were equal – the neuron computes L - M – and that this weighting scheme produces a crossover point at about 540 nm. As shown in Figure 12.1B and D, if we halve the weighting given to the L cones, so that the neuron computes L - 2M, the crossover point shifts to about 590 nm; and so on. Other judicious choices of cone weights will yield other crossover points. Furthermore, other subtractive combinations of cone inputs will yield postreceptoral neurons with different opponent response curves. [Construct a neuron that opposes an S cone input and an M cone input. If the weights are equal, at about what wavelength will the crossover point be?]

12.1.2 Early physiological evidence for opponent coding

But should we be convinced by Hering's arguments? Prior to the early 1950s there were psychophysical data pointing to the presence of opponent processes in color perception, and mathematical models of how opponency could occur. But there was no direct physiological evidence for neurons that actually subtract inputs from different cone types. For this reason, combined with a general bias against the value of physiological inferences drawn from Class B experiments, some some vision scientists remained skeptical about the existence of opponent color processing.

Direct physiological evidence became available in the 1950s, when Gunnar Svaetichin developed techniques that allowed him to record intracellularly from single neurons in carp retinas. He found (as you already know) that early retinal neurons respond with slow potentials rather than spikes. Responses of the kind originally recorded by Svaetichin have come to be called S-potentials (S for slow, or for Svaetichin).

Since Svaetichin could only hold the cells for a minute or two, he needed to present lights of different wavelengths in rapid succession. So he attached a set of color filters to the outer edge of a bicycle wheel, filled in between the filters with opaque material, and rotated the wheel. What Svaetichin found was that some retinal neurons gave the same direction of polarization for all wavelengths of light. These were called L-type (for luminance-type) S-potentials (not shown). But other neurons depolarized to some wavelengths and hyperpolarized to others! Some of Svaetichin's early data are shown in Figure 12.2. These were called C-type S-potentials (C for color), and they provided the first physiological evidence that opponent processes like those postulated by Hering occur in real vertebrate visual systems.

Svaetichin originally proposed that the S potentials were recordings from photoreceptors, but it turned out that fish photoreceptors, like ours, have simple U-shaped spectral sensitivity curves. In fact, Svaetichin's recordings were from the fish's horizontal cells.[1] So, at least in fish and turtles, it turns out that trichromatic theory and opponent process theory are both instantiated, within one synapse of each other in early retinal processing. The zone theorists had it right.

Svaetichin's data, and data from primate LGN cells that followed in the1960s, led most color scientists to believe that opponent coding would turn out to be an important part of primate color processing. Armed with this motivation, in the next few sections of this chapter we will explore the logic of opponent process models.

[1] Thus, fish horizontal cells provide an exception to the claim we made in Chapter 9 that horizontal cells always hyperpolarize to light. In fish (and turtles), some horizontal cells hyperpolarize to one range of wavelengths and depolarize to another.

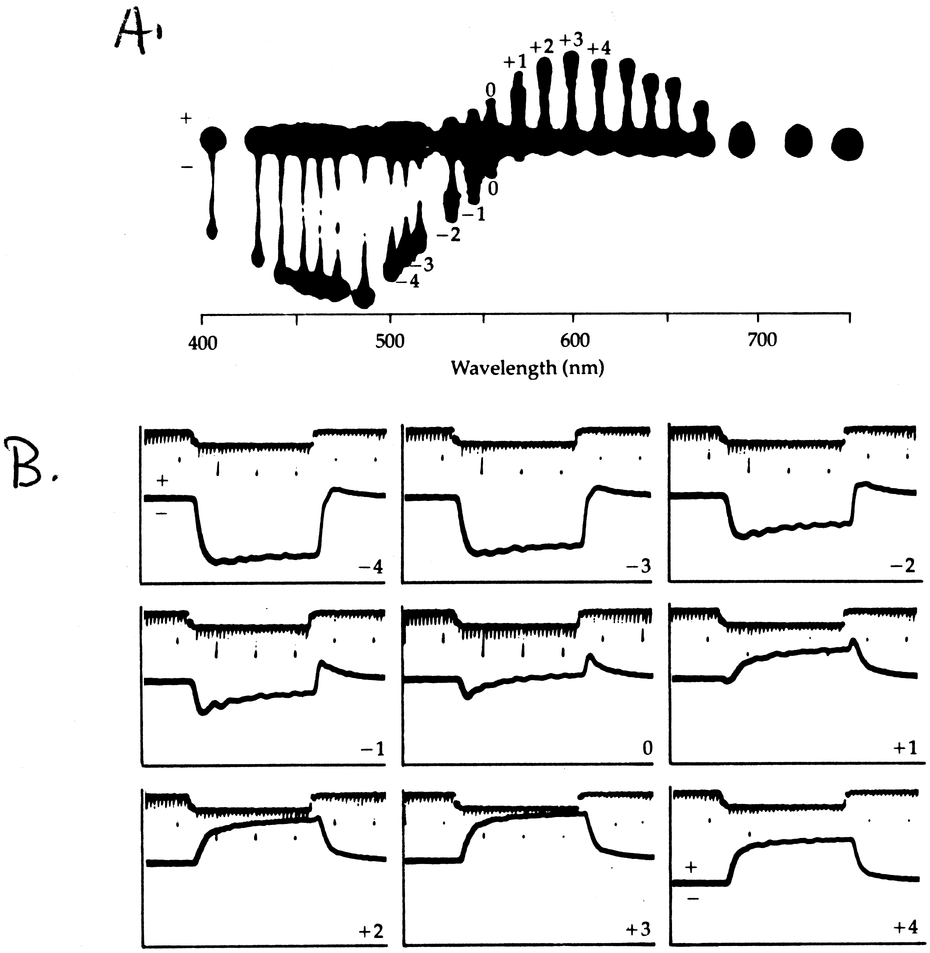


Figure 12.2: A color-opponent neuron in the retina of a fish. A: The neuron gives different polarities of response to different spectral regions, with a crossover point at about 550 nm. The spike-like traces are actually slow potential responses to presentations of lights of a series of different wavelengths. B. The responses on a slower time scale. The sign of the response reverses at around 550 nm. [From Hurvich (1981), Fig. 2, p. 139; after MacNichol and Svaetichin (1958).

12.2 Three-channel linear color models

The above considerations set the scene for introducing an important class of models called *three-channel linear color models.* These models all have three characteristics in common. First, these models all assume the existence of *three and only three postreceptoral channels.* Why make this assumption? Because there are only three cone types – only three degrees of freedom – in the initial code. Therefore, in principle three postreceptoral channels are *sufficient* to carry all of the information contained in the initial code. If more than three kinds of channels are created, their signals will be mutually *redundant.* (Of course, the EDC could have given us more than three channels for other purposes.)

Second, these models assume that the postreceptoral signals are made up of *weighted linear sums and differences* of inputs from the three cone types. All linear color models posit that the signals in the postreceptoral channels take the following general form:

Channel 1 signal: $\mathrm{Ch}_1 = \mathrm{w}_1\mathrm{L} + \mathrm{w}_2\mathrm{M} + \mathrm{w}_3\mathrm{S}$

Channel 2 signal: $\mathrm{Ch}_2 = \mathrm{w}_4\mathrm{L} + \mathrm{w}_5\mathrm{M} + \mathrm{w}_6\mathrm{S}$

Channel 3 signal: $\mathrm{Ch}_3 = \mathrm{w}_7\mathrm{L} + \mathrm{w}_8\mathrm{M} + \mathrm{w}_9\mathrm{S}$

where Ch_1, Ch_2, and Ch_2 are the signals in the first, second, and third putative postreceptoral channels respectively, and the w's are the weights assigned to the various cone inputs in these three postreceptoral channels.

And third, most models have presumed *opponent*, or *subtractive*, coding in two of the postreceptoral channels. That is, in the opponent channels, the weight assigned to the signal from at least one cone type will be *negative* to indicate subtraction. The channel that computes a positive sum is usually termed the *achromatic* or *luminance* channel, and the two channels that subtract cone inputs from each other are usually termed the *chromatic* channels. (DT's mnemonic is that if there's a minus sign in the equation, you're dealing with a chromatic channel.)

12.2.1 An optimal neo-Heringian model

Just as many color theorists used the fact of trichromacy to guess at the cone spectra, many color theorists have also used the identities of the unique and mutually exclusive hues (in combination with other more or less arbitrary constraints) to guess at the cone input weights to the putative postreceptoral color channels. However, the historical guesses were always limited by inaccuracies in the estimated cone spectra. Now that the cone spectra are known (Figure 7.6), it is possible to seek out the set of cone weights that provides the optimal fits to the unique hues.

The optimal cone weightings are shown in Figure 12.3. They produce crossover points of xx and xx for the red/green channel and xx for the blue/yellow channel (remember unique red is extraspectral). These values come very close to the wavelengths of the corresponding unique hues.

In sum, these curves are viable predictions of the spectral response functions of neurons that could code perceived color in the primate visual system. But there's a slight problem – neurons that instantiate the Hering code have not yet been found in the primate visual system at any level. What do we do now? In particular, how convinced should we be that neurons instantiating the Hering model must (or might) exist within the human visual system?

It seems to DT that there are three possible stances a color vision physiologist can take. First, she can accept Hering's linking proposition, and believe that an opponent color code with the right crossover points *must* exist, and will eventually be found. She might even decide to search for

(I hope to receive this from Joel Pokorny).

Figure 12.3: Optimal cone weightings for a neo-Heringian color code. A: Cone spectral sensitivities. B: The responses of one non-opponent and two oppponent channels, produced by optimal weightings of cone inputs to fit the crossover points of the opponent channels to the unique hues, and to fit the spectrum of the non-opponent channel to $V(\lambda)$. xx ?? The optimal neo-Heringian model predicts that neurons with these spectral response curves will be found in the human visual system. [Calcuilations and figure kindly provided by Dr. Joel Pokorny.]

it in visual cortex. Second, she can come up with a new account of the perceptually unique and mutually exclusive hues, and search for the code predicted from that new account. Or third, she can decide that the argument from psychophysical mutual exclusiveness, and the mutual exclusiveness proposition, are just too frail a reed on which to base a scientific career, and study something else. Which choice will lead to triumph and which to obscurity? Only time will tell.

12.2.2 The Boynton model: L + M, L - M, and S - (L + M)

Here's another fundamental question. Even if the Hering code is instantiated at some high cortical level, does the recoding to that coding scheme happen all at once? Or will there turn out to be one or more earlier opponent codes, perhaps with increasing similarity to the Hering code, as we move from retinal to cortical processing levels? In fact, there is considerable evidence for an *early opponent code* that differs importantly from the neo-Heringian code. We will call this putative early opponent model the *Boynton model*[2].

The Boynton model proposes that there are three early postreceptoral channels that take the form shown in Figure 12.4. The three channels are usually called L + M, L - M, and S - (L + M) after their cone inputs. The proposal is that the *L + M channel* sums signals from the L and M cones. It is designed to provide the neural substrate of the photopic spectral sensitivity curve, Vλ, which we modelled in Chapter 7 by summing L and M cone signals (Figure 7.13). The *L - M channel* takes a difference between L and M cone inputs. The *S - (L + M) channel* is the only one with S cone input. It sums the L and M cone signals, and subtracts them from the S cone signal.

What about the cone input weights? To specify the model fully we would need to specify all of the input weights, w, introduced in connection with the general three-channel linear model. But as yet there is no consensus on the exact weights, and at the present time different authors endorse the use of different weights. Moreover, the option is open that different neurons of the same type have somewhat different cone input weights, and/or that the visual systems of different individuals do so. For this reason, in talking about this class of models, vision scientists usually leave the question of weights purposefully open, and just attend to the *signs* of the inputs to the three channels. The whole story is still a work in progress, as trichromacy was prior to the 1980s.

Nonetheless, for the purposes of this textbook, we will adopt some simple working weights. For the L + M channel, a reasonable way of thinking is that the L cone signals are weighted about twice as heavily as the M cone signals, because this combination makes a good fit to Vλ. Thus, the L + M channel becomes 2L + M. For the L - M channel, a reasonable guess is that the L and M cone signals have approximately equal weights. And for the S - (L + M) channel, a reasonable guess is that the positive and negative signals are in approximately equal balance, so the S cone signals are weighted twice as much as either the L or the M cone signals (which we guess are about equally weighted). Thus, the S - (L + M) channel becomes 2S - (L + M).

The neo-Heringian model and the Boynton model start from different premises. Yet both models can be seen as predictions of the kinds of neurons that will be found inthe human visual system. One way to incorporate both the neo-Heringian model and the Boynton model into our thinking is to assume that beyond the photoreceptors there are at least two recodings involved in the perception of color. Perhaps the Boynton model really does describe an early opponent code, instantiated within the retina or at an early cortical level; and perhaps the neo-Heringian model

[2] This class of models is named in honor of Robert Boynton, who explored this and related models in his classic book on color vision (1979). Many variants of this model can be found in the color vision literature.

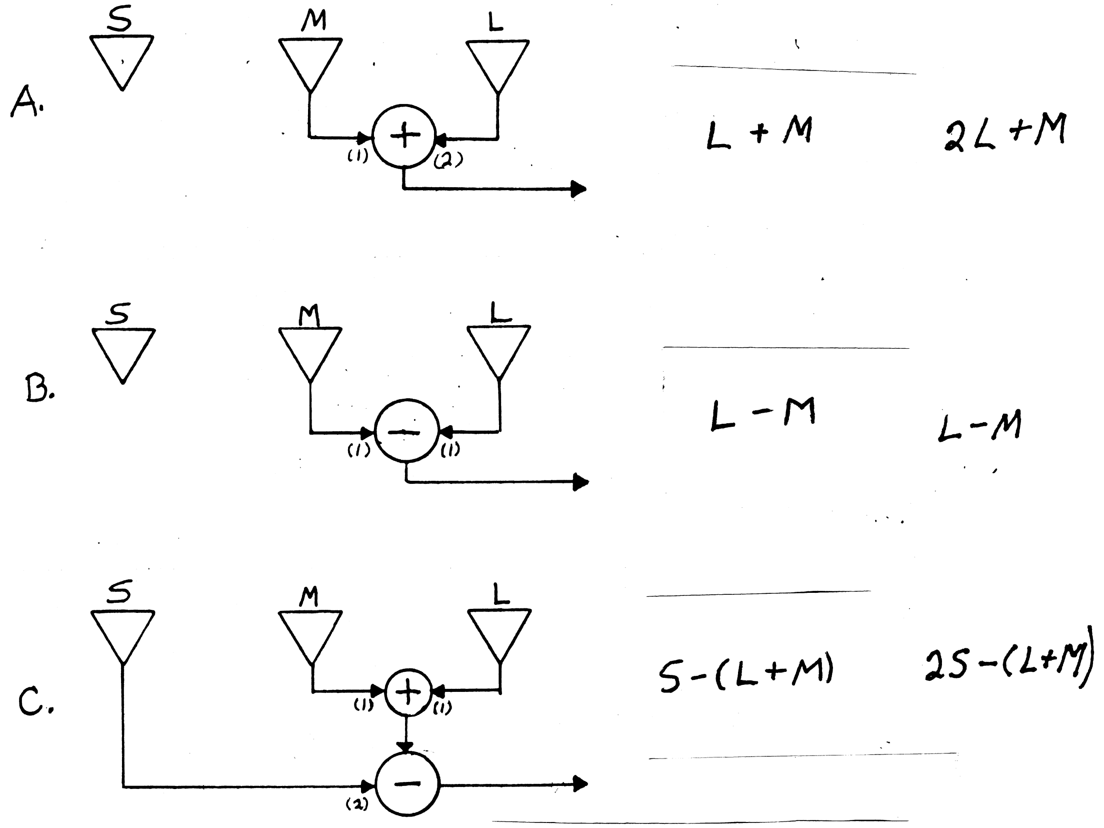


Figure 12.4: The Boynton model. A: The (non-opponent) L + M channel. B: The L - M channel. C: The S - (L + M) channel. The sketches in the left half of the figure show the cone inputs to each channel. The triangles represent cones, and the circles represent sites at which signals from two or more cone types are combined. The small numbers near the combining sites are working approximations of the cone input weights. The middle column shows the names of the three channels, in terms of their cone inputs. The right hand column shows the working cone input weights adopted in this chapter. [Sketches modified from Boynton (1979), Fig. 7.3, p. 212.]

describes a more central code that more directly governs the perception of color – an elaborated three-zone zone theory.

We turn now to algebraic and geometric representations of three-channel linear color theories. We begin with a space in which the L, M and S cone quantum catches are represented on the *cardinal axes*[3] – the x, y, and z axes – of the space, and proceed to more complex variations.

12.3 Cone input space

In Chapter 7 we derived general expressions for the quantum catch rates in the three cone types:

$$L = \sum Q\lambda l\lambda$$

$$M = \sum Q\lambda m\lambda$$

$$S = \sum Q\lambda s\lambda$$

The absorption spectra of the cones give the values of the lλ's, mλ's, and sλ's. Given these values and the spectral composition of the incoming light (the Qλ's), we can calculate the quantum catches L, M, and S – the *cone inputs* – from a stimulus of any spectral composition.

The next step is to construct a three-dimensional graph in which to represent the cone quantum catches. We will call such a graph a *cone input space*. Since three-dimensional drawings can be difficult to draw and even more difficult to "see", Figure 12.5A shows a two-dimensional space representing just the L and M cone quantum catches, or an *L,M plane*. Figure 12.5B shows a three-dimensional graph representing all three cone inputs – an *L,M,S cone input space*.

The first point to appreciate is that light of any given wavelength composition and intensity makes a triplet of cone inputs, L, M, and S. That is, light of any given wavelength composition and intensity can be represented as a *point* in either the two or the three-dimensional space. In the L,M plane, we can represent fully only wavelengths above about 550 nm, as shorter wavelengths yield an S cone quantum catch as well. Long wavelength lights, which make a greater quantum catch in the L cones than in the M cones, will plot to points near the L cone axis, and wavelengths near 550 nm, which make more nearly equal quantum catches in the L and M cones, will plot near the right diagonal of the space. In the three dimensional L,M,S space, wavelengths below 550 come forward out of the plane of the paper, with short wavelengths falling nearest the S cone axis.

Moreover, any given wavelength of light creates a *ratio* of cone quantum catches that remains constant across variations in intensity. Thus, variations in intensity for a fixed wavelength will fall along a specific line, or *ray*, outward from the origin. The rays for wavelengths above 550 nm will lie in the L,M plane, as shown in Figure 12.5C. The rays for wavelengths below 550 nm will rise out of the plane of the paper, as depicted by the arrows that thicken as they leave the origin in Figure 12.5D.

[3] A word about cardinal axes. Imagine that there's a hunk of three-dimensional space in front of you. In order to refer to lines and points in this space, it is necessary to impose a set of reference lines, such as the usual x, y and z axes. These reference lines are called the cardinal axes of the space. However, they are to some degree arbitrary, and different choices of cardinal axes will make different characteristics of the space explicit. The next few sections can be seen as an exercise in choosing and changing sets of cardinal axes, in order to make different aspects of color vision explicit.

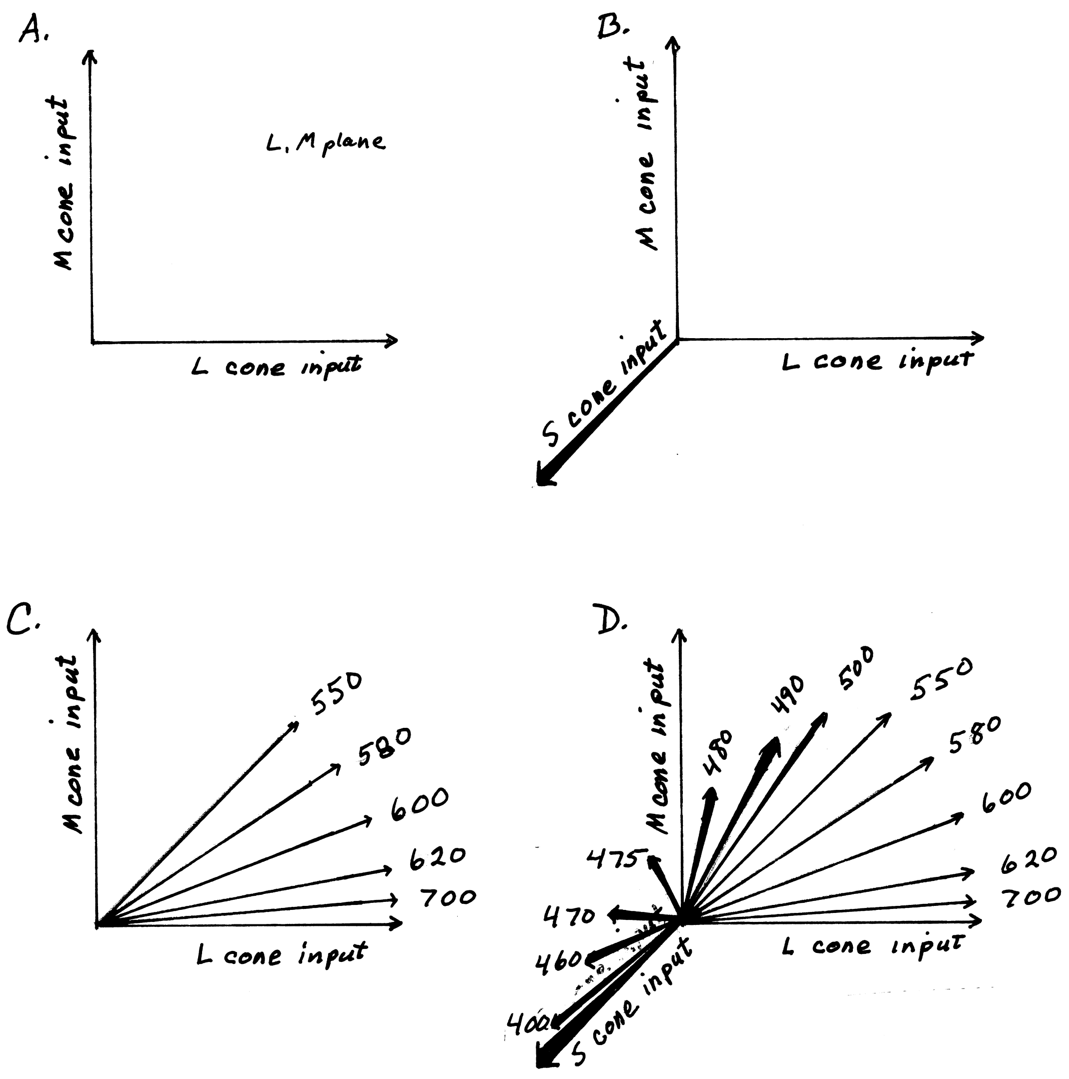


Figure 12.5: The Boynton model. A: The (non-opponent) L + M channel. B: The L - M channel. C: The S - (L + M) channel. The sketches in the left half of the figure show the cone inputs to each channel. The triangles represent cones, and the circles represent sites at which signals from two or more cone types are combined. The small numbers near the combining sites are working approximations of the cone input weights. The middle column shows the names of the three channels, in terms of their cone inputs. The right hand column shows the working cone input weights adopted in this chapter. [Sketches modified from Boynton (1979), Fig. 7.3, p. 212.]

The locus of points that represent individual wavelengths of light is called the *spectrum locus*.[4] In Figure 12.6, the spectrum locus is shown with each ray colored in with the perceived color typical of that wavelength.

Due to the overlapping spectral sensitivity curves of the cones, combinations of cone inputs outside of the spectrum locus do not occur. Desaturated lights, which can be made by mixing spectral lights, will plot to locations inside the spectrum locus. ”White” light will occupy a ray within the spectrum locus, and the ”purples” will plot on the flat surface connecting its two ends. [The mathematically sophisticated will figure out that in cone input space, color mixture can be represented by vector addition.]

Notice that cone input space is another of those interesting hybrids between physics and physiology. Physical stimuli are represented, but in terms of their effectiveness for a physiological system.

Cone input space also captures the essence of our abilities (and failures) to discriminate among lights of different wavelength compositions and intensities. In particular, it nails down the concept of metamers. As discussed earlier, metamers are two (or more) lights that differ in wavelength composition but are perceptually indiscriminable. Trichromatic theory says that they are indiscriminable because they lead to equal quantal absorptions in each of the three cone types ($L_A = L_B$, $M_A = M_B$, and $S_A = S_B$, where A and B are two metameric lights). By definition, both of these lights will plot to the same point in cone input space. They look identical because they are rendered identical at the stage of transduction from quanta to neural signals, and in principle, nothing later in the visual system can sort them back out. In contrast, the fact that two lights that are not metamers can be discriminated is captured by the fact that they plot to sufficiently separated locations in cone input space.

12.3.1 Null planes: Silencing individual cone types

The next important set of ideas, concerning the concepts of *null lines* and *null planes*, is shown in Figure 12.7. In the L,M plane (Figure 12.7A) a line perpendicular to the L axis represents a locus of points for which the L cone quantum catch is constant at a certain value k. This line represents an equivalence class – call it a *null line* – for the L cones. Similarly, in L,M,S cone input space, imagine a plane perpendicular (normal) to the L axis, as shown in Figure 12.7B. This plane represents a larger set of points for which L is constant at the value k. For every point represented in this plane, the L cone quantum catch is identical. In other words, this plane represents an equivalence class – call it a *null plane* – for the L cones. The same is true for any other line or plane perpendicular to the L axis – different lines or planes would represent different values of k. Similarly, M cones have null lines in the L,M plane, and both M and S cones have null planes in three-dimensional L,M,S space.

Now, we have argued before that all neurons must have equivalence classes, but usually we don’t know what they are. The equivalence class of a photoreceptor is special because it is determined solely by the spectral absorption characteristics of the photoreceptor. Since the cone spectra are known, the members of their equivalence classes can be calculated, and a stimulus can be constructed from members of an equivalence class for that cone. Such stimuli can also be called a

[4] Of course, representations like this one break our rules of separation of physiology and perception. Cone input space represents quantum catches, not perceived colors. However, the mixed representation is useful for getting oriented.

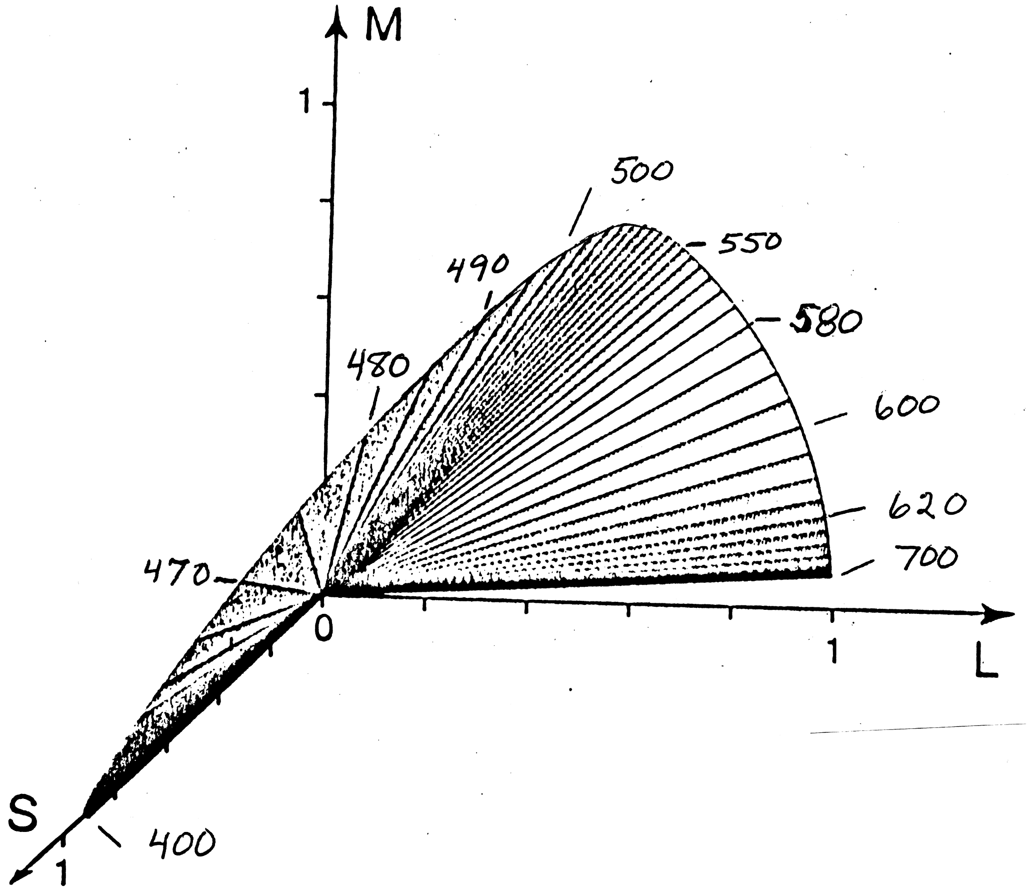


Figure 12.6: COLOR PLATE. L,M,S cone space, with the rays for the different wavelengths shown in their typical perceived colors. The curved surface enclosed by all the rays is called the spectrum locus. The "purple" line connects the two ends of the spectrum locus. Rays for "white" and other desaturated colors fall within the curved surface defined by the spectrum locus and the purple line. [Based on Rodieck, 1998, p. 424.]

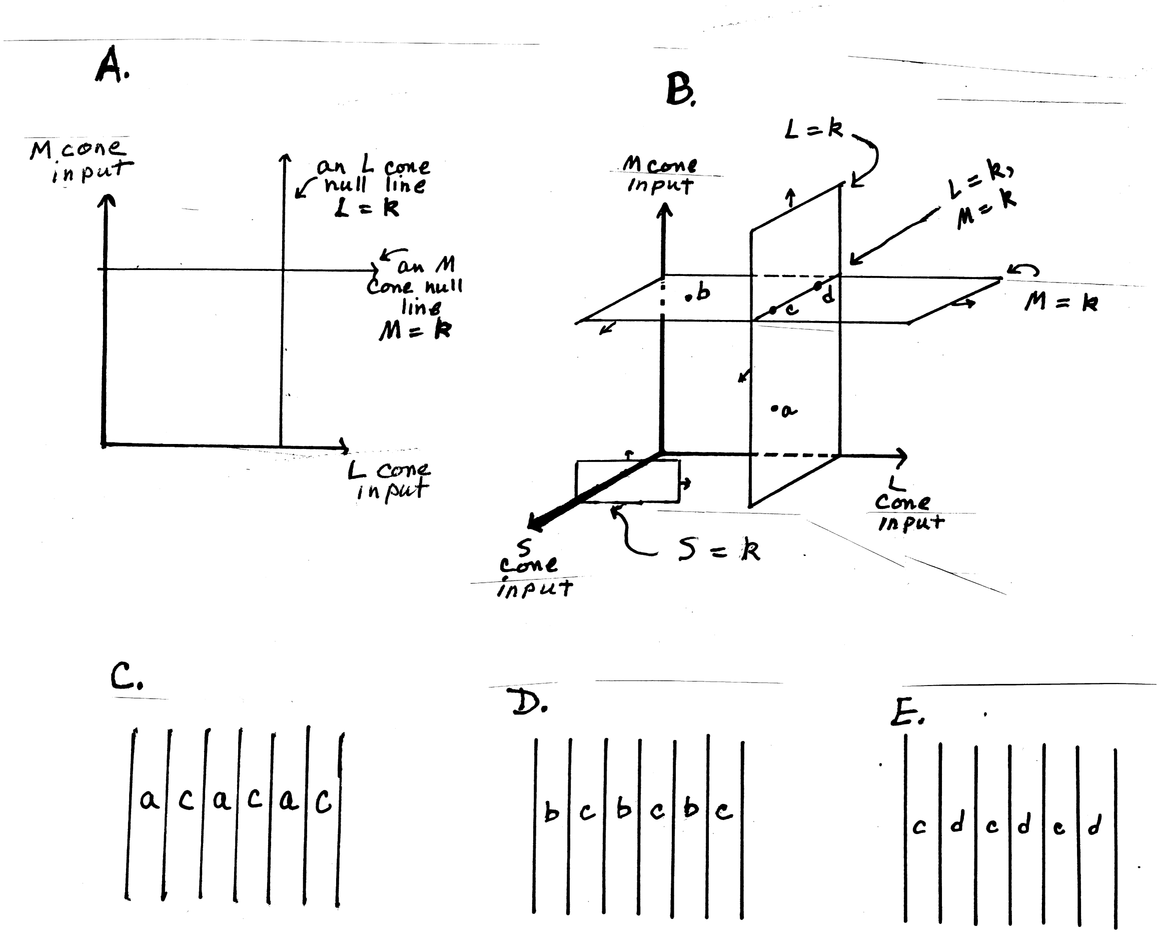


Figure 12.7: Null lines and null planes in cone input space. A: null lines in L,M cone input space. B: null planes and joint null lines in L,M,S cone input space. C: A grating made from two stimuli, a and c, selected from the same L cone null plane. The L cones are silenced (L = k). D: A grating made from two stimuli, b and c, selected from the same M cone null plane. The M cones are silenced (M = k). E. A grating made from two stimuli, c and d, selected from a joint null line for the L and M cones. The S cone input is isolated by this stimulus.

silent substitution set for that cone, because substitution of any one of these stimuli for another should make no change in the signal from cones of that particular type. The exchanging of one for another is called *silent substitution* or *exchange stimulation.*

The concept of silencing individual cone types suggests whole classes of psychophysical experiments, similar to those we discussed in relation to rods vs. cones in Chapter 11. One such set of experiments is shown in Figure 12.7C - E. If we make a spatial pattern (such as the square-wave grating in Figure 12.7C) from two stimuli such as a and c, selected from within a single L cone null plane, the entire pattern will make a constant signal from the L cones; the L cones will be silenced, and will not be able to signal the presence of the pattern! Therefore, if you as a psychophysical subject can detect this pattern, you must be using the signals originating from your M and/or S cones to do so. Similarly, a pattern made from stimuli b and c will silence the M cones, and a similar argument holds for the existence of patterns that silence the S cones.

12.3.2 Joint null lines: Isolating individual cone types

We now come to the related but converse concept of *single-cone-isolating stimuli.* In L,M,S cone input space, the intersection of two null planes – a *joint null line* for two cone types – has a special property. Figure 12.7B shows the intersection of an L cone null plane and an M cone null plane; they define a joint null line. A grating made up of stimuli selected from this line (such as *c* and *d* from Figure 12.7B, shown in Figure 12.7E) will silence *both* the L cones and the M cones. The only signals that vary over space will be the signals from the S cones. Such a pattern can be called an *S-cone-isolating stimulus.* The logic is that if you can see the pattern, you must be seeing it via signals initiated by your S cones. And similar arguments hold for the other two pairings of cone types.[5] [Stop and work out lines that show an L-cone-isolating stimulus and an M-cone-isolating stimulus.]

Note that in this scheme an S-cone-isolating stimulus does not leave the L and M cones with zero quanta caught; instead, it holds the number of quanta caught by each of these cone types *constant* across the spatial or temporal pattern used. These cone types are catching quanta, but cannot help us detect the stimulus *pattern.* This is the meaning of the term S-cone-isolating stimulus; and the same for L- and M-cone-isolating stimuli.

This conceptual framework is important because it allows us to specify whole new sets of diabolically clever stimuli, custom designed to isolate each individual type of photoreceptor or pair of photoreceptors and probe our visual capabilities when visual inputs are so confined. (Thought question – How would you measure grating acuity for signals initiated solely by the L cones? The M cones? The S cones? How about a CSF, or motion perception, or pattern recognition, for patterns initiated by each cone type?)

12.3.3 A little light algebra

We now step back to review some simple concepts from algebra and geometry. Once we review them, we will import them directly into our modelling. We will first treat two-dimensional spaces,

[5] Notice that, although the terms *silencing* and *isolating* sound vaguely similar, the ideas are converses. If you *silence* a channel, you render it incapable of signaling a pattern, and you can study the joint capabilities of the other two. If you silence two channels, you *isolate* the third, and you can study its capabilities when it alone can signal the pattern. For DT, silencing a cone type fits intuitively with selecting stimuli from its null plane. Isolation, then, must be the opposite concept – selecting stimuli from along a single line.

and then generalize to three-dimensional spaces.

In a two-dimensional space with axes x and y, the general equation for a line is:

$$y = mx + k$$

where m is the slope, and k is the y intercept.

If the line passes through the origin, k = 0, this equation reduces to:

$$y = mx.$$

Let m = -a/b. Then y = -ax/b, or

$$ax + by = 0$$

is a second version of the general equation for a line.

Notice that his equation is of the same form as the equations for the postreceptoral channels given earlier in the chapter (but with only two terms so far). Substituting symbols, the constants *a* and *b* become the cone input weights (say w_1 and w_2), and the variables x and y become the quantum catches, (say) in the L and M cones. The signal in the L + M channel is:

$$Ch_1 = w_1 L + w_2 M$$

and the null line for that channel is given by:

$$w_1 L + w_2 M = 0$$

That is, this general equation describes the null line through the origin for *any* postreceptoral channel that signals a weighted sum of L and m cone inputs. As w_1 and w_2 vary, the particular line through the origin will vary, as shown in Figure 12.8A.

Similarly (although less familiarly), there is a general equation for planes through the origin in a three-dimensional space with axes x, y, and z. It is:

$$ax + by + cz = 0$$

where a, b, and c correspond to the cone input weights.

Or, sustituting symbols,

$$w_1 L + w_2 M + w_3 S = 0$$

as the general equation for a plane through the origin in cone input space.

As in the two-dimensional case, the various cone input weights can be either positive or negative. As the weights vary, the particular plane through the origin will vary, as shown schematically in Figure 12.8B. In short, we have derived an expression for the null plane for *any* postreceptoral channel whose input is a linear weighted sum of inputs from the L, M and S cones.

This expression embodies the fundamental reason why three dimensional color spaces help us in understanding color codes and color code transformations. In mathematical terms, this expression means that by choosing the weights properly, we can write an expression for *any* plane in three-dimensional space. Each set of weights defines a unique plane, or conversely, each plane defines a unique set of weights.

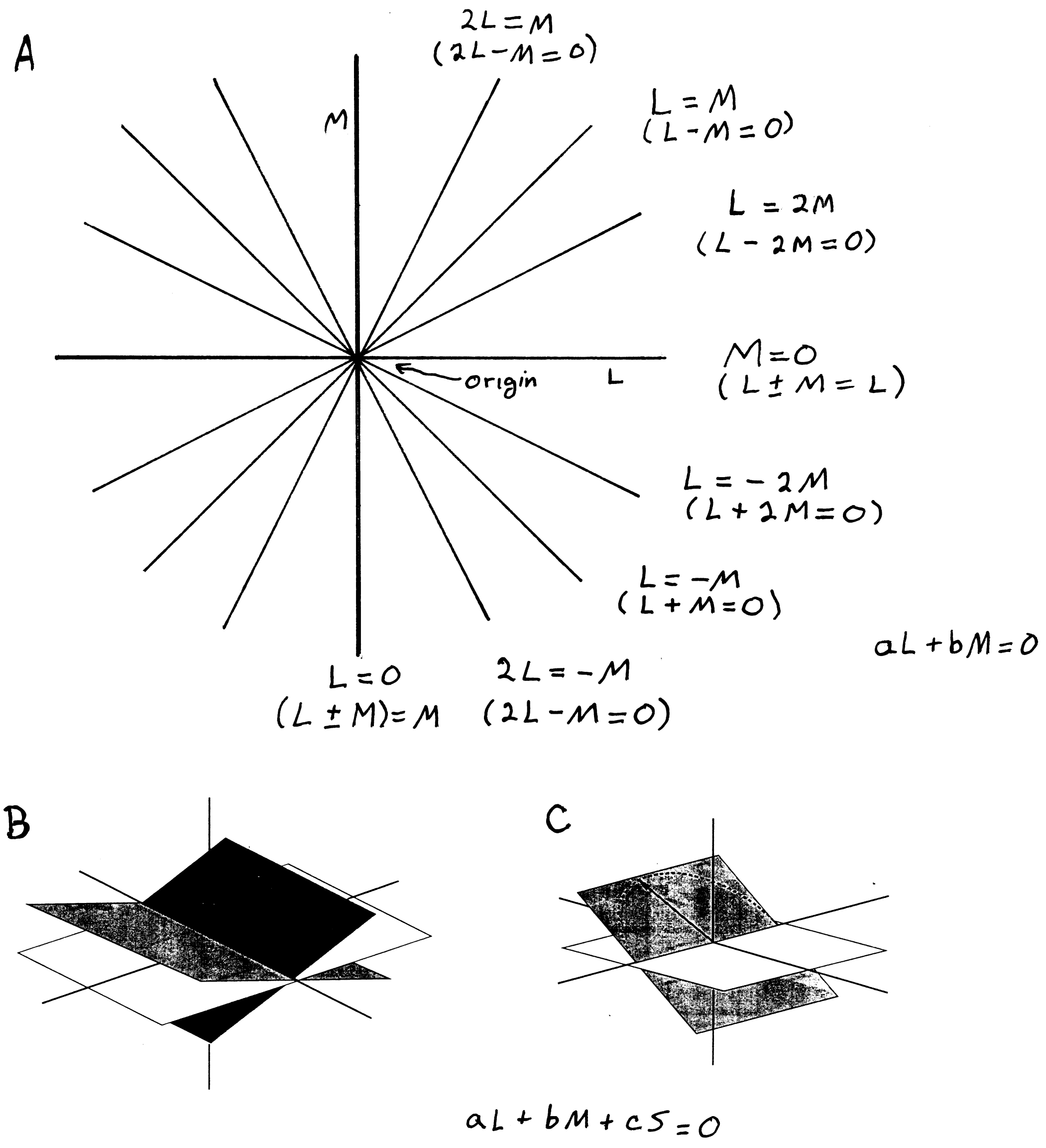


Figure 12.8: Lines and planes through the origin. A: The equations for several different lines through the origin in the L,M plane. Each line can be expressed as aL + b M = 0. That is, the equation for each line can be expressed as a weighted sum of L and M cone inputs, where a and b are the weights. B, C: A sample of planes through the origin. Each plane can be expressed as aL + bM + cS = 0. That is, the equation for each plane can be expressed as a weighted sum of L, M and S cone inputs, where a, b, and c are the weights.

12.3.4 Null planes as the chromatic signatures of neurons

In physiological terms, these equations mean that *any post-receptoral neuron whose inputs are a linear combination of signals from the three cone types will have a unique null plane in three-dimensional color space.* Neurons with different cone input weights will have different null planes, and vice versa. In fact, for any choice of cone input weights for any linear color model, we could calculate the predicted null planes for those cone input weights.

This realization has profound implications for research in postreceptoral processing in color vision. Why? Because it provides us with an important paradigm for exploring postreceptoral color codes at the physiological level. Suppose that a particular theorist proposes a particular linear postreceptoral code – that is, a set of cone input weights for three kinds of putative postreceptoral cells. Given these weights, we can predict the location of each cells null plane in cone input space. Or conversely, given a measurement of the actual null plane of a postreceptoral neuron, we can calculate the cone inputs to that neuron, decode the postreceptoral color code, and reject our theorist's putative color code or let it stand. This line of experimentation can be called *null plane analysis*, and it will be followed up in Chapter 13.

12.4 Cone opponent space

So far, we have seen that cone input space can reveal interesting characteristics of color coding, most particularly the existence of cone-isolating and cone silencing stimuli. But there are two related problems with cone input space as a general way to represent color models. First, cone quantum catches are zero only at absolute threshold. But when we are dealing with cone vision, we are usually dealing with stimulus fields that have average luminances well above zero, with the experimental stimuli consisting of both increases and decreases around the average value. Second and more importantly, most theorists believe that postreceptoral color codes will show opponent coding; that is, some of the cone input weights will be negative. But the axes of cone input space are defined as L, M, and S cone quantum catches, and negative quantum catches do not occur.

12.4.1 Shifting the origin to allow for opponent coding

How shall we represent the activity of opponent channels in color space? As shown in Figure 12.8, we shift the origin to create a second type of color space: *cone opponent space.* We choose some above-zero values of all three cone output signals as the new origin, so that we can represent both increases and decreases in activity in each channel.

In the L, M plane, we can shift the origin out along some ray (for example the 580 nm ray), as shown in Figure 12.9A. In the L, M, S case we will choose to shift the origin out of the L,M plane along the "white" ray, as shown in Figure 12.9B. The values of the L, M, and S cone signals at the new origin can be called L_0, M_0, and S_0 respectively. Although any ray would be OK, the shift of the origin out along the "white" ray is particularly attractive on perceptual grounds, because it allows us to center cone opponent color space on a stimulus that is typically perceived as achromatic.[6] Hering would be pleased, because the space that began with cone inputs has taken a step toward resembling Hering's perceptual color space, which represents an achromatic perception (white) at its origin.

[6] The new origin is usually called the "white" point, but occasionally the "grey" point.

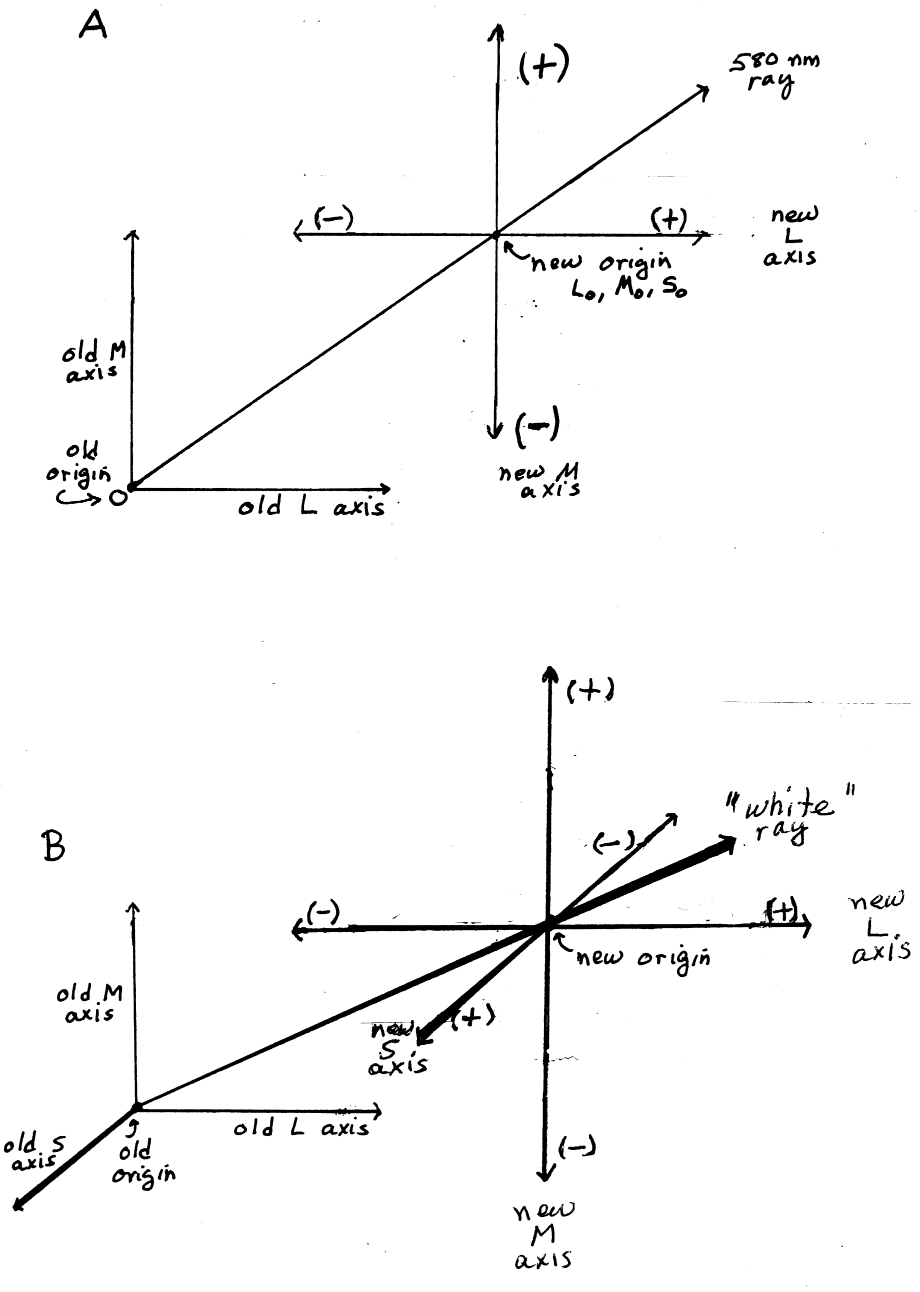


Figure 12.9: Cone opponent space. In order to represent stimuli with a space average luminance above zero, the origin of cone input space is shifted to above-zero values of the cone quantum catches. The new origin is labelled L_0, M_0, S_0. A: Shift of the origin of L,M space out along the 580 nm ray. B: Shift of the origin of L,M,S space out of the page along the "white" ray. Both increases (+) and decreases (-) of each cone input signal can now be represented.

12.5 Postreceptoral channels space

We now come to a third and final kind of three-dimensional color space: *postreceptoral channels space*. In cone input space the L, M,and S cone quantum catches lie on the main, or *cardinal*, axes of the space. In consequence, the corresponding stimuli for silencing or for isolating the L, M, and S cones are *explicit* – easy to find and think about. At the same time, we know from the algebra developed above that null planes for any chosen set of linear postreceptoral channels must also lie somewhere in this space. But the null lines do not coincide with the cardinal axes of cone input space – they are *implicit*, and hard to think about – and the locations of the null planes are similarly obscure, so the silencing and isolating stimuli for the postreceptoral channels are hard to think about.

Is there a way to transform cone opponent space into a space that makes a specific postreceptoral channels model explicit? The answer is yes. We can do so by *rotating* and *skewing*[7] the axes of the space. We move the L, M and S cone signals off the x, y, and z axes, and move the axes representing our chosen postreceptoral channels code onto them.

12.5.1 The Boynton model in a postreceptoral channels space

Figure 12.10 shows a three-dimensional postreceptoral channels space with its isolation axes labelled in accord with the Boynton model. This three-dimensional representation of the Boynton model were first introduced by Derrington, Krauskopf, and Lennie (1984), and the space is sometimes called DKL space after the three authors.

In DKL space, the origin represents the average luminance of the stimuli in use in a particular experiment. The vertical axis represents the ”white” ray. Variations in the luminance of a ”white” light above and below the average luminance are represented along the vertical axis, which is labelled the L + M axis. Importantly, the null plane for the L + M axis is called the *isoluminant* or *equiluminant plane*, and singled out for particular attention. Stimuli that are equal in luminance, but differ in wavelength composition, all plot to this plane. It represents the locus of purely chromatic stimulus variations, and has been the subject of much research interest, as described below. The isoluminant plane with its typically perceived colors is shown in Figure 13.3A. The x and y axes are used to represent the signals in the two opponent channels – the L - M and S - (L + M) channels – respectively.

12.5.2 Why did you teach us this?

Once again – why are these transformations of color space interesting? For three interrelated reasons. First and most fundamentally, color spaces are logically neat. Geometry gets put to use. Different geometrical spaces make different color vision models explicit – they are an aid to thinking within the terms of a particular color model. And color models represent one of the forefronts of precision in visual science.

Second, these representations lead to a pair of remarkable new research paradigms. Physiologically, we can search for the null planes of visual neurons, and from them decode their cone inputs. And psychophysically, remember that a postreceptoral channel is a putative neuron that lies deep

[7] To *rotate* the axes, think of rotating the white ray around the origin until it is vertical – lies on the y axis – dragging the other axes along in a rigid 3-D configuration. To *skew* the other two axes, think of rotating each of them independently, to place two newly chosen axes on the x and z axes of the new space.

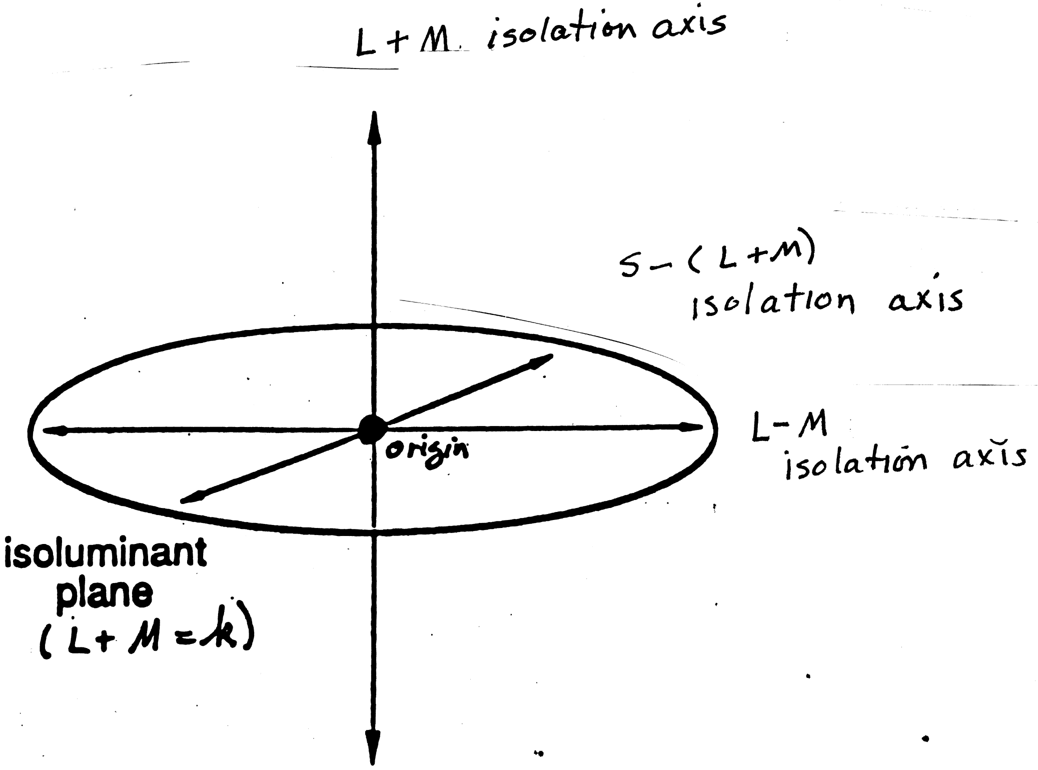


Figure 12.10: The Boynton model represented in a postreceptoral channels space. The vertical axis is the "white" ray, and it represents the L + M isolation axis within the Boynton model. The two horizontal axes represent the two chromatic channels: the L - M and the S - (L + M) isolation axes. The plane perpendicular to the "white" ray is the isoluminant plane, and it represents the locus of purely chromatic stimulus variations.

within the retina or visual system. Yet (within the context of a given model) we know how to custom design a stimulus to silence it, or conversely, to activate it in isolation. So we can ask, what is vision like when it is driven by inputs from each individual cone type, or from each pair of cone types?

And third, more practically, in the modern color vision literature stimuli are often chosen, and data are often presented, within the framework of one or another of these linear three-channel models and stimulus paradigms. The experiments discussed in Chapter 13 and 14 make sense only given an understanding of the concept of channel-isolating stimuli in three-dimensional postreceptoral channels space.

What about the generality of this line of thinking? Remember the mathematical argument from trichromacy, which led to the realization that there must be three and only three cone types, but did not determine their spectral sensitivities. After a century of suspense, the true cone spectral sensitivities were finally defined empirically in the 1980s. Similarly in the case of opponency, the existence of a linear three-channel model does not determine which three postreceptoral channels we actually have. Psychophysical and physiological experimentation are required if we are to find out which code or codes actually describe the workings of the human visual system at its various levels. Perhaps more definitive answers, at least concerning the form of the early opponent code, will emerge within the next few years.

12.6 Modulations through the origin in three-dimensional color space

Now we come to the final set of concepts in our exploration of color spaces. Suppose we wanted to know what vision is like when we can use only a single postreceptoral channel, or only two of the three postreceptoral channels? How do we custom design stimuli to silence each channel, or conversely to isolate each channel represented in this space? The approach is exactly analogous to our earlier use of square wave gratings to silence, or to isolate, the different cone types (Figure 12.7), but now we use stimuli selected from particluar planes and lines in postreceptoral channels space.

To illustrate our point, we will switch briefly to a generic post-receptor channels space, with putative channels 1, 2, and 3. To silence Channel 1 – to make a grating detectable only by Channels 2 and 3 – we make the grating from two stimuli selected from the null plane of Channel 1. Conversely, to isolate Channel 1 – to make a grating detectable only through Channel 1 – we make the grating from two stimuli selected from along the Channel 1 isolation axis. And the same logic governs the silencing or the isolation of each of the other two channels. Moreover, any stimulus can be used – not just a square wave grating. So we can ask, how good is the grating acuity, or contrast sensitivity, or the perception of motion, or the recognition of faces, or any other perceptual function, when it is carried out with signals that traverse any particular postreceptoral channel or pair of channels?

Now let's return to sinusoidal modulation and contrast sensitivity functions (CSFs). In Chapter 5 we defined the concept of luminance-modulated sinusoidal gratings and variations in their contrast (Figure 5.1), and we introduced the measurement of contrast sensitivity functions (CSFs). We now generalize these concepts to modulation along any axis in postreceptoral channels space.

Figure 12.11 shows the strategy schematically, in a generic postreceptoral channels space. Sup-

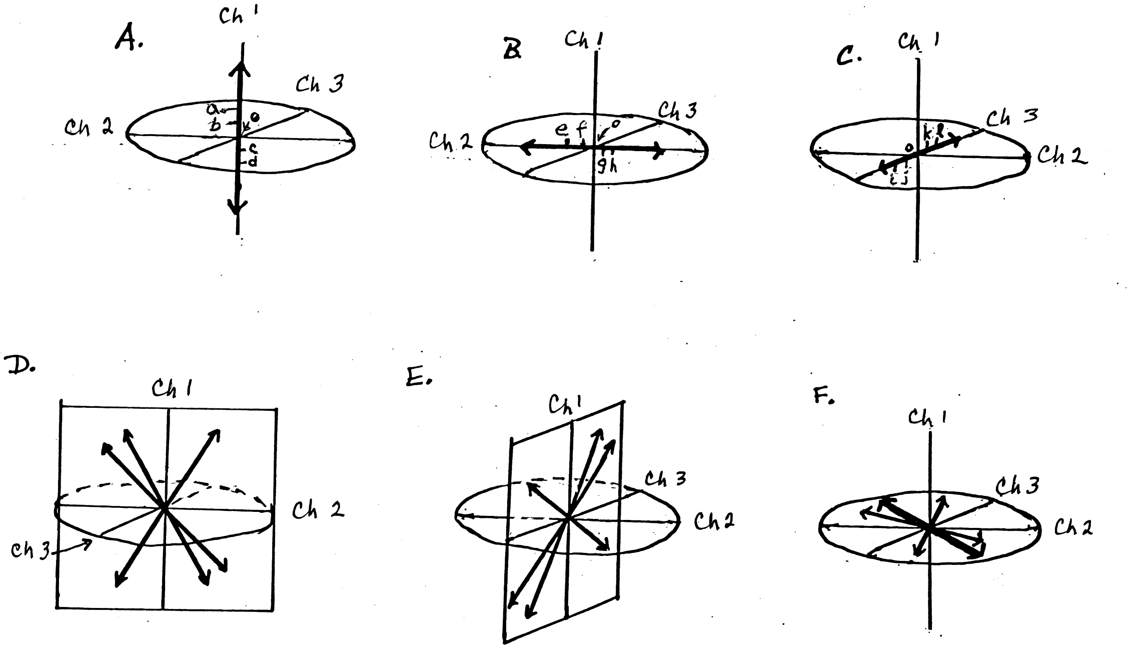


Figure 12.11: Modulations in a generic three dimensional channels space. A. Modulations along the Channel 1 isolation axis – o, a, b, a, o, c, d, c, o, etc. B. Modulations along the Channel 2 isolation axis – o, e, f, e, o, g, h, g, o, etc. C. Modulations along the Channel 3 isolation axis – o, i, j, i, o, k, l, k, o, etc. D. Joint modulations of Channels 1 and 2. E. Joint modulations of channels 1 and 3. F. Joint modulations of Channels 2 and 3. Modulation along any axis that is outside all three null planes will provide modulation to all three postreceptoral channels.

pose we want to modulate Channel 1 in isolation (Figure 12.11A). We make a spatially sinusoidal grating from stimuli selected from along the Channel 1 isolation axis – a continuous spatial variation represented by stimuli o, b, a, b, o, c, d, c, o, and so on. To modulate Channel 2 in isolation (Figure 12.11B), we make a grating from stimuli o, f, e, f, o ,g, h, g, o and so on. To modulate channel 3 is isolation (Figure 12.11C), we make a grating from stimuli o, j, i, j, o, k, l, k ,o and so forth. These modulations are represented by the double-headed arrows along the axes. To increase or decrease the *contrast* of the stimulus, we increase or decrease the excursion – lengthen or shorten the arrow – along the chosen axis.

Similarly, we could choose to silence any one channel and modulate the other two, by modulating along any axis through the origin confined to the null plane of the channel we wished to silence. Modulations of this kind are shown in Figure 12.11D, E, and F. And finally, modulation along any axis outside of all three null planes provides a modulation of all three channels simultaneously. Paradigms in which these stimuli are used will be introduced in Chapters 13 and 14.

Finally, to return to our concrete example, how can we we custom design stimuli to silence, or to isolate, the signal in each individual postreceptoral channel of the Boynton model? At this point it's simple. To silence the L + M channel, we modulate the stimulus within the isoluminant plane. To isolate the L + M channel, we modulate the stimulus along the L + M axis. Similar logic holds for silencing or isolating either of the chromatic channels. So we now have the stimulus paradigm that allows us to ask to ask, what is vision like when we are allowed to use only our L + M, or L - M, or S - (L + M) channel, or any chosen pair, or all three of them in combination?

What do such stimuli look like? The isoluminant plane of DKL space is represented in terms of its typical perceived colors in Figure 12.12. Sinusoidal modulations along the L + M isolation axis (Figure 12.12B) are the same as the stimuli we used to define sinusoidal gratings in Chapter 5. They typically look achromatic – black and white. Modulations along the L - M isolation axis (Figure 12.12C) turn out to look reasonably similar to red/green modulations, although not exactly so. And modulations along the S - (L + M) isolation axis (Figure 12.12D) look like modulations between yellow/green and violet, and not between yellow and blue. The stimuli in Figure 12.12, then, summarize again the most fundamental difference between the neo-Heringian model and the early opponent model. Hering's blue/yellow channel and the S - (L + M) channel in the Boynton code are physically and perceptually very different.

As of the early 21st century, the Boynton model is the commonly accepted working model of early color processing, and the faithful believe that we will find neurons that instantiate the Boynton model in early visual processing. The Heringian faithful would also believe that we will find neurons that instantiate the Hering code at a higher visual processing level.

12.7 The design question: Why opponent coding?

Why do we have opponent coding for color vision? Why might the EDC give us a postreceptoral channels code made with one sum and two differences among cone types? There are two non-mutually-exclusive kinds of answers.

First, at the photoreceptor level, wavelength information is coded in the relative quantum catches among the three cone types. Since the absorption spectra of the L and M cones overlap extensively, it can be argued that for most visual scenes the signals from the L and M cones will be highly correlated; and the same is true for the M and S cone signals. From a computational perspective correlated signals are redundant, and it would be useful to decorrelate them early in

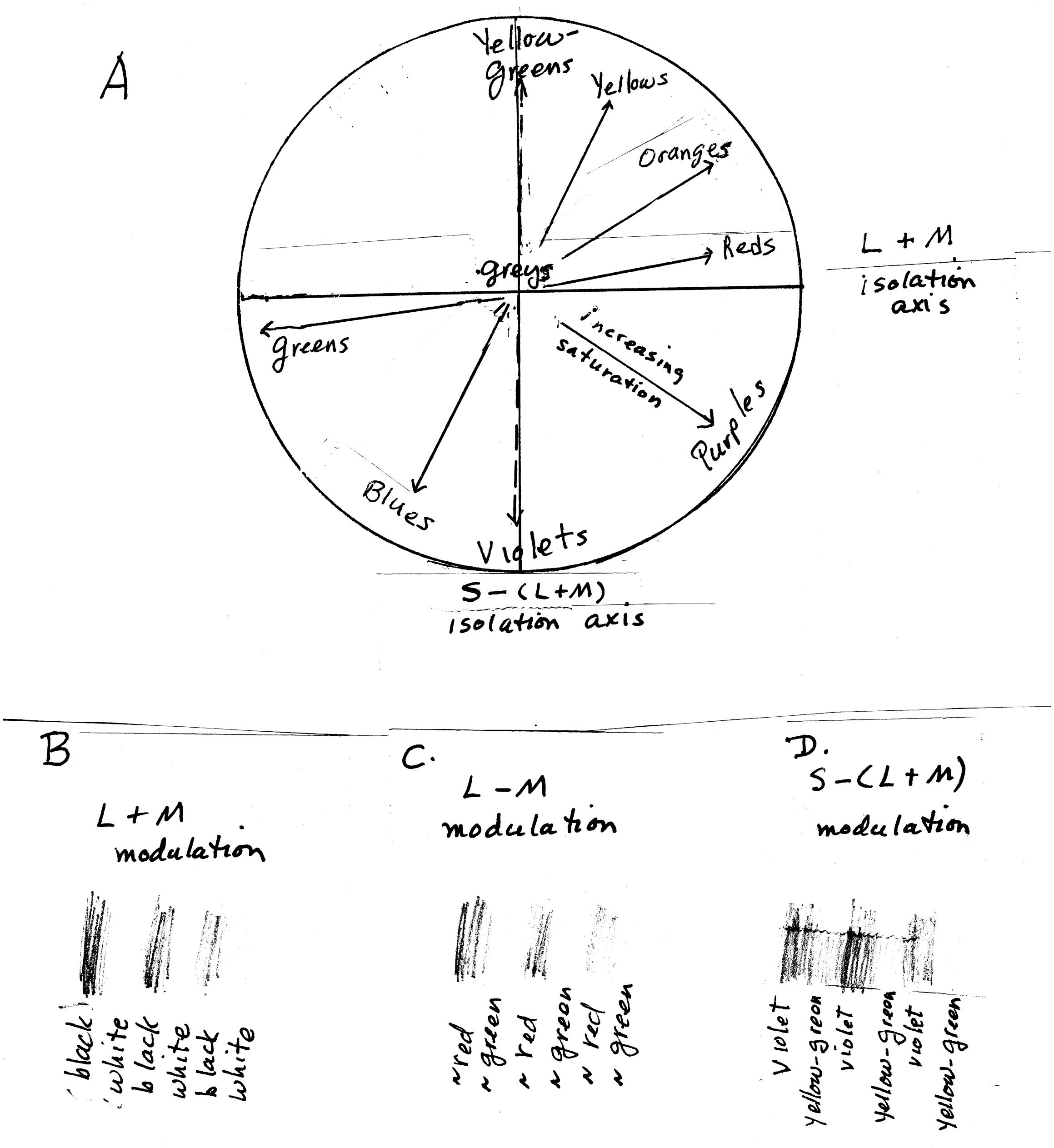


Figure 12.12: COLOR PLATE. The Boynton model and color appearances. A. The appearance of the isoluminant plane in the Boynton model. Notice that the chromatic axes do not coincide with the unique hues. B. The appearance of a sinusoidal grating modulated along the L - M isolation axis. C. The appearance of a sinusoidal grating modulated along the L - M isolation axis. D. The appearance of a sinusoidal grating modulated along the S - (L + M) isolation axis.

visual processing to make the most efficient use of the limited capacity of the set of ganglion cell outputs. It turns out that mathematical transformations that make use of weighted sums and differences of the cone signals can minimize this redundancy. There are many possible schemes, but a variety of codes with one sum and two differences work well.

Second, vision begins with the cone input code: quantum catches in three different cone types (plus the rods). But the EDC chose this code only because catching quanta was the first necessity, and the only reasonable way to catch quanta was with photopigments; not because the triplet of signals S, M, and L was a particularly happy choice of code for other purposes. Once transduction has been accomplished, the EDC is free to change the form of the code by recombining the signals from the three cone types to make a new color code.

In particular, the quantum catch of each cone type confounds two very different features of the world. The first factor is the overall intensity of the light – if the illumination from the sun on an object changes by six orders of magnitude, the quantum catches in each of the three cone types will also change by six orders of magnitude. The second factor is the wavelength compositions of objects in the world. Each object has a *spectral reflectance function*, which describes the percentage of the incident light reflected from the object at each wavelength. We will see later that the spectral reflectance function of an object largely determines its perceived color, and different spectral reflectance functions lead to different *ratios* among the quantal catches of the L, M and S cones. So the effects of illumination levels and spectral reflectance functions are confounded in the cone quantal catches. The perceptual need to sort out changes of illumination from changes in object color probably guided the EDC in coming up with a recoding, or a series of recodings, that moves toward deconfounding these two different sources of quantal catch variations.

12.8 Summary: Linear models of postreceptoral color coding

In this chapter we returned to the problem of color appearance. We began with a reminder of Hering's perceptual observations of unique vs. binary hues and mutually exclusive hue pairs, and a neo-Heringian opponent process model designed to account for them. We argued that subtractive interactions among signals originating in different cone types provide a simple way of modelling unique and mutually exclusive hues, and we introduced some early data from Svaetichin's studies of fish horizontal cells that provided the first evidence of physiologically opponent coding in vertebrate visual systems.

We then discussed three-channel linear models for postreceptoral color processing. We introduced a working model – the Boynton model – that many color scientists find useful at the present time. Current thinking is that the Bounton model describes an "early" opponent code that might well be instantiated, with more or less precision, as early as the retinal output.

We then introduced two- and three-dimensional color spaces. We began with cone input space, and developed the geometrical ideas of null planes and joint null lines. We used the algebraic definition of a plane to argue that any linear postreceptoral channel will have a null plane in three-dimensional cone input space, and that the particular null plane exhibited by a neuron provides the signature for its cone inputs. We then introduced a second color space – cone opponent space – produced by shifting the origin of cone input space to allow a representation of opponent processes. Finally, we introduced postreceptoral channels spaces, and the concepts of silencing, and isolating, activity in particular postreceptoral channels with particular sets of custom-designed stimuli.

If these ideas are correct, these custom-designed stimuli allow us to reach right inside the organism and silence, or isolate, particular types of neurons. They thus allow us to explore what vision is like based on any one channel or pair of channels, as well as all three channels together. We will see such stimuli put to use in physiological and perceptual experiments in the next two chapters.

Chapter 13

Psychophysics of Postreceptoral Color Codes

In Chapter 12 we introduced the concept of postreceptoral processing in color vision – the idea that the initial signals from the L, M, and S cones are recombined early in the visual system into a new color code. We illustrated the concept of recodings at the theoretical level by discussing linear three-channel color models, and the color spaces in which they can be represented, with a change of color code being represented by a change in the axes of a three-dimensional color space.

In the present chapter, we follow through these ideas in the psychophysical realm. The question is, are there ways in which system properties determined from psychophysical experiments can provide any evidence about postreceptoral neural codes? At first glance this challenge seems an impossible one. From a psychophysical perspective the visual system seems like an impenetrable black box, and there might seem to be no way to deduce the wiring without opening the box.

In fact, several psychophysical paradigms have been developed and used, within specific theoretical contexts, to glean information concerning postreceptoral code changes. In the present chapter we will introduce two of these paradigms, and show how they play out in the context of color vision. Later (Chapter xx), we will use the same two paradigms to explore some surprising characteristics of postreceptoral processing in spatial vision.

13.1 Psychophysical paradigms that reveal postreceptoral codes

In her influential book, *Visual Pattern Analysers*, Norma Graham (1989) provided an in-depth analysis of four psychophysical paradigms that provide information about postreceptoral visual processing. We will discuss two of these paradigms: *summation-near-threshold* and *adaptation-near-threshold*. In each case, after a brief general discussion, we will show how the paradigm is applied in the context of color vision.

13.2 Paradigm # 1: Summation-near-threshold

13.2.1 The summation square: Summation, subtraction, or independence?

Suppose you have measured detection thresholds for two stimuli, A and B. (For example, A and B might be test spots of 540 and 650 nm respectively.) Suppose you now present the two stimuli simultaneously, superimposed at the same location. We will call A and B the *component stimuli*, and A +O B [a plus with a circle around it; the superposition of A and B xx] the *compound stimulus.* And suppose you are interested in whether or not the signals SA and SB, set up by stimuli A and B, interact within the visual system, and if so, in what way.

A rather dramatic way to express the possible outcomes of such an experiment is in a *summation square*, as shown in Figure 13.1. In the summation square, the x and y axes represent the intensities of the stimuli A and B. The units on both axes are normalized to detection thresholds, so that 1 unit on the x axis denotes the detection threshold for A, and 1 unit on the y axis denotes the detection threshold for B. With this normalization, the thresholds for A alone and B alone are represented by the solid circles at the points 1,0 and 0,1 respectively.

Now, imagine superimposing components A and B, and measuring thresholds for the compound stimulus. Let's set the normalized intensities of A and B to a fixed ratio – say 1:1. As we vary the intensity of the compound, the stimulus will vary along a ray out from the origin, as shown in Figure A. The subject 's task is to vary the intensity of the compound to find its detection threshold. The experiment is repeated with different fixed ratios between A and B, yielding a set of detection thresholds, represented by the black dots. A line connecting these threshold values can be called a *detection contour*. The question is, what pattern of threshold values – what detection contour – will be traced out for the compound stimulus?

The three solid lines in Figure 13.1B represent detection contours predicted from three very different rules for recombining the signals from A and B. First, if the signals are processed by a channel that exhibits linear *summation*, the threshold for the compound will be reached whenever the sum of the signals, $S_A + S_B$, reaches a value of 1. The higher the intensity of A, the lower the intensity of B that will be needed for the compound to be at threshold. The detection contour predicted by linear summation model is represented by the solid diagonal line that runs from 1, 0 to 0,1, and represents the equation $S_A + S_B = 1$.

Second, what would happen if the signals from the two stimuli were combined with opposite sign, or *subtracted* from each other? The two signals should cancel. The higher the intensity of A, the higher the intensity of B that would be needed for detecting the compound stimulus. Detection contours predicted from a *subtractive* (or *opponent*, or *antagonistic*, or *cancellative*) model are represented by the two parallel diagonal lines that originate from the points at 1,0 and 0,1, and rise outside the summation square with slopes of 1. (The two lines actually represent predictions from two different subtractive rules, S_A - S_B = 1 and S_B - S_A = 1, but we will suppress the difference between the two.) [Figure out the predictions for weighted differences; e.g., 2L - M, 3L - M, etc.]

And third, what would happen if the two stimuli were detected *independently* in a system with two independent detection channels? Neither stimulus should affect the threshold for the other. The compound stimulus would be detected whenever either S_A *or* S_B reaches its individual detection threshold value of 1. So the detection contour predicted by a two channel, independent detection model is a square, defined by the line segments from 0,1 to 1,1 and from 1,0 to 1,1.

In addition to summation, subtraction, and independence, there are options that fall in between. A general class of models used to predict outcomes that fall between the summation and

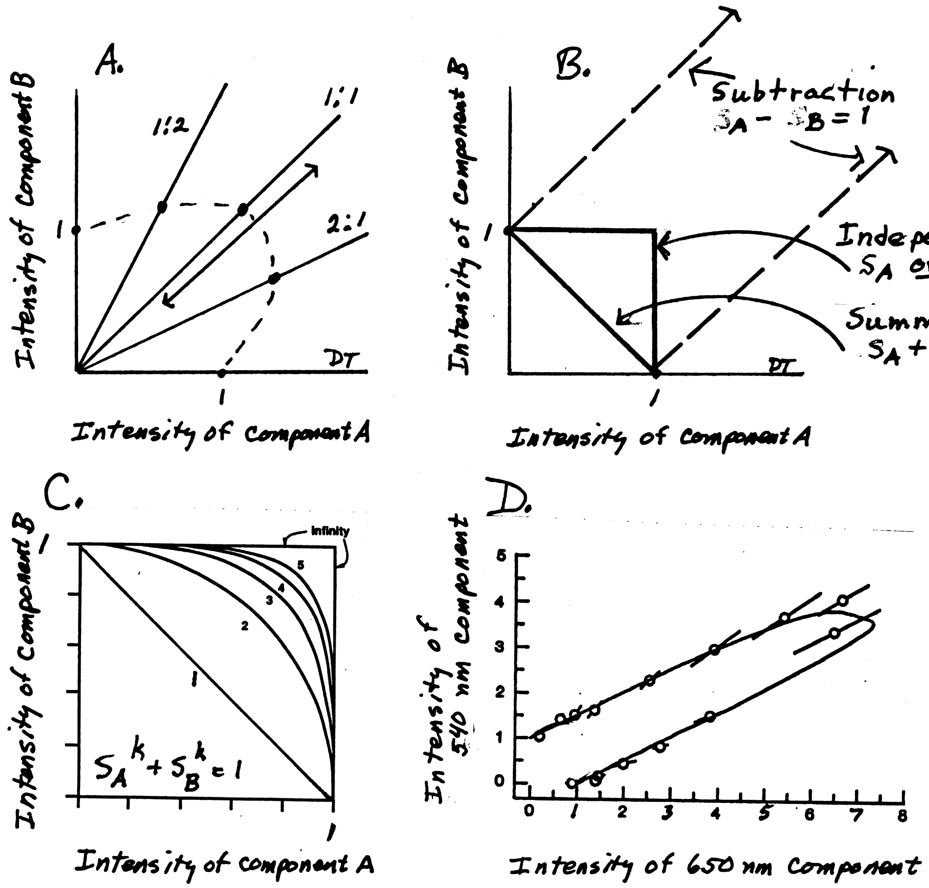


Figure 13.1: Summation squares. The axes represent the intensities of two stimulus components, A and B, normalized so that the detection threshold for each component is set to the value of 1. A: Detection thresholds are measured for various compounds (mixture ratios) of A and B. The detection threshold for each A:B ratio is marked with a dot. The line joining the threshold values is called a detection contour. B. Detection contours predicted from three major recombination rules: summation, subtraction, and independence. C. A set of detection contours predicted from the rule $S_A{}^k + S_B{}^k = 1$, for various values of k. D. Data from a summation-near-threshold experiment using 540 and 650 nm stimulus components. [A,B: DT. C: modified from Graham, 1989, Fig. 4.9, p. 173. D: modified from Thornton and Pugh, 1983, Fig. 2, p. 192.]

independence contours posits that the signals S_A and S_B will be raised to the exponent k before being combined, so that the signal from the compound stimulus is ${S_A}^k + {S_B}^k$. As shown in Figure 13.1C, k = 1 yields the prediction of linear summation, already discussed; k = 2 predicts a quarter circle; and larger exponents yield contours that bow out to approach the independence prediction more and more closely[1]. These mathematical models are not closely tied to physiological models, but they are often used to describe the results of summation-near-threshold experiments.

An important detail used in interpreting summation-near-threshold data is an effect called *probability summation.* Assume we are doing a two-alternative forced-choice experiment, so the subject's percent correct in detecting component A at threshold is 0.75, and similarly for component B. By analogy, imagine we are tossing two biased coins, each with the probability of a head set at 0.75. The probability of detecting the compound corresponds to the probability of detecting either or both of the components – getting at least one head. This probability is 1 minus the probability of getting two tails, or 1 - (.25)(.25) = .94. That is, when the probability of detecting each component is 0.75, the probability of detecting the compound would be greater than 0.75, and to find the threshold for the compound the subject would need to turn down the intensity of the compound below the intensity needed for either component alone. In graphical terms, the effect of probability summation is to round off the corner of the independence prediction, and make it look like a summation contour with a high value of k.

Finally, if the two gratings *facilitated* each others' detection – a low intensity of stimulus A makes stimulus B much more detectable, and vice versa – then the data would bow inward from the linear summation line. And subtractive models with various weightings of different cone signals yield diagonals of various slopes outside the square.

In summary, summation-near-threshold experiments are a powerful paradigm for exploring the recombination of neural signals within the visual system. A major advantage is that, as illustrated in Figure 13.1B, the predictions from the three major recombination rules – summation, subtraction, and independence – differ enough to be distinguished readily with experimental data. Thus, this paradigm could potentially give us an important tool for attacking the question of how signals from the L, M and S cones are recoded into a new postreceptoral code.

Notice that models that posit detection of the signals generated by the stimuli A and B by two independent channels, predict detection contours aligned with the original axes of the summation square. In contrast, models that posit interactions – addition or subtraction of the signals from A and B – predict two different kinds of diagonal detection contours. Thus, by comparing detection thresholds for the two component stimuli vs. the compound, we should be able to determine the combination rules for the signals from the two components. This pattern of predictions – contours aligned with the axes signalling independence, and contours diagonal to the axes signalling interactions – will be with us throughout our treatment of summation squares.

13.2.2 Color codes: Evidence from a summation-near-threshold experiment

In the context of color vision, an interesting example of data taken with the summation-near-threshold paradigm is a study by Thornton and Pugh (1983). In this experiment, the authors first determined detection thresholds for two component stimuli – 650 and 540 nm test spots. They then

[1] The exponent k is sometimes called the *Minkowski metric.* The case of k = 1 is (linear) summation, as described above. The case of k = 2 is sometimes called *Eucidian summation* ($x^2 + y^2 = c$ is the formula for a circle, so k = 2 predicts a quarter circle).

measured detection thresholds for compounds of the two[2]. The test stimuli were spatially blurred and ramped gradually on and off, and all thresholds were measured against a high intensity, 580 nm background.

Thornton and Pugh's data are shown in a summation square format in Figure 13.1D. As before, the axes of the summation square are normalized to the detection thresholds for the two component test lights. The detection contour for the compound stimulus clearly shows subtraction – adding 540 nm light makes the 650 nm stimulus much harder to see than it was when it was presented alone, and vice versa. Because only wavelengths above 540 nm were used, and S cones are virtually unresponsive in this spectral region, we can guess that the subtractive interaction is between L- and M-cone-initiated signals. In sum, the results point strongly to an L - M channel, but they do not directly reveal the cone input weights.

13.3 Theoretical elaboration of summation squares in color space

13.3.1 Detection contours in L,M cone input space

A more sophisticated approach would be to work in a cone input or cone contrast space[3]. Let's go to a two-dimensional L, M cone input space (Figure 12.5). Rather than representing two arbitrary stimuli A and B, the x and y axes will now represent stimuli custom designed to isolate the L and M cones respectively. Therefore – and here is the main point – the different possible threshold contours will correspond, not to different rules for combining arbitrary signals S_A and S_B, but to different rules for combining L-cone- and M-cone-initiated signals. The predicted detection contours are shown in Figure 13.2A. Remarkably, the threshold contours – the predictions for summation, independence,and subtraction – should now directly reveal the combination rules employed in the postreceptoral code.

13.3.2 Detection contours in L,M cone contrast space

Now let's take one more step up in sophistication. Rather than just measuring thresholds for increments of light as Thornton and Pugh did, let's measure modulation thresholds. That is, the stimulus will be a simusoidal modulation through the origin along some particular axis of L, M cone contrast space, as shown in Figure 13.2B. For example, if the stimulus is a temporal modulation, the stimulus will modulate smoothly from, say, a high to a low intensity 'white' light, or a high intensity 'red' to a low intensity 'green', or between isoluminant 'red' and 'green'. The subject's task is to vary the modulation depth along the chosen axis until the modulation is just barely visible; that is, to measure a modulation or contrast threshold. The experiment is repeated on many axes through the origin, to generate a set of contrast thresholds for modulations in many

[2] The experiment was actually done in a slightly different way. Instead of combining 540 and 650 nm lights, Thornton and Pugh measured thresholds for a series of lights of intermediate wavelengths. This substitution is legitimate because we have only two cone types sensitive in the mid- to long-wavelength region of the spectrum. Thus, there exists a combination of 540 and 650 nm lights that is metameric to, say, a 560 nm light; and the same is true for any wavelength between 540 and 650 nm. From the perspective of the photoreceptors (and the rest of the visual system), the experiment can be done equivalently with either a series of wavelengths between 540 and 650, or a series of mixtures of 540 and 650 nm lights.

[3] Problem. The current chapter on color spaces (Chapter xx) leaves out cone contrast space – must reconcile in next draft. xx

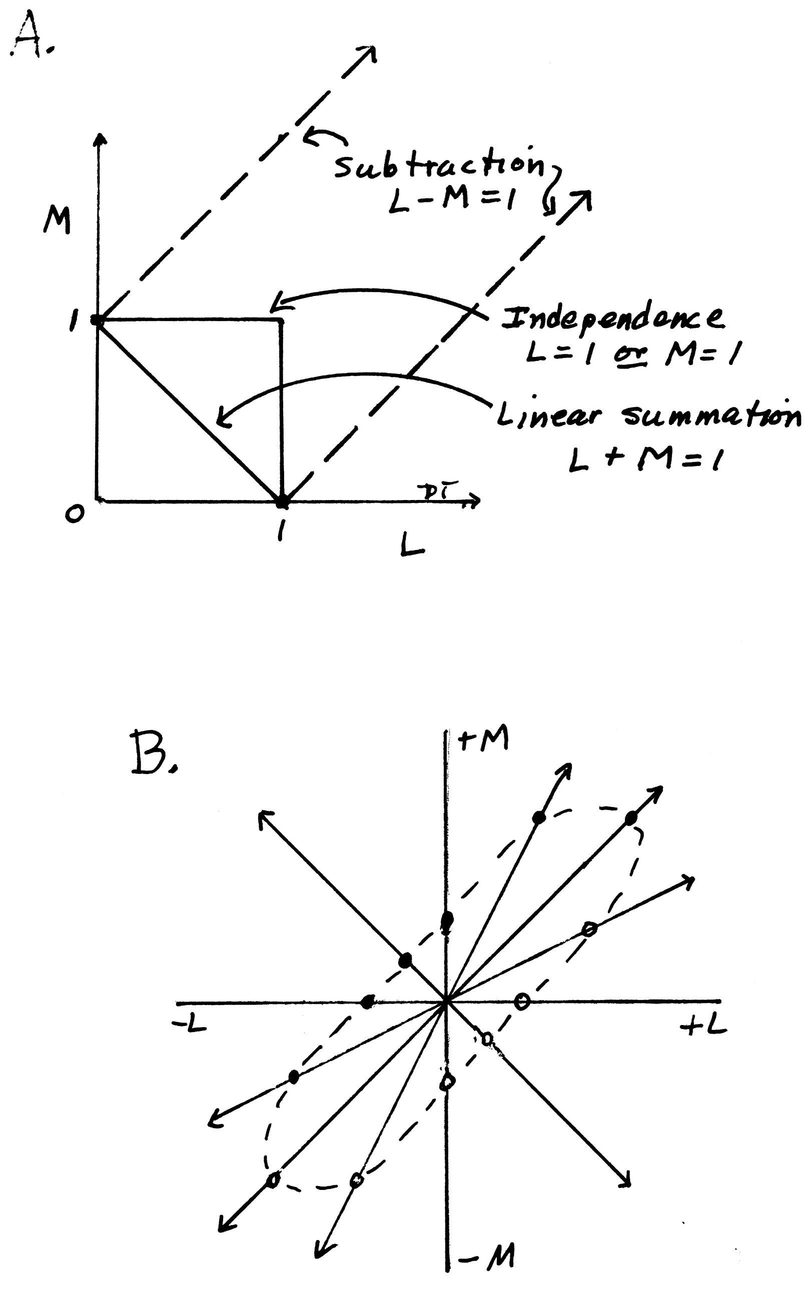


Figure 13.2: Summation squares in two L,M cone spaces. A. Predicted detection contours in L,M cone input space. Predictions are shown for summation, subtraction, and independence. B. Method for measuring modulation thresholds in L,M cone contrast space. Thresholds are measured for modulations along many axes through the origin. Each threshold is indicated twice, by the open and closed circles on each axis. A line connectiong the data points forms a detection contour in cone contrast space.

directions in L, M cone contrast space. By convention, each contrast threshold is recorded with a pair of dots, one on each side of the origin on the axis in question.

What shape of detection contour should be traced out by contrast thresholds in L,M cone contrast space? Different combination rules predict different contours, as shown in Figure 13.3. For a channel that sums L and M cone signals – an L + M channel – the predicted detection contour is given by points that fit the equation L + M = 1. This prediction is a pair of parallel lines passing through the points 1, 0 and 0,1 with negative slope, as shown by the dashed lines in Figure 13.3A. For a channel that subtracts L vs. M signals – an L - M channel – the prediction is given by points that fit the equation L - M = 1. The prediction is a pair of parallel lines passing through the points 1, 0 and 0,1 with positive slope, as shown by the dashed lines in Figure 13.3B. But if (contrary to the Boynton code) the L and M cone signals do not interact, but rather feed into separate, independent channels, the prediction is that the set of modulation thresholds should trace out a square, aligned with the L and M axes, as shown by the dashed lines in Figure 13.3C.

Moreover, these predictions can be extended to cover variations in the sensitivities of the L + M and L - M channels. That is, we expect that as we vary (say) the temporal frequency of the stimulus, the sensitivities of the different channels will vary. The less sensitive the L + M channel (say) is under the conditions tested, the farther apart will be the two parallel lines in Figure 13.3A. This effect is illustrated by the two dotted lines for the case in which the sensitivity of the L + M channel has been reduced by a factor of two, so that a signal of the value 2 is needed for detection. The same argument holds for the L - M channel (Figure 13.3B), and for independence (Figure 13.3C).

What if the system has both an L + M and an L - M channel? This case is shown in Figure 13.3D. The L + M channel will contribute two lines with negative slopes, and the L - M channel will provide two lines with positive slopes. In that case we should see a diagonally oriented, rectangular contour, with its dimensions determined by the sensitivities of the L + M and L - M channels. Figure 13.3D shows two such rectangles, one expected with stimulus conditions in which the L + M channel is more sensitive than the L - M channel, and the other for conditions for which the L - M channel is more sensitive than the L + M channel.

To summarize, extending what we said earlier, the sizes and orientations of detection contours in cone contrast space give us information both about the identities of the postreceptoral channels and about their relative sensitivities. For a system made up of two independent channels L and M, we should always see a square or rectangle that maintains its orientation parallel to the L and M axes. The less sensitive the L and M channels the larger the observed square or rectangle should be. But for a system made up from the joint presence of L + M and L - M channels, the observed contour should be oriented diagonally with respect to the L and M axes. The more sensitive the L + M channel and the less sensitive the L - M channel, the more the detection contours will be elongated parallel to the negative diagonal. In contrast, the less sensitive the L + M channel and the more sensitive the L - M channel, the more they will be elongated parallel to the positive diagonal.

One final but important step. Under the theoretical approach we have been using, the exact orientations of the detection contours will be determined by the cone input weights to the L + M and L - M postreceptoral channels. [Work out some numerical examples.] But notice that if the cone input weights are constant and only the sensitivities of the two channels change with stimulus parameters, the axes of empirical detection contours will only lengthen and shorten, as shown in Figure 13.3D, but never change their orientations. Or reversing the logic, sets of diagonally oriented

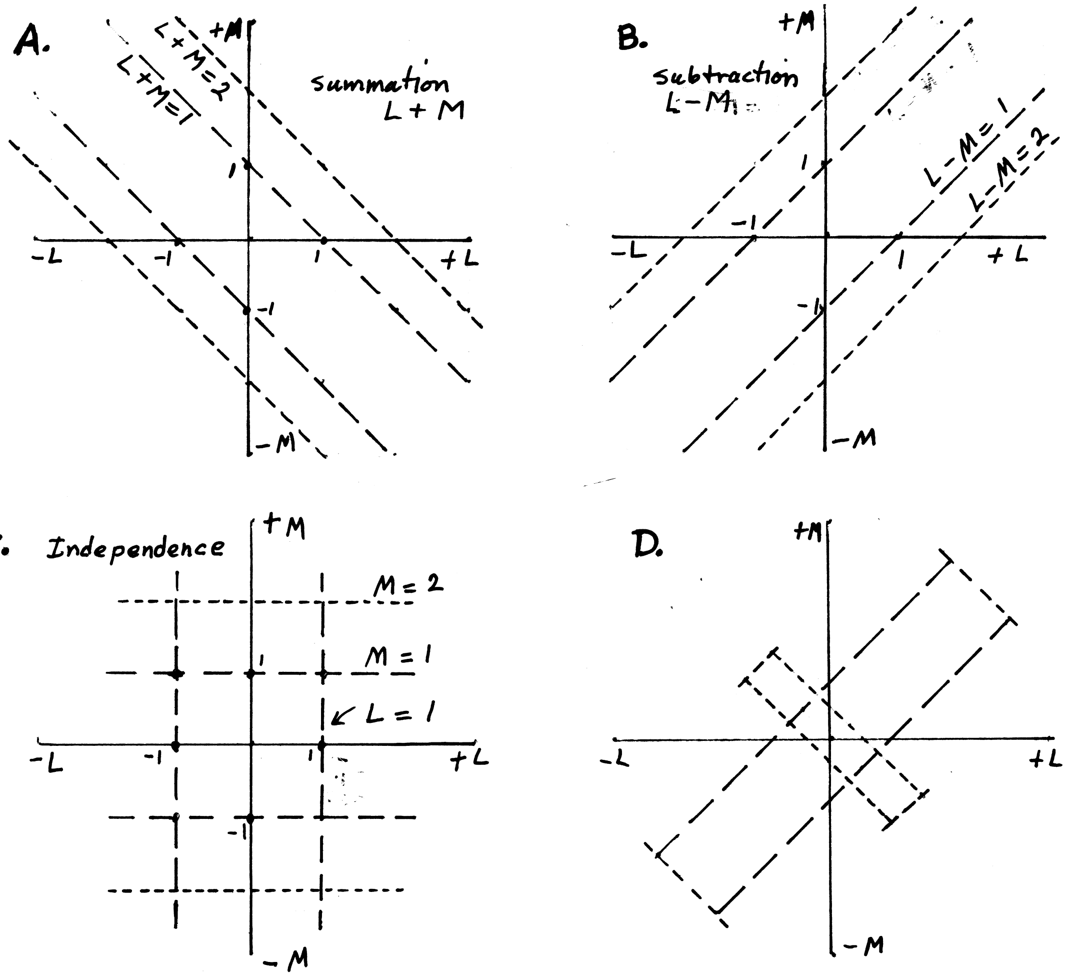


Figure 13.3: Predicted detection contours in L,M cone contrast space. Predictions are shown for summation (A), subtraction (B) and independence (C). The less sensitive the channel, the farther apart the predicted contours, as shown by the dotted lines. D: Predictions for a system composed of the combination of two channels, L + M and L - M. The L + M channel contributes the regions of negative slope, and the L - M channel contributes the regions of positive slope. As the sensitivity of the two channels vary, the contour changes size, and changes orientation from the positive to the negative diagonal.

detection contours that stretch and shrink, but never change their orientation, provide evidence for a set of postreceptoral channels with fixed cone input weights – a consistent postreceptoral code. Sets of detection contours that rotate reveal that the cone input weights vary with stimulus parameters, and no single weighting scheme prevails across stimulus conditions.

13.3.3 Curve fitting: ellipses and hyperellipses

In mathematical models, detection contours like those predicted in Figure 13.3D are often fitted with ellipses, like that shown in Figure 13.4A. The formula for an ellipse is:

$$\frac{x}{a_x} + \frac{y}{a_y} = 1$$

where the variables a$_x$ and a$_y$ represent the lengths of the major and minor axis of the ellipse. In addition, the ellipse has an orientation parameter, ϕ, that describes the angle that the major axis makes with the horizontal axis of the graph.

If the data are too square at the corners to be well fit by an ellipse, hyperellipses can be used, as shown in Figure 13.4B. For a hyperellipse, the absolute values of the two terms are taken, and both terms are raised to the power β before being combined:

$$\left|\frac{x}{a_x}\right|^\beta + \left|\frac{y}{a_y}\right|^\beta$$

In Figure 13.1C, notice that the quarter-circle is a piece of an ellipse (k = 2), and the more squared off contours are pieces of hyperellipses ($2 < k < \infty$). Similar equations can be written for ellipsoids or hyperellipsoids in the three-dimensional case[4].

When ellipses and hyperellipses are used to fit empirical detection contours, the parameters have ready theoretical interpretations. The major and minor axes of the best-fitting ellipse are identified with the isolation axes of two channels of the postreceptoral color code. The major axis will coincide with the isolation axis along which the contrast threshold is highest – the less sensitive channel – under the conditions tested. The minor axis will coincide with the isolation axis of the less sensitive channel. The lengths of the axes will correspond to the sensitivities of the two channels. And a code with constant weights predicts ellipses that change size but not orientation as stimulus parameters are varied.

Thank you for being interested enough to follow through this long section on theory! Sometimes it happens that the theoretical treatment takes a long time to work out, but it's worth it because when the alternate predictions are clear enough, the data discriminate very directly among theories. This is the case with detection contours. In sum, and referring back to Figure 13.3, detection contours aligned with the axes of a particular space reveal independent detection by the channels plotted on those axes, whereas diagonally oriented contours reveal combinations of channels that perform summation and subtraction of the signals represented on the axes.

[4] Since detection contours are often well fit with ellipses, these contours could be called *detection ellipses*. They are more often called *discrimination ellipses*, although this is something of a misnomer because the measured thresholds are detection thresholds – nothing about discrimination is being measured.

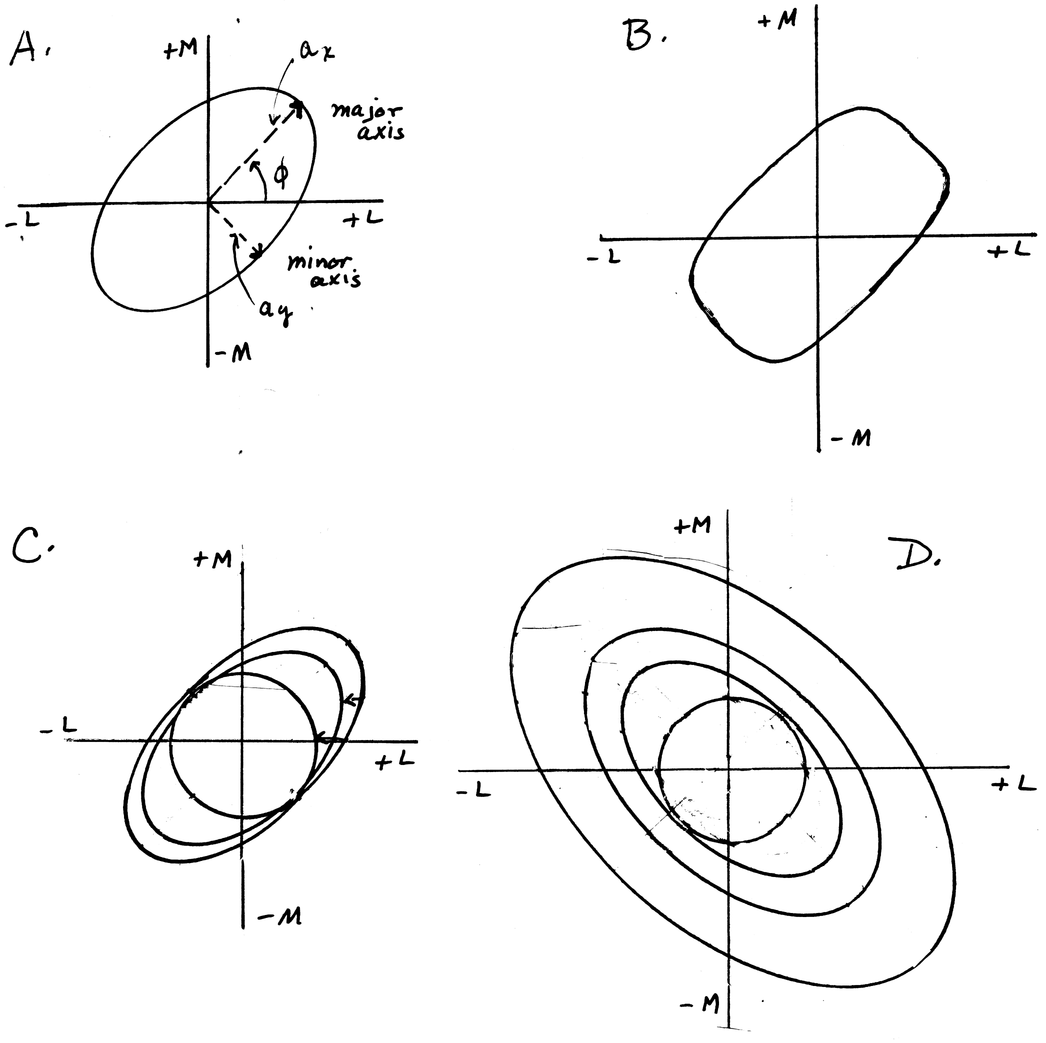


Figure 13.4: Detection contours predicted from (A) an ellipse and (B) a hyperellipse with an exponent of 4. C, D: Sets of detection contours predicted for stimuli with different spatial and temporal frequencies, given fixed cone input weights for the L + M and L - M channels. If the cone input weights are fixed, the detection contours can elongate only along two fixed axes; they cannot rotate. [Must get exact hyperellipse for B.]

13.3.4 Empirical detection contours

We are now ready to look at some empirical detection contours. In 1996, Marcel Sankeralli and Kathy Mullen presented an extensive study of detection contours, which they plotted in two-and three-dimensional cone contrast spaces. They varied the spatial and temporal parameters of the stimuli, in order to see whether or not the orientations remained consistent over these variations. Three combinations of spatial and temporal frequencies were used: 1 cy/deg, 0 Hz; 0.125 cy/deg, 0 Hz, and 1 cy/deg, 24 Hz. (For 0 Hz, the stimulus was ramped slowly on and then off in time.) We will use their data to illustrate the shapes and orientations of real discrimination contours, and interpret them in terms of the models outlined above.

13.3.5 Plots in 2-D and 3-D cone contrast spaces

Detection contours plotted in an L,M cone contrast space are shown in Figure 13.5. For the data in Figure 13.5A, the spatial frequency was 1 cy/deg, and the temporal frequency was 0 Hz. The result is an elongated, ellipse-like contour, oriented approximately along the positive diagonal. These data provide evidence that for these stimuli, modulations in L,M cone contrast space are detected by a combination of two postreceptoral channels – a relatively more sensitive L - M channel and a relatively less sensitive L + M channel. The orientation of the detection contour close to the positive diagonal means that the L - M channel determines the detection thresholds along the long sides of the contour, and argues for an L - M channel with nearly equal weights for L and M cone inputs. This result is consistent with the Boynton code. (Since the L + M channel determines only a point or two at the ends of the contour, the weights of the L + M channel cannot be estimated from this data set.)

For the data in Figure 13.5B, the spatial frequency was reduced to 0.125 cy/deg, and the temporal frequency was kept at 0 Hz. The result is another, larger ellipse-like contour, still oriented approximately along the negative diagonal. Both channels are less sensitive to the lower spatial frequency, but the L - M channel is still the more sensitive of the two. The unchanged orientation of the contour reveals unchanged weights of the L and M cones in the L - M channel across the change in spatial and temporal parameters.

For the data in Figure 13.5C, the stimulus was again 1 cy/deg, but the temporal frequency was increased to 24 Hz. As a result, the detection contour gets very large, suggesting that the sensitivitiesof both channels are greatly reduced under these stimulus conditions. The contour also flips to an orientation along the negative diagonal, suggesting that now the L + M channel is the more sensitive. (Surprisingly, the cone input weightings required to fit these particular data are a weighting of three or four for the L cones to one for the M cones in the L + M channel. These cone input weightings depart more than usual from the 2-to-1 weightings incorporated in the Boynton code.)

To look for an S - (L + M) channel, Sankeralli and Mullen shifted to another plane of cone contrast space. Since they were looking for the interaction of an S cone signal with a summed L + M cone signal, they looked in a plane whose axes are defined by S and L + M. Alignment of the detection contour with these axes would suggest the presence of both kinds of signals, but argue against their interaction, whereas alignment of the contour along the positive diagonal would suggest the subtractive interaction of these two signals. A detection contour for a spatial frequency of 0.125 cy/deg and a temporal frequency of 0 Hz,, plotted in an S vs. L + M space, is shown in Figure 13.6A. The orientation of the contour along the positive diagonal reveals an S - (L +

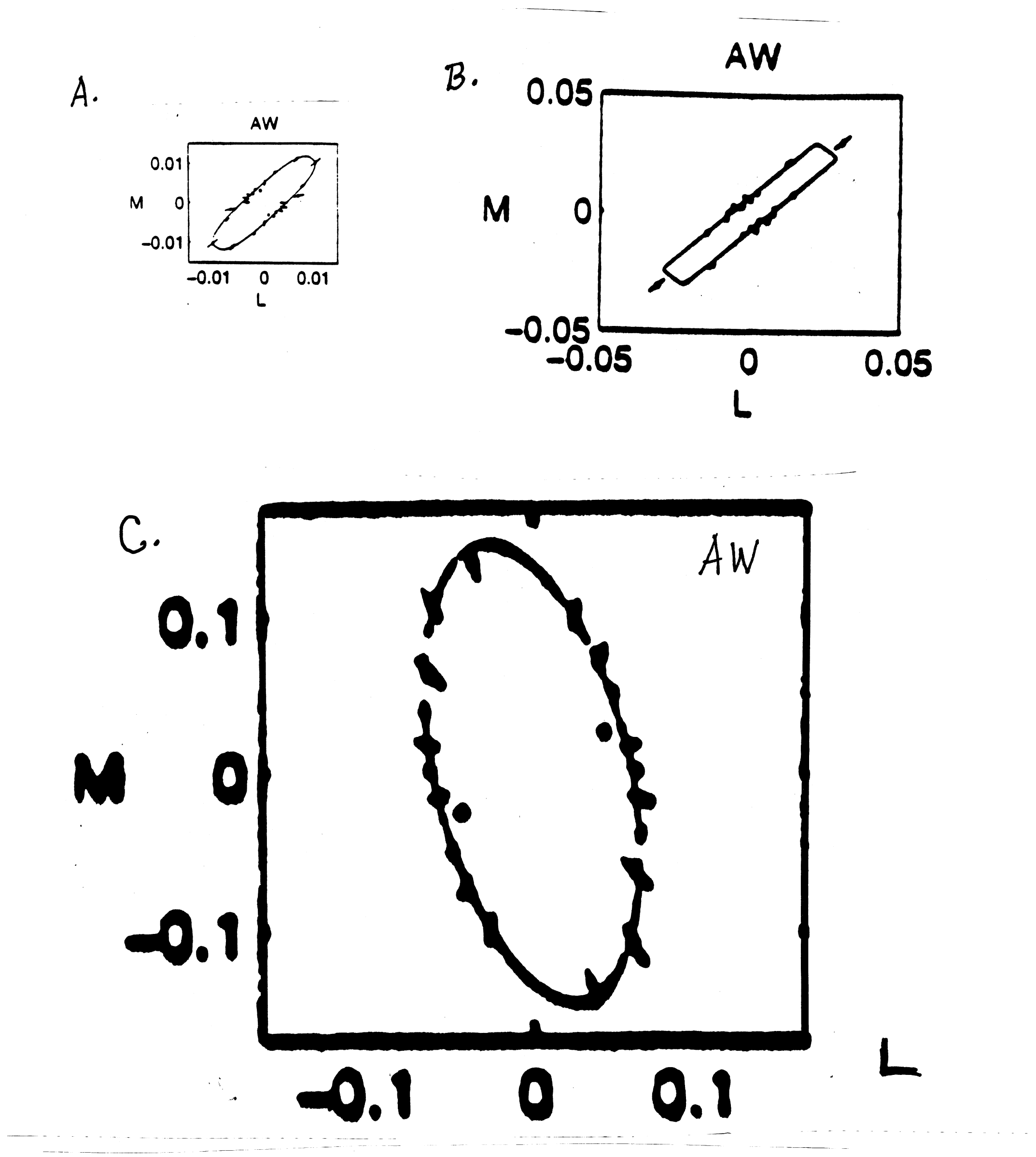


Figure 13.5: Empirical detection contours in L,M cone contrast space. The three contours were measured with stimuli of different combinations of spatial and temporal frequency. A: 1 cy/deg, 0 Hz. B: 0.125 cy/deg, 0 Hz. C: 1 cy/deg, 24 Hz. A and B consistently suggest a more sensitive L - M channel with approximately equal weights for L and M, and a less sensitive L + M channel with indeterminate weights. C suggests that both channels have lost sensitivity with the shift to a high temporal frequency stimulus. Also, the L + M channel has the higher sensitivity, with L/M cone input weights of about 3 or 4 to 1. The L - M channel has the lower sensitivity, with indeterminate weights. [NOTE: Need to ask Sankeralli and Mullen to replot these data on commensurate axes (I have done a rough approximation with magnification on the Xerox).] [Modified from Sankeralli and Mullen, 1996. A: from Fig. 4, p. 909; B: Fig. 6, p. 911; C: Fig. 8, p. 912.]

M) channel with approximately equal weights for its two components, S and (L + M), exactly as stipulated in the Boynton code.

And just to show you that not all interactions happen: Suppose you predicted the presence of a S - (L - M) channel. You would look in an S vs. L - M plane, and predict that the empirical detection contour would be oriented along the major diagonal. Data for a spatial frequency of 0.125 cy/deg and a temporal frequency of 0 Hz are displayed in this plane of cone contrast space in Figure 13.6B. The detection contours are oriented virtually parallel to the axes – these two channels do not interact. and no S - (L + M) channel is seen.

Finally, some examples of three-dimensional detection contours are shown in Figure 13.7. For a 1 cy/deg, 0 Hz stimulus, the contours are longest on the L + M axis, and shortest on the L - M axis – the L - M axis is the most sensitive. And for an 0.125 cy/deg, 0 Hz stimulus, the contours are longest along the L + M axis, shortest along the L - M axis, and intermediate along the S - (L + M) axis. (For the 1 cy/deg stimulus modulated at 24 Hz, some thresholds could not be measured, and no three-dimensional contour could be plotted.)

In summary, Sankeralli and Mullen's data show how the summation-near-threshold paradigm can be used to constrain models of post-receptoral color codes. Most of the available data on detection contours, from their lab and others', are in good general agreement with the description provided by the Boynton model. This is one of the major reasons that many visual scientists now take the Boynton code as a useful model of the early postreceptoral color code.

13.3.6 The shift to postreceptoral channels space

If the Boynton model provides a valid description of the postreceptoral code, it is useful to conceptualize our experiments and plot our data in a color space in which the channels specified by the Boynton code are made explicit. We will therefore shift to a post-receptoral channels space based on the Boynton code. That is, the isolation axes of the Boynton code (L + M, L - M, and S - (L - M)) will be used explicitly as the z, x, and y axes of the color space (Figure xx in prior chapter).

Sankeralli and Mullen's three-dimensional detection contours are replotted in postreceptoral chanels space in Figure 13.8. And finally we have the representation we have been looking for: plotted in this space, the detection contours align themselves with the axes of the space. Following the familiar interpretation rules, this alignment visually captures the argument that the signals plotted on the axes of the space are carried by three independent postreceptoral channels.

In sum, postreceptoral channels space represents model and data in a satisfying esthetic correspondence. The second paradigm for determining the characteristics of postreceptoral channels – adaptation-near-threshold – will be presented in this space, and we will see whether the correspondence holds.

13.4 Paradigm #2: Adaptation-near-threshold

13.4.1 Adaptation and cross-adaptation

A second paradigm used to reveal the characteristics of postreceptoral coding is that of *adaptation-near-threshold* (Graham, 1979). The adaptation-near-threshold paradigm has been widely used in many areas of visual science. To introduce it, we will use the hypothetical example of stimuli composed of 540 and 650 nm lights.

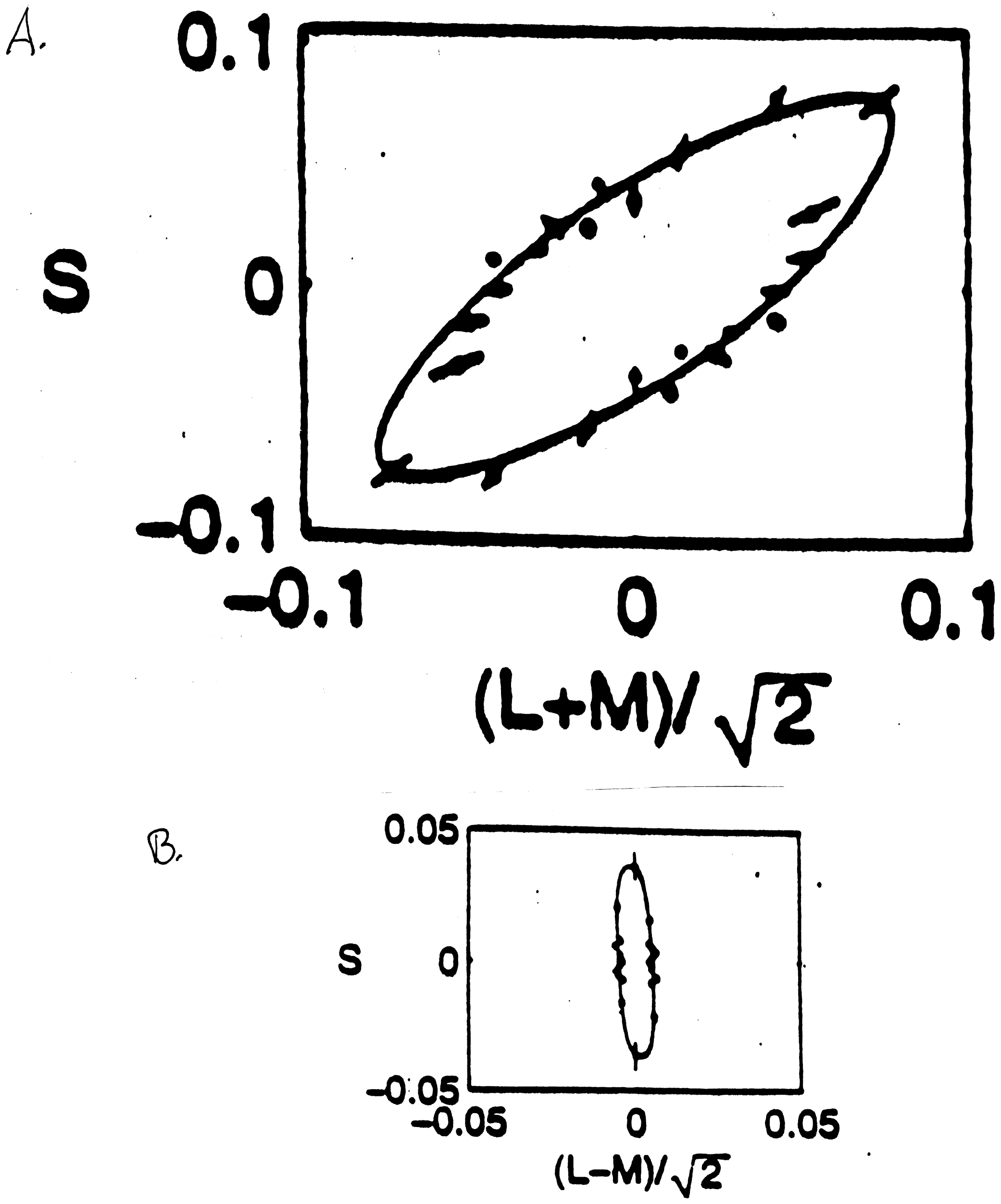


Figure 13.6: Empirical detection contours involving S cone inputs. A: Data collected in an S vs. L + M plane, designed to detect an S - (L + M) channel. The data fall along the positive diagonal, indicating an S - (L + M) channel with approximately equal weights of S vs. L + M. B. Data collected in an S vs. L - M plane, designed to detect an S - (L - M) channel. The data align with the axes of the plot, revealing two independent detection channels, and no interaction between them. Thus, no S - (L -M) channel is seen. (The square root of two on the abscissae is a scaling factor that can be ignored for our purposes.) [Modified from Sankeralli and Mullen, 1996, Fig. 6, p. 911.]

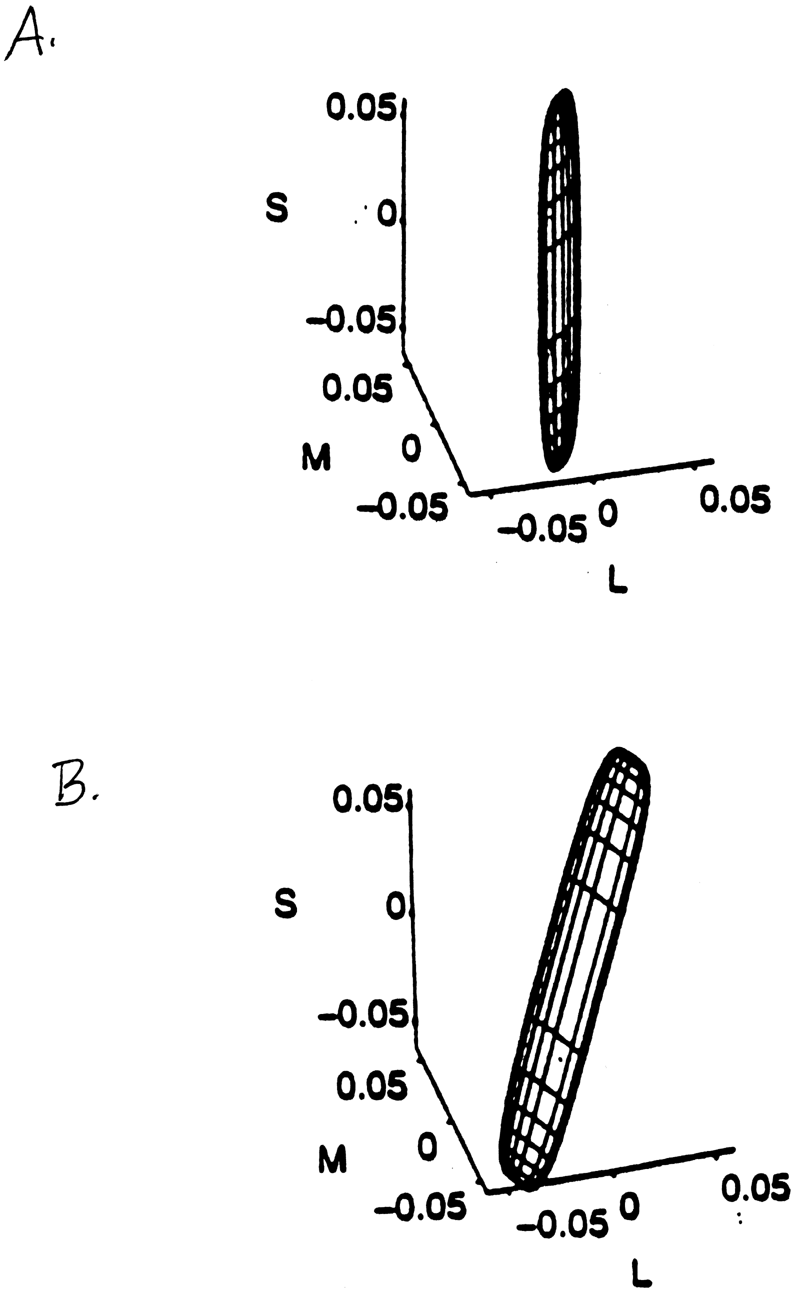


Figure 13.7: Empirical detection contours in three-dimensional cone contrast space. A: 1 cy/deg, 0 Hz. B. 0.125 cy/deg, 0 Hz. [Modified from Sankeralli and Mullen, 1996. A: Fig. 5, p. 910; B: Fig. 7, p. 911.]

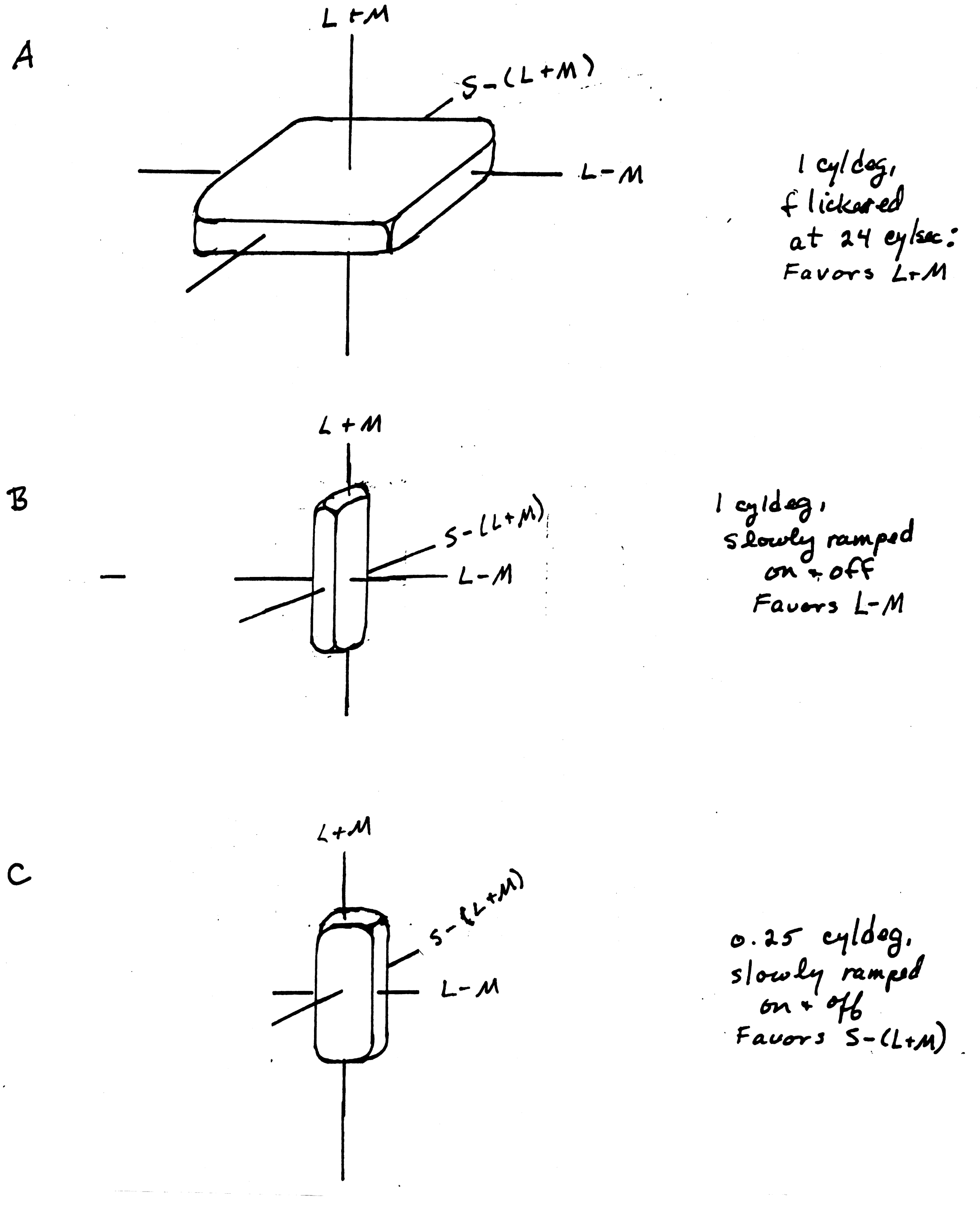


Figure 13.8: Three-dimensional detection contours plotted in postreceptoral channels space. The elongations of the contours are aligned with the three isolation axes of the space. In other words, the space accurately describes the three independent channels that contribute thresholds to the detection contours. [Sketches only so far. I hope Sankeralli and Mullen will replot their data in this space for me.]

Our schematic adaptation-near-threshold experiment is shown in Figure 13.9. The experiment consists of two steps. First, as shown in Figure 13.9A, we set up a pre-adaptation condition, and measure detection thresholds for two different stimuli – here, test spots of 540 and 650 nm light against a dim 'white' background field. Second, as shown in Figure 13.9B, we superimpose each of two *adapting fields* – here, fields of either 540 or 650 nm – on the 'white' background field, and re-measure the thresholds for the 540 and 650 nm test spots. This gives us two *same-wavelength conditions* – 540 on 540 and 650 on 650 – and two *cross-adaptation conditions* – 540 on 650 and 650 on 540.

Various possible outcomes are schematized in Figure 13.9C. The left and right panels shows adaptation to 540 and 650 nm respectively. The thresholds measured against the white background field, normalized to a value of 1, are shown by the dark symbols. Now, since the chromatic adapting fields increase the total background intensity, and we know that higher intensity backgrounds elevate detection thresholds (Chapter 10), we would expect threshold elevations when the chromatic adapting fields are introduced. Thus, we would routinely predict threshold elevations for the *same-wavelength* conditions.

A wider range of predictions are appealing, however, for the *cross-adaptation* conditions. Three different predictions are shown by the set of three open symbols in each of the cross-adaptation conditions in Figure 13.9C. At one extreme, we might believe that all of the stimulus fields are detected by a single visual channel (for example, a single photoreceptor type). In that case, wavelength can't matter, and we predict *complete cross-adaptation*, as shown by the top symbols. But at the other extreme – and here's the point – we might believe that the 540 and 650 lights are processed by two entirely separate and independent channels within the visual system. If so, then an adapting field processed by one channel would have no effect on the detection threshold for test stimuli processed by the other channel. In that case we predict that *no cross-adaptation* will occur, as shown by the bottom symbols. And if stimuli of the two wavelengths are processed by two different channels that are only partially independent, *partial cross-adaptation* might also be predicted, as shown by the middle symbols.

To summarize: In an adaptation-near-threshold experiment, two different stimuli (or stimulus dimensions) are used. The pattern of cross-adaptation between the two kinds of stimuli yields information about the independence or interdependence of processing of the two stimuli. In particular, cross-adaptation is consistent with the conclusion that the two stimuli are processed by the same or only partially separable visual channels. But more interestingly, a failure of cross-adaptation is consistent with the conclusion that the two stimuli are processed by two separate processing channels – physiologically, separate kinds of neurons – within the visual system.

13.4.2 Flicker adaptation

Next, we need to introduce the phenomenon of *flicker adaptation.* Again for concreteness, suppose that a subject views a homogeneous 'white' adapting field, upon which a small 'white' test spot – a brief luminance increase – can be superimposed. In the first phase of the experiment, the subject sets a detection threshold for the test spot. In the second phase, the adapting field is *flickered (modulated in time)* at a high contrast, above and below the mean luminance present in phase 1. The subject is asked to stare for a minute or two at the flickering adapting field. And finally, in the third phase, the subject is asked to reset the detection threshold for the test field. The question is, will the threshold for the test field be changed by exposure to the high contrast flicker of the

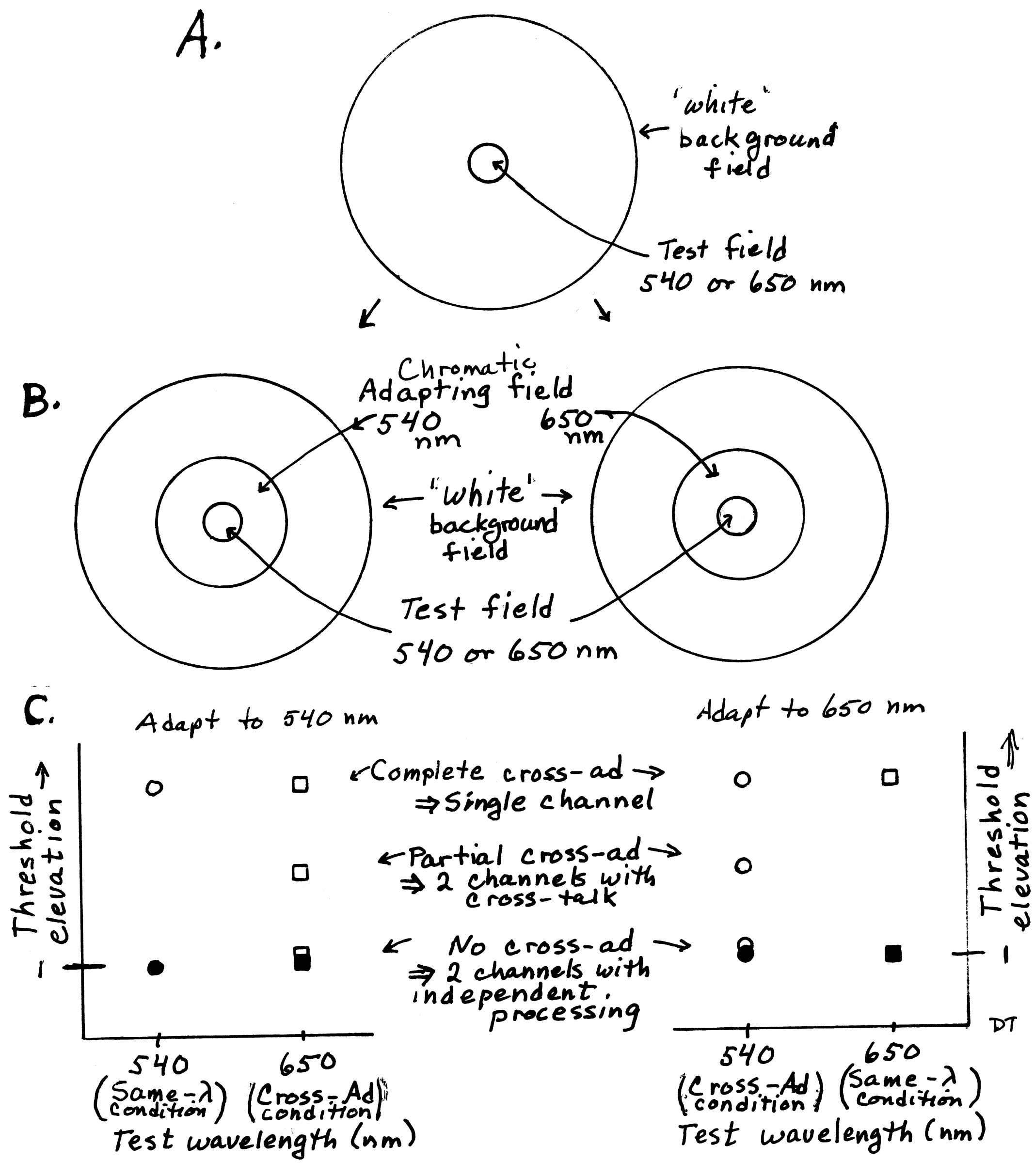


Figure 13.9: The adaptation-near-threshold paradigm: Adaptation and cross-adaptation. A: In the first phase of the experiment, thresholds are measured for 540 and 650 nm test spots against a dim 'white' background field. B: In the second phase, 540 and 650 nm chromatic adapting fields are introduced, and thresholds are measured for both test spots against both chromatic adapting fields. C: Predictions. Thresholds should be routinely elevated in the same-wavelength conditions. But three different outcomes are possible for the cross-adaptation conditions. Complete cross-adaptation suggests that both wavelengths are processed by a single processing channel; partial cross-adaptation suggests two partially independent channels; and an absence of cross-adaptation suggests two independent channels.

adapting field?

Typically, the exposure ("adaptation"[5]) to the modulating adapting field will elevate the detection threshold for the test stimulus, sometimes by as much as 0.5 log units (a factor of 3). The subject's sensitivity recovers only gradually, perhaps over the ensuing 20 to 30 seconds. A common way to think about this result is to assume that somewhere within the visual system, neurons that respond to the flickering field become fatigued (or are reduced in sensitivity, or have their dynamic ranges shifted upwards) during the adaptation interval, so that their detection thresholds are elevated for some time after exposure to the flickering field.

Now the next interesting question is, is there *cross-adaptation*? That is, does flicker adaptation generalize across stimulus dimensions? For example, if the adapting flicker is a luminance modulation, would it elevate the thresholds only for luminance-modulated stimuli, or would it also elevate the thresholds for stimuli defined solely by chromatic changes, such as a change of the test field from 'white' to an isoluminant 'red' and back again? If the luminance-modulated adapting field elevates the threshold for the 'white-to-red-to-white' test pulse, and vice versa, one can argue that these two stimuli must affect the same neural channel (otherwise they could not interact). Conversely, if there is no cross-adaptation, one can argue that luminance-modulated and chromatically modulated stimuli are detected and processed by two separate, independent channels within the visual system.

13.4.3 Color codes: Evidence from differential cross-adaptation

A classic flicker adaptation experiment in color vision was carried out by John Krauskopf, David Williams, and David Heeley in 1982. Their experiment is schematized in Figure 13.10. Krauskopf and his colleagues reasoned that if the adapting light is modulated along one isolation axis in post-receptoral channels space, detection thresholds on that axis should be elevated; but by hypothesis, that adapting light creates no variation at all in the signal in the other two channels. Thus, if the choice of the model is right, these channels cannot become adapted, and their contrast thresholds cannot be changed. And conversely, adaptation along a second or third isolation axis should have no effect upon the threshold along either of the other two isolation axes. In other words, there should be *no cross adaptation* between any two true isolation axes.

On the other hand, suppose the modulation of the adapting light occurs along an axis in between two isolation axes. The adapting light would create modulated signals in *both* of the two channels. In consequence, one should see *cross-adaptation*: since both channels should adapt, the subsequent thresholds for the test pulse should be elevated on both axes, and all axes in between. And an adapting light that modulates the signals in all three channels should elevate thresholds for pulses on all axes of color space.

In sum, this experiment adds another layer to the logic of the adaptation-near-threshold paradigm.

[5] You will notice that we are recycling the term *adaptation* with a shift of meaning. In Chapter 10, this term was used to describe the very large changes in detection thresholds that occur in the presence of background fields of various luminances (light adaptation), and for many minutes after the termination of a high intensity background field (dark adaptation); and also to refer to the physical and physiological mechanisms that cause these changes in thresholds. Here, we again use the term to indicate an elevation of a detection threshold. But the mechanisms for the two kinds of adaptation cannot be the same, because here the background field is flickered around a constant mean luminance. That is, it provides no change in time-average luminance, and cannot (light) adapt the visual system in the old sense of the term. Rather, some mechanism that responds to temporal modulation per se must have been reduced in sensitivity by being exposed to the flicker of the adapting field, and must recover only gradually over time.

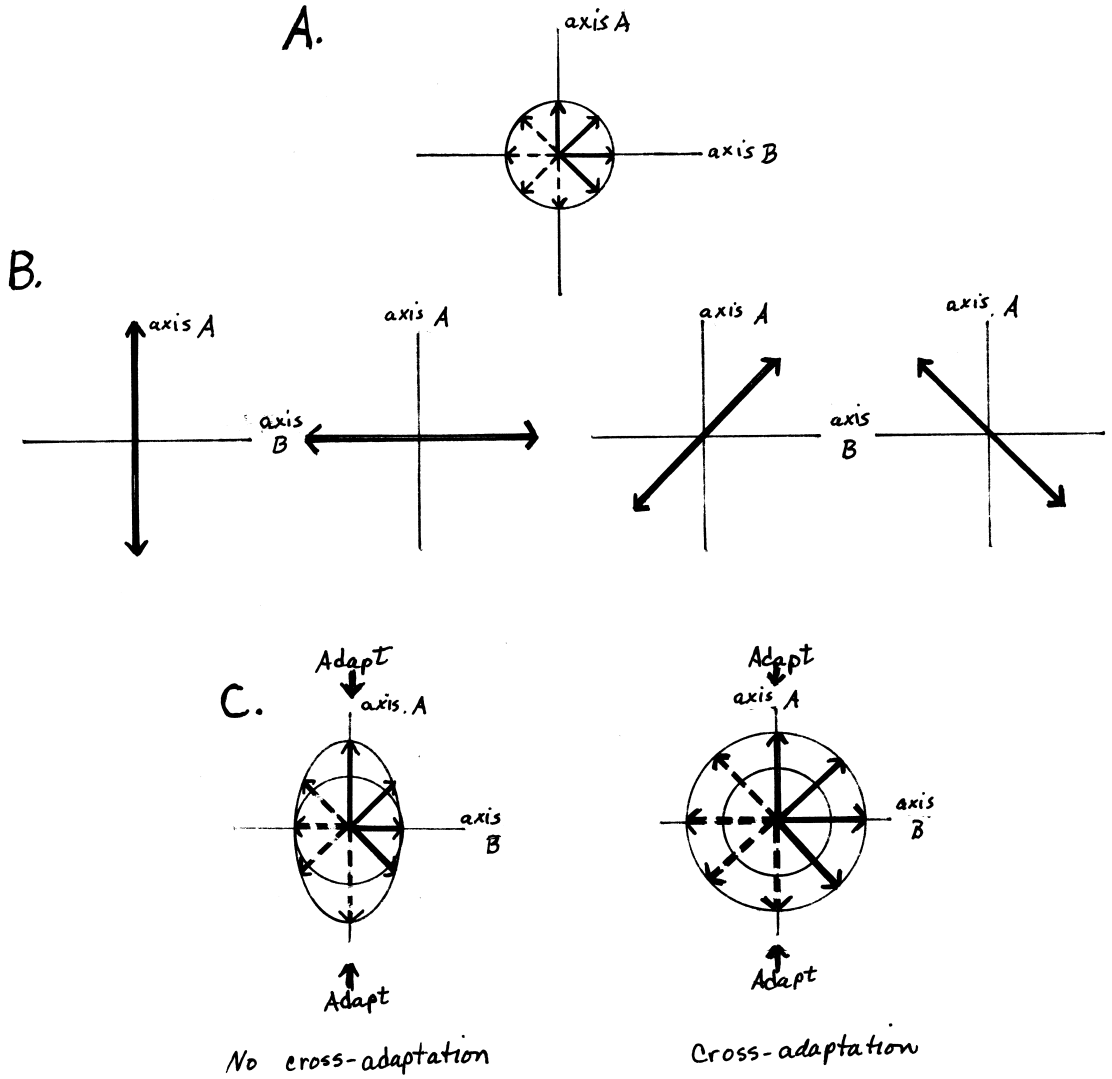


Figure 13.10: Paradigm for Krauskopf et al's experiment. A. First, a plane is defined by choosing two axes, A and B. Initial threshold measurements are made for pulses away from the origin along many axes in this plane. These thresholds are normalized to trace out a circular detection contour, as shown. B. Next, the subject is adapted to high contrast modulation (flicker) along any one of four axes in the chosen plane, as shown. Finally, thresholds are remeasured along all of the original axes. C. Possible outcomes. Assume that thresholds are measured after adaptation to flicker along Axis A, as shown by the arrows above and below the figures. The data could show cross-adaptation – threshold elevations on all axes (right panel); or no cross-adaptation – threshold elevations across only a narrow range of axes, specifically excluding Axis B (left panel).

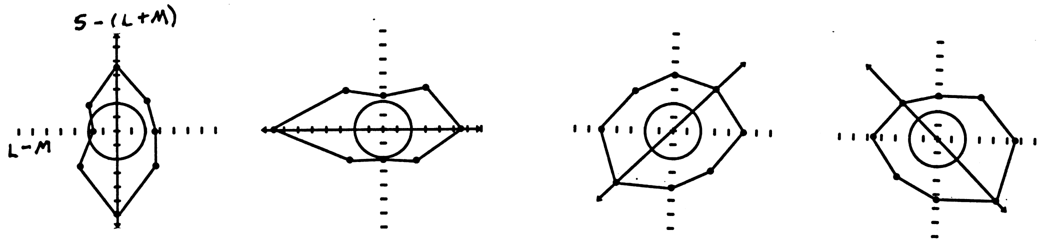


Figure 13.11: Krauskopf et al's results: Differential cross-adaptation. In each panel the long solid arrow shows the axis of flicker adaptation. There is no cross-adaptation between the L - M and S - (L + M) axes (two left panels), but there is cross-adaptation between intermediate axes (two right panels). The same result was found in the L + M vs. L - M and the L + M and S - (L + M) planes. [Modified from Krauskopf (1999), Fig. 16.3, p. 305.]

It is designed around the premise that if there is a three-channel code, there will be a *differential cross adaptation* among different sets of axes. That is, there should be a failure of cross adaptation among a unique set of three and only three axes in color space, and cross-adaptation among all other sets of axes. Turning the argument around, a particular pattern of results – the *absence* of cross-adaptation among a unique set of three axes, and the *presence* of cross-adaptation among all other sets of axes – suggests that the three axes that do not cross-adapt are the isolation axes of the postreceptoral color code. Importantly, then, this experiment provides our second potential line of psychophysical evidence concerning the post-receptoral color code.

How did it come out? The results of Krauskopf et al's experiment were that no cross-adaptation occurred among the L + M, L - M, and S - (L + M) axes, whereas cross-adaptation did occur among sets of intermediate axes. Sample data are shown in Figure 13.11 for the case of cross-adaptation between axes within the isoluminant plane. The results thus suggest that three channels, L + M, L - M, and S - (L + M) exist as a postreceptoral color code, and that this code is in force at the physiological locus at which flicker adaptation has its effect[6]. This experiment, published in 1982, in fact provided one of the major early sources of support for the Boynton model.

13.5 Evidence for the Boynton code: The state of the art

In this chapter, we have examined two psychophysical paradigms – summation-near-threshold, and adaptation-near-threshold – designed to reveal the postreceptoral color code. Both paradigms yield data consistent with the Boynton code. Two very different psychophysical paradigms thus reinforce each other, as do the ganglion cell and LGN cell physiology. In short, the good news is that vision scientists have the retinal output color code nailed, to a very good first approximation.

We should emphasize at this point, however, that much untidiness remains. The results from different laboratories on different individual subjects, with the use of different experimental param-

[6] Krauskopf et al called the independent axes discovered in their experiment the *cardinal axes* of color space. A second useful term, not tied to Krauskopf et al's particular choice of axes, is *privileged axes.*

eters and stimulus paradigms, are more variable that would be ideal. In general, the paradigms that reveal specific coding schemes, reveal the Boynton code; but other paradigms reveal less unique sets of privileged axes. The data can be reconciled *post hoc* by assuming that some of the experiments are controlled by a level of the system at which the Boynton code is in force, whereas others are controlled at other levels.

In DT's view, this situation is not as bad as it might seem – it's just that the experiments we are discussing are at the current forefront of science. Remember that trichrmacy was known for a century or so, during which many creative but incompatible estimates of the spectral sensitivities of the three cone types were proposed. Eventually the data from different paradigms converged closely on a single set of cone spectra (Figure 7.2).

History repeats itself. We are now in the messy process of coming to consensus on the cone inputs to the three channels of the early postreceptoral code. The visual system presumably contains at least a few later levels with additional color code transformations, and different levels will control the data when different paradigms and stimulus parameters are used. The task ahead is to make a consistent multistage model that incorporates all of the data. This is an area of ongoing research and theory in color science.

13.6 Photometry revisited

13.6.1 Degenerate detection contours: Tubes and pancakes

Finally, let's return to the question of detection contours, and reconsider them in postreceptoral channels space. These ideas will lead us to propose an explanation for the facts about photometry, introduced in Chapter 3.

Sankeralli and Mullen's empirical detection contours were shown in postreceptoral channels space in Figure 13.8. Now let's look at some theoretical examples of degenerate cases of detection contours in three-dimensional postreceptoral channels space, shown in Figure 13.12. Suppose we were to try to measure a detection contour, but we used stimuli to which one of the channels – say the S - (L + M) axis – is very insensitive, as shown in Figure 13.12A. The detection contour would elongate along the axis of insensitivity. In the extreme, suppose the subject simply could not detect the stimulus modulation along that axis, even at the highest modulation available with the equipment being used to make the measurements. In that case, there would be no way to measure the contrast threshold on that axis – all we could do would be to indicate that at the highest available contrast, the stimulus remained undetectable. We will indicate such an option with outward pointing arrowheads, as seen on the S - (L + M) axis in Figure 13.11A. Assuming the contrast thresholds on the other two axes can be measured, the detection contour would take the form of an open ended *tube*, oriented along the S - (L + M) axis. And conversely, an empirical detection contour that degenerates into a tube suggests that the visual system has only two rather than three functional channels available under the conditions tested.

Now suppose the visual system is even more limited. Suppose that we use stimuli that (say) only the L + M channel can detect. In that case, there would be no measurable threshold along the L - M axis, nor along the S - (L + M) axis, nor along any other axis in the isoluminant plane. In fact, the stimulus would be detected only when it made a detectable signal in the L + M channel. In that case, as shown in Figure 13.12B, the detection contour would degenerate to a *pancake* – two planes parallel to the isoluminant plane, one above and one below it, separated by a distance equal

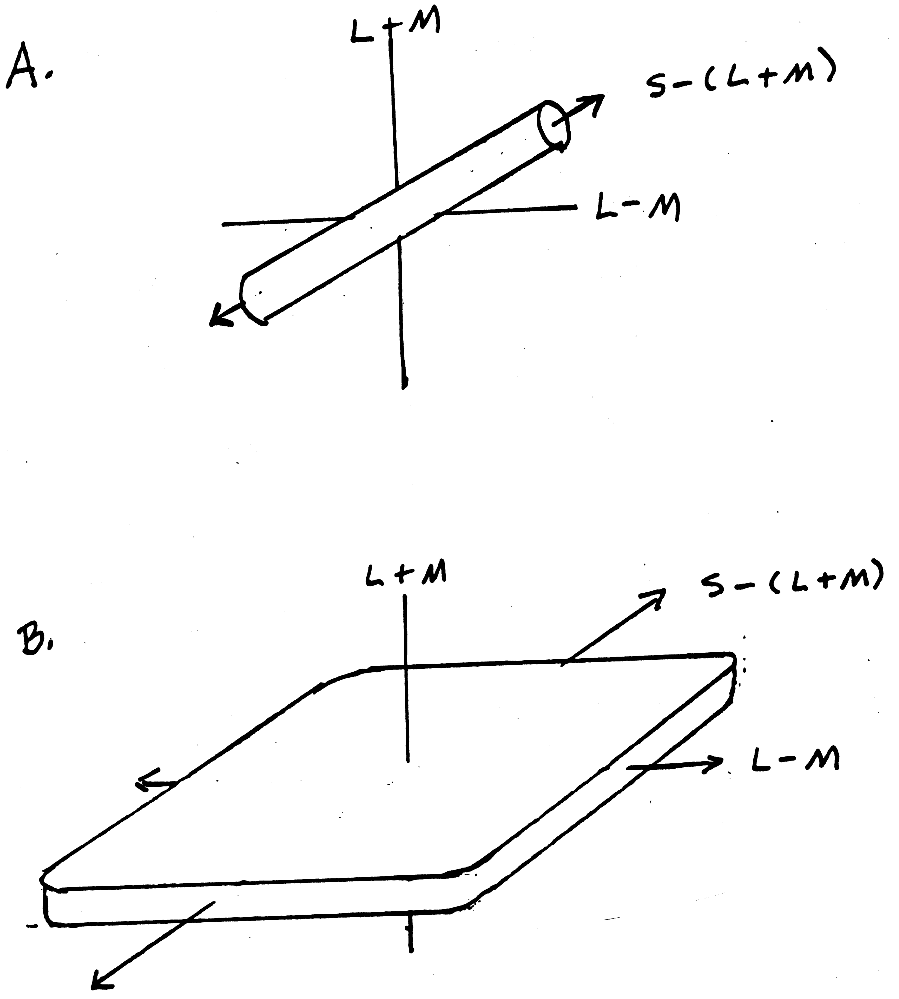


Figure 13.12: Degenerate cases: Tubes and pancakes. A: If one channel (say the S - (L + M) channel is so insensitive that no detection threshold can be measured, the detection contour degenerates into a tube along that axis. B: If two channels (say the two chromatic channels) are so insensitive that their thresholds cannot be measured, the contour degenerates into two parallel planes (a pancake). The orientations of the degenerate contours reveal the postreceptoral color code.

to the contrast threshold on the L + M axis. And conversely, a detection contour that degenerates into a pancake suggests that only a single functional channel is present under the conditions tested. Moreover, the detection contours bracket the null plane of that channel.

An example of some data that approximate pancakes is provided by an experiment in which DT served as a subject. In this experiment, Delwin Lindsey and DT (1993) used moving sinusoidal grating stimuli, modulated through 'white' along many different axes within and tilted out of the V?-defined isoluminant plane. On each trial, the grating moved either upward or downward, and the subject's task was to report its *direction of motion* in a forced-choice task. The experiment was repeated along four different chromatic axes.

The results for subject DT are shown in Figure 13.13. The motion threshold could not be measured along a small range of axes with tilts very close to the isoluminant plane. It became measurable for slightly greater tilts of the axis, and then decreased rapidly as the tilt approached the achromatic axis. For each of the four different chromatic axes tested, the data traced out two parallel lines, making a pancake overall. Lindsey and Teller called the space within the pancake the *motion dead zone* – the region of three-dimensional color space within which no motion was seen. Their data suggest that for the stimuli and task used, there is only one functional channel – the L + M channel – and that DT's isoluminant plane differs slightly from the standard isoluminant plane.

13.6.2 Photometry and the isoluminant plane

In Chapter 3, we discussed the topic of photometry. We noted that three different methods – flicker photometry, minimally distinct borders, and motion photometry – all yield the same or highly similar photopic spectral sensitivity curves, and we wondered why this might be so. At this point we are ready to unite our earlier discussion of methods of photometry with the present discussion of three-dimensional postreceptoral channels spaces, and provide a simple model of why photometry works. Take flicker photometry as an example. Let us assume (as is true) that the two chromatic channels are limited in temporal resolution, and cannot follow 15 Hz flicker. In consequence 15 Hz flicker is detected solely by the luminance channel. Let's carry out a thought experiment on the detection of 15 Hz flicker, along all of the different axes through the white point at the origin of DKL space.

What we expect to find is a pancake, with its crusts parallel to the isoluminant plane. There are two factors at work here. First, neither of the two chromatic channels can detect the flicker, so the L + M channel is the only available possibility. Second, when we find the isoluminant plane – the null plane for the L + M channel – and flicker two lights from within it against each other, the L + M channel no longer detects the flicker either. By hypothesis it was the only channel that could follow fast flicker, so when that channel is nulled, no flicker can be seen. Or reversing the logic, when the subject reports that he perceives no flicker, we have put the stimulus modulation within the null plane for his L + M channel.

A similar argument can be made in the spatial domain in the case of the minimally distinct border technique. Let us also assume (as is true) that the chromatic channels cannot signal the presence of high spatial frequencies. The L + M channel, as we know, can signal spatial frequencies up to 60 cy/deg. Now, make the reasonable assumption that the perception of a sharp border depends on having activity in a channel that responds to high spatial frequencies. Then sharp borders can be perceived whenever there is a signal in the L + M channel – that is, when we

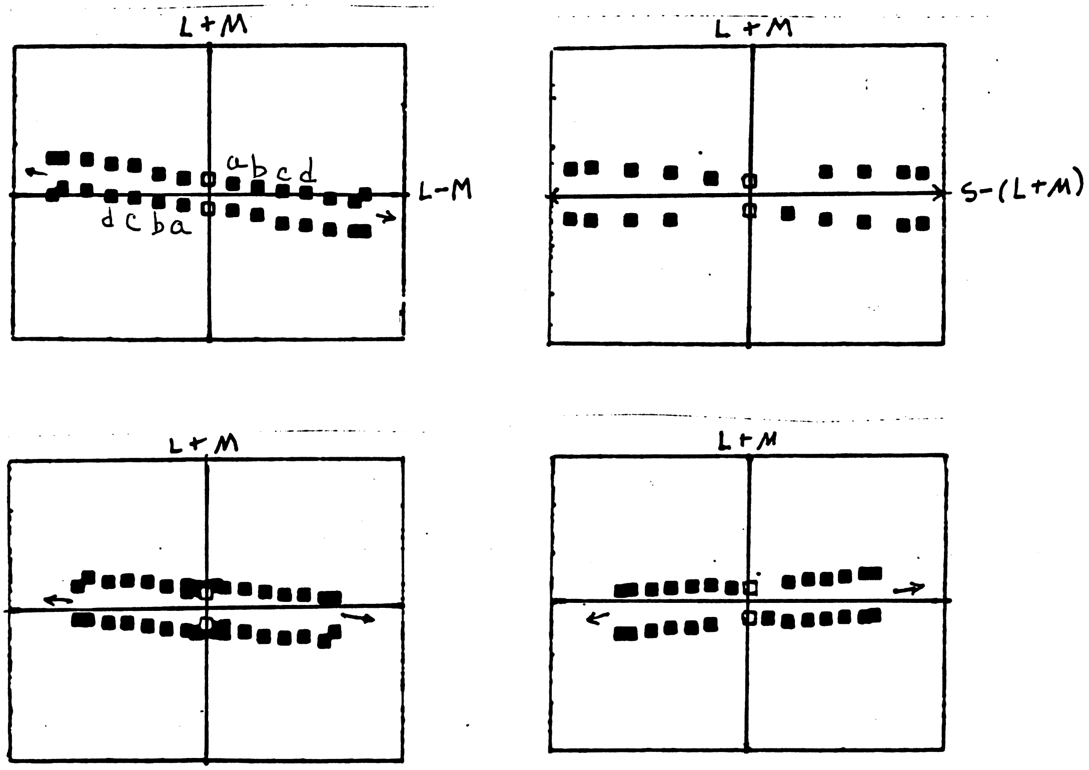


Figure 13.13: A pancake measured in a direction-of-motion task. Stimuli were modulated along axes through the origin. In the top left panel, the matching letters (a,a; b,b; etc.) show the two representations of the same contrast threshold. The subject's task was to judge the direction of motion of the stimulus. The four panels show four cuts through the L + M axis, along the L + M axis (top left), the S - (L + M) axis (top right), and two intermediate axes (bottom panels). In each case the contrast threshold became too large to measure at a small range of axes of modulation near the isoluminant plane, as indicated by the arrows. Thus, all four data sets together trace out a pancake that nearly coincides with the isoluminant plane. This subject's isoluminant plane is tilted slightly from the isoluminant plane predicted from Vλ. [Modified from Lindsey and Teller (1993), Fig. 3, p. 1328]

modulate out of the isoluminant plane. But when the only channel that can do the analysis – the L + M channel – is nulled by modulating within the isoluminant plane, the perceptual sharpness of the border is lost.

A similar but more imaginative story can also be told about the motion minimization technique. The analysis of motion is a complex computational task, which we will discuss further in Chapter xx. But for the moment, let's adopt the assumption that the analysis of motion is done solely on signals in the L + M channel. In that case, the perception of motion, like the perception of fast flicker or sharp borders, would be lost for stimuli confined to the isoluminant plane.

From these examples we can extract a general theoretical argument. The argument is that there are psychophysical tasks – for example, following fast flicker, or processing the fine details of a border, or analysing the direction of motion – that for whatever reason are accomplished by calculations on only the signal in the L + M channel, and cannot be sustained by signals in the chromatic channels. The subject's capacity to do these tasks will therefore fail when the only channel that can do the task – the L + M channel – is disabled, by using stimuli that fall in its null plane.

We can now see that in the theoretical approach we have been developing, the science of photometry has to do with finding tasks that only the L + M channel can do, and then nulling that channel. The photometric system thus defines stimuli of different wavelength composition that fall in the null plane for the L + M channel. Thus, in principle, any task that the L + M channel is good at and the chromatic channels are bad at, could potentially be used as a basis of photometry. Flicker, border distinctness, and motion are the three most commonly used photometric tasks, but there are others as well.

In sum, at the the beginning of this chapter we stated that since the whole human subject can monitor all of her channels at once, and detect a stimulus with any channel that responds, null planes in three dimensional color spaces do not usually reveal themselves for the subject as a whole. But we now see that photometry points out the exception to this statement. If we design or encounter stimulus conditions in which only one of the available channels can respond to the stimulus and do the analysis required by the task, and then adjust the axis of stimulus modulation to fall within the null plane of that channel, our subjects should fail to detect that stimulus, or to analyse its properties. Such cases are probably rare outside the laboratory, but in the laboratory our diabolical cleverness can be put to use, and our stimuli can be custom designed for the purpose. The practice of photometry illustrates a set of cases in which, in the context of this theory at least, we do determine a null plane with psychophysical procedures.

13.7 Summary

In this chapter we have pursued two goals: to introduce two general psychophysical paradigms used to explore the properties of postreceptoral coding, and to work these two paradigms through for the case of the postreceptoral color code. The first paradigm, summation-near-threshold, reveals summations and subtractions of cone inputs that conform rather closely to the Boynton code. The second paradigm, adaptation-near-threshold, reveals a unique independence of processing between stimuli that isolate the three channels of the Boynton code, and it therefore similarly provides evidence that the early postreceptoral code conforms quite closely to the Boynton code.

Finally, we returned to the topic of photometry introduced in Chapter 3, and placed it within the context of detection contours in three-dimensional color spaces. In this context photometric tasks are seen as tasks that the visual system, for whatever reason, can do only with signals in the L + M channel, and not with signals in the L - M or S - (L + M) channels. In the cases of flicker photometry and minimally distinct borders, these limitations come about because the high temporal and spatial frequencies involved exceed the spatial and temporal resolution of the chromatic channels. In the case of motion photometry, we introduced a novel argument – that the analysis of motion might be carried out only on signals in the L + M channel. In all three cases, our new interpretation of why photometry works is that photometric techniques correspond to circumstances in which only the L + M channel is sensitive enough to detect the stimulus, and photometric matches reveal the null plane of the L + M channel.

Chapter 14

The Primate Retina

In Chapter 8 and 9 we used data from cat and mudpuppy to lay out the overall plan of vertebrate retinas, and the five major types of neurons of which they are composed. We also introduced the four major recodings that take place in vertebrate retinas – formation of receptive fields with center/surround antagonism; receptive fields with ON-center vs. OFF-center configurations; neurons with sustained vs. transient temporal properties; and the ”smooshing” or multiplexing of rod-initiated with cone-initiated signals. We also used cat X and Y cells to set up the idea of parallel processing of different aspects of the incoming visual information by different ensembles of neurons within the retina.

In the present chapter we turn our attention to the primate retina – *our* retina. Anatomical analysis of the primate retina began in the xx, and physiological recordings from primate ganglion cells have been possible since the late 1960s. But it was only in the late 1980s that it became possible to record the physiological responses of primate retinal interneurons. These new techniques caused great excitement, because they meant that our reliance on mudpuppy interneurons could be left behind, and we could finally begin to work directly on our own unique mechanisms of information recoding.

In the present chapter we will introduce the newly emerging picture of primate retinal processing. Our treatment will be centered around the three major types of ganglion cells – midgets, parasols, and small bistratified cells – that project to the primate LGN. These three types of ganglion cells occupy a pivitol position in visual processing, because they initiate three parallel processing channels and a coding scheme that persists from the retina all the way to the early stages of the visual cortex. Their axons provide the output from the retina, and thereby control the inflow of information that ultimately underlies our perception of the visual world.

After introducing these neurons and the circuitry that provides their inputs, we will return to the question of spatial processing. We will pay special attention to the fovea, and the midget system – midget bipolars and midget ganglion cells – that makes our highest acuity possible. We will then consider the striking variations in retinal anatomy and physiology that occur with variations in retinal eccentricity, and their probable consequences for spatial vision. We will go beyond the properties of individual receptive fields, to consider the question of processing by ensembles of neurons spread across broad retinal regions.

Finally, we will look at the most unique aspect of primate retinal processing – the processing of wavelength information. We will see how the combination of anatomical techniques and null plane analysis is revealing step-by-step the changes in the color code that occur within the primate retina. Our journey will be guided by the Boynton model, and we will see that, interestingly, the Boynton code seems to be partially but not completely instantiated in the ganglion cells of the primate retina.

As always, we use the term *primate* as a shorthand to refer to humans and old world monkeys, and not to all primates. The distinction is particularly important in this chapter, because the color vision of macaque monkeys and humans is highly similar, as are the retinal recodings that deal with wavelength information. On the other hand, the color vision of new world primates is typically dichromatic rather than trichromatic, but varies in fascinating ways from one species to another.

14.1 Overview: A light micrograph of the human retina

Figure 14.1 shows a classic light micrograph of the human retina, about 1.25 mm from the center of the fovea. Comparing this micrograph to those from cat (Figure 8.1) and mudpuppy (Figure **??**), it is clear that there is much in common among vertebrate retinas. In particular, the alternation of nuclear and plexiform layers persists across species. The only obviously novel feature of the primate retina is the layer of diagonally oriented fibers that form the inner part of the outer nuclear layer. This specialization arises from the presence of the fovea, to which we will return below.

14.2 Ganglion cells that project to the LGN

14.2.1 Methods

How do we know what kinds of ganglion cells project to the LGN? The earliest approaches used to trace neural pathways were lesion techniques. Destruction of cell bodies leads to degeneration of the axons from those cells, and allows one to trace the pathways taken by them. It can also lead to trans-synaptic degeneration of neurons in the structures to which the axons project.

More recently, anatomical *tracers* have come into use. That is, a substance that is readily taken up by neurons is injected at a particular level of the system. The substance is transported away from the injection site along the axons of the affected cells. *Anterograde* tracers are carried toward higher levels of processing, whereas *retrograde* tracers are carried toward lower levels of processing. The animal is then sacrificed, and the various areas of the retina and brain are examined for the presence of the tracer.

When anterograde tracers are introduced into the retina, they are found in many brain structures, but most especially in the LGN. When retrograde tracers are introduced into the LGN, three major kinds of neurons reveal the presence of the tracer: the midget, parasol, and small bistratified ganglion cells.

Studies of primate retinal neurons have also been aided by refinements of anatomical techniques. Primate retinas can now be studied *in vitro* (alive but outside the eye). The excised retina can be stained with a vital dye (a dye that leaves the neurons alive), and placed on the stage of a light microscope. With practice, it is possible to recognize and select neurons of individual types, and inject them with a second dye that spreads to all of the processes of the cell. Thus, as with Golgi stains, one can study the individual cell as a whole. Moreover, one can select many examples of each chosen cell type, and trace its systematic variations and its population properties across the retina. Both human and macaque retinas have been studied in this way.

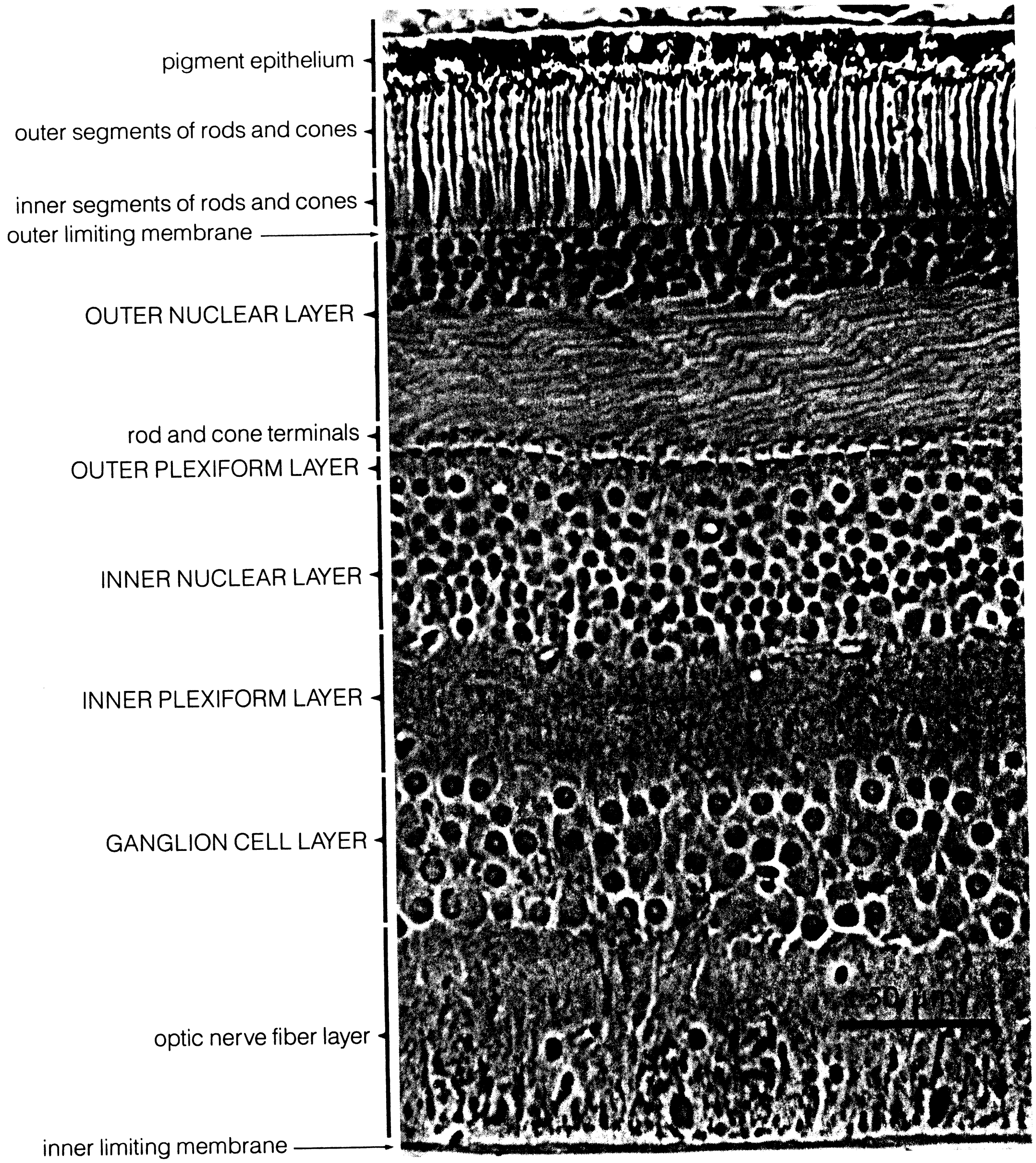


Figure 14.1: Light micrograph of a vertical section through the human retina, just outside the fovea. The overall plan is very similar to that seen in cat and mudpuppy. The only unusual feature is the striking set of diagonal fibers in the inner part of the outer neuclear layer. They are called the fibers of Henle, and they occur because near the fovea the terminals of the photoreceptors (rods and cones) are displaced to the side, as discussed below [From Dowling, 1987, p. 14, Fig. 2.1.]

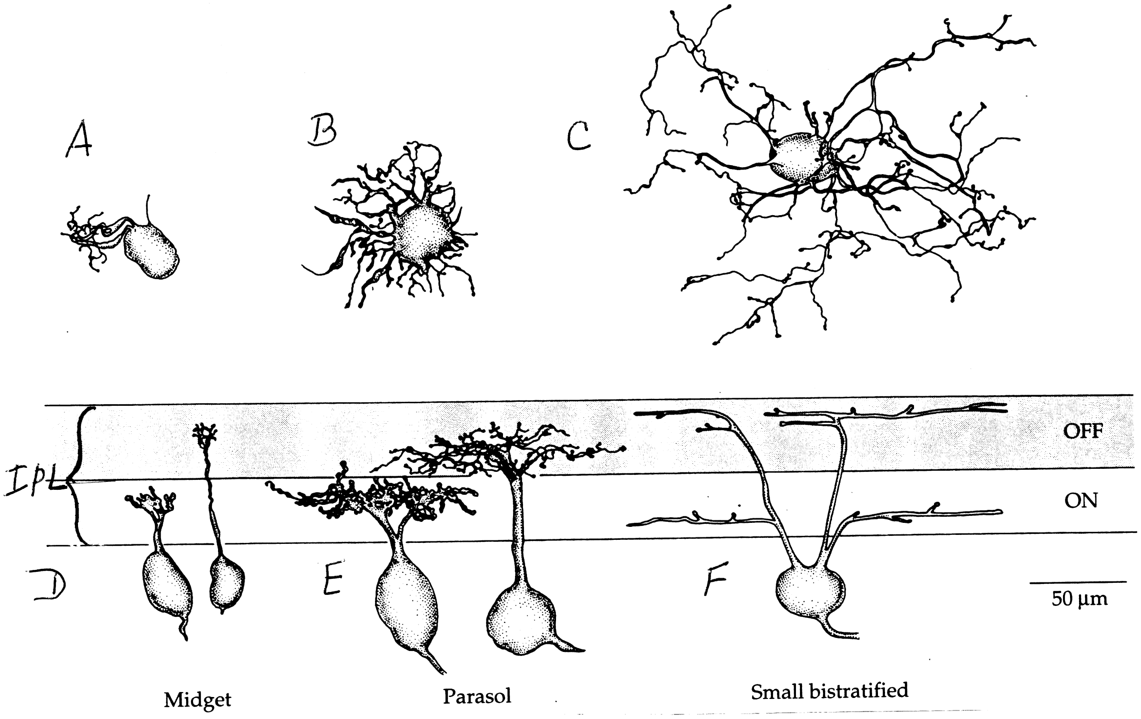


Figure 14.2: Portraits of midget, parasol and small bistratified ganglion cells. A-C: flat mounts; D-F: vertical sections. The appearances of midgets (A, D) and parasols (B, E) are similar, except that at any given retinal location the midget cells are distinctly smaller than the parasol cells. Both midgets and parasols come in two anatomical types, branching in either the OFF or the ON layer of the IPL. The small bistratified cell (C, F) has a sparser dendritic tree when looked at in flat mount. In vertical section, it reveals two distinct layers of branching, one in the OFF and one in the ON layer of the IPL. [Modified from Oyster, 1999, Fig. 14.15, p. 625.]

14.2.2 Midgets, parasol, and small bistratified cells

Three major types of ganglion cells – *midgets, parasols*, and *small bistratified cells* – project to the LGN in primates. As illustrated in Figure 14.2, the three types of cells are named for the appearances of their dendritic fields in the inner plexiform layer (IPL): midget ganglion cells have the smallest dendritic trees; parasol ganglion cells have dendritic trees that spread out like umbrellas; and small bistratified ganglion cells are *bistratified* – their dendritic trees branch at two distinct levels within the IPL. Midgets make up about 80% of the total primate ganglion cell population; parasols, about 10%; small bistratifieds about 5%; and all other types together, about 5%.

Midget ganglion cells are shown in Figure 14.2A and D. They have very small, densely branched dendritic trees confined to a single layer of the IPL. Except for their small size, they have many characteristics in common with the ganglion cells of other species. Like cat X and Y cells, some midgets spread their dendrites in the OFF (outer) and others in the ON (inner) layer of the IPL.

Midgets have very small receptive fields with center/surround antagonism, and as expected from their anatomy in the IPL, both ON-center and OFF-center subtypes occur. Responses to light are relatively sustained rather than transient.

Parasol ganglion cells are shown in Figure 14.2B and E. They have densely branched dendritic trees that are larger than those of the midgets, and larger receptive fields. They also have center/surround antagonism, and both ON-center and OFF-center types occur. Their responses to light are relatively transient.

Small[1] *bistratified ganglion cells* have some novel characteristics. Unlike midgets and parasols, the small bistratified cell spreads two distinct layers of dendrites, one within the ON and the other within the OFF sublamina of the IPL. Small bistratified cells receive inputs from both ON-center and OFF-center bipolar cells, and show antagonism in that they show both ON and OFF responses to light. However, the two dendritic fields are approximately the same size, and so are the sizes of the ON and OFF regions of the receptive field. In other words, small bistratified cells show no spatial antagonism – if you will, "center" and "surround" regions are both about the same size. Their responses, like those of midgets, are relatively sustained.

14.2.3 Primate interneurons

We turn now to the question: By means of what retinal circuits are these three ganglion cell types cooked up? But before we can work out recipes, we need to sort out several new subtypes of retinal interneurons, beyond those introduced in Chapter 9.

Let's begin with horizontal cells. In Chapter 9 we mentioned only a single generic horizontal cell. But in mammalian retina, two distinct types of horizontal cells are now recognized (usually called Type B and Type A). In primates, the two corresponding types are called *H1* (or *HI*) and *H2* (or *HII*) respectively. Flat mounts of H1 and H2 cells are shown in Fig14.3. In all cases, including primates, horizontal cells spread their dendrites broadly within the OPL, and sum the signals from many photoreceptors. As we saw in Chapter 9, they provide the receptive field surrounds of bipolar cells.

H1 and H2 cells differ in their synaptic contacts with the different types of photoreceptors. The H1 cell has two distinct parts – a so-called axonal[2] arbor that receives input only from rods, and a dendritic arbor that receives inputs only from L and M cones. The H2 cell, in contrast, contacts all three cone types, but has particularly frequent and intensive synaptic contacts with S cones. We will discuss these differences in cone inputs more fully in the section on color processing below.

What about bipolar cells? The primate retina contains four classes of bipolar cells: *rod bipolars, midget bipolars, diffuse* (or *diffuse cone*) *bipolars*, and *S-cone* (or *blue-ON*, or *blue-cone*) bipolars. The primate *rod bipolar* closely resembles the generic rod bipolar (Figure xx), and will not be discussed further. The other three kinds of bipolars receive inputs only from cones.

The *midget bipolars*, like the midget ganglion cells, are unique to the primate retina. As can be seen in Figure 14.4A, the dendritic fields of midget bipolars are remarkably small and in fact,

[1] The small bistratified cells are not small in relation to midget or parasol cells. In fact, the dendritic trees of small bistratified cells are similar in size to those of parasol cells. Their name comes from the fact that they are the smallest type of *bistratified* cells seen in primate retina. Small bistratified cells are not unique to primate retina – they are broadly distributed across mammalian species. They were omitted from Chapter 8 solely for simplicity.

[2] Neuroanatomists sometimes use the term *axon* to refer to any long, thin neural process, whether or not it carries action potentials. In fact, the tiny process that connects the two arbors of the H1 cell is probably too small to conduct any neural signal, and the two parts are thought to function quite independently.

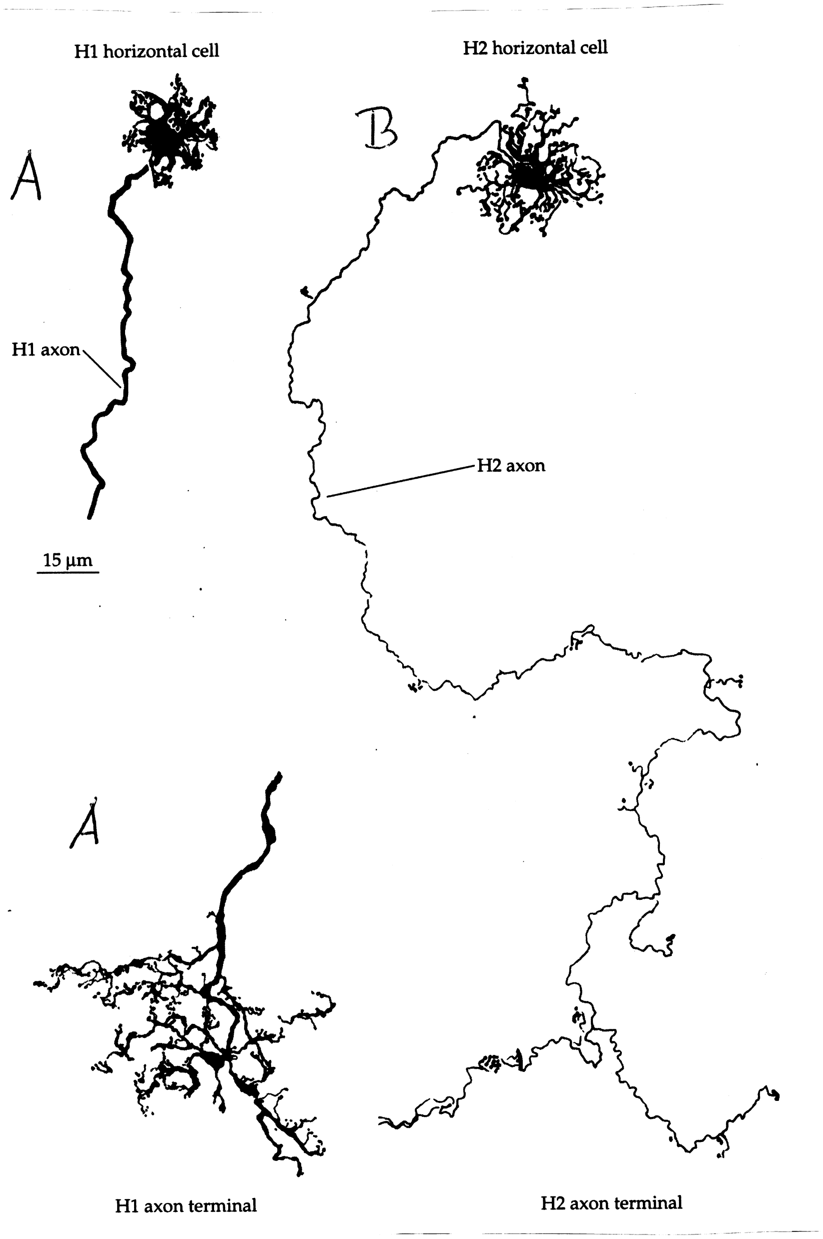


Figure 14.3: Horizontal cells in a macaque monkey retina. Flat mounts, Golgi stain. A: H1; B: H2. The two arbors of the H1 cell are drawn unconnected because they can be widely separated across the retina. The dendrites, at the top, contact L and M cones; the axonal arbor, at the bottom, contacts rods. The H2 cell contacts all three cone types, but makes its densest synapses with S cones (see Figure 14.10 below). [From Kolb, Mariani, and Gallego, 1980, via Oyster, 1999, Fig. 14.1, p. 598.]

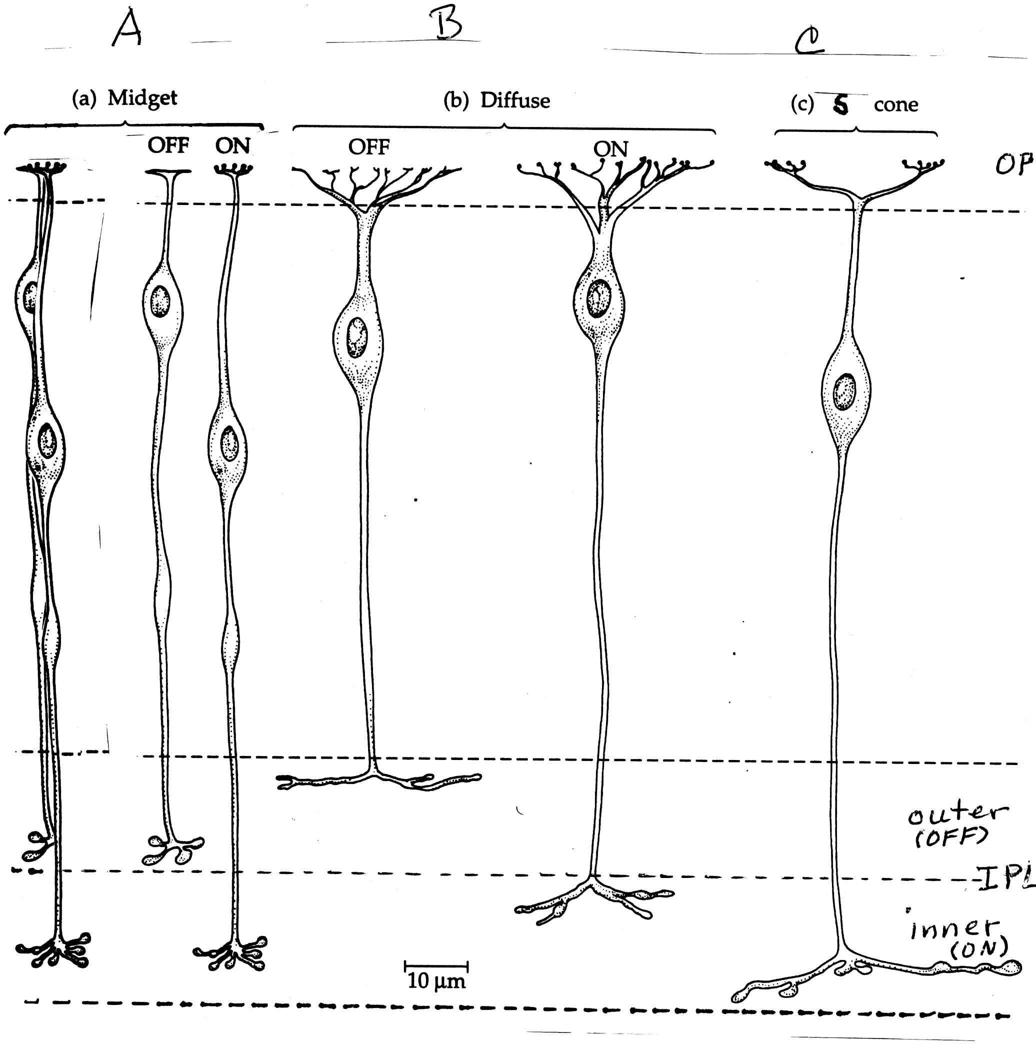


Figure 14.4: Primate cone bipolar cells. A: Midget bipolars. Over the central 50^o of the retina, each midget bipolar contacts only a single cone. Some midget bipolars make flat contacts with the cone, and synapse with a midget ganglion cell in the OFF sublayer of the IPL. Others make invaginating contacts with the cone, and synapse with a midget ganglion cell in the ON sublayer. At the left are two midget bipolars contacting the same cone. B: Diffuse bipolars. Each diffuse bipolar contacts several cones. As with midget bipolars, different diffuse bipolars spread their dendrites in different sublayers of the IPL. C: An S-cone bipolar. The S-cone bipolar makes invaginating synapses with several S cones. [Modified from Oyster, 1999, Fig. 14.7, p. 607.]

over the central 50^o of the retina each midget bipolar contacts only a single L or M cone. Even more remarkably, each cone is contacted by *two* midget bipolars, one invaginating (ON-center) and one flat (OFF-center); so each cone is represented by signals in two different midget bipolars (a remarkable concession to redundancy). These two subtypes of midget bipolars make synaptic contacts with the midget ganglion cells whose dendritic trees lie in the outer and inner sublayers of the IPL respectively.

The *diffuse bipolar*, shown in Figure 14.4B, is a lot like the bipolars described for the generic retina (Figure 9.3xx). Diffuse bipolars spread their dendritic trees widely in the OPL to contact many L and M cones. They come in two subtypes – invaginating and flat – that synapse as expected in the outer and inner sublayers of the IPL.

The fourth type of primate bipolar, the *S-cone bipolar*, is shown schematically in Figure 14.4C. S-cone bipolars contact only S cones, and each S-cone bipolar contacts several S cones. Interestingly, as is the case with rods, the contacts to S cones are all made with invaginating synapses, and all S-cone bipolars make their output synapses in the inner sublamina of the IPL. Thus, all S-cone bipolars provide ON signals. (It is presumed that there is also an S-OFF pathway, but its anatomical basis is not yet known.)

Finally, what about amacrine cells? As many as 40 different types of amacrine cells have been distinguished anatomically in the primate retina. Of these, we will discuss only two: the *A1* and *A2 amacrine cells*, both of which show patterns familiar from Chapter 9. The *A1 amacrine* has a very large dendritic field, covering several square degrees of visual angle. It is a transient amacrine, responding to both the onset and the offset of light. The *A2 amacrine*, like the one shown in Figure 9.12xx, receives inputs largely from rods via rod bipolars, and feeds these inputs into cone bipolars, allowing rod-initiated signals their major route out of the retina.

14.2.4 Recipes

Now we can ask, in parallel to our question in Chapter 9, how are the ganglion cell signals made up from the photoreceptor inputs, as processed through the retinal interneurons? The proposed circuits are shown in Figure 14.5.

The circuits for midget ganglion cells (Figure 14.5A) should be quite familiar, except for their midget-ness. As noted previously, the receptive fields of midget ganglion cells have a center/surround organization. In the fovea, the center response starts from a single cone, and proceeds via a single midget bipolar to a single midget ganglion cell. The surround response originates from the H1 horizontal cells, and proceeds through the same midget bipolar to the midget ganglion cell. In fact, each cone is contacted by both a flat and an invaginating midget bipolar, and initiates both an ON-center and an OFF-center signal in the population of midget ganglion cells. Each foveal cone is thus often said to have a *private line*[3] –actually two private lines – through the retina.

The circuits for parasol ganglion cells (Figure 14.5B) are also familiar, as they are similar to the circuits for transient ganglion cells in the cat (Figures 9.11xx and 9.12xx). The centers of their receptive fields probably arise from photoreceptor contacts with diffuse bipolar cells, and the surrounds from H1 horizontal cells. The transient nature of parasol cell responses suggests that a transient amacrine cell influences the cell's response in the IPL; the A1 would do the job.

[3] In the early days of telephones, there were "party lines" shared by several households, and more expensive "private lines" serving a single household. Foveal cones have the privelege of private lines through the retina.

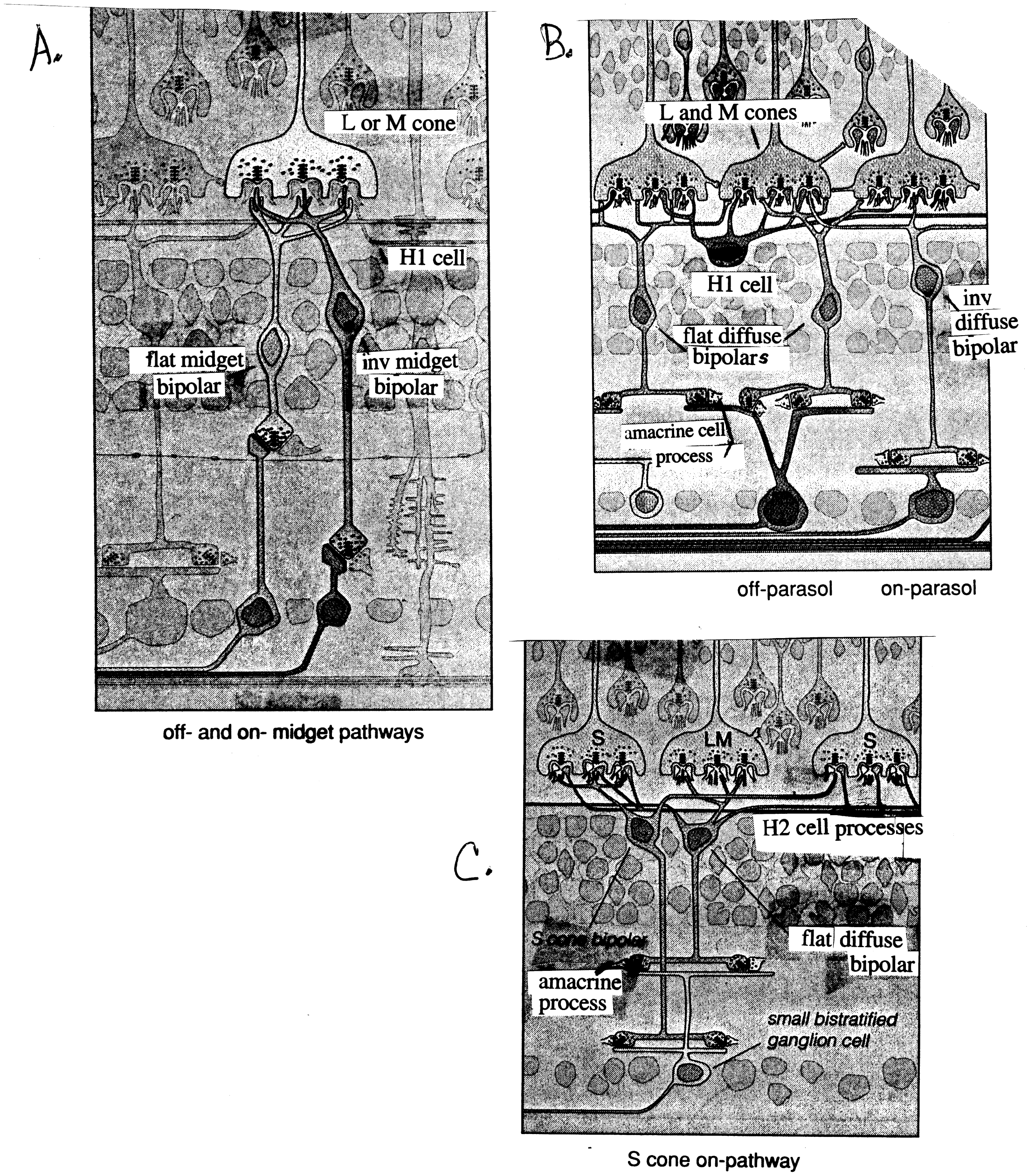


Figure 14.5: Circuits for primate ganglion cells. A: OFF and ON midget ganglion cells. B: OFF and ON parasol cells. C: A small bistratified cell. [Modified from Rodieck (1998), pp. 55, 236, and 346.]

Finally, the circuit for small bistratified ganglion cells (Figure 14.5C) has some novel features. As shown earlier (Figure 14.2C), the small bistratified cell has two sets of dendrites, which stratify respectively in the inner and outer sublayers of the IPL. Not by coincidence, the axon terminals of the S cone bipolars (Figure 14.4C) also terminate in the inner layer of the IPL, and synapse with the inner set of dendrites, providing an ON response to S cone input. But in addition, the axon terminals of the OFF-center (flat) diffuse bipolars (Figure 14.4B) terminate in the OFF layer of the IPL, and synapse with the outer set of dendrites, providing an OFF response to L and M cone input. We will analyse this circuit further inthe section on color codes below.

14.3 Spatial processing: Variations with eccentricity

As mentioned above, the receptive field characteristics of primate midget and parasol ganglion cells closely resemble those of the generic retina. Thus, rather than reviewing the data on primate receptive fields in detail, we will place our emphasis on variations of receptive field size with retinal eccentricity. These variations are seen in most vertebrate retinas, but were introduced only briefly in Chapter 8 and 9. They are particularly marked in primates because of the existence of a major retinal specialization: the fovea.

14.3.1 The primate fovea

A classical sketch of the foveal region, drawn through a light microscope by Steven Polyak in 1941, is shown in Figure 14.6.

The foveal region is marked by four major anatomical characteristics. First, there are no rods within the central $1/2^o$ or so. Second, the cone outer segments are packed more tightly together in the central fovea than anywhere else in the retina. Third, most of the cone nucleii and all of the other retinal layers are swept to the side, so that very few cell bodies lie in the optical pathway between the pupil and the cone outer segments.

And fourth, foveal cones are served by the midget system: the *midget bipolars* and *midget ganglion cells*. As discussed above, each foveal cone is contacted by two midget bipolars, one flat (OFF-center) and one invaginating (ON-center). Moreover, each midget ganglion cell contacts only a single midget bipolar. In consequence, midget ganglion cells have very small receptive fields: so small that the receptive field center approximates the size of a single foveal cone. Clearly the fovea is a system set up for seeing spatial details, and in all probability the EDC designed the midget system to give primate vision the highest possible level of spatial resolution.

14.3.2 Dendritic fields and receptive fields

Of course, it would be impossible to provide such high levels of spatial resolution across the whole visual field, and peripheral regions of the retina provide decreasing spatial resolution. One way of characterizing these chages in spatial properties is to measure the sizes of the dendritic trees of the different kinds of ganglion cells. Dendritic field diameters of human midget, parasol, and small bistratified ganglion cells are shown in Figure 14.7B. As can be seen, the sizes of dendritic fields of all three ganglion cell types change with eccentricity by roughly an order of magnitude, but the three ganglion cell types maintain their relative sizes across the whole extent of the retina.

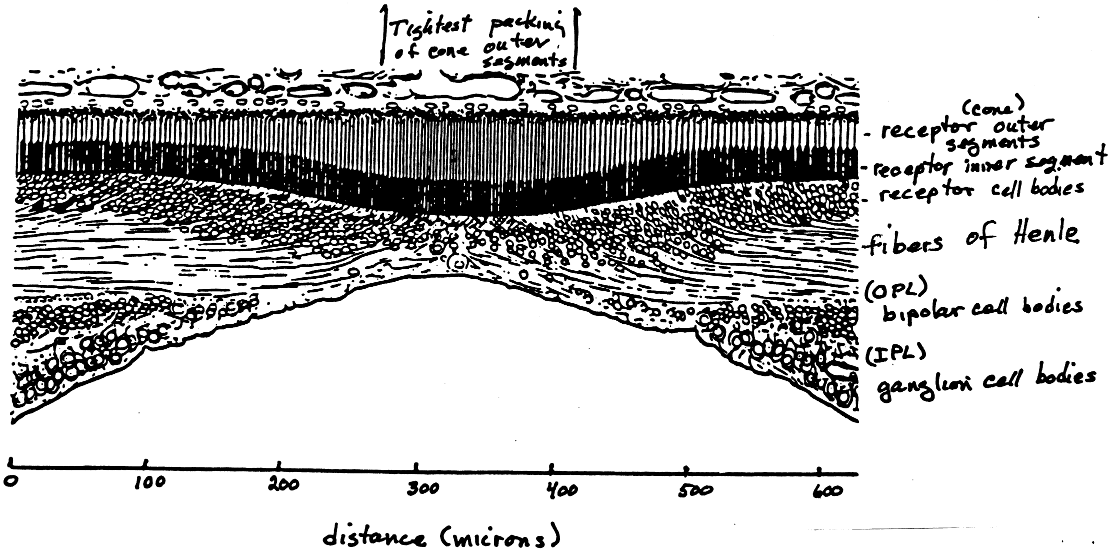


Figure 14.6: The primate fovea. The cone outer segments are most tightly packed at the center of the fovea. The cell bodies and synaptic terminals of the photoreceptors, and all of the other retinal neurons, are swept to the side, leaving the cleanest possible optical pathway for image formation. The elongated connecting elements between the photoreceptor cell bodies and their synaptic terminals are called the fibers of Henle, and they make the distinctive band of diagonal fibers, unique to the primate retina, seen in Figure 14.1. [After Polyak (1941)]

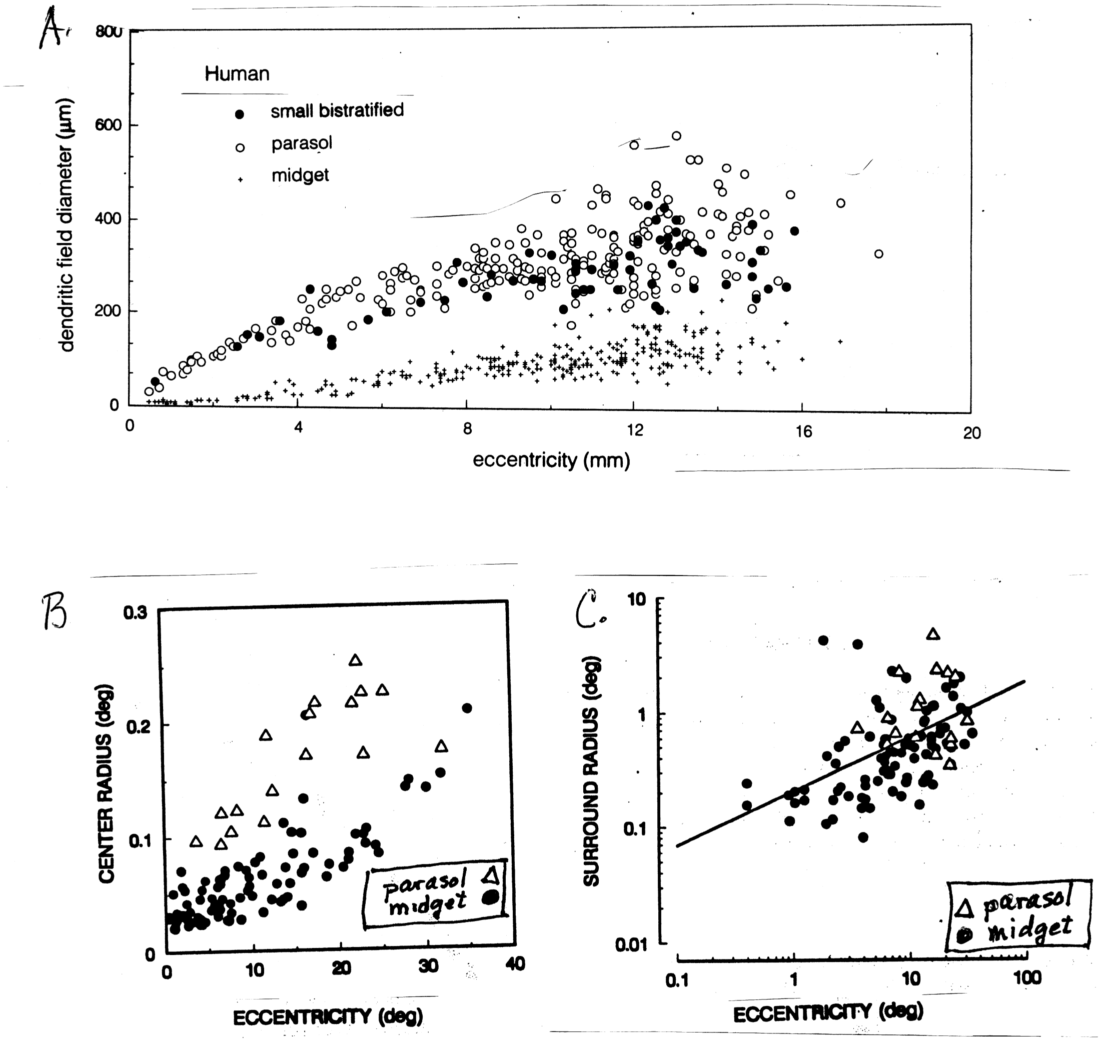


Figure 14.7: Variations of dendritic fields and receptive fields with retinal eccentricity. A: Variations of dendritic field sizes of midgets, parasols, and small bistratified ganglion cells in human retina. Dendritic field sizes vary enormously with eccentricity, but midgets, parasol, and small bistratified cells retain their relative sizes. B, C: Variations of the sizes of receptive field centers (B) and surrounds (C) of midgets and parasol cells in macaque monkey retina. As is the case with dendritic fields, the sizes of receptive field centers vary with eccentricity, but maintain their relative sizes. The surround sizes also increase with eccentricity, but the surrounds of midgets and parasols differ relatively little. [A modified from Dacey, 1993, Fig. 1, p. 1082; B and C modified from Croner and Kaplan, 1995, Fig. 4, p. 12. For the final text these graphs will be on commensurate axes.]

A more laborious but more functionally motivated approach is to measure the sizes of receptive fields. Variations in the sizes of centers and surrounds for primate midget and parasol cells are shown in Figure 14.7A. For both midget and parasol cells, the sizes of both the centers and the surrounds of receptive fields increase with retinal eccentricity, again by roughly an order of magnitude. (Small bistratified cells have been characterized only recently, and no comparable data are available on their receptive fields.)

The total numbers and percentages of midget, parasol, and small bistratified cells also vary with eccentricity, as shown in Figure 14.8. The densities of all three cell types fall off with eccentricity. In terms of percentages, midget ganglion cells constitute over 90% of the ganglion cells that serve the fovea. In the periphery their numbers drop off to less than 50%. In functional terms, what the midget cells have, the fovea needs.

14.3.3 Coverage

Finally, there is an interesting design feature that we have not addressed previously: that of retinal *coverage*. The question is, at each point on the retina, how many different cells of a particular type overlap? The general answer is that for each subtype of ganglion celll, the dendritic tree of each cell seems to exclude the dendrites of its nearest neighbors. As a consequence, the matrix of dendritic fields of a particular subtype of cell just exactly covers the retina, without overlap, as shown in Figure 14.9. This arrangement is said to provide a *coverage factor* of 1: each point on the retina is represented by one and only one cell of each type. As the dendritic fields and receptive fields get larger with eccentricity, the numbers of ganglion cells of each type are reduced in proportion, and the coverage factor of 1 is maintained across the retina.

So, for five different types of ganglion cells – ON-center midgets, OFF-center midgets, ON-center parasols, OFF-center parasols, and small bistratified cells – there are five different overlapping ensembles of cells, each one covering the retina with a coverage factor of 1. Each local region of the retina will be represented by five ganglion cells – a single cell in each of the five ensembles. The farther peripheral, the fewer the numbers of cells and the larger the receptive fields, but these two factors trade off precisely so that the coverage of 1 is maintained. These ideas give us our first glimpse of the full spatial panorama of the retinal output code.

14.3.4 Final reprise on grating acuity

We can now return for a final look at a central issue treated in earlier chapters – consideration of the various limits that retinal processing places on grating acuity.

As we argued in earlier chapters, several specializations of the primate fovea work together to optimize foveal acuity. With the tight packing of cone outer segments in the central fovea, each cone outer segment subtends about 30” of arc. This is the region of the retina at which the Nyquist limit is about 60 cy/deg. In addition, since the bipolar cells and ganglion cells are swept to the side, the fovea is also the region in which the optical image is sharpest, and the optical MTF has its highest cut-off frequency, also near 60 cy/deg.

But what about the problem of neural convergence? As stated in Chapter 1, in the human retina as a whole there are about 100 times more photoreceptors than there are ganglion cells. In consequence, on average, inputs from at least 100 photoreceptors must converge onto and be processed by each ganglion cell. Without some special provisions, the fine photoreceptor spacing and the fine optical quality would be wasted, because neural convergence would necessarily degenerate

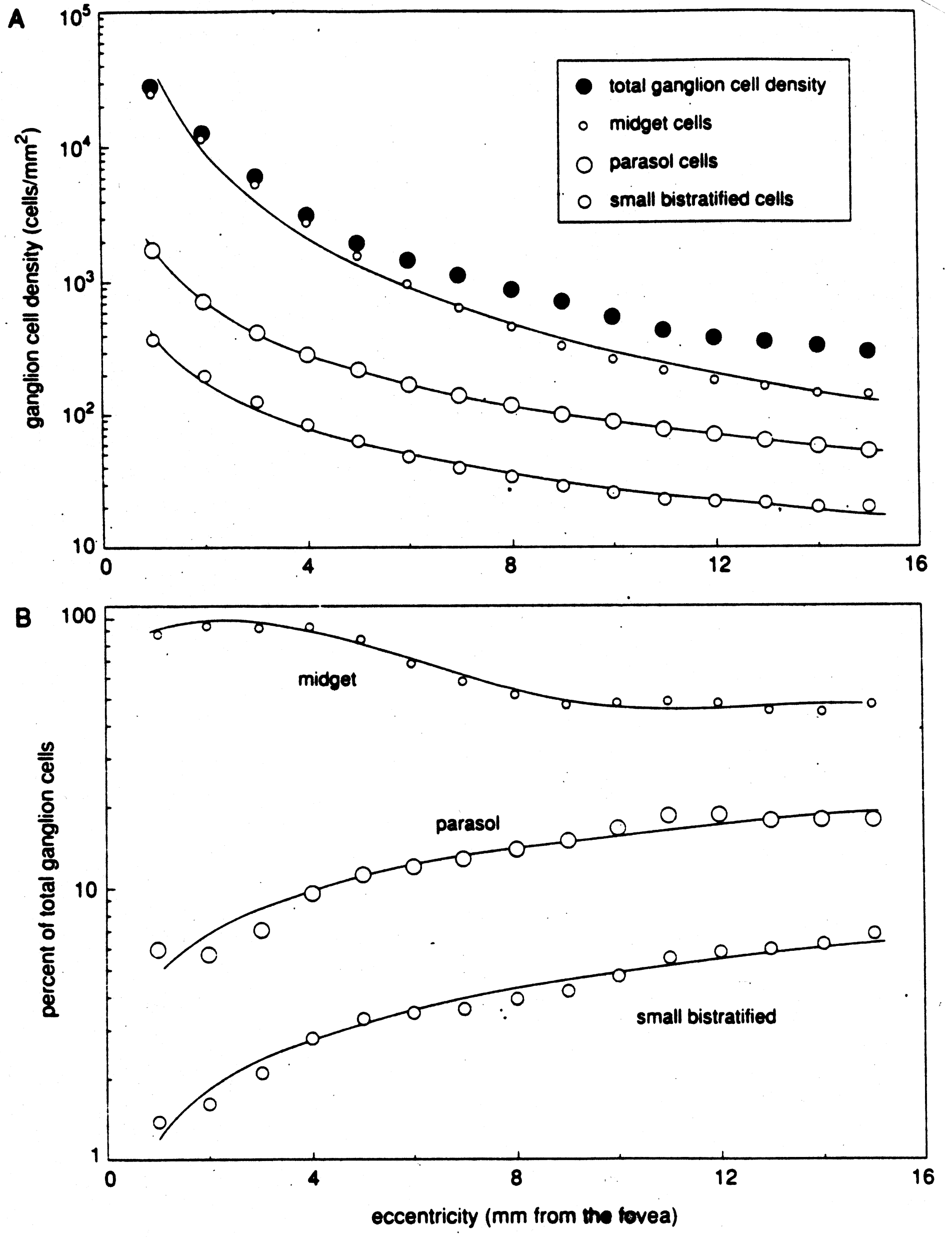


Figure 14.8: Variations in spatial densities with eccentricity. A. Variations in the densities of midgets, parasols, and small bistratified cells. The numbers of all cell types fall off with eccentricity. B. Variations in the proportions of each cell type. Midgets make up over 90% of foveal ganglion cells, but only about 50% of ganglion cells in more peripheral retina. [From Dacey, 1994, Fig. 3, p. 18.]

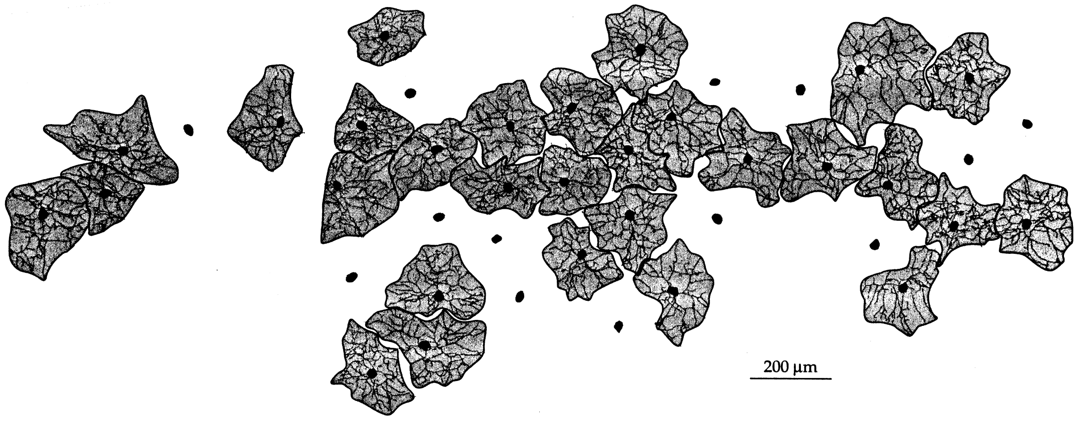


Figure 14.9: Coverage. This figure shows the dendritic fields of ON-center midget ganglion cells in the peripheral retina. The dots represent cell bodies, and the grey areas, dendritic trees (a few of the dendritic trees were unstained). The dendritic trees just cover the retina, without overlap, for a coverage factor of 1. OFF-center midgets, ON-center parasols, OFF-center parasols, and small bistratified cells each form matrices with coverage factors of 1, and this pattern is maintained across the retina. [From Oyster, 1999, Fig. 15.31, p. 688, after Dacey, 1993.]

grating acuity. The EDC resolves the problem posed by neural convergence by making a fovea: a small, central retinal region within which a special new class of ganglion cells – midgets – provide private lines with no convergence, specialized to preserve fine spatial detail.

But what limits acuity outside the fovea? We can address this question by extending the concept of the Nyquist limit. Since ON-center and OFF-center midget ganglion cells each have a coverage factor of very close to 1, the centers of the receptive fields of each cell type cover the retina without overlap, very much as the photoreceptors do. Thus, we can extend the concept of the Nyquist limit, from discrete sampling by photoreceptors to discrete sampling by closely juxtaposed but nonoverlapping receptive field centers. In the central fovea, there is one (actually two) midget ganglion cells for each cone, and grating acuity follows the Nyquist frequency of the cone mosaic. But beyond about 7^o eccentricity, as we have said, inputs from several midget bipolars converge onto each midget ganglion cell, and the midget ganglion cells no longer keep up with the cones in spatial resolution. Grating acuity then follows the Nyquist frequency imposed by the midget ganglion cells.

In sum, foveal and peripheral retinas serve different, complementary functions. The fovea provides the best spatial resolution. Other objects appear in peripheral vision, where fair spatial vision but excellent temporal and motion processing are retained. When an object of potential interest is sensed via peripheral vision, we rotatate our eyes to place the image of that object on the fovea.

14.4 The retinal color code: Is it the Boynton code?

14.4.1 Methods: The primate retina in vitro

Physiological recordings have been made from mammalian retinal ganglion cells in the intact animal (*in vivo*) for many years, beginning with Kufflers early recordings of the center-surround properties of retinal ganglion cells in the cat. In addition, the retinal interneurons of the mudpuppy have been studied in the excised retina (*in vitro*) with intracellular recordings, in the hands of Werblin and Dowling and others. But the retinas of warm-blooded animals have been difficult to keep alive outside the animal.

In the late 1980s, these problems were solved, and the *in vitro* primate retinal preparation described earlier was modified by Dennis Dacey and his colleagues to allow physiological recording from primate retinal neurons. With this technique, investigators are now able to look directly at the living retina on the microscope stage, select neurons of particular types, stainthem, and penetrate them with an electrode for intracellular recording. Thus, rather than just having to record from whatever cell one happens to encounter, one can select a particular cell type and study it systematically across the retina. The primate retina and its interneurons are finally yielding up their secrets.

In addition, Dacey and his colleagues have implemented an optical system that provides the stimuli needed to study the retinal neurons with null plane analysis. A set of three light emitting diodes (LEDs) of different colors are electronically controlled, in such a way that the stimulus can be modulated along any chosen axis of three dimensional color space, and the null planes of cells sought out. Thus, in principle we can determine the pattern of cone inputs to any retinal cell, and thus analyse the retinal color code at each processing level and in each type and subtype of retinal neuron. It is hard to exaggerate the excitement generated by this work – following on beyond the search for the three kinds of photoreceptors, we are about to know the color code transformations that take place within the human retina.

A limitation on these techniques to date is that they are more readily used on larger neurons. The smaller the neuron, the more difficult it is to penetrate it and keep it alive for recording. Accordingly, work in the fovea is the most difficult. For this reason, data collection so far has centered on neurons in the extra-foveal retina.

14.4.2 Horizontal cells and their cone inputs

The use of null plane analysis, combined with detailed anatomical analysis, is nicely illustrated by a recent study of primate horizontal cells (Dacey, Lee, Stafford, Pokorny and Smith, 1996). It had been previously thought that both H1 and H2 horizontal cells had non-selective inputs from all three cone types. However, Dacey and his colleagues showed that H1 and H2 cells in fact have different, selective cone input patterns, as shown in Figure 14.10.

Figure 14.10A shows the results from injecting a neural stain into all of the H1 cells in a small region of retina, and drawing the resulting patterns of cells and synapses. These techniques show that H1 cells contact only L and M cones – they skip a few cone terminals, thought to belong to S cones. And physiologically (Figure 14.10C), null plane analysis shows that H1 cells respond to L-cone-isolating and M-cone isolating stimuli, but not to S-cone isolating stimuli. They also respond to modulation of both L and M cones together (L + M isolating); but not to L - M isolating stimuli. Thus, the two kinds of data agree exactly in excluding an S cone input, and the physiological data

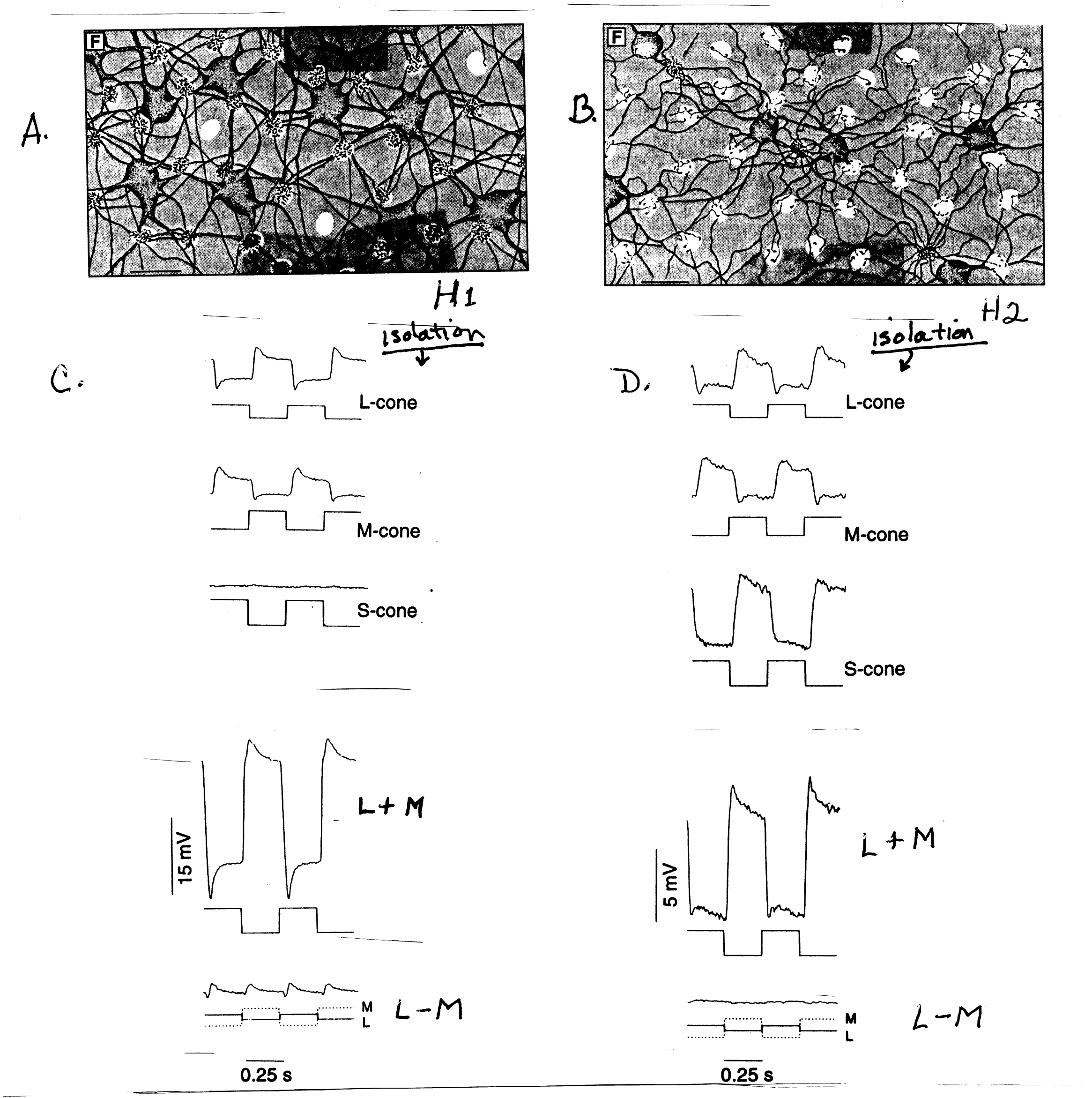


Figure 14.10: Cone inputs to H1 and H2 horizontal cells. A, B: Contacts between horizontal cells and cones. The white areas show cone synaptic terminals, and the black dots within them show synaptic contacts from horizontal cells. A: H1 horizontal cells. There are three blank cone terminals, thought to belong to S cones. B: H2 horizontal cells. There are three terminals that are densely contacted (one indicated by the arrow in the upper left). These are thought to belong to S cones. The other cones are sparsely contacted. C, D: Responses to various modulations in color space. The solid line under each graph shows the stimulus modulation. C: The H1 cell responds to L- and M- cone, but not to S-cone isolating modulation. It also responds to L + M, but only weakly to L - M isolating modulation. D. The H2 cell responds strongly to S-cone isolating modulation, and more weakly to L- and M-cone isolating modulation. It also responds strongly to L + M, but not at all to L - M isolating modulation. [Modified from Dacey et al, 1996, Figs. 3 and 4, p. 658.]

extend the anatomical data by showing that the L and M cone inputs are both of the same sign.

In contrast, H2 cells contact all cone types within their dendritic fields, but with particularly dense connections to S cones, a shown in Figure 14.10B. And physiologically (Figure 14.10D), H2 cells respond strongly to S-cone isolating stimuli, but only weakly to L- and M-cone isolating stimuli. They also respond strongly to L + M isolating, but not to L - M isolating stimuli.

In short, H1 and H2 horizontal cells have different cone input patterns, different null planes, and different roles in the recoding of cone signals within the retina. The most surprising feature of primate horizontal cells is that neither the H1 nor the H2 creates an L - M opponent signal. As described in Chapter 12, the first recordings of color-opponent neurons were reported by Svaetichin in fish retina, and they came from horizontal cells. In consequence, for many years the best available guess about the origins of color opponency in primates attributed it to horizontal cells. But Dacey et al's null plane analysis forces us to discard this speculation. In primates, opponency arises later in the retinal circuitry. But where?

14.4.3 Probable cone inputs to bipolar cells

Unfortunately, as this book is being written, there are as yet no reports of null plane analysis on bipolar cells. We can, however, use the anatomical properties of bipolars to speculate about their cone inputs.

Let's begin with the midget bipolars. Remember that in the central retina, each midget bipolar contacts only a single cone. So the center of the receptive field of a midget bipolar should have the spectral sensitivity of a single cone – either L or M. The surround, on the other hand, probably arises from H1 cell inputs, and if so should contain a mixture of L and M cone signals. Thus the spectral sensitivities of center and surround will differ, creating a bipolar cell whose center and surround are spectrally opponent. On this basis we can speculate that spectral opponency between the L and M cones – something akin to the L - M opponent channel of the Boynton code – originates at the midget bipolar cells. [Whenever DT asks Dacey about opponent coding in midget bipolars, he tells her to be patient!]

Notice the complications, however. If the centers of some midget bipolars arise from a single L cone and others from a single M cone, two subtypes are produced. And in addition, if each type also has ON-center and OFF-center varieties, we are left with four different variants of midget ganglion cells: L-ON-center, M-ON-center, L-OFF-center, and M-OFF-center. Each of these four cell types would be a candidate for the L - M opponent channel, as would various combinations of them. But it remains a mystery how or whether these signals are combined to form a single L - M channel.

The speculative stories for diffuse bipolars and S-cone bipolars can be more simply told. Anatomy suggests that the diffuse bipolars probably sum signals from L and M cones in both their centers and their surrounds, and subtract the two signals spatially, creating spatial but not spectral opponency. ON-center and OFF-center varieties would provide two candidates for a non-spectrally-opponent L + M channel. And the S-cone bipolars contact S cones with invaginating synapses, and probably carry an S-ON signal (but the S-OFF signal is nowhere to be seen).

14.4.4 Midgets, parasols, and small bistratified cells

If the Boynton code were instantiated in the primate retina, we would expect to see three types of ganglion cells. One type would respond well to L + M isolating stimuli, but have no response in a null plane that includes L - M and S - (L + M) modulations. The second cell type would respond well to L - M isolating stimuli, with a null plane that includes L + M and S- (L + M) modulations. And the third cell type would respond well to S - (L + M) isolating stimuli, with a null plane that includes L + M and L - M modulations.

Figure 14.11 shows the first use of null plane analysis in a study of primate ganglion cells (Dacey and Lee, 1994). The three rows of the figure show the responses of a parasol, a midget, and a small bistratified cell respectively. The three columns show the responses to L + M isolating, L - M isolating, and S - (L + M) isolating stimuli respectively. We can see immediately that the fit between Dacey and Lees results and the Boynton code is good, but not perfect. The best news is that parasol cells respond as we want them to to all three kinds of modulations. They *do* respond to L + M, but *not* to L - M nor S - (L + M) modulation, so they exactly fit the characteristics needed for a L + M channel.

The midget and small bistratified cells are more problematic. Within the chromatic domain their responses fit the predictions of the Boynton model. The midgets respond to L - M modulation, and not to S - (L + M) modulation. And the small bistratified cells do the opposite – they respond to S - (L + M) modulation and not to L - M modulation. So these two cell types seem to sort out the two chromatic signals as called for by the Boynton model.

However, notice that both midgets and small bistratified cells also respond to L + M modulation, which is forbidden in the Boynton code. That is, rather than carrying a pure L - M signal, the midget cells respond to (and thus either smoosh or multiplex) both L + M and L - M signals. And the same problem occurs in small bistratified cells – these cells apparently either "smoosh" or multiplex both L + M and S - (L + M) signals.

Moreover, it turns out that individual midget and small bistratified cells vary in the degree to which they trade off their chromatic vs. luminance responses; that is, there is quite a lot of variation in null planes from one midget or small bistratified ganglion cell to another. Unfortunately, if these results are right, we will have to give up a strict Boynton model of the color code at the level of the retinal output, in favor of a model that allows some variability in the null planes of cells we hoped would correspond to the two chromatic channels.

14.4.5 Reconciliations of theory and data: Four options

What are we to conclude about the Boynton code? It seems to DT that there are four major options. The first is to argue that the Boynton code is simply not instantiated within the human visual system – and give up looking for it. The second is to argue that the transformation to the Boynton code is only partially accomplished within the retina, and that it will be more fully instantiated at a more central level of visual processing – and look for it there. The third is to argue that it is instantiated in the retina, but that these are not the cell types that carry it – and look for other retinal ganglion cell types that do.

The fourth and perhaps most attractive option is to remember that the Dacey and Lee recordings are actually from peripheral rather than foveal retina. Perhaps the responses of midget and small bistratified cells to luminance variations, and the variability of their null planes, come about when these neurons have relatively large receptive fields summing the responses of many cones. And perhaps when it becomes possible to record from foveal midgets and small bistratified cells, the results will fall more closely into line with the Boynton code. [Want to lay a bet? DT bets on the fourth option, but only time will tell.]

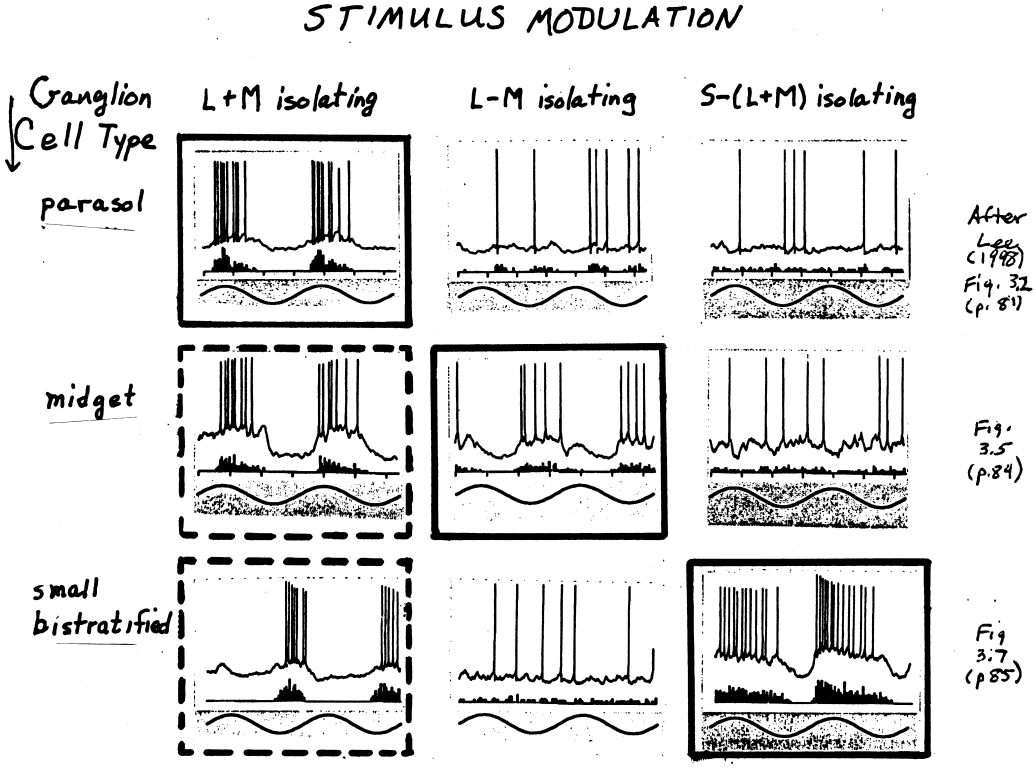


Figure 14.11: Ganglion cells: Results of null plane analysis. The three rows show the responses of a parasol cell, a midget cell, and a small bistratified cell. The three columns show responses to L + M, L - M, and S - (L + M) isolating stimuli. The sinusoidal line under each graph shows the modulation of the stimulus. If the Boynton model were right, each of the three kinds of cells should respond in synchrony with one and only one of the three modulations. The black outlines show the cases in which each cell is driven by a different kind of modulation, in accord with the Boynton model. The non-outlined panels show failures of neurons to respond, also in accord with the Boynton model. The dashed outlines show responses that should not have occurred, but did. In sum, the Boynton model is partially but not fully instantiated in these three ganglion cells. [After Lee (1998), Figs. 3.2, 3.5, and 3.6, pp. 81, 84, 85]

A problem of a different kind also arises, concerning the parasol cells as the neural basis for the L + M channel. Since human subjects can resolve very fine black and white gratings, we would expect the luminance channel to have very fine spatial resolution. But the parasol cells have rather large dendritic fields, even in the fovea. They are constrained by neural convergence, and do not look as though they would serve acuity well. For this reason, some theorists have suggested that a second L + M signal is indeed multiplexed into the midget cell signals, and that this signal is demultiplexed into a L - M signal and a L + M signal at a later level of the visual system. Since this second L + M channel would be served by the midget cells, it could have the fine spatial resolution needed for foveal acuity.

14.5 Is the L - M channel piggybacked on a system designed for acuity?

It is often argued that the two chromatic channels in the old-world primate retina have very different evolutionary histories. Most mammalian species have a maximum of two cone types, an S cone with a peak absorption at short wavelengths, and a single ”LM” cone with a peak absorption in the mid- to long- wavelength spectral range. Moreover, small bistratified ganglion cells are seen in many mammalian retinas. For these reasons, it is suggested that the S - (L + M) system is an ancient one, and is the only chromatically opponent mammalian channel in the retinas of most mammals.

In contrast, the midget system is novel in the primate retina. Where did it come from? It is sometimes argued that the foveal midget system – the private lines from photoreceptors through midget bipolars to midget ganglion cells – evolved to serve high acuity. The argument continues with the claim that the evolutionary separation of L from M cones occurred after the midget system was in place.

The receptive field center of a midget bipolar samples only a single cone, and that single cone must necessarily be only of one type. But after the separation of L from M cones, two types of midget bipolars were created: L-center and M-center. In contrast, a surround process that sampled all the neighboring cones would now contact more than one cone type – both L and M cones. Thus, the spectral sensitivities of center and surround would necessarily differ, and the spatial subtraction process that makes *spatial* center/surround antagonism would then automatically create a *spectrally* opponent cell. Apparently the primate visual system was able to exploit this spectral difference and use it to create a new chromatically opponent system – the L - M system. The multiplexing or ”smooshing” of luminance and r/g signals still seen in midget ganglion cells could arise from the evolutionary origin of the L - M color code in the L + M system originally designed for high acuity.

Support for this speculation is provided by the some details about coverage. Our earlier story was that for both midget and parasol cells, ON-center and OFF-center varieties form separate matrices, each with a coverage of 1. But with the separation of L from M cones, there must be four types of midgets: L-ON-center, M-ON-center, L-OFF-center, and M-OFF-center. It turns out that in terms of coverage, these four types of midgets are treated as only two. That is, the L-ON-center and M-ON-center midgets form a single matrix, with either type of ON-center cell excluding the dendrites of both ON-center types; and similarly for the two OFF-center types. This arrangement suggests that the two ON-center types may have diverged only recently in evolutionary history, and similarly for the two OFF-center types.

14.6 Limitations on three channel linear color models

We began our discussion of post-receptoral color codes in Chapter 12, with an account of three-channel linear models of color vision. Historically, such models have been very popular, because they are the simplest sufficient models for preserving all of the information initially encoded by the three types of cones. However, the EDC is not bound by mathematical parsimony. The primate retina serves more than color vision, and the retinal output code is more complex than the simplest three-channel linear models allow.

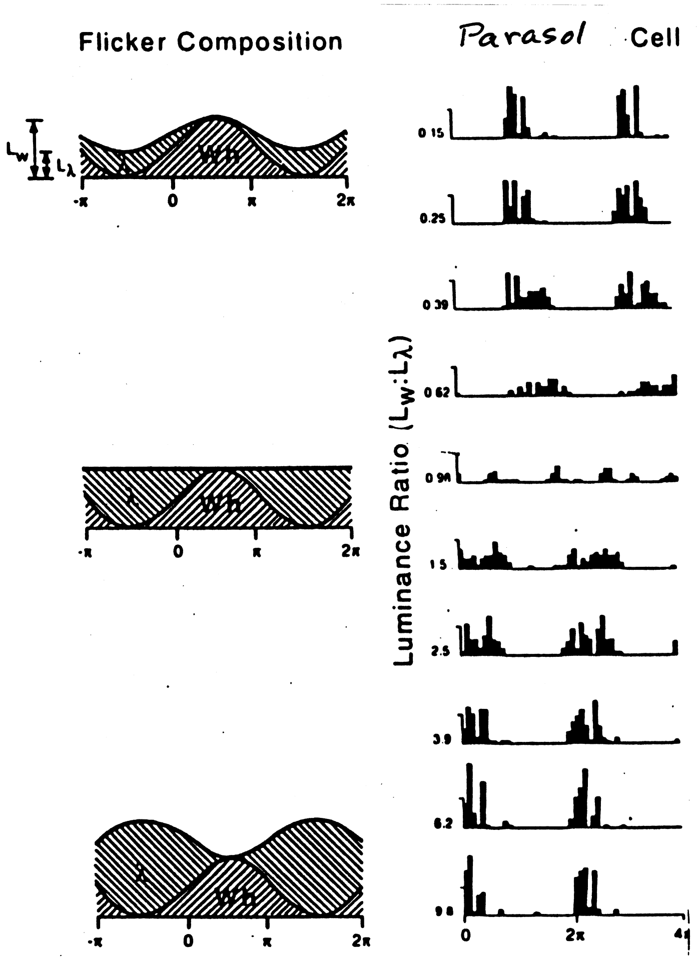
We have already noticed some deviations. First, there are three ganglion cell types that provide approximations to the three channels postulated by the Boynton code. But these cell types have subtypes. In particular, there are four types of midget ganglion cells – L-ON-center, L-OFF-center, M-ON-center, and M-OFF-center. Second, there is more variability in null planes among individual neurons of each type than would be ideal from the simplest model.

And third, there is the problem of non-linearities. The above analyses have suggested that primate parasol cells should have a null plane in the isoluminant plane. That is, we should be able to substitute a light of one wavelength composition for a light of a different wavelength composition without seeing a response in the parasol cell. This is true as a good first approximation, but not in detail. The problem is shown in Figure 14.12. Presented with a ”white” light, and an equal luminance light of a single wavelength λ, this parasol cell gives a small burst of spikes – a transient – at *each* change of color; that is, a response at twice the frequency of the stimulus modulation. This response pattern is called a *frequency-doubling nonlinearity.* The same phenomenon can be seen in the response of anH1 cell to L - M isolating modulation (Figure 14.10xx).

The presence of such nonlinearities limits the eventual range of application of all linear models of color vision. In particular, suppose that a psychophysical subject can detect a stimulus generated from a supposedly L - M isolating modulation. In the Boynton model we would argue that the L + M cell cannot detect this modulation, and that the channel that passes on this information at the ganglion cell level must be the L - M channel. But if the L + M neuron has a frequency-doubling nonlinearity, the frequency doubled signal could allow the subject to detect the so-called L - M isolating stimulus with his L + M channel. These issues are hotly debated in the color vision literature.

The scientific lesson is this. A simple model can be a very useful tool. It forces the scientist to be explicit in her predictions, and provides a backbone for keeping track of the data as they accumulate. It is useful as long as it embodies the most fundamental aspects of the subject, and as long as the backbone it provides, plus the variations and exceptions that accumulate, remain the most satisfying available description of the facts at hand. Only when another model weaves together the facts more attractively is the first model abandoned.

Three-channel linear models have the great advantage that they emphasize the fundamental limitation imposed on our color vision by the presence of only three kinds of cones. Moreover, much of the sorting of cone inputs within the retina can still be reasonably approximated by positing linear combination rules. Although they clearly fail in detail, three-channel linear models will continue to provide an important backbone of our understanding of color theory.



← Frequency-doubling non-linearity

Figure 14.12: Frequency doubling non-linearity in a primate parasol cell. Ideally the cell should have a null plane, and not respond to modulation within the null plane. In this case, a "white" light is being flickered against a light of a narrow wavelength band, λ. The vertical sequence of graphs shows the response of the cell to a series of relative luminances of the white vs. λ light. When the luminance of the λ light is low (top panels), the ganglion cell responds to the onset of the "white" light (note the timing of the frequency histograms). When the luminance of the λ light is high (bottom panels), the cell responds at the onset of the λ light (note the shift in timing of the histograms). In other words, the cell responds to an increase in luminance in each instance. But at the balance point of the white and λ lights, the neuron is not quite nulled, but instead produces a frequency doubled response, responding at the onsets of both the λ and the "white" lights (note four clusters of spikes in the frequency histograms). This occurrence is called a frequency-doubling non-linearity, and it represents a deviation from the predictions of any linear model. [After Lee, Martin, and Valberg (1989)]

14.7 A terminological merry-go-round

Finally, its time for a terminological confession. Earlier in this book, we have noted the fact that two groups of scientists sometimes both study the same thing and name it with different names. For example, the anatomists flat bipolar is the physiologists OFF-center bipolar, and the anatomists invaginating bipolar is the physiologists ON-center bipolar.

The names we have been using for ganglion cell types – midgets, parasols and small bistratified cells – are anatomists names. It turns out that physiologists also have names for these different primate ganglion cell types. As we will see, these names derive from the layers of the LGN to which these cell types project. The LGN has *parvocellular* (small-cell) layers to which the midget ganglion cells project, and *magnocellular* (large-cell) layers to which the parasol cells project. Thus, physiologists call the midget cells *P cells* (P for parvo), and the parasol cells *M cells* (M for magno). Note the reversal of initials set up to trap the unwary: Midget = P, and Parasol = M.[4]

The small bistratified ganglion cells are now thought to project to the koniocellular (very small cell) layers of the LGN, and some authors now call them *K cells.* Both sets of names are included in the summary table (Table 14.1) below. As we move beyond the retina to the LGN, the physiologists' nomenclature, which derives from LGN anatomy, will serve us better.

14.8 Summary: The retinal output code

The primate retina contains about 100 million photoreceptors, and each of them initiates a signal about quantum catches in a tiny retinal region. But not all of this information need be preserved. We don't care about quantum catches in photoreceptors; we care about physical objects and their properties. The retinal recoding must replace the photoreceptor code with a code that at once provides information to the cortex in the most useful possible form, and is compact and efficient enough to be pumped through the one million fibers of the optic nerve. And there are theoretical arguments that codes of the kind we have in fact provide optimal codes for these purposes.

At the end of Chapter 8 we asked, what is explicit inthe retinal output code? We argued that in shorthand, the what the ganglion cells make explicit is *local contrast, in space and time.* In the primate retina, things are more complex because we have included wavelength as well as space and time. But color-opponent neurons can be seen as signaling wavelength differences – being active when one wavelength band but not another is present in the retinal image. Thus, we will modify but not abandon our shorthand slogan: what primate ganglion cells make explicit is *local contrast, in space, time, and wavelength.*

The issue, however, is complicated, because rather than being signalled independently, these three factors are multiplexed in the signals of midgets, parasols, and small bistratified ganglion cells. The properties of these three types of ganglion cells are summarized in Figure 14.13 and Table 14.1. Based on these properties, what we can speculate about which aspects of our spatial, temporal, and spectral vision are served by each type of ganglion cell?

Midget (P) cells have very small receptive fields that are both spatially and spectrally opppo-nent. There are four subtypes: L-ON-center, M-ON-center, L-OFF-center, and M-OFF-center. In the spatial domain, their small receptive field sizes suggest that they provide us with high acuity and our best sensitivity to *high* spatial frequencies. In the temporal domain they are sustained, and

[4] Confucius says, all is confusion!

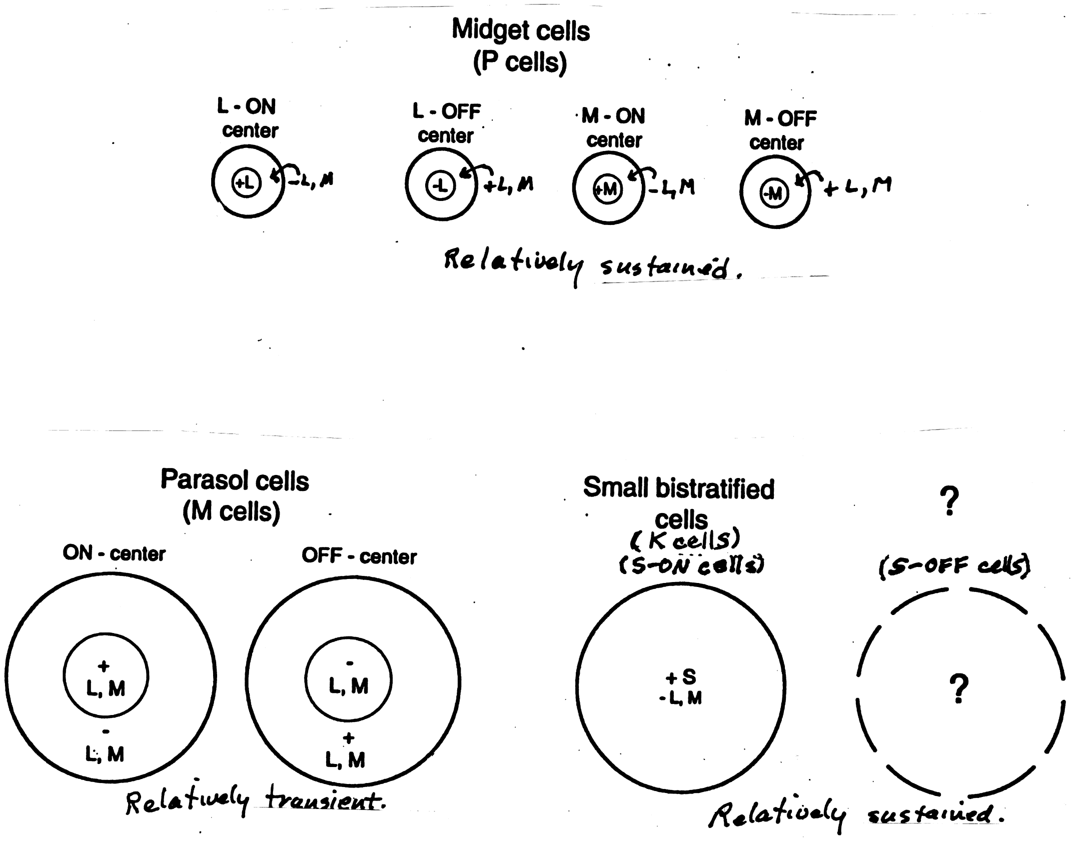


Figure 14.13: Summary of the code in the major retinal outputs bound for the visual cortex. Seven (or eight) types of neurons encode all of the spatial, temporal, and chromatic information that leaves the retina.

Anat Type	Physiol. Type	Cone Input	Spectrally Opponent?	RF Size	Spatially Opponent?	Temporal Property
Midget	P	L, M	Yes	Small	Yes	Sustained
Parasol	M	L, M	No	Large	Yes	Transient
Small Bistrat.	K	S, M, L	Yes	Large	No	Sustained

Table 14.1: Properties of midget, parasol, and small bistratified cells.

probably provide us with our best sensitivity to *low* temporal frequencies. In the spectral domain, they probably provide L - M opponent signals, but carry an L + M signal as well.

Parasol (M) cells have relatively large receptive fields that are spatially but not spectrally opponent. There are two subtypes: ON-center and OFF-center. In the spatial domain, they probably provide our best sensitivity to *low* spatial frequencies. In the temporal domain, they are relatively transient, and probably provide our best sensitivity to *high* temporal frequencies. In the spectral domain, they sum L and M cone inputs, and probably provide a second L + M signal.

Small bistratified (K) cells have receptive fields about the same size as those of parasol cells. They are spatially non-opponent but spectrally opponent, with S cone inputs providing an ON response and mixed L and M cone inputs providing an OFF response. In the spatial domain they are probably most sensitive to low spatial frequencies. In the temporal domain they are relatively sustained. In the spectral domain they provide S - (L + M) opponent signals.

Because of their various combinations of cone inputs and ON vs. OFF signals, these three types of ganglion cells make seven subtypes: four types of midget (P) cells, two types of parasol (M) cells, and one type of small-bistratified (K) cell. These seven cell types are summarized in Figure 14.13. In addition, we expect that an S-OFF system will be discovered, and a hypothetical S-OFF ganglion cell is also included in Figure 14.13. Because the four types of midgets make only two ensembles (ON-center and OFF-center) with coverage fctors of 1, we can argue that these eight cell types produce only six ensembles.

In thinking about the retinal output, then, we can think of the retina as covered by six superimposed ensembles of ganglion cells. Within each ensemble, receptive fields get larger with increasing distance from the fovea. Six ensembles, with a total of a million output lines, charged with encoding the most important spatial, temporal, and spectral information from an infinity of possible retinal images, in enough detail to keep us abreast of events in the physical world.

Finally, on the basis of the cone inputs, receptive fields, and sustained/transient properties of midgets, parasols, and small bistratified cells, we have speculatively assigned probable combinations of visual functions to each of these three kinds of neurons. But how could we test these assignments further? One approach would be to make differential lesions – to eliminate the different ensembles one by one, and see what visual functions are disrupted. But at the retinal level, the ensembles of neurons are all intermixed, and differential lesions are hard to come by.

Occasionally the EDC is kind to vision scientists. As we will see in Chapter 15, these different cell populations are sent to partially separate spatial locations within the LGN. Thus, small LGN lesions, in combination with behavioral assessments of the effects of the lesions on visual function, will allow us to explore these functional assignments more directly.

Chapter 15

The Lateral Geniculate Nucleus (LGN): Gateway to the Visual Cortex

In the first half of this book, we developed the concept that the visual system carries out a series of code transformations. We have divided up the visual processing so far, somewhat arbitrarily, as transformations from the world to the retinal image (the First Transformation); from the retinal image to the quantum catches of four kinds of photoreceptors (the Second Transformation); and from the photoreceptor quantum catches to the ganglion cell outputs (the Third Transformation).

In the present chapter we will finally escape from the retina, and embark on an exploration of the parts of the visual system that lie beyond it. In this and subsequent chapters we will emphasize the pathways, connections and physiological responses of visually driven neurons within the brain. The journey takes us first to the two lateral geniculate nucleii (LGN), the major nucleii to which the axons of retinal ganglion cells project. LGN neurons in turn project to the visual cortex, initially to the first or primary visual area V1, and beyond V1 to a wide range of cortical areas and processing stages.

In the present chapter we describe the patterns of projections of retinal ganglion cell axons to the LGN, and the profound anatomical rearrangements that are carried out by this anatomical pathway. We then turn to what we have chosen to call the Fourth Transformation: The transformation of the visual code between the input and the output of the LGN. As it turns out, surprisingly, the physiological transformations are relatively minor, as though despite the huge anatomical rearrangements, very little in the way of computation occurs within the LGN. In fact, historically the LGN has often been termed a ”relay station”, as though it relays signals from retina to cortex without changing the physiological code. We will evaluate this historical view at the end of the chapter.

In the meantime, it’s time to reintroduce some larger issues. By now you have probably come to expect that beyond the retina we will encounter a continuing series of code transformations, and you are right. But here we arrive at the toughest and most fundamental questions in all of visual science. Why are these code transformations useful? Is there a trend? What do you expect will be the ultimate form of the visual code? How do our brains ultimately represent scenes and objects, locations and distances, brightnesses and colors, and allow us to see and respond to them? Might a different cell be active for each different object we perceive? Or might the pattern of responses in a much larger set of neurons be needed to mediate the perception of a particular object? At which level of the visual system are neural signals mapped to conscious perception, and by what mapping rules? And how are neural signals mapped to action? What are the alternatives? Give these question some thought and keep them in mind as we go along, as we will be exploring them throughout the remainder of the book.

The subject matter of this chapter also provides some grist for the mill of linking propositions. We will point out a few examples as we go along.

15.1 Retinal outputs: Where do retinal ganglion cell axons go?

15.1.1 The optic disk and the blind spot

Since the retina is in "backwards" – the ganglion cells lie on its inner surface – there is no elegant way for the ganglion cell axons to get out of the eyeball. The axons of all types of retinal ganglion cells travel across the inner surface of the retina, join together in a bundle and plunge inelegantly through the retina. The place at which the axons exit the eyeball is called the *optic disk.* The optic disk subtends an angle of about five degrees – five thumbnails at arm's length! – centered at about 15 degrees nasal to the fovea in each eye.

Because there can be no photoreceptors at this location, the optic disk causes a perceptual phenomenon called the *blind spot.* Remarkably, most people have never noticed that they have blind spots, partly because the blind spots in the two eyes do not coincide – each eye loses a different piece of the incoming scene, and each covers for the other. Figure 15.1 provides a demonstration of the blind spot in your left eye. [Notice that explaining the blind spot (a perceptual fact) by means of the optic disk (an anatomical fact) is a causal story, and ferret out its linking proposition involved in it.]

15.1.2 A brief tour of subcortical destinations

Once they leave the retina, the axons from retinal ganglion cells go to several destinations. For our friends the M, P and K ganglion cells, the most important of these are the two (*dorsal*) *lateral geniculate nucleii (LGN).* The LGN in turn project to the visual cortex. The signals carried by this pathway underlie most of the functions that we think of as vision, including object recognition and conscious visual perception. We will return to this pathway immediately below.

Axons of other less common types of ganglion cells project to a wide variety of subcortical destinations. Some of these destinations are shown in Figure 15.2. These circuits serve a wide range of reflex-like visual functions.

The structure that receives the second highest number of fibers from the retina is called the *superior colliculus.* In addition to its visual input, the superior colliculus receives spatially ordered arrays of inputs from other sensory systems, including auditory and tactile maps. This nucleus thus combines information from many senses to create a single common map of visual space. Output from the superior colliculus is critical to controlling eye movements and orienting responses, and serves to orient the organism to regions of interest in the visual world.

A few retinal ganglion cells also project to a region called the *pretectum.* The pathway through the pretectum serves various light reflexes, including the pupillary light reflex (the constrictions and dilations of the pupil in response to increases and decreases in light levels). Other subcortical destinations include the *accessory optic nucleus*, which contributes to the control of eye movements

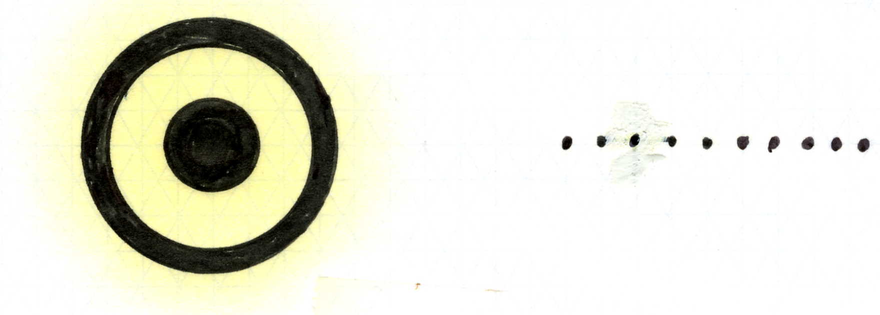


Figure 15.1: The blind spot. Hold your head about 10 inches from the figure. Fixate one of the dots at the right of the figure, and close your right eye. By changing your fixation along the row of dots and varying the distance of your head from the page, you should be able to find a location at which the central dot in the figure disappears. When this happens, the dot is falling on the blind spot in your left eye. If you open your right eye, the spot will reappear.

and image stabilization; the *suprachiasmatic nucleus*, which contributes to the control of circadian rhythms, and the *pregeniculate* or *ventral LGN*, which participates in additional visuomotor functions.

15.2 Anatomical projections: From two retinas to two lateral geniculate nucleii

Up until now, we have been considering one eyeball and one retina at a time. But it turns out that the inputs from the two retinas converge at the two LGN. Before we can make sense of these projections, we need to consider the optics and anatomy of the two eyes as they work together, and establish some landmarks and some terminology.

15.2.1 One visual field, two retinal images

Optical mappings from the visual world to the two retinal images are shown in Figure 15.3. Figure 15.3A shows that, as stated earlier, the part of the world that you can see – the *visual field* – covers about 180 degrees of visual angle. We further subdivide the visual field into the *left* and *right visual hemifields*, located to left and right of the fixation point F.

Now suppose that you are fixating an object located at F. The point F will be imaged on the foveas of both your left and right eyes, and inverted images of objects in the visual field will be formed on your two retinas. The central 120 degrees or so of the two images overlap, and this

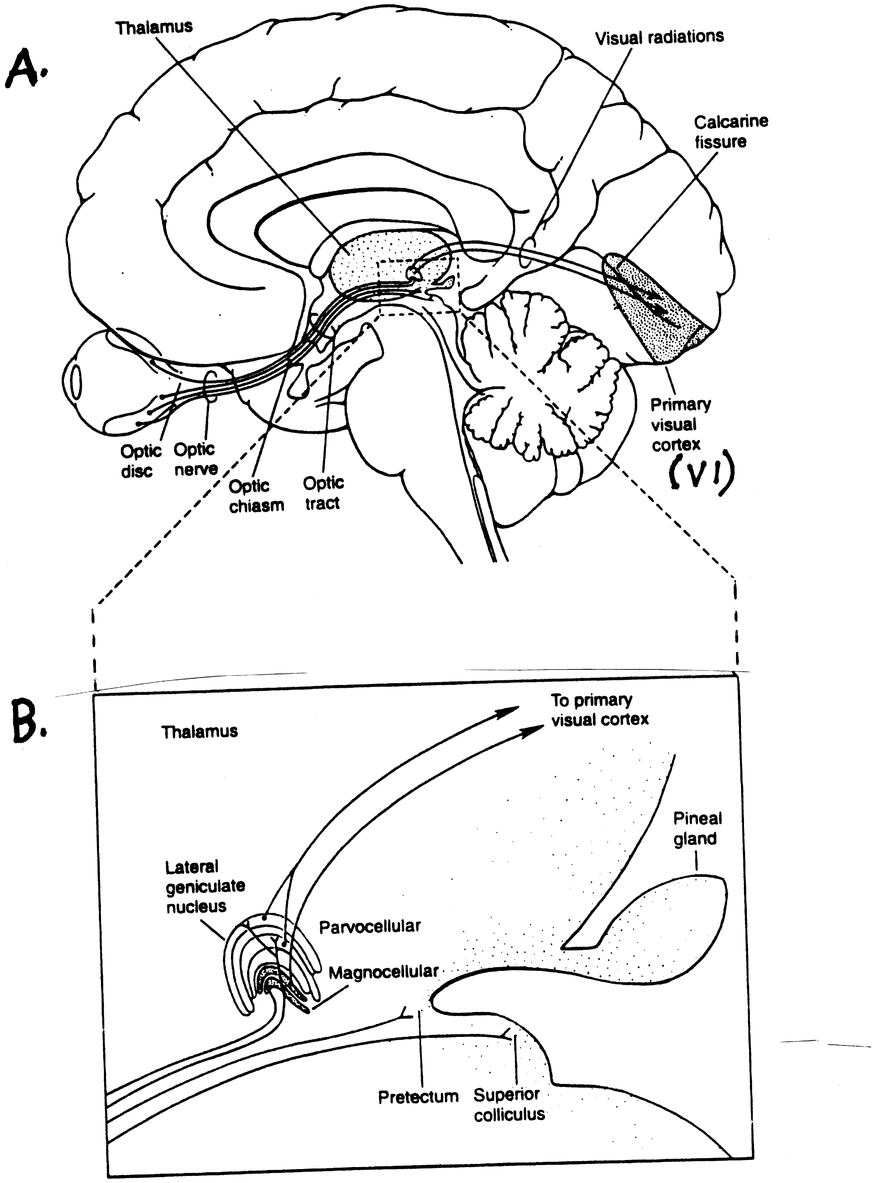


Figure 15.2: Projections of ganglion cell axons. A: A schematic diagram showing some of the anatomical locations to which ganglion cell axons project. The major projection is to the lateral geniculate nucleus, and from there to the primary visual cortex (area V1). B: Destinations in the lateral geniculate, pretectum and superior colliculus are shown at higher magnification. (After Kandel, Schwartz, and Jessel, 1991, Fig. 29-4, p. 423.)

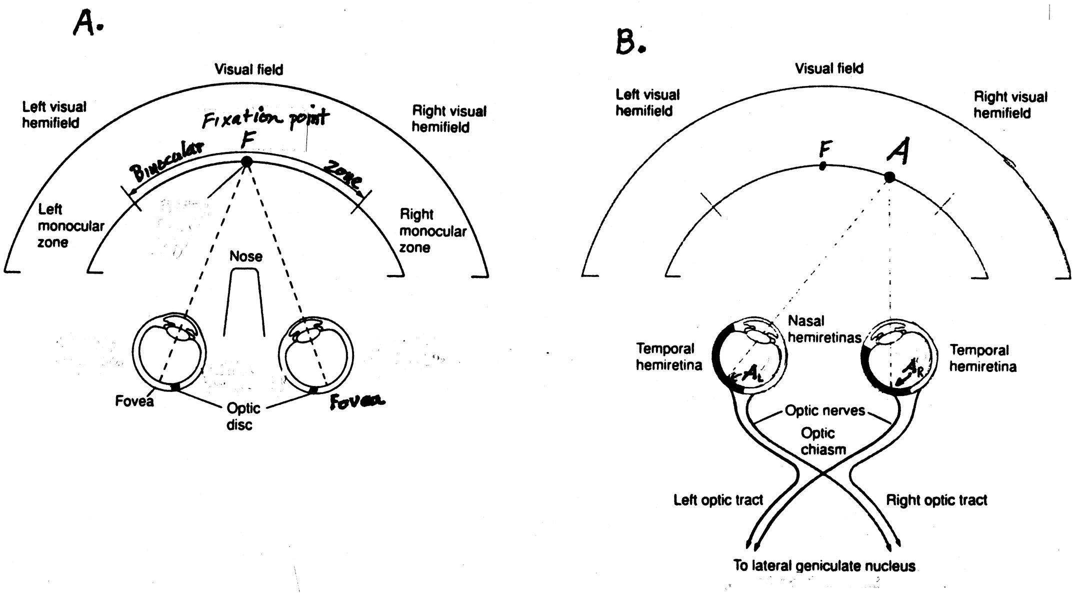


Figure 15.3: The visual field and the two retinal images. A: The visual field, showing the left and right visual hemifields, the binocular zone and the two monocular zones. Two nearly identical images of the binocular zone are formed on the two retinas. B: Projections from the two eyes to the two lateral geniculate nucleii. Ganglion cell axons from the two nasal hemiretinas cross at the optic chiasm, while those from the two temporal hemiretinas do not cross. In consequence, the left LGN gets two neural images of the right visual hemifield (symbolized by the two images of point A, A_L and A_R), and the right LGN gets two neural images of the left visual hemifield. (After Kandel, Schwartz and Jessel (1991), p. 421.)

region is called the *binocular zone*; that is, you have two nearly identical pictures of the central 120 degrees of the visual field, one on each retina. However, your nose cuts off your view of the far right side of the visual field for your left eye, and the far left side of the visual field for your right eye. Thus, in the far peripheries of the two eyes there are two *monocular zones*, each subtending about 30 degrees of visual angle.

You can readily demonstrate these monocular zones to yourself. Fixate an object in front of you. While continuing to fixate, put your left index finger somewhere in your left visual field, and move it peripherally until just before it disappears from view. Now close your left eye – your finger should disappear. Now move your finger back toward the center of your visual field until it reappears. The difference between the two finger positions defines the left monocular zone.

When we come to describing the projections from the eye to the brain, we will also need terms to refer to the left and right sides of the two retinas, as shown in Figure 15.3B. Fixation is still at F. The inner sides, toward the nose, are called the *nasal hemiretinas*, while the outer sides, toward the temple[1], are called the *temporal hemiretinas*. Because the retinal images are inverted by the optics of the eye, the left side of the visual field is imaged on the right side of each retina – on the nasal retina in the left eye, and the temporal retina in the right eye; the reverse holds for the right side of the visual field. Thus, an object at point A in the right visual field will be imaged on the left halves of both retinas at points AL and AR.

The main point here is that across the binocular zone the optics of the two eyes create two highly similar optical images on the two retinas. Moreover, the two retinas create two highly similar *neural* images of these two optical images. Common sense suggests that it would be useful to bring these two similar images to the same place, and use them together in one way or another at later stages of processing.

15.2.2 The optic chiasm: Cutting the world in half

The bundle of axons that exits each eyes creates one of the two *optic nerves*, as shown in Figure 15.3B. The sorting out of the left and right hemifields is accomplished at the *optic chiasm* (chi = greek letter χ; a crossing). At this location, each optic nerve splits: axons from the nasal retina cross to the *opposite (contralateral)* side of the brain, while those from the temporal retina do not cross, but remain on the *same (ipsilateral)* side of the brain. This half-crossing puts the two neural images arising from the left visual field together on the right side of the brain, and the two neural images arising from the right visual field together on the left side of the brain. The two axon bundles that emerge from the optic chiasm are called the *optic tracts*, and each carries two neural images of the same visual hemifield. This simple anatomical maneuver sets the stage for later joint processing of the images from the two eyes.

[Notice, though, that this arrangement has a potential cost: the left and right halves of our visual field are represented separately, on opposite sides of the brain. Here's a question to ponder: do the two halves of the neural image need to be reunited physiologically, within the brain, to give us a single perceptual world? If not, how does the seamless, unified perceptual world come about? What linking propositions are involved in the possible answers to this question?]

[1] Notice that in visual science the term *temporal* is used with two entirely different meanings: a) toward the temple (the side of your forehead, at the outer edge of your eye), as in *temporal retina*; and b) having to do with time, as in a *temporal variation* of a visual stimulus.

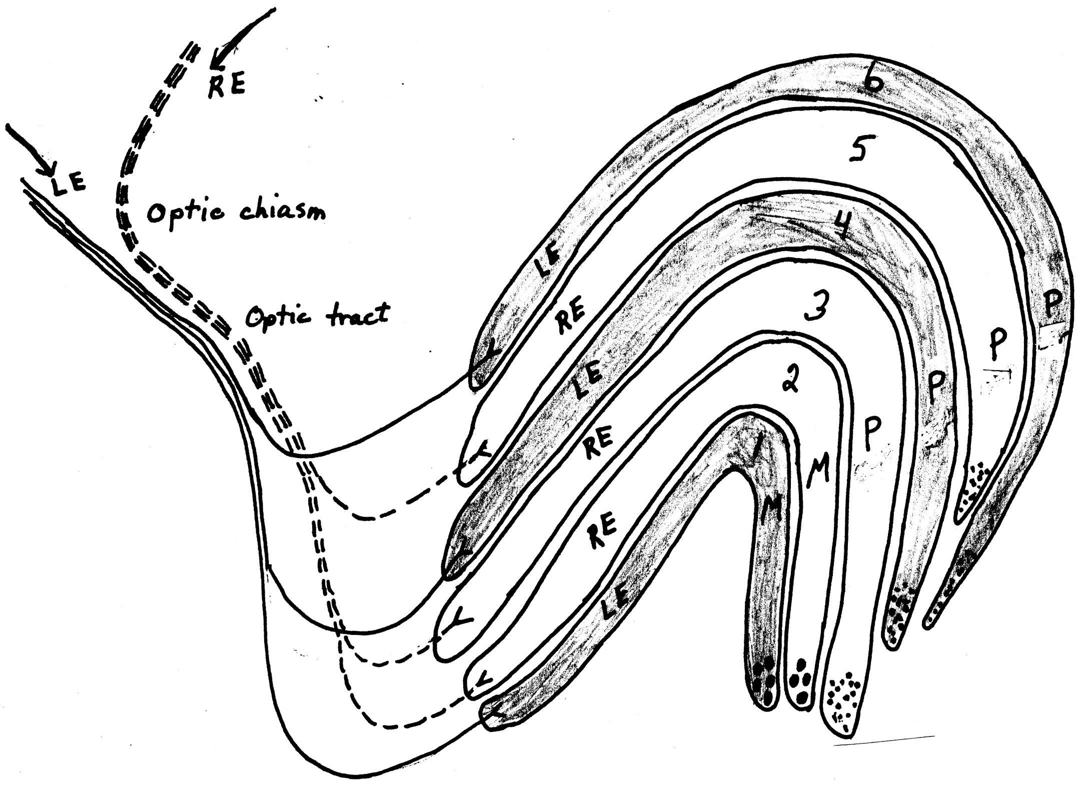


Figure 15.4: Classical textbook description of the right LGN of a macaque monkey (schematic). The layers of the LGN are numbered 1-6 from bottom to top. Layers 1 and 2 are the M (magnocellular) layers, with large cells; layers 3-6 are the P (parvocellular) layers, with small cells (the cell size difference is shown exaggerated at the right hand end of each layer). The left (contralateral) eye projects to layers 1, 4, and 6 (shaded), and the right (ipsilateral) eye to layers 2, 3, and 5.

15.3 More anatomy: Layering and mapping in the LGN

As shown in Figure 15.3, each optic tract projects to its corresponding LGN – the left optic tract to the left LGN and the right optic tract to the right LGN.

15.3.1 The M and P layers: Six retinotopic maps

A classical schematic cross-section of a primate (macaque or human) right LGN is shown in Figure 15.4. The LGN is a layered structure that becomes bent during embryology (*genu* = knee, as in *genuflect*, to bend the knee). Like the retina, the remarkable thing about the anatomy of the LGN is its orderliness: of its layers, its inputs, and its maps of the visual world.

Most impressively, the LGN is a layered structure. The classical description of the primate LGN is that it has six layers, numbered 1-6 from bottom to top. The two bottom layers, 1 and 2, contain neurons with large cell bodies, and are called the M or *magnocellular layers* (magno = large). The four upper layers, 3 through 6, contain neurons with small cell bodies, and are called the P or *parvocellular layers* (parvo = small). These names finally reveal the origins of the physiologists' names for the two main types of ganglion cells that project here: the M (parasol) ganglion cells project to the *m*agnocellular layers of the LGN (layers 1 and 2), and the P (midget) ganglion cells project to the *p*arvocellular layers (3-6). Notice that the P layers are on the outside of the curved LGN structure, and that there are four P layers but only two M layers. This arrangement nicely accommodates the much larger number of P as opposed to M ganglion cell axons coming from the retina.[2] (We will come to the SBS (K) ganglion cells later.)

In addition to its impressive layering, the inputs to the different layers of the LGN are also highly organized. Remember that each LGN receives ganglion cell axons from both eyes – from the nasal retina of the contralateral eye, and from the temporal retina of the ipsilateral eye. But the inputs from the two eyes remain anatomically segregated within the LGN, as shown in Figure 15.4. The contralateral eye projects to layers 1, 4, and 6, whereas the ipsilateral eye projects to layers 2, 3 and 5.

15.3.2 Maps and "magnifications"

A third aspect of order in the LGN lies in its maps of the visual field. Each layer of the LGN contains an orderly *topographical map* – also called a *retinotopic* or *visuotopic map* – of one half of the visual world, and one half of the retina. The maps in the six layers of the LGN preserve the nearest neighbor relationships of retinal ganglion cells of like types – two M-cell maps and four P-cell maps. Notice, though, that an M cell and a P cell that were initially nearest neighbors in the retina do not remain nearest neighbors in the LGN, but rather project to different layers.

A fourth aspect of order in the LGN is that the maps in all six layers are aligned with each other. This fact is captured by the *club sandwich analogy*: if we think of the LGN as a club sandwich (layers 1 and 6 are the bread, etc.), and run a toothpick down through all six layers, we would run through six visuotopic maps. A properly oriented toothpick would go through six representations of the same point in the visual field. That is, a single object in one location in the visual field creates activity in six vertically (or radially) aligned locations in the LGN. [Think about it – the representation of a single object is now distributed among at least six different and spatially separated sets of neurons. We will have to bid goodby to the notion that a single object is represented by a single neuron or a single neural location, at least at this level of the visual system.]

Finally, although the six maps in the LGN preserve nearest neighbor relationships, they are grossly distorted overall, as shown in Figure 15.5. The representations of the fovea and the central retina are greatly stretched, or "magnified", and the representations of more and more peripheral areas increasingly shrunk. This differential *magnification* of the fovea probably comes about for a simple reason: there are many more ganglion cells per unit area in the fovea than in the periphery. If the projection from each retinal ganglion cell took up an equal region in the LGN, one would predict at least approximately the observed pattern. Again, the fovea is represented by many neurons, allowing for the preservation of fine spatial detail. [But why don't we perceive an object

[2] It has sometimes been suggested that two of the P layers receive inputs from ON-center P ganglion cells, and the other two from OFF-center P ganglion cells, but this observation has been challenged. If it is not true, the reason for four P layers is unknown. In fact, the primate LGN only has six layers within the central 17 degrees or so of the visual field. In more peripheral regions, the two P layers from the same eye merge, reducing the total to just two P layers.

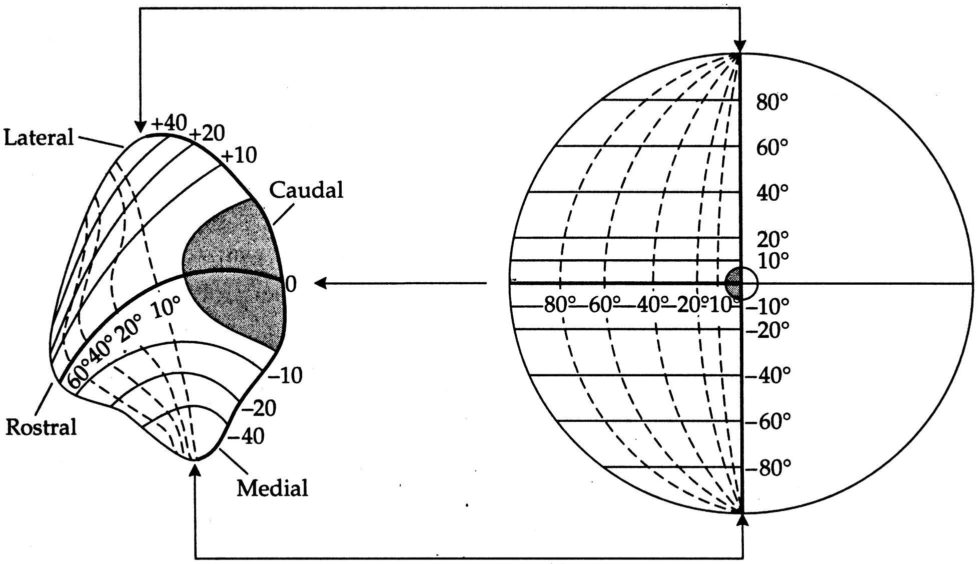


Figure 15.5: The retinotopic map in the right LGN. The right panel shows a face-on view of the left retina, calibrated in degrees of visual angle with respect to the fovea. The left panel shows a top view of the right LGN. The fovea projects to the back (caudal) part of the LGN, and its input is spatially "magnified" in the LGN map. More and more peripheral retinal regions are represented more and more toward the front (rostral) part of the LGN, and take up less and less space. Thus, the right LGN carries a representation of the left half of the visual field, "magnified" in the fovea and compressed in the periphery. [From Oyster, 1999, Fig. 5.9, p. 203.]

as changing size as it moves from fovea to periphery? What false linking proposition can we reject here?]

15.3.3 The K layers: Six more maps

Up to this point, we have been telling the classical story of the LGN. But you may have noticed a discrepancy between our account of the retina, with its three major types of LGN-projecting ganglion cells, and our account of the LGN, with its two kinds of layers. In fact, the P ganglion cells project to the LGN P layers, and the M ganglion cells project to the LGN M layers; but where do the small bistratified, or K ganglion cells, go?

In recent years it has become more and more clear that there is actually a third kind of layer in the LGN – the *koniocellular* or *K layers*, containing neurons even smaller than those in the P layers. As shown in Figure 15.6A and B, classical staining techniques did not reveal the K layers, but recent studies using immunoreactive stains reveal them clearly. In macaque monkeys, six K layers occur – one just beneath each of the M and P layers.

What kinds of retinal ganglion cells project to the K layers? In macaques, the K layers are thin and so interspersed with the M and P layers that they are difficult to study. But another primate species, the marmoset, has two major K layers that are well segregated from the M and P layers. As shown in Figure 15.6C, one of these layers lies between the M and P layers, and the other lies below the M layers.

Paul Martin and his colleagues (Martin, White, Goodchild, Wilder, and Sefton, 1997) recorded from neurons in the marmoset LGN using null plane analysis. Whenever they encountered a neuron with a null plane appropriate to a small bistratified ganglion cell, they made a small lesion in the LGN. As shown in Figure 15.6C, most of the lesions were located in the two major K layers. Thus, at least in the marmoset, it turns out that signals from small bistratified ganglion cells project to the K layers of the LGN[3].

In summary, anatomical studies tell us that in the projection from the two eyes to the two LGN in primates, the incoming neural image is split in half, and each half is sorted into twelve separate, retinotopically organized neural half-images – four P layers, 2 M layers, and six K layers. From the anatomist's perspective, the LGN accomplishes a profound rearrangement of the proximities of cells, grouping cells of different kinds into twelve separate two-dimensional sheets. Nearest-neighbor relationships among like cells (M, P, or K) are preserved by the layering of the LGN, and nearest neighbor relationships across different kinds of cells are preserved by the radial alignment of the layers.

15.4 Physiology

We now turn to the single cell physiology of the LGN. If we put a recording electrode next to a LGN cell, and present various stimuli on the retina, what will we find? What will the receptive fields and

[3] Calling the small bistratified ganglion cells K cells, as we did in Chapter 14, is probably a mistake, because the K layers of the LGN almost certainly also receive input from other, less common types of ganglion cells, and it would be confusing to call them all K cells. The different K layers may also have different specializations; for example, some sources suggest that in macaques it is only layers K3 and K4 (the K layers that lie below layers 3 and 4) that receive input from small bistratified ganglion cells. Nonetheless, the practice of calling small bistratified ganglion cells K cells is catching on, and we will not resist it.

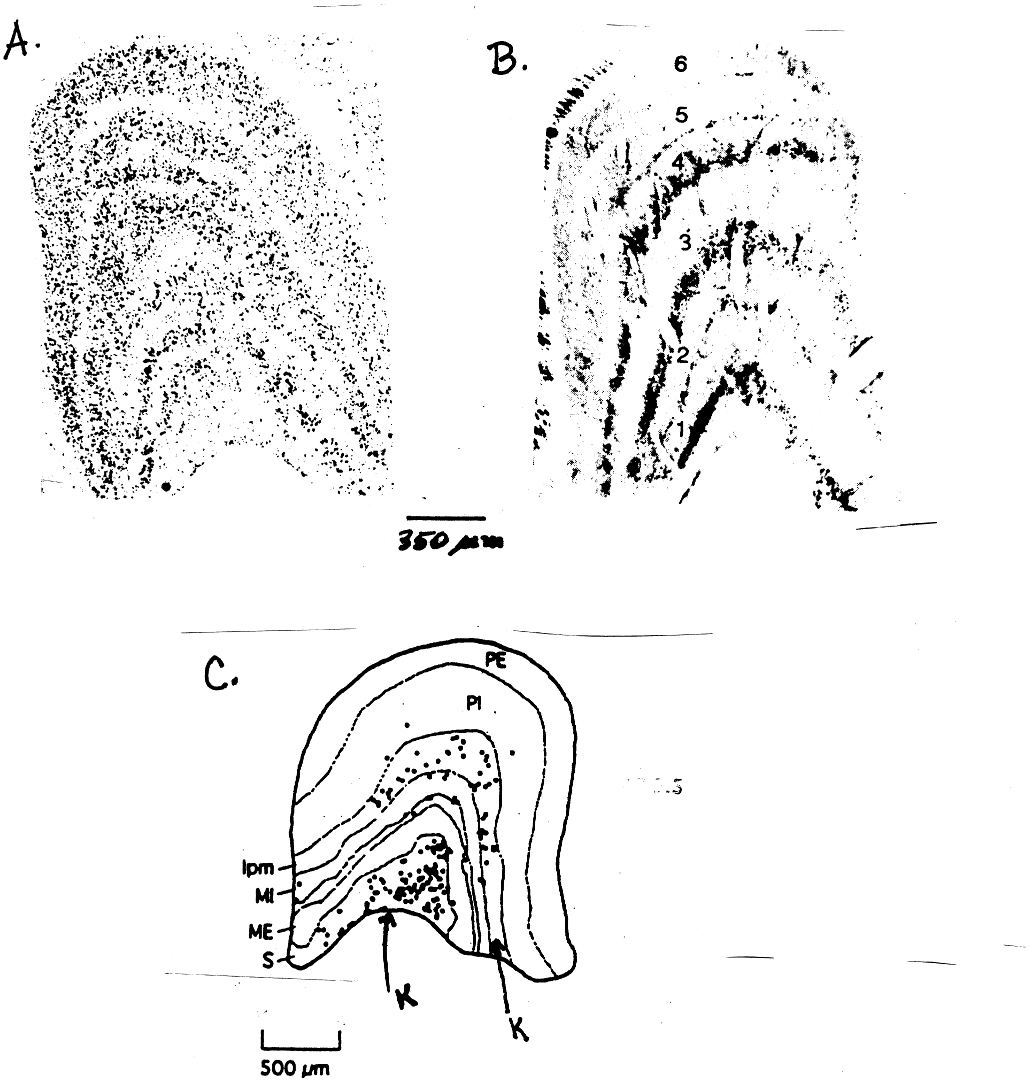


Figure 15.6: The K layers. A: A classical light micrograph of the macaque monkey LGN stained with a stain called thionin, showing M and P layers, with bare spaces between the layers. B: What a difference a stain makes! An immunoractive stain (CaM II kinase) reveals the K layers. C: Marmoset LGN, showing the locations of neurons with null planes that correspond to those of retinal K cells. These neurons are largely confined to the K layers in the LGN. (A and B from Hendry review; C from Martin et al, 1997, Fig. 1, p. 1537).

other properties of LGN neurons be like? As it turns out, each LGN cell receives major input from only one or a very few retinal ganglion cells. Thus, we might guess that the physiological responses of LGN cells are very much like those of the retinal ganglion cells from which they receive their input; and this is indeed the case.

15.4.1 Receptive fields

Like retinal ganglion cells, LGN M and P neurons have receptive fields that show center/surround antagonism. Receptive field sizes correspond quantitatively to those of retinal ganglion cells – smaller for P than for M cells at each retinal eccentricity, and increasing with eccentricity for both cell types. The biggest change in receptive field properties is that LGN cells may have a closer balance between center and surround processes, so that they respond somewhat less than do ganglion cells to homogeneous fields of light.

Additional characteristics of LGN neurons vary with the layer in which they are found (see Table 14.1 for a summary). Cells in the magnocellular layers, like the M ganglion cells that provide their input, give transient responses at the onset and/or offset of light within their receptive fields, and respond only minimally to isoluminant chromatic patterns. Responses of cells in the parvocellular layers are more sustained, like the P cells that provide their input, and these neurons respond well to chromatic modulation. Responses in the K layers are more heterogeneous, and have been less studied, but so far many LGN K cells mirror the properties of their retinal K (small bistratified) cell input.

15.4.2 Chromatic codes

In line with our emphasis on color vision, the chromatic properties of LGN P cells are of particular interest. In 1984, Andrew Derrington, John Krauskopf, and Peter Lennie carried out a pioneering study of the chromatic properties of these cells. These authors were the first to use null plane analysis to sort out the cone inputs to postreceptoral neurons. They showed that LGN P neurons came in two homogeneous types that corresponded well with the two types of chromatically opponent cells called for in the Boynton code – L - M and S - (L + M) cells. Derrington et al's data were thus another major early source of support for the Boynton model. Later recognition of the K pathway allows us to argue that LGN cells that carry an L - M signal would still be indentified as LGN P cells, whereas those that carry an S - (L + M) signal would now be identified as LGN K cells. The three channels of the Boynton code can thus be traced through the LGN via the M, P, and K cells, and from there on upward to visual cortex.

15.4.3 Segregation of left eye and right eye signals: So near and yet so far

Finally, what about monocularity vs. binocularity? Up until now, all of the neurons we have studied have been within the eyeball, and naturally *monocular* – driven by inputs from one eye only. But as we will see, many individual cortical neurons are *binocular* – they receive inputs from both eyes. The question then arises, what is the first level of the visual system at which individual neurons are binocularly driven?

The EDC seems to have taken great pains to juxtapose the retinotopic maps from the left and right eyes in precise register in adjacent layers in the LGN (remember the club sandwich analogy). On this basis, one might guess that individual LGN neurons would receive inputs from both eyes. However, surprisingly, single unit recordings show that there is little physiological interaction between neurons in the left eye and right eye layers. Although there are exceptions, most LGN neurons are monocular under most conditions. We will speculate on this seemingly odd design feature – juxtaposed layers but segregated signals – near the end of the chapter.

15.5 Lesion-and function studies and parallel processing

In the 1980s, the concept of parallel processing provided a novel and exciting perspective in visual science. In the early excitement, some vision scientists proposed very strong versions of parallel processing, arguing that the neural substrates of different visual functions are strongly segregated among different cell types or different neural channels (see Chapter 19). Many vision scientists were attracted to experiments designed to support or probe the limits of this view, and some of these studies took the form of lesion-and-function studies at the level of the LGN.

Why the LGN? In the retina, M, P, and K ganglion cells are intimately interspersed, and it is impossible to make a localized lesion of one cell type without destroying the others. But in the primate LGN, the EDC laid a gift at the feet of vision scientists. M and P LGN cells are segregated conveniently into separate groups of layers, with the result that it is possible to create an experimental lesion confined to magnocellular or parvocellular neurons[4].

It is also possible to test the experimental animal psychophysically before and after making the lesion, to find out which visual functions are lost after a particular lesion, and which are retained. This elegant experimental approach will be called the lesion-and-function paradigm.

15.5.1 The logic of lesion-and-function studies

Historically, lesion studies were extensively used in early neuroscience, but they subsequently went out of favor for several decades. The reason is that before the anatomy and physiology of the brain were understood with much precision, and before our ideas of visual function were well sorted out, lesions were a remarkably blunt instrument. However, over the decades have come an increasingly detailed understanding of both vision and the visual system, and increased precision in our capacity to lesion certain minute structures and spare others. In consequence, lesions can now be used with remarkable precision, and they have come back into the toolbags of visual neuroscientists. Thus, it is worthwhile to take a few paragraphs and examine closely the logic involved in the lesion-and-function paradigm.

The LGN has been a particularly attractive site for lesion-and-function studies, for two reasons. As mentioned previously, the first reason is its layered structure and topological maps. These features make it possible, with a combination of skill and luck, to make a lesion confined to just the P layers, or just the M layers, in one small region of visual space. One can test a visual function (such as color vision or flicker resolution) before the lesion, and then again after the lesion, in just one small region of the visual field, reserving the other regions as within-animal control sites. Changes in function confined to the part of the visual field corresponding to the lesion would implicate the lesioned cells as having a major role in supporting that particular visual function.

[4] The studies we will describe were carried out before the K layers had attracted attention, and K cells are not mentioned in the original papers described below. Following the original papers, and for the sake of simplicity, we omit consideration of the K layers until the end of this section.

The second reason is a deeper logical one, and centers on the philosophical gambit of defining causality in terms of necessary and sufficient conditions. Compared to most other levels of the visual system, the LGN has a charming simplicity. All of the information that goes to the visual cortex must pass through one of the layers of the LGN – M or P (again ignoring the K layers and a few other complicating factors for the sake of simplicity). Thus, at least one of these cell types must be *necessary* to provide the neural basis for any given visual function (if both channels were interrupted, all visual functions would be eliminated). Moreover, in principle the same cell type can be *sufficient*[5] to support a particular visual function at the level of the LGN. In such a case, the other cell type is neither necessary nor sufficient. Alternatively, it could be that either cell type is sufficient, and neither is necessary; or that neither cell type is sufficient, and both are necessary.

A layout of the logic of LGN lesion-and-function studies in terms of necessary and sufficient conditions is shown in Figure 15.7. You are invited to sort out the results of lesion studies in these terms as we go along.

15.5.2 Lesion-and-function studies and the LGN

In 1990, Peter Schiller, Nikos Logothetis, and Eliot Charles published the results of an extensive set of lesion-and-function experiments. Schiller and his colleagues first trained macaque monkeys to perform in an eight-alternative forced-choice task. As shown in Figure 15.8A and B, the monkey was shown a visual display in which a target could be presented at one of eight different locations arranged around a fixation point. The monkey was trained to fixate the fixation point between trials, and to move its eyes to the location of the target when the stimulus were presented. Once they had been trained, the animals could generate several thousand trials per day, and many different detection and discrimination tasks could be tested in the same animal.

The next step was to make the lesions. A small drug delivery tube was lowered into the monkey's brain. A chemical called *ibotinic acid* was injected into a tiny region intended to lie within either the M or the P layers of the LGN, in order to destroy a small region of cells confined to that layer. To make the maximal use of each animal, a lesion was typically made in one type of layer at one eccentricity in the left LGN, and in the other type of layer at a different eccentricity in the right LGN. For example, the monkey whose brain is shown in Figure 15.8C and D had a P layer lesion near the fovea in the left LGN and an M layer lesion in a more peripheral location in the right LGN.

The monkey was then retested behaviorally with the same set of visual tasks, with the separations and locations of the stimuli chosen to place one of the targets within a given lesioned region. If the paradigm is working properly, the monkey should show losses of M-cell-mediated functions only at the location corresponding to the M cell lesion, and normal M-cell-mediated functions at all other retinal locations; and similarly for P-cell-mediated functions.

[Take a minute to predict the outcomes of this experiment. Remember that only half of the visual field is projected to each LGN. In which half of the visual field should the animal show losses of M-cell-mediated functions? P-cell-mediated functions? Take a guess at which functions will be lost when LGN P vs. M cells are lesioned.]

[5] The notion of sufficiency becomes more complex if we think *across* processing levels. Clearly no part of the LGN is sufficient for vision in the broader context, because the retina and cortex are also necessary elements of the causal chain. DT is tempted to coin the term "locally sufficient" – sufficient with respect to the other kinds of neurons at the level of the visual system under discussion.

		Status of visual function A	
Lesioned cells ↓	Intact cells ↓	Lost	Retained
P	M	1 P is necessary for A M is not sufficient for A	2 P is not necessary for A M is sufficient for A
M	P	3 M is necessary for A P is not sufficient for A	4 M is not necessary for A P is sufficient for A

Combinations of outcomes:

1 and 4: P is necessary and sufficient for A; M is neither necessary nor sufficient
2 and 3: M is necessary and sufficient for A; P is neither necessary nor sufficient
1 and 3: Both M and P are necessary for A; neither M nor P is sufficient
2 and 4: Either M or P is sufficient for A; neither M nor P is necessary

Figure 15.7: The logic of LGN-level lesion-and-function studies. The experimenter tests the animal behaviorally to establish the presence of visual function A; makes a lesion in either the M or the P layers of the LGN; and then retests the animal to see whether visual function A is lost or retained. The left columns show which cells are lesioned (no longer available to support visual function), and which are intact (still available to support visual function). The left and right sets of boxes specify possible behavioral outcomes, in terms of whether a visual function, A, is lost or retained after the lesion. The entries in the four boxes show the conclusions that can be drawn, in terms of necessary vs. sufficient conditions. The entries below the table show the possible combinations of outcomes and their interpretations.

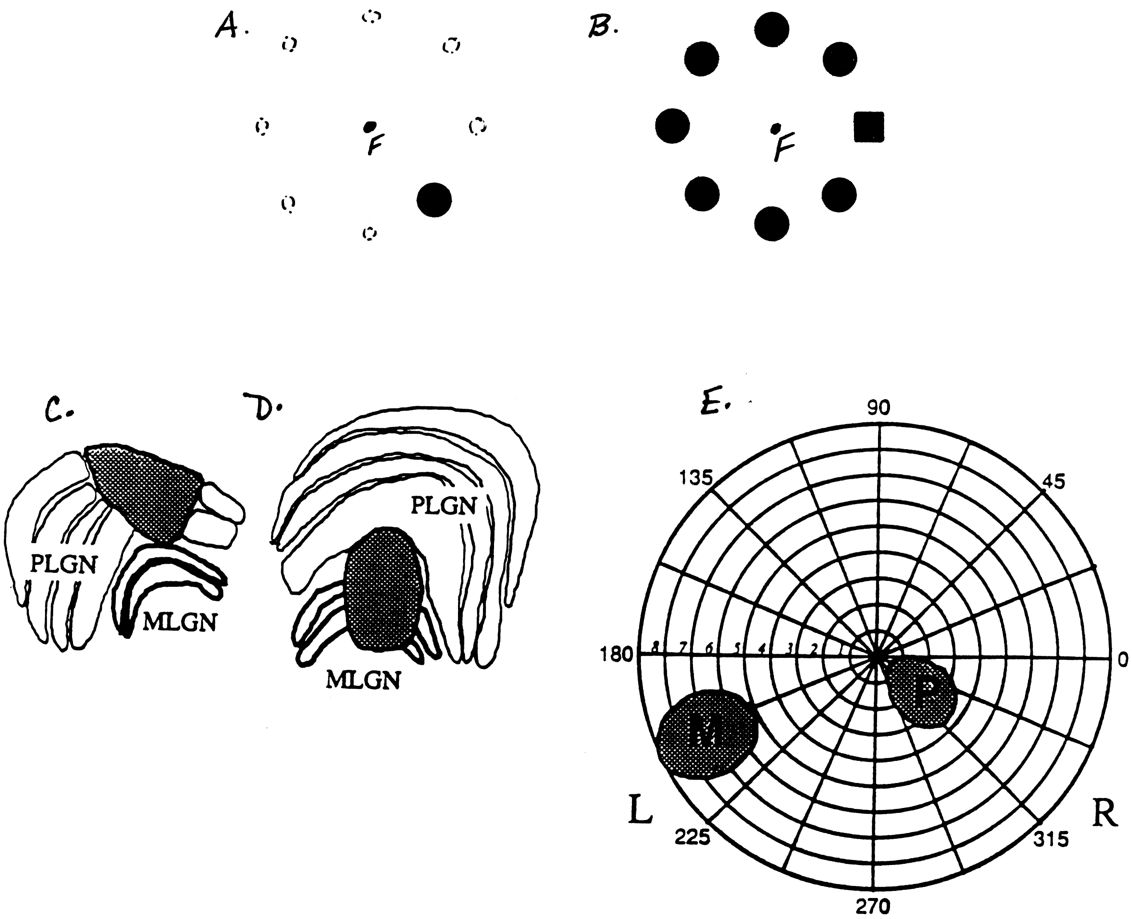


Figure 15.8: The Schiller et al paradigm. A and B: The behavioral tasks. A shows the detection task, and B shows the discrimination task. The distance from the fixation point to the targets was varied to locate one of the targets in the lesioned area under study. C and D: Lesions in the left and right LGN of one of the monkeys. The lesion in the left LGN was confined to the P layers (PLGN) in a near-foveal region, and the lesion in the right LGN was confined to the M layers (MLGN) at a slightly more peripheral location. E. The sizes and locations of the visual field deficits in the left visual field (caused by the lesion in the right LGN) and in the right visual field (caused by the lesion in the left LGN). (A and B after Schiller et al, 1990a, Fig. 1, p. 323. C, D, and E modified from Schiller et al, 1990b, Fig. 1, p. 68.)

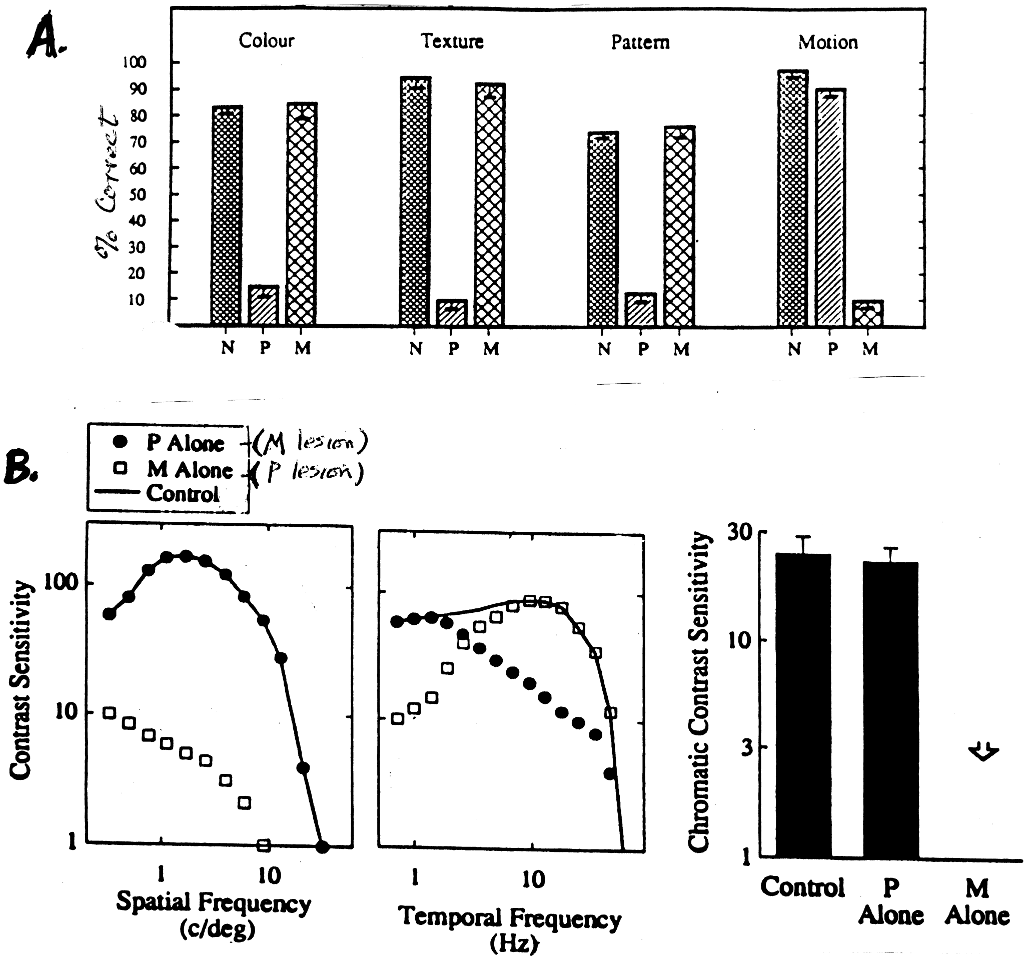


Figure 15.9: Results of lesion studies. A: The performance of Schiller et al's monkeys on several visual tasks: color, texture, pattern, and motion discrimination. In each case, the bars denote the monkey's percent correct. 12.5% correct is chance in an eight alternative task. N = normal (before the lesions), P = stimulus located within the P lesion (M cells alone available); M = stimulus located within the M lesion (P cells alone available). B: Results of the studies of Merigan and his colleagues. In the left and middle panels, the solid lines show spatial and temporal contrast sensitivity functions (CSFs) in normal monkeys. The symbols show CSFs after M lesions ("P alone" – solid symbols) and after P lesions ("M alone" – open squares). Both M cells and P cells are sufficient to process most spatial and temporal frequencies, but with differing levels of sensitivity. The right panel shows the results of color testing. Sensitivity to isoluminant color differences is high both in normal animals and in animals with M LGN lesions ("P alone"), but unmeasurably poor in animals with P LGN lesions ("M alone"). [A modified from Schiller et al, 1990b, Fig. 3, p. 70; B modified from Merigan and Maunsell, 1993, Fig. 2, p. 375.]

A summary of some of the results of Schiller et al's experiment are shown in Figure 15.9A. The experiments came out largely as we might expect. Chromatic discrimination (”colour” in Figure 15.9A) was lost with the P-cell lesion, but retained with the M cell lesion. A similar pattern was seen for the discrimination of fine textures and high spatial frequencies (”texture” and ”pattern” in Figure 15.9A). These outcomes support the notion, mentioned at the end of Chapter 14, that an intact P cell pathway is necessary and sufficient both for chromatic tasks and for tasks requiring fine spatial resolution.

In the M-cell lesioned area, the monkey showed losses of sensitivity to high temporal frequencies (fast flicker), and in the perception of motion (”motion” in Figure 15.9A). In the P-cell lesioned area these functions were intact, suggesting that the M cell system is necessary and sufficient for tasks that require high temporal resolution.

At about the same time, William Merigan and his colleagues (Merigan, 1989; Merigan, Katz, and Maunsell, 1991) also carried out a set of quantitative lesion-and-function studies, measuring detection thresholds for sinusoidally modulated stimuli. Spatial and temporal CSFs for normal and lesioned monkeys are shown in Figure 15.9B. The left panel of the figure shows spatial CSFs for stationary gratings (a temporal frequency of zero) in normal animals and following M- and P-cell lesions. The data suggest that for low spatial frequency stationary gratings, P cells are necessary and sufficient for the high contrast sensitivity seen in the normal animal, whereas M cells are not. But notice that the loss of sensitivity is relative rather than absolute: M cells are sufficient for a residual level of contrast sensitivity under these conditions – lower than normal by a factor of 10 or more, but still above zero.

The middle panel of Figure 15.9B shows temporal CSFs measured at a low spatial frequency in both normal and lesioned animals. Interestingly, under these conditions the results are different in different temporal frequency ranges. At low temporal frequencies, P cells are necessary and sufficient for normal performance, whereas M cells are sufficient for reduced but still non-zero performance. At temporal frequencies near 5 Hz, the two systems are about equally sensitive – either one is sufficient to provide near-normal temporal contrast sensitivity in this range. At a range of higher temporal frequencies, M cells are necessary and sufficient for normal performance, but P cells are sufficient for detection if the contrast is high enough. And near the high-frequency cut-off, at which high contrast stimuli are necessarily used, both systems again appear to be about equally sensitive.

The right panel of Figure 15.9B shows Merigan's (1989) data on the detection of chromatic gratings. The monkeys were tested with isoluminant gratings modulated along several chromatic axes, including both L - M and S - (L + M) isolation axes. The result is straightforward – M-cell lesions left both dimensions of color discrimination intact, whereas P-cell lesions destroyed both dimensions.

What are we to conclude, then, from these two sets of lesion-and-function studies? Over all, the simplest and most consistent outcome has been that chromatic discrimination is little affected by M-cell lesions, but completely lost with P-cell lesions. Quite simply, at the level of the LGN, color vision is mediated by P cells. This result provided one of the strongest encouragements to the notion of parallel processing with strong segregation of functions among physiological channels.

But other functions are not so cleanly separated, and studies like Merigan's suggest a mellower view. In the spatiotemporal domain, the losses induced by lesions can be relative rather than absolute, and can vary in complex ways with the particular choices of spatial and temporal frequency. In such cases, no simple statement of which system is responsible for normal spatial or temporal processing can be made. The simplest and most frequent generalization is that both M and P LGN cells contribute to our sensitivity to low temporal and spatial frequencies. Beyond that, the P pathway extends our sensitivity to high *spatial* frequencies, and is necessary for high acuity in the normal animal. The M pathway extends our sensitivity to high *temporal* frequencies, and is necessary for the perception of fast motion.

Finally, as noted above, these lesion studies were carried out before the K layers were systematically described. The sorting out of the K layers forces us to reconsider the conclusions from the LGN lesion studies. That is, lesions thought to be confined to the P layers must have in fact been lesions of both the P and the adjoining K layers, and lesions thought to be confined to the M layers must have been lesions of the M and the adjoining K layers. We leave it to you to rethink the logical conclusions that can still be drawn from these studies. [Hint – the phrase "and associated K layers" will come in handy.]

15.6 Is the LGN just a relay station?

What is the function of the LGN in visual processing? What if any physiological calculations are performed here? There are major anatomical reshufflings. But from the physiologist's perspective, the neural code changes remarkably little from the retinal ganglion cells to the LGN output. We could even say that from a physiological perspective, in comparison to the First, Second, and Third Transformations described earlier, the Fourth Transformation is a bust. For this reason the LGN is often called a *relay nucleus*, and considered to be of relatively little interest. But could it really be true that at this seemingly important site, where visual information leaves the retina and is relayed to the cortex, there is no functional change in the visual signal?[6]

To addess this question we need to step back and take a look at the internal circuitry of the LGN. A schematic circuit is shown in Figure 15.10. The most prominent and readily recorded neurons in the LGN are the *relay cells* – the cells that receive direct excitatory input from retinal ganglion cells, and send their axons on to visual cortex. As we have already seen, relay cells come in three types, M, P and K, and their physiological properties are similar to those of their respective retinal inputs. If this were the whole story, we could indeed characterize the LGN as just a relay nucleus. However, when the internal circuitry is examined in detail, it turns out that only about 20% of the synaptic inputs to LGN relay cells come from the retina! Clearly, something happens.

The rest of the inputs to LGN relay cells come from many sources. First, the LGN contains intrinsic cells called *intrinsic neurons* or *interneurons* – cells whose processes remain within the LGN. The incoming retinal axons contact local interneurons, which in turn form inhibitory feed-forward synapses onto the relay cells. Second, LGN relay neurons form reciprocal connections with neurons in a nearby region called the thalamic reticular nucleus (TRN), forming local inhibitory feedback loops. Approximately eight other sources of input to LGN relay neurons are also known in detail, but we will not go into them further.

Finally and most importantly, a massive input to LGN relay cells is provided by a feedback loop back from the visual cortex. This feedback loop, like the input to the cortex, is tightly topographically organized, and provides excitatory input back to the original relay neurons and their nearest neighbors. It also provides inhibitory input via the TRN neurons and the interneurons.

[6] DT once saw a sign on a house in Washington DC: "At this site on April 19, 1776 nothing happened." Could this really be the fate of the LGN?

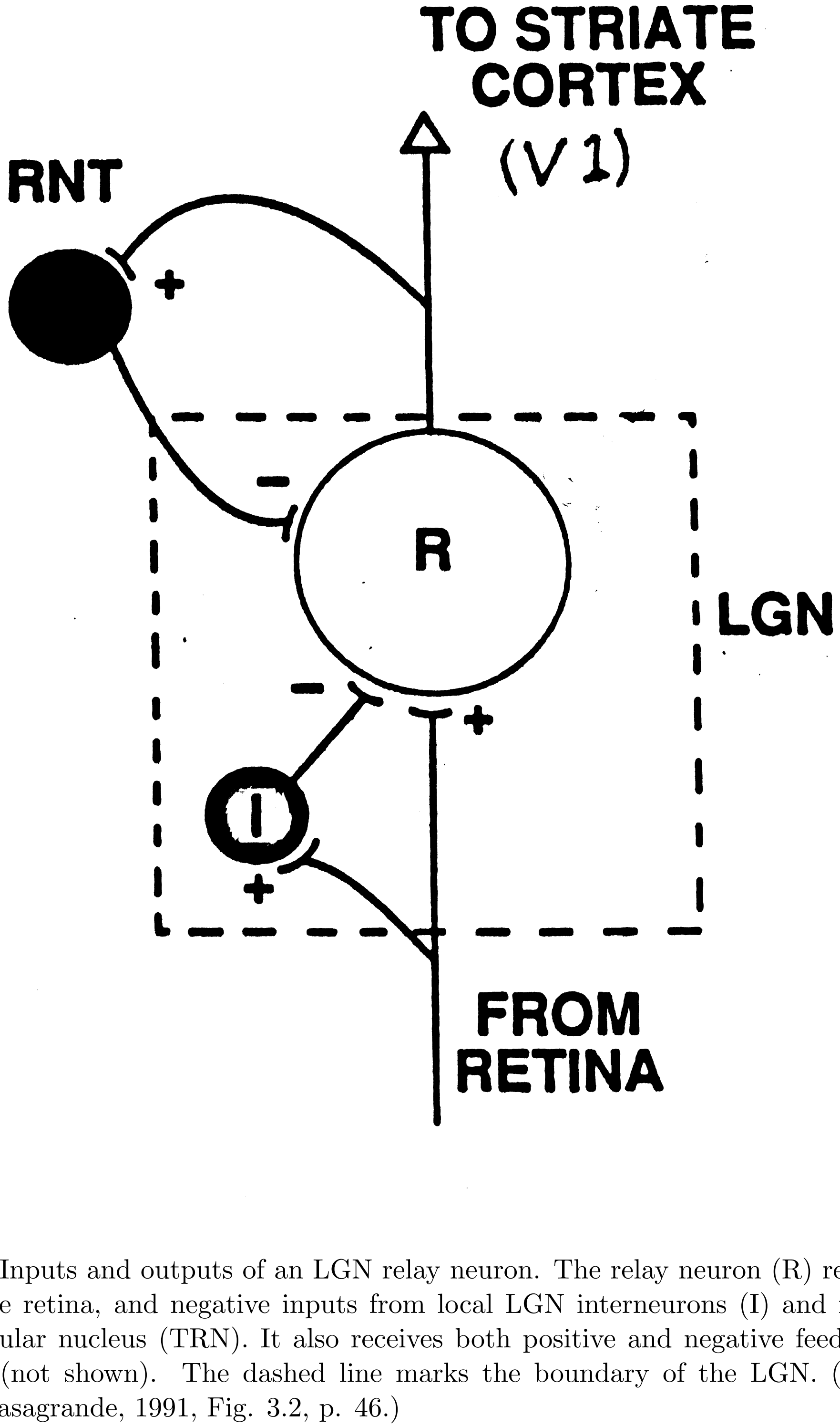


Figure 15.10: Inputs and outputs of an LGN relay neuron. The relay neuron (R) receives positive input from the retina, and negative inputs from local LGN interneurons (I) and neurons in the thalamic reticular nucleus (TRN). It also receives both positive and negative feedback from the visual cortex (not shown). The dashed line marks the boundary of the LGN. (Modified from Norton and Casagrande, 1991, Fig. 3.2, p. 46.)

The extraretinal inputs to the relay cells are believed to perform a *gating* or modulating function, turning the signals in the relay neurons on or off, or modulating their strength. Moreover, the details of the cortical feedback circuits vary between relay cells in the M, P, and K layers, allowing the possibility that different portions of the retinal code can be modulated selectively and independently, so that different aspects of the incoming information can be emphasized at different times and in different locations, in the service of different visual tasks.

But might these feedback circuits also bring about a code change? The influence of cortical feedback on LGN activity can be studied by heating and cooling the cortex, temporarily suppressing and restoring the cortical feedback loop. When the cortex is cooled, the activity of LGN relay neurons is changed to some degree, supporting the idea of a cortical control circuit. Visual stimuli that maximize the firing rates of the cortical cells involved in the feedback loop can minimize the firing rates of the relay cells, and vice versa; and the synchrony of firing of neighboring relay cells can also be increased by cortical feedback. On the other hand, the receptive field properties of individual neurons are not drastically changed by cortical feedback – for example, cells don't change from having concentric receptive fields to responding to some strikingly different stimulus configuration. In sum, cortical feedback appears to modulate LGN activity in relatively subtle ways, rather than changing the fundamental form of the information carried by individual neurons.

The lesson learned is that different anatomical structures perform different kinds of functions. In fact, it can be argued that one of the functions of the LGN is to *preserve* the retinal output code, while allowing differential modulation of its signals. The calculations performed at the LGN have to do with regulating the flow of signals from retina to cortex, without changing the basic visual code.

15.7 The design question: Why juxtaposed layers but segregated signals?

Finally, we raise the design question. In the LGN, the EDC has produced a structure that is at once layered and replete with juxtaposed spatial mappings of inputs from the two eyes, and that at the same time maintains the monocularity of inputs to individual cells. Why isn't binocularity set up in individual neurons in the LGN, when the juxtaposed maps would seem to make it such an easy task to accomplish? It has been often noted that there are major species differences in the lamination pattern of the LGN, yet the juxtaposition of maps and the absence of binocular interactions are conserved across species. This conservation suggests that the continued segregation of the signals from the two eyes at the LGN may serve an important design function. But what function? Three kinds of speculations have been offered.

The first idea is that the LGN is layered to provide a *staging area*[7] for the visual cortex. That is, it sets up a proximity of cells that will need to interact later. Perhaps the neurons that stay together (in the LGN) will play together (at later processing stages).

A second line of speculation is that the anatomical layering at the LGN serves a developmental purpose. Perhaps it's easier for neurons that carry signals from corresponding points on the two

[7] A good illustration of a *staging area* would be a grocery store. All of the different cooking ingredients – lettuce, tomatoes, etc etc – are laid out in an orderly fashion on the shelves. With such a good staging area, it is easy to assemble the separate items one needs in the shopping cart, to be transported home together. They interact only later, when the cooking begins.

retinas to reach the same cortical point if they are next to each other in the LGN. Perhaps one of the two axons guides the growth of the other.

A third line of speculation is that perhaps the continued segregation of signals from the two eyes allows the cortical feedback circuit to effect separate fine tunings of the left and right eye signals, before the two signals are irretrievably smooshed together. For example, cortical feedback might fine tune the left and right eye signals in some way designed to optimize the analysis of small differences between the two retinal images, and thus serve the perception of depth (see Chapter xx).

15.8 Summary: The Fourth Transformation

In summary, we have two eyes and two retinal images of the world. Thus, after retinal processing, we have two neural images of the same visual stimuli. These two images are spatially juxtaposed at the LGN. The necessary anatomical realignment is accomplished at the optic chiasm, where the retinal ganglion cell axons from the two nasal retinas cross, so that the two neural images of each half-field project together to the same LGN: left visual field to right LGN, and right visual field to left LGN.

At the LGN, further order is imposed. M ganglion cells from the two eyes project separately to the two M layers, P ganglion cells to the four P layers, and K ganglion cells to some or all of the six K layers. Thus, there are twelve neural images of each half of the visual field, aligned in a stack in the twelve layers of the LGN. In other words, the neural image of a single object in space is distributed among a stack of twelve locations in the LGN. Moreover, the foveal representations are greatly magnified while the more and more peripheral regions are alotted less and less space. In terms of anatomy, the Fourth Transformation is profound.

On the other hand, there is apparently no major change in the visual code. Even though left eye and right eye neurons that serve corresponding points in the visual world are juxtaposed in their separate maps in the LGN, they interact little if at all: LGN neurons remain mostly monocularly driven. The characteristics of single neurons mirror those of their retinal input neurons, and to an excellent first approximation a description of the M, P and K ganglion cells that project from the retina can also serve as a description of the LGN output code carried by the M, P and K relay cells to the cortex.

What calculational function, then, does the LGN perform? Unlike retinal ganglion cells, LGN neurons have inputs from many regions of the brain, and form feedback loops with their targets in visual cortex. We will see in later chapters that feedback circuits routinely influence the form and flow of visual information through the cortex. The LGN is the first place at which neurons from the rest of the brain have a chance to modify the retinal input – the very first opportunity for such control circuits to regulate the form and strength of the incoming visual signals. This control function will doubtless be unfolded in greater and greater detail as our understanding of the LGN evolves over the next few years.

Chapter 16

Spatial Frequency Analysis

When DT's son was in 5th grade, he had a science textbook with a section entitled "How we see". It said that the lens of the eye makes a picture of the world on the back of the eye, and the photoreceptors turn the picture into nerve impulses, and the nerve impulse make a picture of the world in the brain, and that's how we see. DT calls this view a picture-in-the-brain theory – the notion that even deep within the visual system, two-dimensional space is still represented by a point-for-point two-dimensional neural image.

In the present chapter we consider a revolutionary alternative. We begin with the idea that, rather than representing two-dimensional space in terms of a point-for-point or local-region-by-local-region map, the visual system might use an altogether different code. In technical terms, the idea is that the visual system might carry out something akin to a Fourier analysis of the visual scene, and represent the visual scene in terms of its spatial frequency components. Such a code transformation might be achieved by creating a set of filters, or channels, tuned to different spatial frequencies. This idea leads to the concept of multiple channel models of visual processing.

We begin the chapter with an attempt to make the concept of multiple spatial- frequency-tuned channels accessible via an historical and conceptual introduction. We then review some psychophysical evidence for the existence of multiple channels in the visual system, and some evidence concerning the numbers and bandwidths of the channels. Next we consider psychophysically based, quantitative models of spatial-frequency-tuned channels, and ask how such channels might be created. We then look briefly at an alternative view – coding by edges or other local features – and pose a reconciliation of these views. Finally, we pose the Design question: Why might the EDC choose to create a visual processing system based on multiple spatial channels? In the next chapter, we explore the properties of neurons in V1, and ask whether or not V1 neurons might form the basis of a multiple channels model.

In earlier chapters of this book, color vision was used to introduce the interplay of psychophysics, physiology, and mathematical modelling, and the idea of coding and recoding in the visual system. In the present chapter we develop a second example – spatial vision. There are many parallels between color vision and spatial vision, and we will point out some of them as we go along.

16.1 The Fourier revolution: Spatial frequency coding and the visual system

16.1.1 Fouriers theorem and Fourier analysis

In Chapter 5 (Figures 5.1 and 5.2), we introduced sinusoidal gratings in connection with modulation transfer functions (MTFs) and contrast sensitivity functions (CSFs). But why do optical engineers and visual scientists use such seemingly weird and unnatural stimuli? Historically, the earliest reason stems from a mathematical discovery made in the late 1700s by the French mathematician Jean Baptiste Fourier. Fourier showed that *any signal that varies in space or time can be described mathematically as the sum of a set of sinusoids that vary in frequency, amplitude, and phase.* By starting with sinusoids of the requisite frequencies, manipulating amplitudes and phases, and summing these patterns appropriately, any more complex function can be generated. Breaking down a complex pattern into its spatial frequency components is called *Fourier analysis*, and recombining them to make the original pattern is called *Fourier synthesis.*

In visual science, Fourier's theorem is important because it implies that all visual stimuli – any pattern of light in the two-dimensional visual field – can be described with a common set of components. We are in the business of trying to understand how the visual system codes and recodes two-dimensional visual patterns. Natural visual stimuli – objects and scenes – vary in an infinitie number of ways, and before the arrival of Fourier analysis vision scientists had no way to reduce them to a common description. But Fourier's theorem tells us that all visual scenes can be described in common if we represent them in terms of their spatial frequency, amplitude, and phase components. Perhaps such a common description will come in handy in understanding how visual stimuli are processed.

Figure 16.1 shows two classic example of Fourier representations. The two leftmost columns show the synthesis of a square wave grating from a set of sinusoidal grating components. Rows 1 and 2 show gratings of spatial frequencies F and 3F respectively. In each of these rows, the first column represents each sinusoid with a photograph. The second column represents the grating graphically, plotting luminance against spatial position.

The third column introduces a new concept – the *amplitude spectrum* of the grating. In these graphs, the abscissa represents spatial frequency, and the ordinate represents the amplitude of the sinusoid. Since each grating – F or 3F – consists of only a single spatial frequency, the amplitude spectra are particularly simple – in each case the amplitude is zero except for a single non-zero value at the spatial frequency of the grating (F in the first row, 3F in the second). Notice that in the amplitude spectra, the amplitude of 3F is set to 1/3 the amplitude of F, for reasons that will become immediately apparent.

Now, suppose we wish to synthesize the square wave grating in row 6. Let F be the *fundamental* spatial frequency – the spatial repetition rate – of the square wave grating. We begin the Fourier synthesis with a sinusoidal grating of frequency F. We then add a grating of frequency 3F, with the contrast of 3F set to 1/3 the contrast of F, as they are in the amplitude spectra in column 3. The third row of Figure 16.1 shows the combination of these two gratings, combined so that the peaks of F coincide with troughs of 3F, and the troughs of F coincide with peaks of 3F (this phase relationship is referred to as ”peaks subtract”). Notice that the 3F component begins to square off the corners in the combined pattern.

The fourth row shows a grating of frequency 5F at a contrast of 1/5 the contrast of grating

Figure 16.1: Fourier synthesis of square wave and triangle wave gratings. The synthesis of a square wave grating is shown in the two leftmost columns, and the synthesis of a triangle wave is shown in the two rightmosot columns. The middle column shows the amplitude spectra for both sets of gratings. [After Levine and Shefner, 1991, Fig. 10-6 and 10.7, pp. 217 and 219.]

F. The fifth row shows the 5F grating added to the combined pattern, again in a peaks subtract location, resulting in a further squaring off of the corners of the pattern and smoothing out of ripples. Finally, the sixth row shows the result of adding higher spatial frequencies – 7F, 9F, etc, in contrasts 1/7, 1/9 etc. of the contrast of F. If this series of spatial frequencies is combined in the proper phase relationships, the limit of the series is the square wave grating of frequency F.

Finally, the two rightmost columns in Figure 16.1 show a new kind of spatial grating – a triangle wave – and its Fourier components. The triangle wave was chosen because, as it turns out, a triangle wave and a square wave grating with the same fundamental spatial frequency have exactly the same spatial frequency components and exactly the same amplitude spectra. In fact, the third column represents the amplitude spectra for both kinds of gratings! Comparison of columns 2 and 5 will convince you that the difference between square waves and triangle waves is just that the components are combined in different phase relationships – peaks-subtract for the square wave and peaks-add for the triangle wave. Figure 16.1 thus makes the important point that phase relationships are critical in specifying the Fourier components of visual stimuli.

As a more fanciful analogy to the concept of Fourier analysis, consider the view a lecturer has when he is standing in front of the class. In his retinal image, the heads and bodies in the front row make a repeating pattern of a relatively low spatial frequency, and those in the middle and back rows provide patterns of middle and higher spatial frequencies. While the true situation is of course much more complex than this, a Fourier analysis of the lecturer's retinal image might well reveal amplitude maxima at these three spatial frequencies in the horizontal dimension. [Think about some other visual scenes, and imagine how they might break down into Fourier components. For example, what about the Parthenon, or the facade of a large building with many identical windows? At the other extreme, what about a single vertical edge? A vertical bar? What about a person?[1]]

16.1.2 Linear systems theory: A lens as a linear system

Linear systems analysis and *linear systems theory* are concepts that originated in electrical and optical engineering, and were imported into vision science in the 1950s. Suppose that we are interested in a complex system such as an audio amplifier, or a lens, or the human visual system as a whole. Also suppose that our goal is to be able to predict the output of the system for any arbitrary input. Since measuring the response of the system to every possible input would take a very long time, we might instead try to accomplish the task by measuring the responses of the system to just a few inputs. If the inputs were well chosen, maybe we could use just a few measurements and use them to derive a general description of the system. Perhaps this description could be used in combination with a theory or a set of algorithms, to predict the system's response to any arbitrary input. If so, we would have carried out a successful systems analysis, and developed a successful systems theory, of the system we are dealing with.

If the system we are dealing with is linear, the task is very much simplified. The first question is, what small set of inputs should we use? Although there are other alternatives, a common approach is to use sinusoidally modulated stimuli such as sinusoidal gratings. Why? Because Fourier's theorem tells us that any spatial pattern can be represented as the sum of a set of sinusoidal gratings. If we know the responses of a linear system to sinusoidal gratings of all of the spatial frequencies in the pattern, we can just sum these responses to predict the system's response to the pattern.

As an example, let's walk through an application of the systems approach to the characterization of the spatial imaging properties of a lens. First, to characterize the lens, we present sinusoidal gratings of different spatial frequencies, and image each one separately with the lens. We present each spatial frequency at (say) 100% contrast, and measure the contrast in the retinal image – the *output*. For each spatial frequency, we calculate the *gain* (output contrast/input contrast). We have now specified the *modulation transfer function (MTF)* of the lens (as we did for the optics of the eye as a whole in Figure 5.3).

Now we are ready to combine Fourier analysis of the stimulus with the MTF of the lens to predict the image of a scene. The process is shown schematically in Figure 16.2. Panel A represents the input to the lens: any arbitrary visual scene. The arrow from A to B represents the Fourier analysis of the scene – breaking it down into its spatial frequency components. Panel B symbolizes the resulting amplitude spectrum for the scene. We also need to keep track of the phases of the different Fourier components, but these have been omitted for simplicity.

Panel C shows the previously measured MTF of the lens. As always, the lens is low pass – it passes all of the low frequency components with a gain near 1, but increasingly attenuates the contrasts of the higher frequency components. Between Panels C and D, we multiply the amplitude of each spatial frequency component in the scene by the appropriate gain at that frequency (specified by the MTF). Panel D shows the result – an output amplitude spectrum that reflects the input

[1] When this approach to vision was first introduced in the 1960s, we irreverant young students were reminded of an old quip about Freudian theory: ”You shouldn't criticize psychoanalysis until you've been psychoanalysed.” We provided an update: ”You shouldn't criticise Fourier analysis until you've been Fourier analysed!” We proceeded to amuse ourselves at meetings by Fourier analysing the various senior scientists in our minds – the slim ones perhaps having high pass amplitude spectra, and the round ones being particularly well endowed with low spatial frequency components.

Figure 16.2: Linear systems analysis. A schematic illustration of how Fourier analysis and synthesis, in combination with the MTF of a lens, allow us to predict the image of any object as formed by the lens. Panel A represents any scene. Panel B represents the analysis of that scene into its Fourier components (the amplitude spectrum of the scene). Panel C represents the MTF of the lens in question. The solid arrow shows the spatial frequency at which the gain of the MTF falls below 1; the lens reduces the amplitudes of the spatial frequencies above that value. The open arrow shows the spatial frequency at which the gain falls to zero; the lens eliminates all spatial frequencies above that value. Panel D shows the amplitude spectrum of the image formed by the lens. Each of the spatial frequencies represented in Panel B is multiplied by the gain at that spatial frequency as represented in Panel C. Finally, Panel E represents the resynthesis of the amplitude spectrum of the image, to predict the image of the original scene.

amplitude spectrum weighted by the MTF of the lens. Finally, between Panels D and E, the component spatial frequencies are recombined (in the appropriate phases) to synthesize the output – the optical image of the original scene. If the system is linear, as lens systems are, the above sequence of operations will provide an accurate prediction of the image of the original scene.

The practical consequence of the above analysis is that we no longer have to measure the image made by the lens for each new object or scene. Thus, linear systems theory gives optical engineers a powerful and efficient tool for describing the properties of lenses and lens systems.

16.1.3 A technical aside: Two-dimensional Fourier spectra

A technical tangent: The above description is intended to convey the conceptual basis and the practical usefulness of using Fourier analysis to describe visual stimuli. However, as a matter of honesty, we need to point out two complications. First, as shown in Figure 16.1, a stimulus cannot be specified unambiguously by its amplitude spectrum alone – the phases of the components must also be specified.

And second, grating patterns are a particularly simple class of stimuli. They are *one-dimensional*, in the sense that luminance varies along only one spatial dimension (the horizontal in Figure 16.1), and is constant along the other (the vertical in Figure 16.1). In contrast, a scene is *two-dimensional* in the sense that its luminance varies along both vertical and horizontal dimensions. As it turns out, Fourier analysis of two-dimensional spatial patterns reveals spatial frequency components in an infinite number of different *orientations*: vertical, horizontal, left diagonal, right diagonal, and every orientation in between. Thus, the true amplitude spectrum of a scene would include spatial frequency components and their amplitudes at all possible orientations. [I hope to have an illustration of a 2-D amplitude spectrum for the next draft. xx]

16.2 Multiple spatial frequency channels

The idea of Fourier analysis brings us back to the question of possible visual codes, and here a novel and powerful insight arose. Could it be that early stages of the visual system carry out a Fourier analysis of the visual scene? Could it be that at some processing level our visual systems represent the visual scene not in terms of its point-for-point luminance values, but in terms of its spatial frequency components? Perhaps there is a set of neurons somewhere within the visual system, each one tuned to a different spatial frequency in the retinal image. One neuron might respond only to a spatial frequency of 1 cy/deg, others to 2, 3,... 60 cy/deg. The pattern of activity in such a set of neurons could represent a visual scene not point-by-point, but spatial-frequency-by-spatial-frequency, in terms of its amplitude spectrum. In the most general terms, perhaps the EDC and the systems engineers have both evolved the same analytical tools.

This conceptual framework was thrilling to many vision scientists when it was first introduced. Although few vision scientists thought it was literally true, many thought it could be almost true, and many Fourier-like models of visual processing were proposed. More generally, it enabled many of us to break free from our implicit picture-in-the-brain theories of the visual code and ask with an open mind, how are scenes and objects actually represented at the higher levels of the visual system?

Moreover, these ideas led immediately to the next question. Might the visual system as a whole, like the optics of the eye, be amenable to some variant of systems analysis? Might it be possible to use subjects' responses to sinusoidal grating stimuli to predict the output of the whole visual system for *any* input? If so, then visual psychophysicists could do calculations rather than measurements, and new psychophysical studies would no longer be necessary.

At the psychophysical level, the first step in pursuing this analogy would be to measure something akin to an MTF, but for the whole visual system. We could present the subject with sinusoidal gratings of different spatial frequencies, and measure a psychophysical threshold for each one. In fact, we did exactly this in Chapter 5, and we called the resulting function a contrast sensitivity function (CSF). Historically, then, the original reason for testing human vision with sinusoidal grating stimuli was to measure a CSF, and use it to test the usefulness of linear systems theory for describing the human visual system as a whole.

16.2.1 The insight: Campbell and Robson (1968)

In 1968, Fergus Campbell and John Robson published a classic paper asking whether the CSF could be used, in the context of linear systems theory, to predict the response of the visual system to complex stimuli. They noted that a variety of complex gratings – square waves, triangle waves, and others – could be represented as sums of sinusoidal components (Figure 16.1). Campbell and Robson asked, can the detection thresholds for these complex gratings be predicted from a theory in which we pass each of the component spatial frequencies through the CSF, and then sum the outputs?

Alas, the answer was no. Campbell and Robson found that the detection thresholds for complex gratings were much more closely predictable from the detection thresholds for their fundamental spatial frequencies alone, than from sums of the responses to the fundamental and the higher frequency components. The higher frequency components seemed to contribute remarkably little to the detectability of the complex gratings, as though the visual mechanism that detected the grating were narrowly tuned to the fundamental spatial frequency of the grating, and unresponsive to the rest.

The failure of the predictions from linear systems theory, however, was not without its benefits. In accord with the best scientific traditions, Campbell and Robson rose to the challenge, and produced a remarkable new theory to account for their findings. They suggested that the visual system does not process visual stimuli through a single CSF covering the whole range of visible spatial frequencies. Instead, they argued that *the visual system contains several spatial-frequency-tuned-channels, each tuned to different spatial frequency range.* They argued that these channels were so narrowly tuned that the channel that detected the fundamental frequency F of a squarewave grating was insensitive to frequencies as similar as 3F – that's why the 3F component didn't contribute much to the detection of the square wave grating. Moreover, rather than summing their outputs, the channels seemed to act quite independently. Campbell and Robson further speculated that the overall CSF of the visual system is just the upper boundary, or *upper envelope*, of the set of underlying channels.

A schematic illustration of Campbell and Robson's proposal is shown in Figure 16.3. Theories of this kind became known as *multiple channel models*[2] of visual processing.

[2] When we defined the term *channel* in the context of color vision (Chapter xx), we used it to refer to photoreceptors and their differential responses to different wavelengths of light. In more general terms, a channel is *an entity that responds selectively along some stimulus dimension.* A photoreceptor fits the definition, in that it absorbs a limited range of wavelengths, and filters out the rest. Similarly, a *spatial frequency channel* is an entity that responds to a limited range of spatial frequencies, and filters out the rest. By this definition a ganglion cell is a spatial

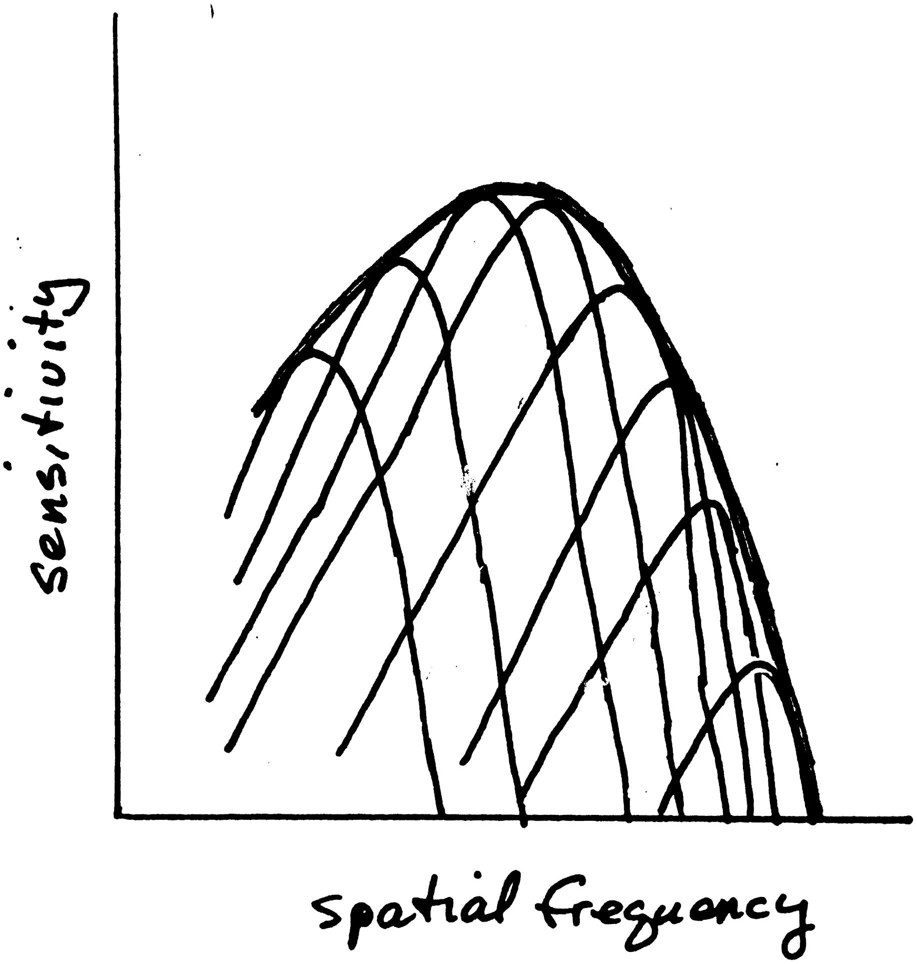


Figure 16.3: A schematic multiple channel model. The figure shows eight different channels, each rather narrowly tuned to a different range of spatial frequencies. The heavy line shows the the overall CSF. Each of the eight channels has been shifted vertically to model the overall CSFas the upper envelope of the channels.

16.2.2 A crude sort of Fourier analysis?

Given Campbell and Robson's speculations, an analogy immediately arose between a system whose output can be predicted from its Fourier components (like an optical system) and a system that might represent stimuli by the pattern of activity in a set of spatial-frequency-tuned channels. However, clearly, the visual system does not do a mathematically pure Fourier analysis. The analogy is limited, and there are many theoretical and practical differences between the two systems.

First, a real physiological system cannot have an infinite number of infinitely narrowly tuned channels, as would be required for a true Fourier representation. Second, the systems analysis shown in Figure 16.2 will only work perfectly if the system being described is linear. In particular, the principle of superposition must hold: the response to one spatial frequency must be independent of the presence of other spatial frequencies. If this were not so, then by definition, the CSF (measured for each spatial frequency component separately) would not predict the response to the scene, in which many spatial frequencies are presented simultaneously. But the visual system is nonlinear in many ways, not all of them necessarily benign.

And third, it can be argued that in any case a strict Fourier analysis would probably not provide a good visual code. Notice that Fourier's theorem applies to the incoming spatial pattern *as a whole*. That is, in Fourier analysis the amplitude of a particular spatial frequency component is determined by the presence of that spatial frequency *over the entire scene*, and represented as a single line in the amplitude spectrum. We have traded a purely spatial representation – point-by-point – for a purely spatial-frequency-based representation – spatial frequency-by-spatial frequency. But a purely spatial-frequency- based representation might well not be a good one to use in vision, because we will want to keep track of the spatial locations as well as the identities of objects, and location information is lost, at least in the amplitude spectrum. Instead of a purely spatial-frequency-based representation, it would probably be better to use a representation that preserves information about both spatial frequency and spatial location. (We will return to this point below.)

In sum, the visual system does not (and probably should not) do a mathematically pure version of Fourier analysis. But it might carry out what came to be called *a crude sort of Fourier analysis*[3], encoding visual scenes by the magnitudes of the signals in a set of spatial-frequency-tuned tuned channels. This idea captured the imagination of many visual scientists. Suddenly the use of sinusoidal gratings as visual stimuli, the search for psychophysical indications of spatial-frequency-based representations within the visual system, and the invention of specific mathematical and physiological models of spatial-frequency-tuned channels, became hot topics in visual science.

frequency channel (see Figures xx and xx in previous chapters).A slightly different definition of the term channel is also sometimes used. Graham (1989, p. 26) defines a channel in physiological terms as *an array or ensemble of cells with the same kind of receptive field, varying only in location on the retina.* Or, in mathematical terms, a channel is *a set of analysing units, or analysers, with the same spatial characteristics, varying only in location.* Notice that this definition counts a whole ensemble as a single channel.

[3] Of course, to a mathematical purist "a crude sort of Fourier analysis" is an oxymoron – a phrase that contradicts itself – and some mathematically erudite vision scientists have argued that the analogy is too strained to be useful. Tolerance for the conceptual wrenching imposed by imperfect analogies differs from one scientist to another. (DT confesses to finding the analogy conceptually useful and even charming, despite its limitations.)

16.3 Psychophysical evidence for multiple spatial frequency channels

Now we ask, as we asked in the case of color vision, can psychophysical experiments provide evidence of the form of postreceptoral coding in the visual system? In particular, can psychophysics provide evidence for spatial-frequency-based representations? The answer is yes, particularly in the processing of near-threshold stimuli. In fact, we can recycle the two psychophysical paradigms introduced in Chapter xx: summation-near-threshold and adaptation-near-threshold.

16.3.1 Summation-near-threshold

A classic summation-near-threshold experiment in the spatial frequency domain was performed by Norma Graham and Jacob Nachmias in 1971. Graham and Nachmias' paradigm is shown in Figure 16.4. Four stimuli were used, as shown on the left side of the figure. The first two stimuli were component gratings of spatial frequencies F and 3F – a grating of a particular spatial frequency and another of three times that frequency (e.g., 2 and 6 cy/deg). The other two stimuli were compound gratings created by combining F and 3F in either a peaks-add or a peaks-subtract phase relationship. Subjects varied the contrast (the peak-to-trough difference) in each stimulus to mesure a set of detection thresholds.

Notice that when the two components are combined, the contrasts in the compound gratings are higher than the contrasts of the original component gratings. If the contrast required for detection threshold in each component grating is normalized to a value of 1, it turns out that peaks-subtract and peak-add combinations have contrasts of 1.4 and 2 respectively. The experimental question was: What will the thresholds be for the compound gratings in comparison to the thresholds for the original components? Will the components sum their signals, so that the compounds are more readily detectable than the components? Or will the components be detected independently, so that the compounds are no more readily detectable than the components?

Figure 16.5 shows these two hypotheses in the context of a summation square (cf. earlier Figure xx). The hypothesis of summation is shown for the peaks-add stimuli by the familiar negative diagonal. The hypothesis of independence is shown by the square contour. (The dotted curve corresponds to the hypothesis of summation for the peaks-subtract stimuli. We ignore this prediction for simplicity, but you may be able to derive it.)

Graham and Nachmias carried out this experiment with several different values of F and 3F, at several different relative contrasts. Their results are shown in the summation square in Figure 16.5. To the surprise of many, the data followed the square contour – the compound gratings were no more detectable than their more detectable component! Thus, the data were consistent with independent detection of the two component gratings within the compound, and hence with the multiple channel model.

16.3.2 Adaptation-near-threshold

The adaptation-near-threshold paradigm (Chapter xx) provides a second line of evidence for the existence of multiple spatial channels. The classic experiment was carried out by Colin Blakemore and Fergus Campbell in 1969. The experiment consisted of three steps. First, a subjects contrast sensitivity function (CSF) was measured. Second, the subject looked at a stationary, high contrast

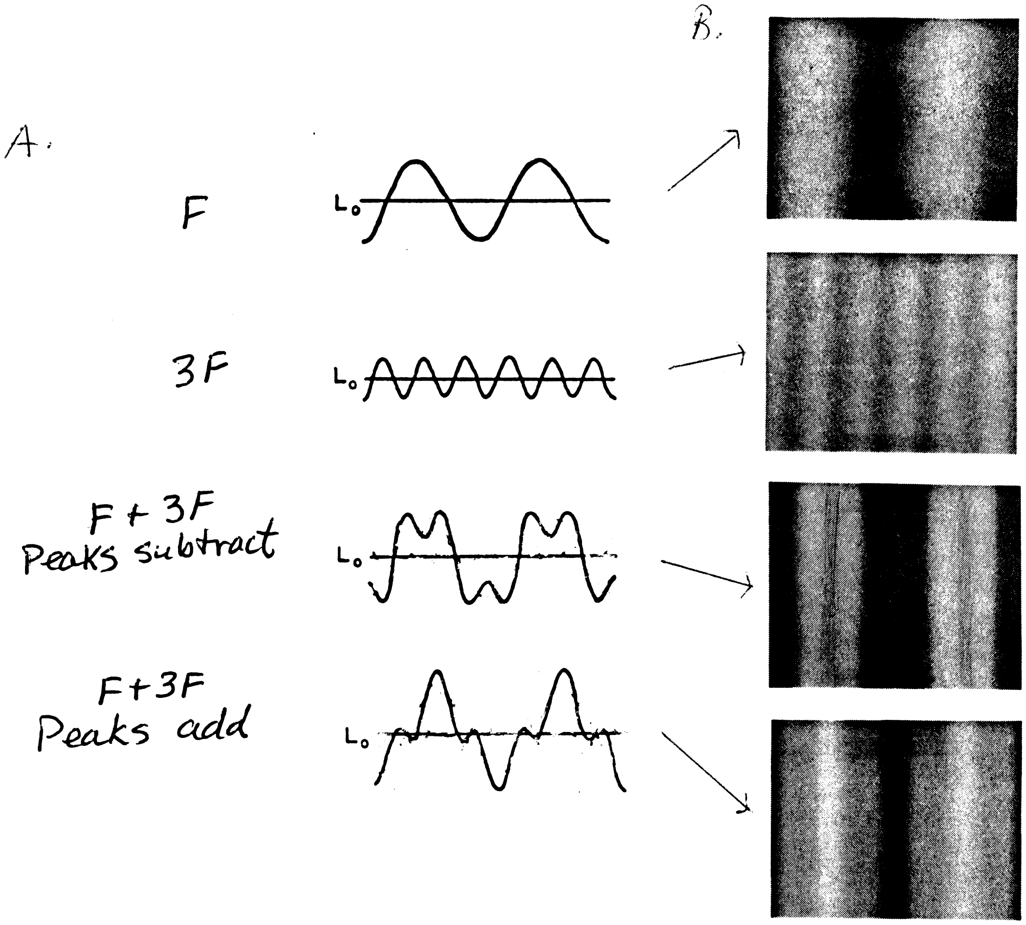


Figure 16.4: The summation-near-threshold paradigm applied to compound gratings. Panel A shows the stimuli used in Graham and Nachmias' experiment. The subject is tested with stimuli of spatial frequencies F (top row) and 3F (2nd row), and two combinations of F and 3F – peaks-subtract (3rd row) and peaks-add (4th row). The right column shows photographs of the four stimuli. Notice the increased contrast in the compound stimuli, particularly in the peaks-add case.

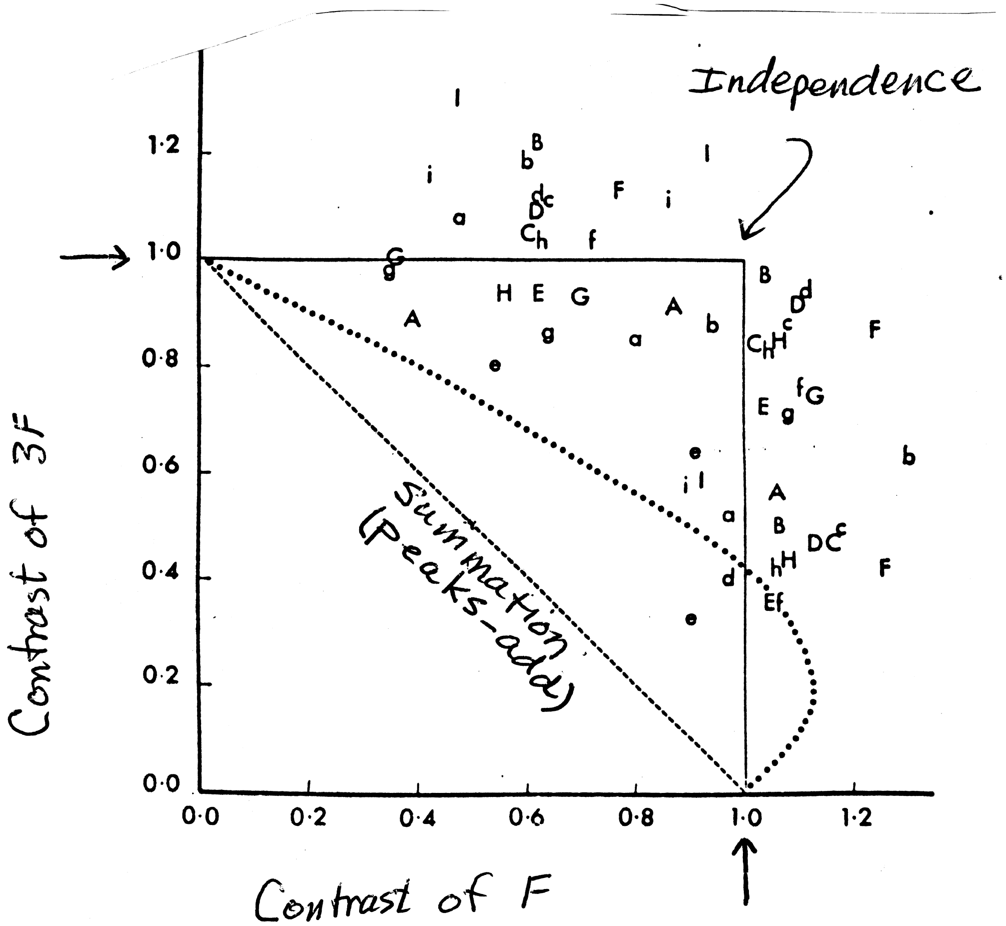


Figure 16.5: The summation-near-threshold paradigm applied to compound gratings. Panel A shows the stimuli used in Graham and Nachmias' experiment. The subject is tested with stimuli of spatial frequencies F (top row) and 3F (2nd row), and two combinations of F and 3F – peaks-subtract (3rd row) and peaks-add (4th row). The right column shows photographs of the four stimuli. Notice the increased contrast in the compound stimuli, particularly in the peaks-add case.

adapting grating of a fixed spatial frequency for about 1 minute. And third, the adapting grating was turned off for brief intervals, and in those intervals the CSF was remeasured.

What are the predictions? If the CSF were determined by only a single broad spatial-frequency-tuned channel, then if sensitivity is changed at all by the adaptation stimulus, it must be changed equally at all spatial frequencies. Just as a single photoreceptor has no means of preserving wavelength information, a single spatial- frequency-tuned channel has no means of preserving spatial frequency information, and only a uniform vertical shift of the whole CSF can occur. In contrast, a multiple channel model predicts that an adapting grating of a particular spatial frequency should desensitize only channels that are sensitive to that spatial frequency. Thus, sensitivity should be lost selectively – there should be a notch in the CSF at the adapting frequency. The results supported the multiple channel model, as shown in Figure 16.6.

In sum, both the summation-near-threshold and the adaptation-near-threshold paradigms, along with Campbell and Robson's original experiments, provide data consistent with a multiple channels model of visual processing. Let us therefore provisionally adopt the multiple channels view, and move to the next set of questions: What are the bandwidths of the channels, and how many are there?

16.4 Bandwidths of the channels

16.4.1 Spatial frequency bandwidths

If an individual channel responds to less than the whole spatial frequency range, then how much of the range does it respond to? Are the individual channels broadly or narrowly tuned? This is the question of *spatial frequency bandwidth.* Examples of possible spatial frequency tuning functions are shown in Figure 16.7. Spatial frequency bandwidths are traditionally specified by their full width at half height (or, on a logarithmic ordinate, their full width 0.3 log units down from their maximum sensitivity). These widths are typically specified in *octaves* (factors of 2).[4]

Each of the paradigms just described can be used to estimate the bandwidths of the channels. Using the summation-near-threshold paradigm, we know from Graham and Nachmias' experiment that F and 3F are detected independently. We now move the two spatial frequencies closer together – F and 2F, F and 1.5F, and so on – until we find that they start to sum. Similarly, using the adaptation-near-threshold paradigm, we can estimate bandwidth from the width of the notch in the post-adaptation CSF.

The estimates of bandwidth vary somewhat from one paradigm to another, but center around a value of about two octaves (a factor of 4). That is, a channel centered at 2 c/deg would fall to half of its maximum sensitivity at about 1 and 4 c/deg. We will return to this question later. [If each channel spans two octaves at half height, what is the minimum number of channels required to span the overall CSF?]

16.4.2 Orientation bandwidths

We now return to the question of orientation. Remember that in our discussion of the complexities of Fourier representations (Figure 16.2), we stated that a two-dimensional spatial pattern has

[4] In music, an *octave* is a factor of two in temporal frequency (e.g. 256 vs. 512 Hz). By analogy, in vision an octave is a factor of two in spatial frequency – from 1 to 2 or from 2 to 4 cy/deg.

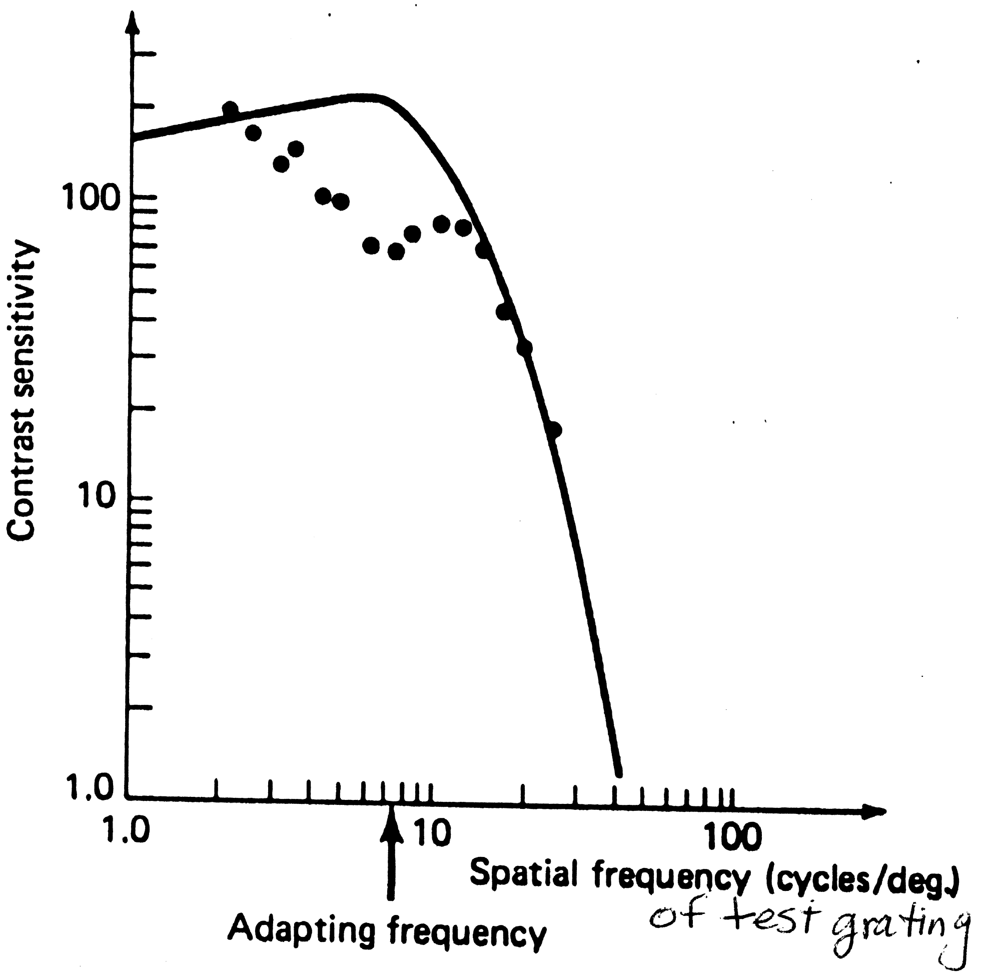


Figure 16.6: The adaptation-near-threshold paradigm. The solid line shows the CSF measured before adaptation to a large, 7.5 cy/deg grating (arrow). The data points show contrast thresholds remeasured after adaptation. The CSF is distorted – sensitivity is lost for spatial frequencies near the adapting frequency, but not at frequencies far away from it. [After Blakemore and Campbell, 1969; via Levine and Shefner, 1991, Fig. 10-13, p. 227.]

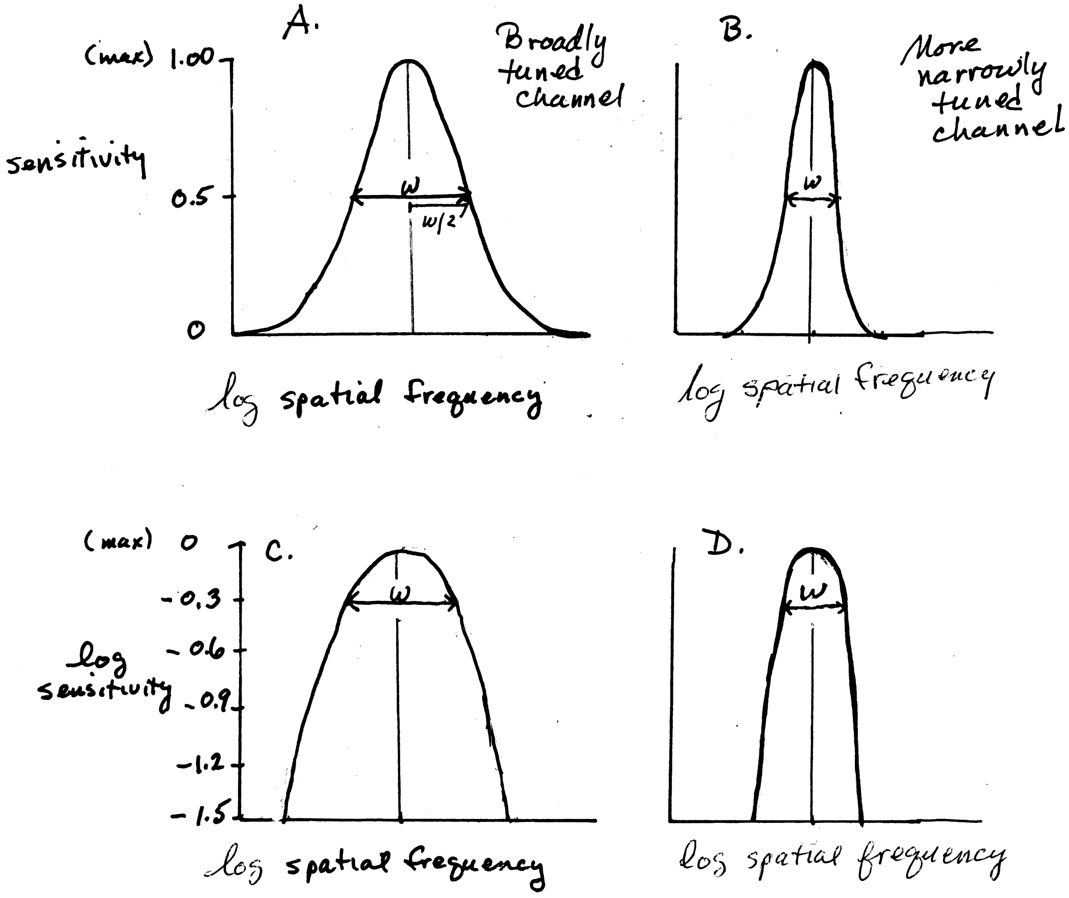


Figure 16.7: Spatial frequency bandwidths. The bandwidth of a spatial channel is a description of the breadth of its spatial frequency tuning curve. The customary measure of bandwidth, w, is the full width of the channel at 1/2 the maximum sensitivity – the full-width at half-height. (A common alternative measure is the half-width at half height (w/2,often stated as ±w/2). Panel A shows a channel of bandwidth w, plotted on a linear sensitivity axis. Panel C shows the same channel plotted on a log sensitivity axis. (Remember that 0.3 log units is a factor of 2, so a reduction of sensitivity by 0.3 log units takes sensitivity down to 1/2 of its maximum value.) Panels B and D show a channel that is more narrowly tuned.

Fourier components at many orientations. But if this is so, and if we are designing a system to represent two-dimensional spatial patterns, it seems likely that at each spatial frequency there will be channels selectively tuned to different orientations. Thus, in addition to their spatial frequency bandwidths, these channels will have orientation bandwidths.

Suppose, for example, that we have done our adaptation-near-threshold experiments with vertical gratings. What happens if we use vertical gratings for the adapting stimuli but oblique or horizontal gratings for the test stimuli? What about pairs of adapting and test gratings that vary in orientation by 10, or 20, or 30^o? It might be that orientation doesn't matter – that all 5 cy/deg gratings cross-adapt equally with each other. Or it might be that stimuli very close together in orientation would cross-adapt, but that the cross-adaptation would fall off as the difference in orientation increases. Similar questions can be asked about summation at threshold. How large must the orientation difference be before two stimuli show signs of independence? Or might the orientations of the stimuli not matter at all – might a set of 5 cy/deg gratings be equally independent regardless of their orientations?

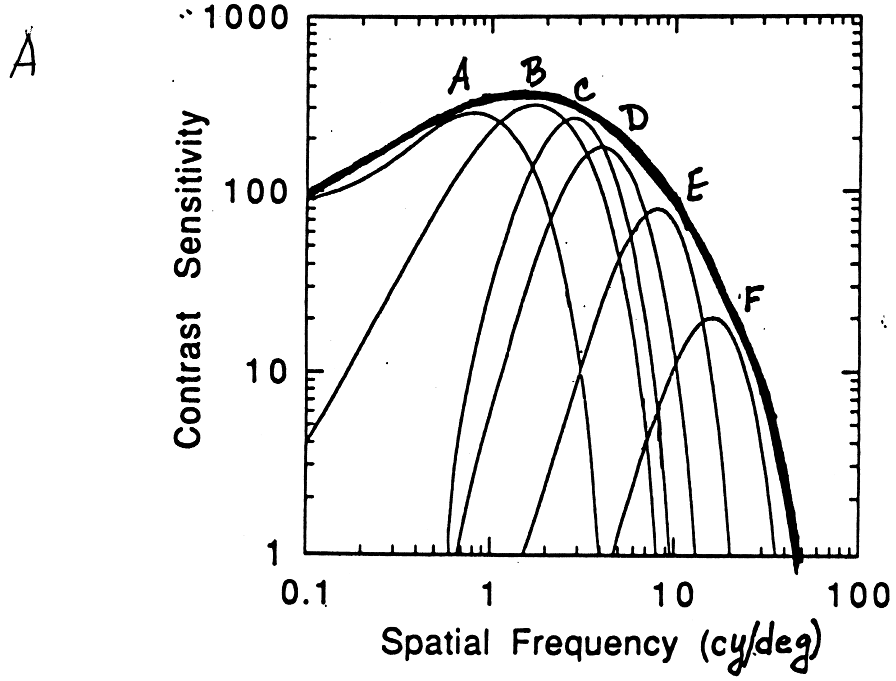
Orientation bandwidth is conventionally specified in terms of half width at half height. Again, the answers are not fully consistent across paradigms, but a good rule of thumb is 15 to 30^o. If the orientation difference between two stimulus components is less than 15^o, summation and cross-adaptation typically occur; if it's more than 30^o, no summation or cross-adaptation is typically seen.

16.5 A specific multiple-channels model: Wilson and Gelb (1984)

Another related paradigm – *selective masking* – has been used to discriminate between single-channel and multiple-channel models. In selective masking, rather than turning on the adapting grating and then turning it off while measuring thresholds, the adapting stimulus (now called a *masking* stimulus) is left on throughout the experiment (or sometimes flashed simultaneously with the test field). The experimenter first measures a traditional CSF. He then turns on a masking stimulus of a given spatial frequency. The subject sets the thresholds for test stimuli of many different spatial frequencies in the presence of the masking stimulus. A single channel model would again predict a constant effect of the masking stimulus on thresholds at all test frequencies, whereas a multiple channel model would suggest that the masking stimulus should selectively affect only a narrow range of test frequencies. The results of simultaneous masking experiments again support the multiple channels model.

Based on extensive masking data, Hugh Wilson and xx Gelb (1984) developed a quantitative multiple channel model of spatial vision. In building their model they asked, what is the *smallest* number of spatial frequency and orientation channels necessary to fit the data? And, what are the spatial frequency and orientation bandwidths of the required channels? The data required a minimum of six channels, whose characteristics are shown in Figure 16.8.

The derived channels have several interesting characteristics. First, the channels are discrete – six and only six channels are sufficient to model the data. Second, as peak frequency increases, spatial frequency bandwidths decrease (on a logarithmic abscissa), and orientation bandwidths also decrease – as their peak frequency increases the channels get more finely tuned in both respects. And third, although not shown in the table, Wilson and Gelb modeled the known changes in the CSF with eccentricity by assuming that the peak frequencies of all six channels decrease together as the stimuli move farther and farther into the peripheral retina.



B

	A	B	C	D	E	F
Peak frequency (c/deg)	0.8	1.7	2.8	4	8	16
Bandwidth (octaves)	2.5					1.5
Orientation bandwidth (deg)	30					15

Figure 16.8: Wilson and Gelb's six-channel model. Panel A shows the six channels of the model, labeled A through F. The heavy line shows the overall CSF that would be predicted from these channels set to these sensitivities. Panel B specifies the spatial frequency and orientation bandwidths of the channels. In this model, the higher the spatial frequency, the narrower the bandwidths. [A modified from Banks, 1993; B not yet found in the literature.]

The model also seems flexible enough to handle the variations in CSFs with light and dark adaptation (see Figure xx in earlier Chapter xx). Changes could be modelled by increasing the relative sensitivities of the high frequency channels with light adaptation and decreasing them with dark adaptation.

16.6 Models of spatial-frequency-tuned neurons

Up until this point, we have tried to confine our discussion to the psychophysical and theoretical realm, and not mix in physiological concepts. Yet in historical terms, one of the factors that fed into the enthusiasm for spatial frequency coding was the striking qualitative resemblance between the band-pass CSFs measured in human subjects, and the band-pass CSFs of individual neurons in the cat visual system. This similarity led to a strong intuition that modelling the visual system with ensembles of neurons with bandpass CSFs – center-surround receptive fields – would be fruitful. At this point we explicitly reintroduce physiological concepts in order to be ready to make bridges between the two realms. We begin by pulling together the concepts and terminology.

In Chapter 8, we introduced the concept of the *receptive field* of a ganglion cell – an essentially empirical and physiological concept. We characterized ganglion cell receptive fields by means of their *spatial weighting functions* – a corresponding but more abstract and mathematical concept. We modelled the spatial weighting functions of ganglion cells with a circularly symmetrical difference of gaussians (DOG) model.

We also introduced the *contrast sensitivity function (CSF)* of a ganglion cell – the detection thresholds for sinusoidal gratings of varying spatial frequencies. In the present chapter we introduced the highly similar concept of the *spatial frequency tuning function* of a channel. If a cell is the physiological instantiation of a channel, its spatial frequency tuning function is its CSF.

We also argued in Chapter 8 that to the extent that a neuron is linear, its receptive field (or spatial weighting function) and its CSF (or spatial frequency tuning function) are two alternative descriptions of the same thing – either can be predicted from the other. A neuron with a DOG-like receptive field has a band-pass CSF – it responds to only a specific, relatively narrow range of spatial frequencies. The peak frequency and the high frequency cut-off of its CSF are determined by the size of the center of its receptive field, whereas the low spatial frequency fall-off results from the presence and properties of its antagonistic surround (Chapter xx, Figure xx).

16.6.1 Spatial weighting functions and spatial frequency bandwidths

By way of further illustrating these points, Figure 16.9 shows examples of six spatial weighting functions and their corresponding CSFs. To work through these examples, imagine sinusoidal gratings of various spatial frequencies falling on receptive fields with each of the various spatial weighting functions, with a bright bar of the grating centered on the receptive field in each case.

Panel A shows the case of an excitatory neuron with a circularly symmetric, spatially non-opponent receptive field. This neuron will respond equally well to all low spatial frequencies, as long as the whole receptive field is covered by a bright bar of the grating. As the bright bar becomes narrower than the receptive field, the response of the neuron will be reduced, and it will respond less and less well to higher and higher spatial frequencies. The result is a low-pass CSF with a high spatial frequency cut-off determined by the size of the excitatory receptive field.

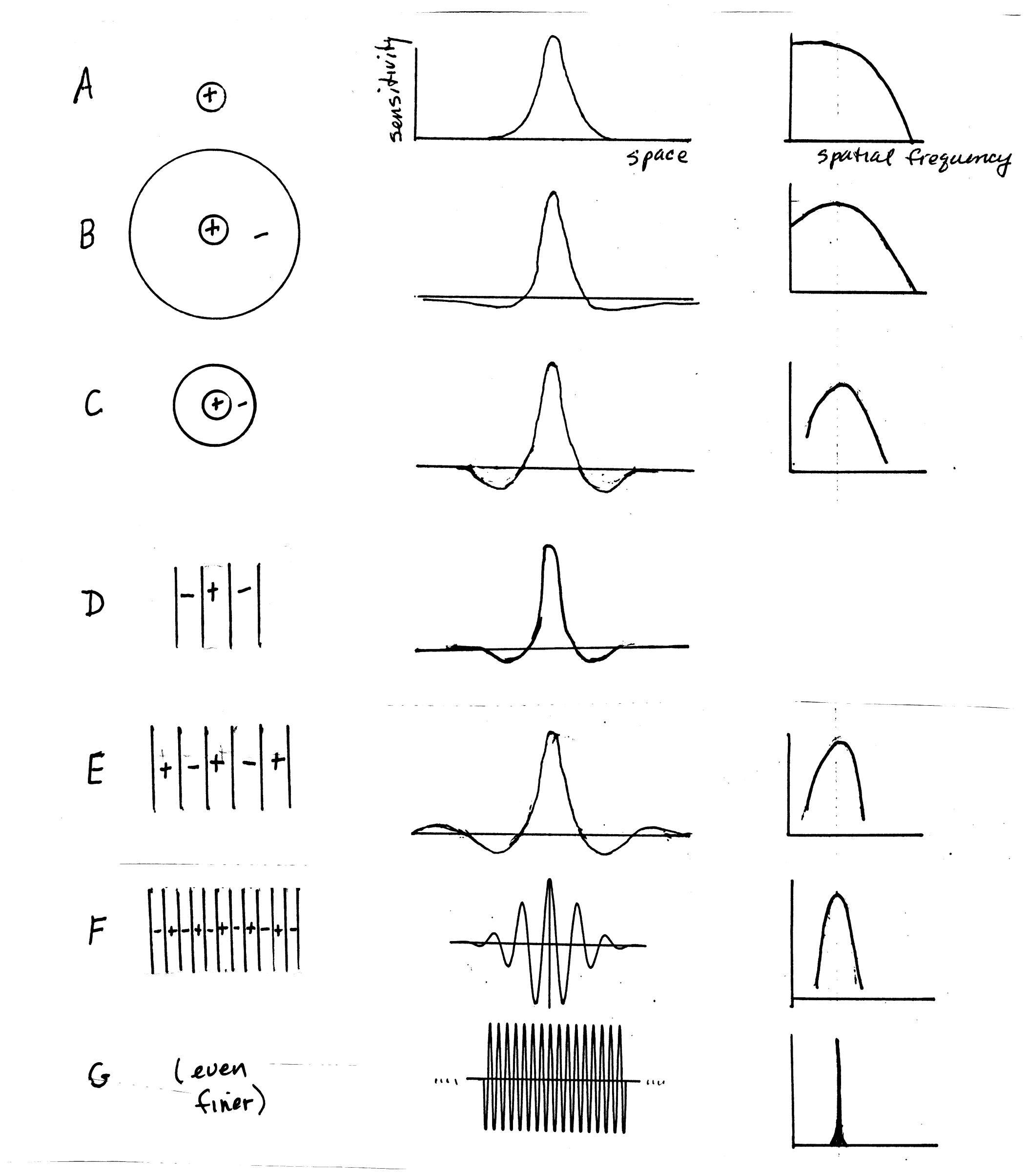


Figure 16.9: Receptive fields, spatial weighting functions and spatial frequency tuning curves (schematic). The left column shows receptive field maps for a series of hypothetical neurons. The middle column shows the corresponding spatial weighting functions, and the right column shows the corresponding spatial frequency tuning curves. When the receptive field has no surround, the tuning curve is low pass (Panel A). A broad, radially symmetrical surround gives a broad bandpass tuning curve, typical of the human CSF as a whole (Panel B). Radial symmetry with the surround width about three times the center width gives a narrower tuning curve, typical of ganglion cells (Panel C). Oriented receptive fields with increaing numbers of sidebands give increasingly narrow tuning curves (Panels D-G). [DT – sketch]

Panels B and C show two more circularly symmetric receptive fields with DOGs as spatial weighting functions. The neuron in Panel B has a very broad inhibitory surround. A spatial weighting function like this one generates a broadly tuned band-pass CSF – one that resembles the psychophysically measured human foveal CSF (Figure xx in earlier chapter). Panel C shows a DOG function in which the spatial extent of the surround process is only about three times that of the center process. The resulting CSF is narrower than the psychophysical CSF, but typical of ganglion cells and LGN cells (see Figures xx and xx in earlier chapters). In fact, the narrowest CSF one can make with a DOG function has a bandwidth of about xx octaves.

In Panels D through F we switch to a new kind of spatial weighting function: an *oriented* receptive field made from a set of elongated excitatory and inhibitory bands. Panel D shows a vertically oriented receptive field with three bands of input, an excitatory band flanked by two inhibitory bands. For the three-band case in Panel D, the predicted response profiles will not be very different from the circularly symmetric receptive field in Panel C.

Panels D and E show oriented receptive fields with *sidebands* – one or more extra excitatory and/or inhibitory regions added on at the periphery of the receptive field. The sidebands allow us to further sharpen the spatial frequency tuning of the neuron. Why? Because a grating with a spatial frequency such that its light and dark bars coincide exactly with the excitatory and inhibitory regions of this receptive field would yield the greatest possible response from the cell. Stimuli of a narrow range of neighboring spatial frequencies would also fit the receptive field profile sufficiently well to lead to a response, but as the bars of the stimulus get out of phase with the sidebands of the receptive field, the neuron's response would diminish.

Panel D shows a neuron with a single set of excitatory sidebands, and a narrower tuning curve. Panel E shows a receptive field with several sidebands, with an even narrower tuning curve. Finally, Panel F shows a receptive field with many precisely spaced sidebands. This neuron requires an exact fit between the spatial frequency of the stimulus and the spacing of the sidebands in its receptive field. Such a neuron will respond only to a very narrow range of spatial frequencies, or in the extreme, a single spatial frequency.

Now, suppose we believed that a multiple channel model – say, Wilson and Gelb's – were actually instantiated by single neurons within the primate visual system. If so, we should be able to find neurons with the characteristics corresponding to Wilson and Gelb's channels somewhere within the visual system.

The quantitative considerations of Figure 16.9 allow us to be more specific. Neurons with receptive fields like those in Panels A and B are too broadly tuned to fit the model. Those in Panel C or D would do for Wilson and Gelb's lower spatial frequency channels, but sidebands akin to those in Panels E or F are needed to create the narrow spatial frequency tuning seen in their higher spatial frequency channels. And if we found a set of neurons like the one in Panel G, we could go back to the idea that processing in the visual system approximates a literal Fourier analysis. Will the right kinds of neurons be found? Wait and see!

16.6.2 Another technical aside: Gabor functions

A mathematical function that has often been used recently in modelling spatial channels is the *Gabor function*, shown in Figure 16.10A. A Gabor function is created from the combination of a spatial sinusoid with a normal (bell-shaped) curve. The normal curve is used to *window* the sinusoid – to vary the amplitude of the grating as a function of space.

Within this research area, Gabor functions are used in two different ways. The first use is in the construction of visual stimuli, known as *Gabor patches*, as shown in Figure 16.10B. Gabor patches are useful because they are spatially local yet contain a relatively narrow range of spatial frequencies. Moreover, the spatial frequency content of the stimulus can be varied while keeping its the overall size constant, as in the left and right panels of Figure 16.10B. Thus, Gabor patches can be used to probe the properties of spatial frequency channels in local retinal regions.

The second use of Gabor functions is for modelling receptive fields. Three important properties of receptive fields can be varied by varying the parameters of the Gabor: the optimal spatial frequency by the spatial frequency of the sinusoid, the bandwidth by the breadth of the Gaussian window, and the orientation by the orientation of the sinusoid within the Gaussian window. Gabor functions, then, allow the modeller to use a single mathematical function to create a set of spatial frequency channels tuned to any desired set of spatial frequencies, each with any desired bandwidth and orientation. In addition, Gabor functions are particularly useful for modelling purposes because they turn out to have very tractable mathematical properties.

Moreover, notice that Gabor functions are larger than points, but smaller than the whole retina. That is, they provide receptive fields with an excellent compromise between strictly local and broadly distributed spatial processing capabilities. The idea is that a set of Gabor-like spatial frequency channels, limited in spatial extent, would serve one local region of the visual field, and be duplicated in a modular fashion across the retina, many times, to cover the visual field as a whole. This scheme has been referred to as *patch-wise Fourier analysis.*

We end this section with a caveat. Gabors and other similar sets of mathematical functions are great modelling tools – there is plenty of flexibility for creating a large variety of multiple channel models. But by the same token, too much flexibility can be a disadvantage, in the sense that it is difficult to imagine a set of data that would reject this class of models, or discriminate among them. A theory that is too flexible can lead to its own demise.

The present state of the art is that there are many different multiple channel models available, differing in the numbers, peak spatial frequencies, spatial frequency tuning functions, orientations, orientation tuning functions, and temporal and chromatic properties of the putative channels. A good example is the model of Singer and D'Zmura (1995), which has a total of 72 channels: channels tuned to each of four spatial frequency bands, at each of six orientations, for each of three chromatic axes. Other models incorporate sustained vs. transient properties for each channel. However, the generation of multiple channels models has diminished in recent times, partly because of the growing realization that the possible varieties of models exceeds the power of empirical data – at least psychophysical data – to sort among them.

At the same time, a multiple channels stage is firmly embedded as a mid-level processing stage in most models of higher-level visual processing. Yet all are agreed that a multiple channels code is not the end of the processing sequence. We fully expect that whatever code exists at intermediate processing levels will itself be recoded in many ways, to serve many different higher-level visual and motor tasks.

16.6.3 Resampling ganglion cell ensembles to create spatial-frequency-tuned neurons

Now, let's return to physiology. Suppose we have adopted a specific spatial channels model, and we believe that neurons that instantiate that model will be found within the visual cortex. For our

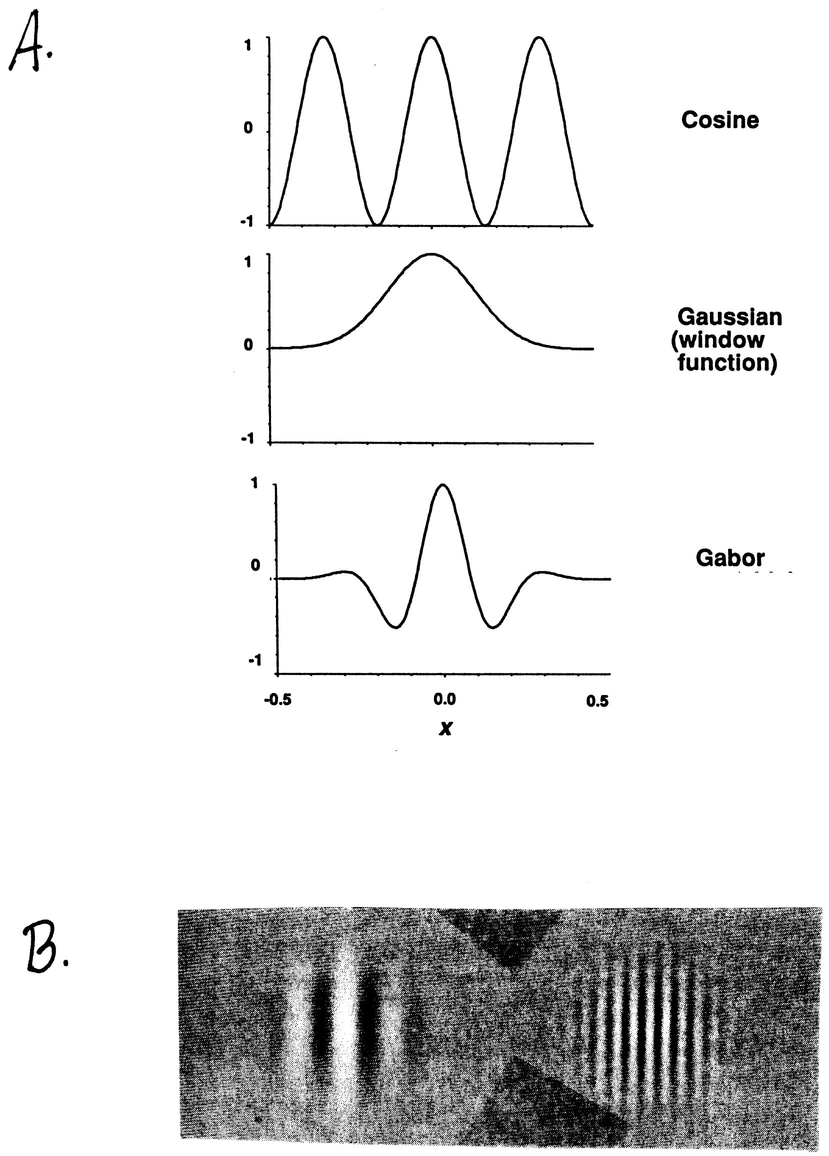


Figure 16.10: Gabor functions. A. A Gabor function (bottom panel) is created by "windowing" a sinusoidal grating (top panel) with a Gaussian (normal, bell-shaped) curve (middle panel). That is, the amplitude of the sinusoid is maximum at some central point, and decreases away from the center in all directions, in accord with the shape of the normal curve. By varying the width of the normal curve, one can produce a Gabor function with any number of sidebands. B. Photographs of two Gabor patch stimuli, both made with the same normal curve, and thus having the same overall size. The left and right panels were made by windowing low and high spatial frequency gratings. [After Graham, 1989, Figs. 2.4 and 2.6, pages 48 and 53.]

physiological model to be dredible, we will have to propose circuitry with which such neurons be created within the real visual system.

The receptive fields needed for spatially tuned neurons with particular tuning properties are shown in Figure 16.9. But there is a problem: remember that by definition, receptive fields are specified in terms of spatial weighting functions *with respect to the retina.* Yet the inputs to higher level neurons can only come from the available ensembles of LGN relay neurons described in Chapter xx – ON-center and OFF-center M cells, ON-center and OFF-center P cells, K (small bistratified) cells, and perhaps a second type of K cell with an OFF or OFF-center receptive field. Moreover, each type of neuron has a coverage factor of 1 – each point in space is served by only a single receptive field center of each type of ganglion cell. How can we constrain our modelling to incorporate the physiological and anatomical facts, and still manufacture the desired set of higher level neurons?

Let's begin with foveal P cells. Remember that each foveal cone subtends an angle of about 30" – two cones per minute of arc. The center of the receptive field of a foveal P ganglion cell samples only a single cone, and P LGN cells sample single P ganglion cells. To make a higher level channel that responds optimally to a 60 cy/deg vertical grating, then, we could resample the ensemble of (say) ON-center P LGN cells, as shown in Figure 16.11A. The higher level neuron could receive alternating excitatory and inhibitory input from closely spaced vertical rows of P cells, each row only one P cell wide.

To make a cortical neuron tuned to one/half that spatial frequency – 30 cy/deg – we could sample the ensemble of ON-center P cells more coarsely, with two P cells per excitatory region and two per inhibitory region, as shown in Figure 16.11B. And more coarsely spaced resampling strategies could be used to manufacture a succession of cortical neurons tuned to frequencies of 20, 15, and10 cy/deg, and so on.

But is it possible to make a cortical neuron tuned in between those shown in Figure 16.11 – say, to 45 or 50 cy/deg? Although the answer is not intuitively obvious, mathematical arguments can be made to the effect that it is in fact possible to construct a neuron tuned to any spatial frequency below the Nyquist limit imposed by the sampling matrix. However, much more computational work is required to make neurons tuned to intermediate spatial frequencies. If neurons tuned to spatial frequences spaced by factors of 2 will do the job, perhaps that is a good design strategy[5].

Similarly, M LGN cells could be used to make cortical neurons tuned to any of many intermediate and low spatial frequencies, with the highest frequency being determined by the spacing of the centers of the M LGN cell receptive fields, and other channels tuned to frequencies of 1/2, 1/3, etc. of the highest frequency. And K cells could be used to manufacture one or more kinds of cortical neurons tuned to specific low spatial frequency bands. These physiological constraints could reduce the number of channels required in a model like D'Zmura and Singer's, since it is not feasible to make M-cell- and K-cell-driven channels at high spatial frequencies. In the extreme, one could combine chromatic and spatial tuning properties, using P cell, M cell and K cell inputs to manufacture spatial channels tuned to high, middle, and low spatial frequencies respectively.

[5] Interestingly, most of the channels proposed by Wilson and Gelb are spaced apart by factors of about two. DT is drawn to wonder whether the fit between this property of Wilson and Gelb's model, and the available simple resampling strategies, is more than coincidental.

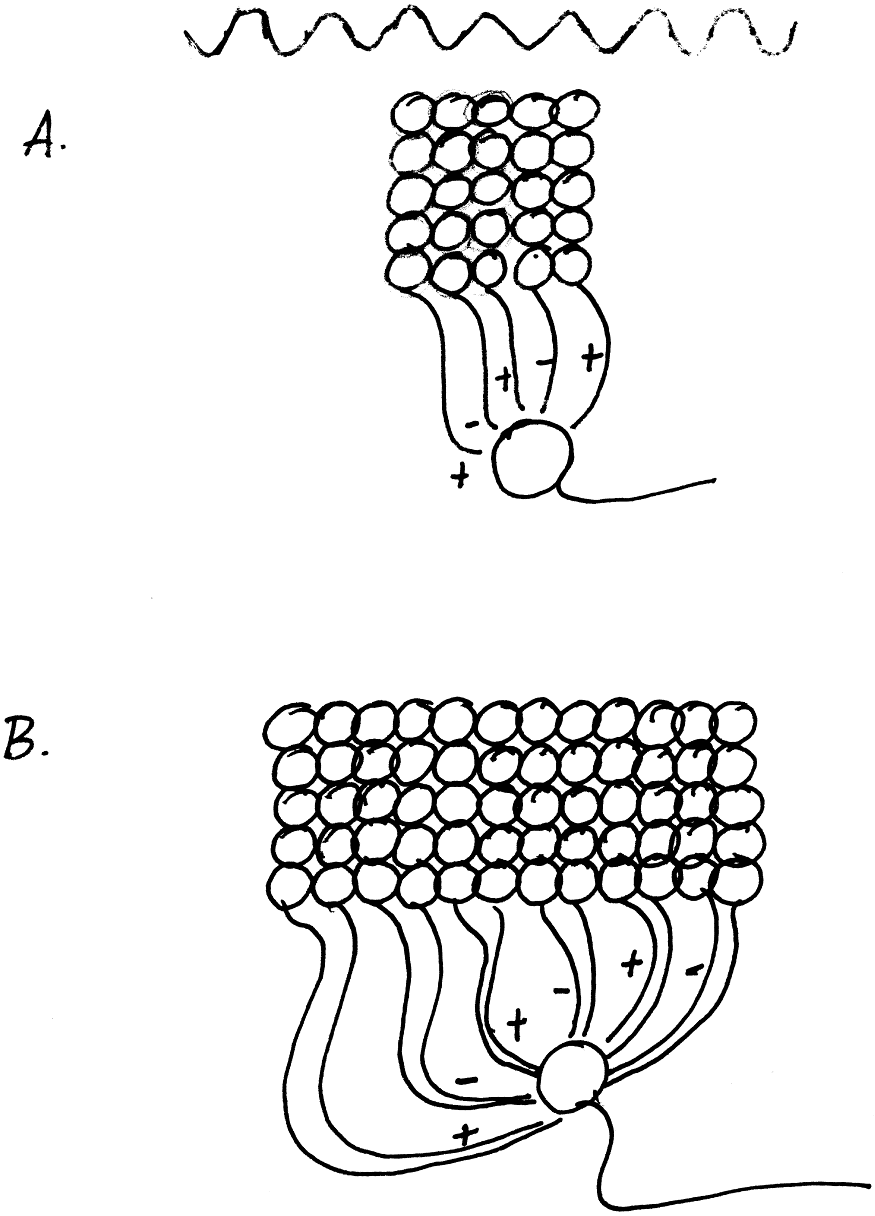


Figure 16.11: Resampling of LGN cells to make oriented, spatial frequency tuned neurons. The matrices of circles show the centers of a set of LGN cells of a single type (say, ON-center P cells) with a coverage of 1. To make a channel tuned to the highest possible spatial frequency, LGN cells could be sampled in vertical rows, with alternating excitatory and inhibitory inputs to the higher-level cell. To make a channel tuned to 1/2 that frequency, pairs of rows could be sampled; and so on.

16.7 Alternative views

16.7.1 Coding by edge locations or other local features

Some visual theorists, however, reject the whole line of reasoning that we have worked through in this chapter. In their view, there is no point in thinking that the visual system uses a spatial frequency based representation of the visual world. One of the major arguments is that even the spatial weighting functions in Panels D and E of Figure 16.9 are actually not very different from those generated from classical center/surround receptive fields, and are not very narrow in terms of spatial frequency coding. In fact, a Gabor function windowed to a width of about 1.5 cycles is virtually indistinguishable from a DOG function. So, the argument continues, why all the fuss about spatial-frequency-tuned neurons?

A major alternative is to return to something much closer to our original point-by-point view. Rather than using spatial-frequency-tuned neurons, perhaps the visual system resamples the ensembles of P, M, and K cells to make neurons tuned to more naturalistic features of the visual world, such as the *locations of edges or bars* in the retinal image. That is, perhaps the visual system just uses *edges or bars and their locations* rather than spatial frequencies as the basis of its coding of visual scenes.

A scheme of this sort was hinted at in Chapter 11 (Figure 11.6), in which we illustrated the neural image of an edge in an ensemble of ON-center ganglion cells. A higher level neuron with the right spatial weighting function – one that resampled the ganglion cells so as to get an excitatory input from all the neurons in a single vertical row, and inhibitory input from all the neurons in a single adjacent vertical row – would be tuned to respond to a sharp vertical edge in a particular retinal location. Higher level neurons of this kind could be considered to be tuned to vertical edges in specific locations, and the argument could be extended to create neurons tuned to edges and/or bars in various orientations.

16.7.2 A synthesis: Coding on multiple spatial scales and orientations

In the old days, the edge-location coding view and the spatial frequency coding view were seen by some vision scientists to be in conflict, and there was some light-hearted name calling among the irreverant. Multiple channel theorists called the edge theorists *feature creatures*, and the edge theorists called the multiple channel theorists frequency freaks, or even *Fourier freaks* if the fur was flying.

But over time a mellower view has emerged. First, as we have seen, the spatial frequency view was modified from models that required a Fourier-like analysis of the scene as a whole, to models with neurons with local receptive fields, reflecting the compromise between spatial frequency and local spatial tuning.

Second, the two views are not mutually exclusive. Neurons tuned for spatial frequency also respond to bars and edges. In fact, a neuron tuned to a particular spatial frequency will respond best to a bar that just covers its receptive field center, and a neuron tuned to respond to a bar of a certain width will have a band-pass spatial frequency tuning curve. Scientific descriptions can sometimes come in alternative, equally preferable forms, and DT knows of no arguments that a crude sourt of Fourier analysis is more or less elegant or parsimonious than a code based on the locations of edges and bars. In such a case, which description one adopts is partly a matter of personal preference, and it makes sense to speak both languages.

And third, even theorists who favor an edge-location rather than a spatial frequency code tend to posit the existence of several sets of edge-tuned and/or bar-tuned neurons, tuned to different spatial scales. For example, the computational scientist David Marr (19xx) favored edge tuning, but postulated that each edge is processed by edge detectors tuned to three different spatial scales – coarse, medium and fine.

16.8 The Design question: Why multiple spatial channels?

Finally, the Design question: Why might the visual sustem have evolved to have the particular coding schemes that it does? In particular, why might it be useful to have a stage at which spatial information is coded in a set of frequency-tuned channels with bandwidths of about 2 octaves?

Two kinds of principles often enter into Design arguments in this field. The first is the general Design principle introduced in Chapter 1: *the visual system evolved to process the stimuli that occur in the natural environment in which the organism evolved.* In the field of spatial frequency analysis, this Design principle has led vision scientists to Fourier analyse sets of natural scenes, and determined their statistically likely amplitude spectra. As it turns out, most natural scenes have broad amplitude spectra – they contain a wide range of spatial frequencies. However, the highest amplitudes occur at the lowest spatial frequencies, and amplitudes decline with 1/f. The argument is, then, that our visual systems must have evolved to process stimuli with amplitude spectra of this kind.

The second kind of Design principle addresses the question: What kinds of coding rules make the most desirable neural codes, and do different rules work best at different levels of processing? At the retinal output level, for example, we introduced the hypothesis that what is needed is a *compact* or *efficient* code – a code in which *the maximal amount of information is packed into the minimum total number of neurons.* But at the cortical level, where (as we will see) compactness is not such an issue, some theorists have proposed the hypothesis that what is needed are *sparse codes*: codes in which many neurons are available to participate, but in which *any given scene is represented by activity in a minimal subset of the avialable neurons.*

The final step in this form of argument is to combine the two kinds of principles. That is, suppose one believes that in natural scenes the amplitude spectrum declines with 1/f, and that a sparse code is optimal at an early cortical level. One can vary the parameters of a multiple channels model – for example, the numbers and bandwidths of the channels – to see which combinations of parameters yield the sparsest coding of the amplitude spectra of natural scenes. The question is not yet settled, but some modellers argue that a few spatial channels, with bandwidths on the order of about two octaves, provide an optimally sparse code for functioning in our natural environment. Such arguments, if they are correct, provide an answer to our Design question.

Other theorists, however, do not buy the argument for sparse codes, and suggest instead that it makes more sense to use efficient codes throughout the visual system. Yet others have argued that much of the machinery of the visual system has evolved to reduce the redundancy in the visual code, and that an optimally efficient code will also be minimally redundant. Modelling of these kinds of Design properties is currently receiving increasingly active attention in some parts of visual science.

16.9 Summary

In summary, we began this chapter with the idea of neural codes and neural coding in the domain of two-dimensional visual space. In the color domain, it would never occur to us to think that color is coded by color – that for us to see red, a literally red neuron must be active. Yet in the spatial domain it is easy to fall into the trap, and assume uncritically that a spatially extended object must be represented in a neural image that is also extended in space.

A major alternative conceptualization, of historical importance, is that at some intermediate level of visual processing a spatial pattern is represented by its amplitude spectrum, with a different high level neuron tuned to each different spatial frequency that Fourier analysis would reveal in the physical scene. Although we have had to back off from the extreme of this perspective – physiological systems cannot do mathematically perfect Fourier analysis – we have come to the idea that perhaps the visual system does a crude and local sort of Fourier analysis, representing the subregions of scenes by the activity in a set of local channels tuned to different spatial frequency bands. Such channels would also respond well to bars of particular widths, and the feature and frequency views are now seen as mutually compatible descriptions. Moreover, we fully expect that whatever the intermediate code, it will itself be recoded in many ways to serve different visual and motor tasks.

We close with a comparison between multiple channels models and models of color vision. Quantitative color theory began over 100 years ago with the psychophysical fact of trichromacy. It was realized early on that trichromacy implies the presence of three and only three univariant "fundamentals", but does not constrain the options enough to allow us to derive their spectral sensitivity curves. Many ancillary assumptions were made, and many mathematical models proposed, with no means of resolution among them. Finally, beginning in the 1960s, a variety of direct techniques allowed us to pin down the spectral sensitivities of the three cone types with greater and greater precision.

At present, there are many multiple channels theories of spatial vision. What this probably means is that multiple channels theory is still in an early stage of its evolution. Psychophysical data like those from the Graham and Nachmias, and Blakemore and Campbell experiments suggest the presence of multiple spatial frequency channels, but do not constrain these models in enough detail. We do not know at this stage whether or not direct physiological evidence will ever allow us to decide among the different mathematical models – whether or not multiple channels theory can ever achieve the status of trichromatic theory. As mathematicians, we can of course be satisfied with a purely mathematical model, with channels defined as a set of mathematical analysers with somewhat arbitrarily chosen properties. But as vision scientists, we would very much prefer a model that incorporates the physiology of the visual system. In the next chapter we will explore the receptive fields and spatial weighting functions of neurons in V1 cortex.

Chapter 17

Area V1 (Primary Visual Cortex)

As discussed in Chapter 15 (Figure xx), most of the ganglion cell axons from the two retinas project to the two LGN, and most of the LGN cell axons project to two symmetrically arranged cortical regions buried deep within the cerebral cortex at the very back of the brain. This region has had several names. It has historically been called the *primary visual cortex*, since it is the cortical area that first receives the incoming visual signals. It has also been called *Area 17* and *striate cortex*. Most recently, in the monkey it has been called *cortical area V1* (for visual area #1). We will use the term *primary visual cortex* when we wish to be general across species, and *V1* when we are referring to monkeys (and by extension to human beings).

Throughout this book, we have been viewing visual processing as a series of transformations of the neural code in which information about the world is carried. The code transformation from LGN to V1, and within V1, is in our counting scheme the fifth transformation. In contrast to the minimal transformation from retina to LGN, significant transformations of the visual code occur between the input and the output of cortical area V1. What might they be like? In particular, will they begin to approximate a code that might form the basis of the perception of objects?

Of all of the various regions of the cerebral cortex, V1 is probably the most studied. It is an attractive target of research because the activity of neurons in V1 is strongly but selectively influenced by visual stimuli, and yet V1 is also part of the broader visual cortical circuit. An enormous amount is known about V1, and we cannot hope to be true to the richness of knowledge in this area. As with other chapters, the strategy of this chapter is to use an historical approach to define the basic concepts, and then to add a few highly selected current studies that illustrate important philosophical principles or changes of perspective.

17.1 Overview of the human brain and the primary visual cortex

We begin with a very brief overview of the human brain. Figure 17.1A shows the brain as it would be seen from the left side of the head. The largest and most prominent feature of the brain is the symmetrical pair of structures called the *cerebral hemispheres*, which form a gnarled but roughly hemispherical mantle that covers the rest of the brain. A thin outer layer, about 2-4 mm thick, on the surface of the cerebral hemispheres is called the *cerebral cortex*. Largely for convenience, neuroscientists divide each cerebral hemisphere into four parts, or lobes – the *frontal, parietal, occipital, and temporal lobes*. As we shall see, all four lobes contain regions that are important to

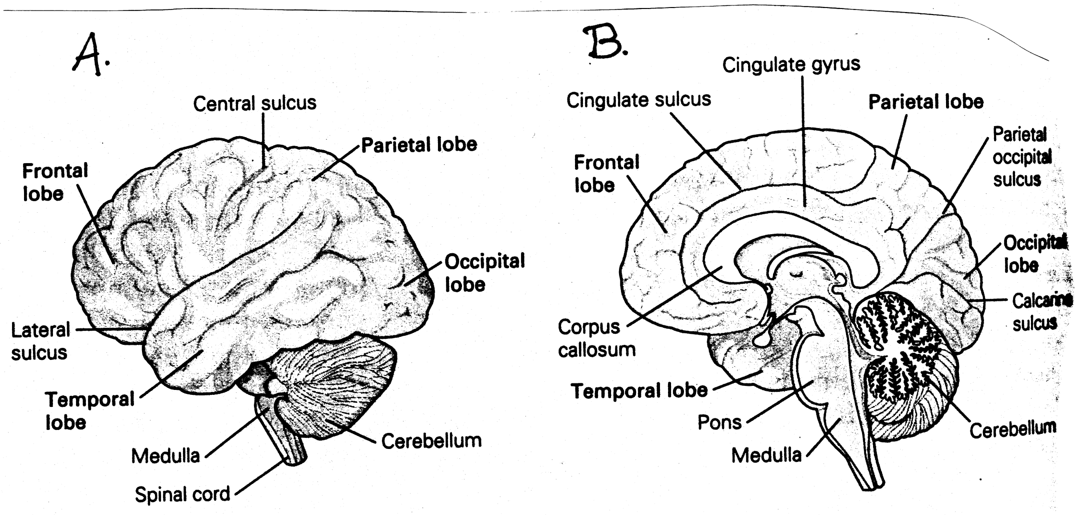


Figure 17.1: The human cerebral cortex and its major divisions. A: A view of the human brain from the left side. The crinkled mantle covering most of the outside of the brain is the cerebral cortex. B: A medial (midline) view of the right side of the brain. Note the calcarine sulcus; visual area V1 lies in this sulcus. [After Kandel, Schwartz, and Jessel, 2000, Fig. 17-4, p. 324.]

vision.

A vertical slice between the two hemispheres reveals the medial (or midline) surfaces of the cerebral hemisphere. A view of the right half of the brain, after the left half has been removed, is shown in Figure 17.1B. The convoluted surface covering the medial surface of the hemisphere is more of the cerebral cortex. Figure 17.1B also shows the *corpus callosum*, a large band of fibers that makes connections between the two hemispheres. It also shows the *calcarine sulcus*, a deep groove in the medial surface of the occipital lobe, within which lies the target of our current interest: the primary visual cortex, or in the monkey, area V1.

17.1.1 Topography: two half maps, one visual world

As discussed in Chapter 15, Figure xx, the optic nerves carry the axons of the retinal ganglion cells from the retina to the LGN. The crossing over of half of the axons from each optic nerve splits the retinal input vertically down the middle. Signals from the left half of the visual field (the right half of each retina) project to the right LGN, and vice versa. Moreover, the topographic maps originating in left and right retinas are aligned within the LGN – remember the club sandwich analogy – such that each LGN contains six superimposed maps of half of each retina, and correspondingly, half of the visual field.

Axons from output neurons in the two LGN project directly to the two areas V1, with the left LGN projecting to V1 in the left hemisphere and the right LGN to V1 in the right hemisphere. Thus, as previously noted, information from the two halves of the visual world ends up in two entirely separate cortical regions. [Again, does it bother you that we nonetheless see the visual world as a seamless whole? Why or why not? There is a linking proposition lurking here.]

Figure 17.2 shows the topographic mapping from the visual field to the two areas V1. The map that starts in the retina and is topographically preserved at the LGN (Figure **??**xx) is also preserved at area V1. The differential increase in size of the foveal and central retinal representation (the *magnification* of the fovea and the central retina) seen in LGN is continued in V1, such that the representation of the central 10 degrees of the visual field occupies almost half of the topographic map in V1. As discussed in Chapter 15, the cortical magnification of the fovea and central retina probably occurs because of the larger numbers of ganglion cells in these regions, as though each ganglion cell claims at least roughly the same amount of processing space in the LGN and visual cortex.

In Figure 17.2, several landmarks have been indicated on the topographic maps. The vertical and horizontal lines through the fovea are called the *vertical* and *horizontal meridia.* The vertical meridian is important because it separates left and right hemifields – that is, regions that will be projected to separate LGNs and to separate (left and right) V1 cortex. The horizontal meridian similarly separates upper from lower halves of each hemifield. The detailed spatial layout of the topographic maps in V1 will be important to the interpretation of data from functional magnetic resonance imaging (fMRI) in Chapter 18xx.

17.1.2 Effects of lesions: Blindness and blindsight

Damage to V1 interrupts the major sensory input from the eyes to the cortex. One would therefore expect that a person with a V1 lesion would be blind in the visual field region corresponding to the lesion, or that a person missing all of V1 would be completely blind. And indeed, people with V1 lesions report that they are perceptually blind in these ways. Clearly, V1 lesions destroy a critical link in the causal chain between the eyes and conscious perception.

However, some V1 lesion patients who have been studied in the laboratory demonstrate an intriguing syndrome known as *blindsight.* Patients with this syndrome report that perceptually they are completely blind in the parts of the visual field corresponding to the cortical damage, yet when carefully tested they can be shown to retain some visual and visuomotor capacities. Pupillary responses are usually retained, and these patients can move their eyes with considerable accuracy to point them toward objects that they report they cannot see. In some cases such patients can even, to their own surprise, point with considerable accuracy to targets in their "blind" visual fields. Some can insert a card in a slot, while not being able to report the orientation of the slot, or indeed its presence in their visual field. An example of pointing behavior by blindsight patient DB (Weiskrantz et al, 1974) is shown in Figure 17.3.

The phenomenon of blindsight shows that at least some visual information must reach the human motor systems through some additional pathway other than through V1. The most probable route is via a pathway from the retina to the superior colliculus and from there to a region of high level visual cortex that lies beyond V1. [This pathway is beyond our scope, but sneak a peek at Figures 18.xx and 18.xx].

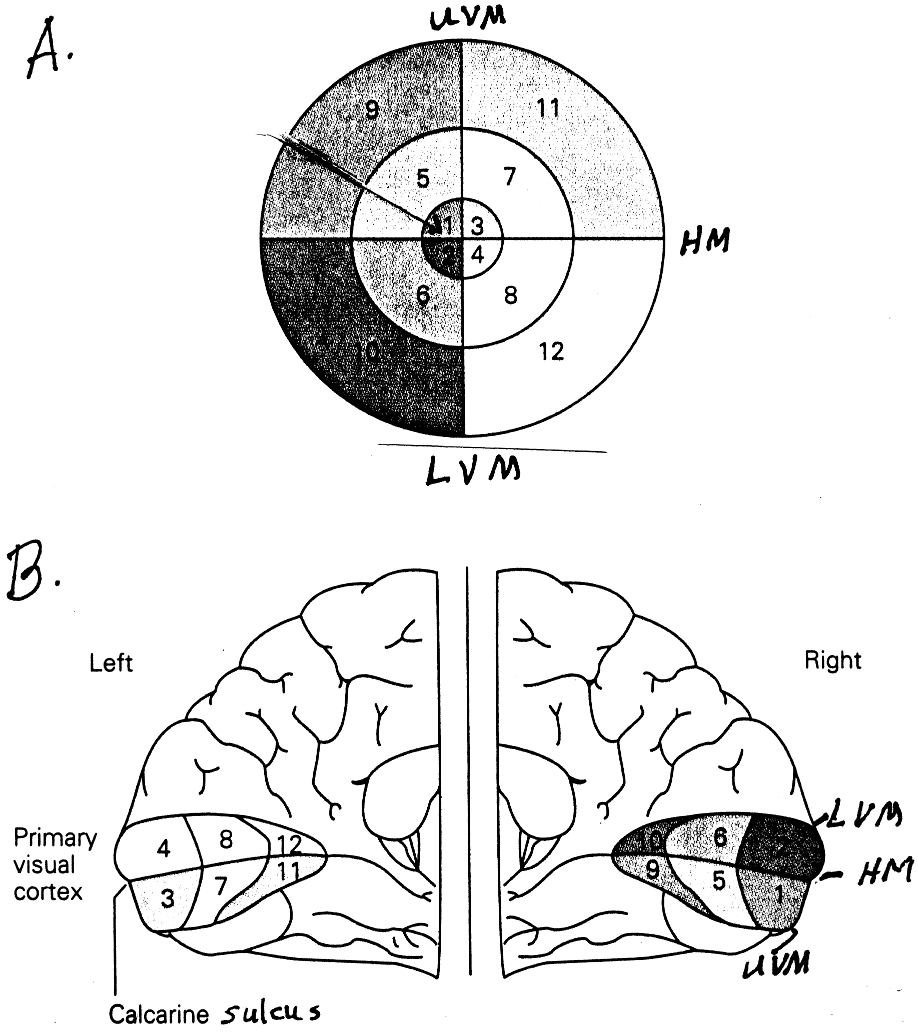


Figure 17.2: The mapping of the visual hemifields to left and right V1 cortex. A: the visual field. The numbers represent different regions – 1, 2, 3, and 4 are sections of the central retina; 5, 6, 7, and 8 are regions of the near periphery, and regions 9, 10, 11, and 12 are regions of the far periphery. B: the brain has been sliced down the midline and opened out, with the medial surfaces of the two occipital lobes laid at left and right of the figure, and the calcarine sulci pulled open for illustrative purposes. The numbers in B correspond to those in A. The central retina in the left hemifield projects to the region closest to the surface of the occipital lobe; the near periphery lies deeper; and the far periphery lies deepest.LVM = lower vertical meridian; UVM = upper vertical meridian; HM = horizontal meridian. [Modifieid from Kandel, Schwartz, and Jessel, 2000, Fig. 27-9, p. 532].

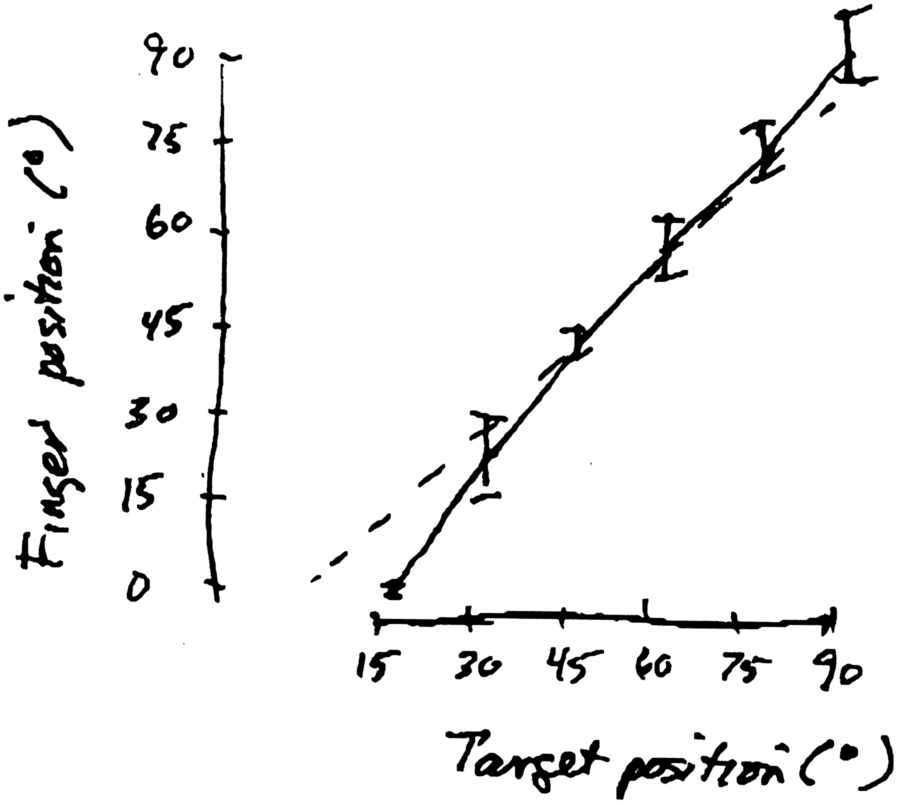


Figure 17.3: Blindsight. Patients with lesions of V1 typically report perceptual blindness in the corresponding region of the visual field. However, when asked to point to targets in the blind field, many of these patients can do so with remarkable accuracy. The data show the correspondence between finger position and target position for small targets in the "blind" field for blindsignt patient DB (adapted from Milner and Goodale, 1995, p. 71).

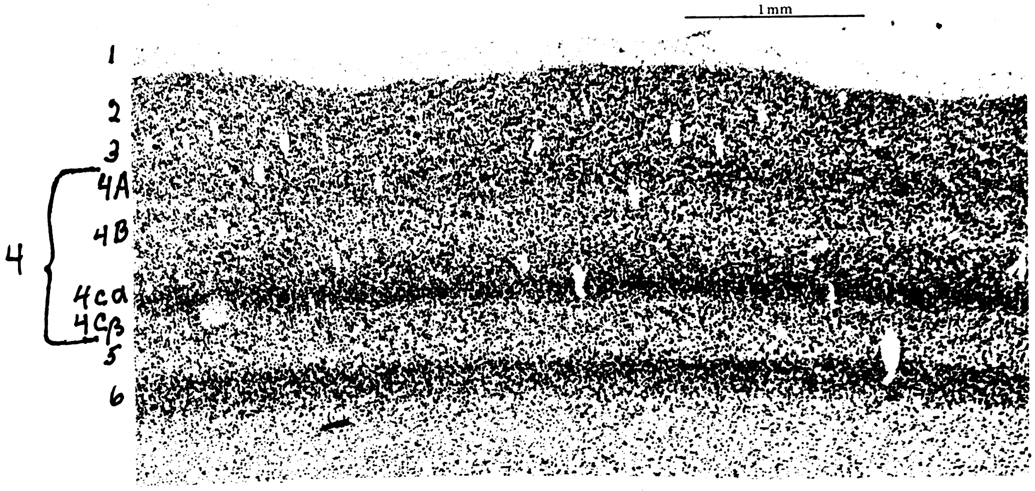


Figure 17.4: A light micrograph of a cross-section of area V1. The section has been stained with a stain called cresyl violet to reveal the cell bodies. The surface of the cortex is at the top. The layers are numbered from 1 to 6, as shown at the left. The dark stripe in layer 4Cα, is visible to the naked eye. It is confined to area V1, and provides one of the names for this area – the striate cortex. [After Hubel and Wiesel (1977), after Levine & Shefner 1991, Fig. 8-5, p. 170].

17.2 Neuroanatomy of area V1

17.2.1 The classic story: Inputs, outputs, and simple circuits

A light micrograph of a cross-section of area V1, stained with a stain that reveals the cell bodies, is shown in Figure 17.4. It is obvious that V1 is richly endowed with cell bodies. In terms of numbers of cells, each V1 contains about 250 million neurons (500 million between the left and right V1 regions together).

These large numbers raise a puzzle. The LGN sends about one million afferents to V1, and all of the incoming information must be packed into the activity of these neurons, presumably in a very compact code. But then why did the EDC provide us with 500 million V1 cells to carry the same information? The numbers alone tells us that some important recoding scheme must be carried out here – the incoming information must be represented in the activity of many more neurons, presumably in some kind of a sparser code.

Figure 17.4 also shows that area V1 is a markedly *layered* structure. By tradition, neuroanatomists divide all regions of the cerebral cortex into six layers (the fact that the primate LGN has six layers is coincidental). The cortical layer that receives most of the inputs from other regions of the brain lies in the middle of the cortex, and this region is traditionally called *layer 4*. In primary visual cortex layer 4 is very thick, and anatomists divide it into sublayers – 4A, 4B, 4Cα, and 4Cβ – for reasons that will become apparent shortly. Above layer 4 lie the superficial layers: layer 1 (which contains no cell bodies), and layers 2 and 3. Below layer 4 lie the *deep layers*, layers 5 and 6.

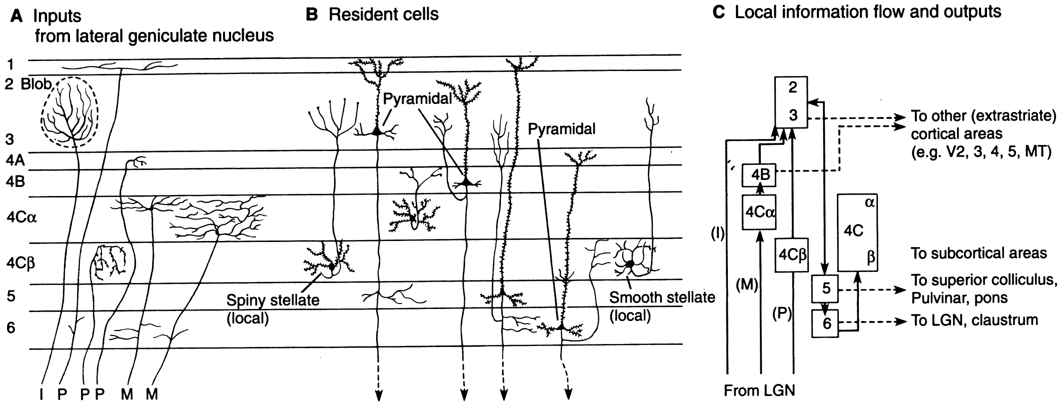


Figure 17.5: Inputs, resident cells, and outputs from Area V1. A: inputs from LGN neurons: M cells project to layers 4Cα; P cells to layers 4Cβ and 4A; and K (I) cells to layers 2 and 3. B: resident cells. Spiny stellate cells receive the input from LGN. Smooth stellate cells are local circuit neurons. Pyramidal cells participate in the internal processing, and send axons out to other regions of the brain. C: some probable internal circuits, traceable in A and B. C also shows the output destinations of the pyramidal cells: layers 2, 3, and 4B project forward to other cortical areas, layer 5 projects to the superior colliculus, and layer 6 projects back to the LGN. [From Lund, 1988, via KSJ 2000, Fig. 27-10, p. 533.]

In 1988, Jennifer Lund published a classic study of the inputs and circuitry for area V1 in the macaque monkey. Figure 17.5 shows Lund's summary diagram. The *inputs* from LGN to V1 are shown in Figure 17.5A. In Lund's analysis, most of the geniculate inputs arrive at layer 4. Interestingly, axons from the magnocellular (M) and parvocellular (P) layers of the LGN remain separate, and deliver their inputs to different sublayers of layer 4: M cells to layer 4Cα, P cells to layers 4A and 4Cβ. The inputs from koniocellular (K) LGN cells – then called interlaminar cells (I) – project directly to cortical layers 2 and 3, thus providing an exception to the general rule that all input comes to layer 4.

Figure 17.5B shows the *resident cells* – those whose cell bodies lie within Area V1. By Lund's account, resident cells in V1 are of three general types. The first two types have cell bodies that are roughly star shaped, or *stellate.* The first type of stellate cells, the *spiny stellate* cells, are covered with protruberances called spines. Spiny stellate cells have their cell bodies in layer 4, and receive and process the inputs from the M and P LGN cells. The second type, called *smooth stellate* cells, are local circuit neurons – they contact only other V1 cells, and process information only within V1 itself.

The *outputs* from V1 are provided by the third type of resident cells, called *pyramidal cells* because of their roughly triangular cell bodies. Pyramidal cells occur in layers 2, 3, 4B, 5, and 6. The axons of pyramidal cells project downward from the cell body. These axons send out branches, or *collaterals*, as they pass through various layers of V1, and doubtless make major contributions to the integration of signals within and across the cortical layers. In Lund's classic sketch, the axons of pyramidal cells project downward through layer 6 and thereby out of area V1.

Figure 17.5C shows Lund's speculative summary of some simple but likely circuits for information flow within area V1. By comparing Figure 17.5A and B with 17.5C, you can trace out the types of neurons that make up some of the circuits. For example, the M and P inputs arrive in layer 4. The fates of their signals can be followed through the cortex, and M, P and K (I) inputs can all be seen to converge in layers 2 and 3. The output destinations of the pyramidal cells are also shown. Pyramidal cells in layers 2 and 3 project to higher cortical areas (see Chapter 18xx). In contrast, pyramidal cells in layers 5 and 6 project to subcortical areas. Layer 5 projects to the superior colliculus, and gives it access to processed information from V1. Layer 6 projects back to the LGN, and forms the feedback projection we have already described in Chapter 15.

In the broader realm of overall cortical neuroanatomy, neuroanatomists have proposed two general rules that describe the connections among cortical areas. First, most of the *inputs* to a given cortical area arrive and synapse in layer 4 of the target area. And second, most of the *outputs* originate from layers 2, 3, 5, and 6. Moreover, outputs from the superficial layers, 2 and 3, typically feed forward to higher cortical areas, whereas outputs from the deep layers, 5 and 6, typically feed back to lower cortical areas or subcortical structures. As described above, most of the connections in V1 follow these rules, with the notable exception that the inputs from K cells project directly to the superficial layers 2 and 3.

17.2.2 Laateral connections and feedback

Lund's summary of V1 connections shown in Figure 17.5 emphasizes the *vertical* connections among V1 neurons, perpendicular to the cortical surface. In fact, Figure 17.5 could leave the impression that V1 carries out only spatially local computations, and this scheme would be consistent with preserving a relatively precise topographic map in V1.

However, it has been shown more recently that the pyramidal cells in V1 also make *lateral, or horizontal*, connections across V1, parallel to the cortical surface. An example of horizontal connections made by the axon collaterals of a pyramidal cell in layer 3 of monkey cortex is shown in Figure 17.6. Notice that the terminations of the bunches of axon collaterals are patchy, as though they were seeking out some specific set of other neurons at regular spatial intervals across the topographic map. We will return to the meaning of this pattern later.

Finally, a catalog of the inputs to V1 must also include *feedback* connections from higher cortical levels. We will return to this topic briefly in Chapter 18xx.

17.2.3 Modern complexities: More cell types, more connections

Since Lund's classic study, much more has been learned about the details of cell types and circuitry in area V1, and the literature is much too extensive to be summarized here. To give you the flavor of the work, Figure 17.7 shows some examples of the complexity of the known cell types. Figure 17.7A shows a set of eight different types of pyramidal cells (which are now known to be excitatory) with cell bodies located in layer 6 of macaque monkey cortex; and Figure 17.7B shows a set of 12 different varieties of local neurons (all known to be inhibitory) with cell bodies located in layers 3B, 4A, and 4B. Each of these different cell types probably enters into different, precisely defined neural circuits that do different visual processing tasks within V1. Of course, defining all of the circuits in V1 is a formidable goal, but vision scientists have taken it on, and work on these circuits is a field of very active investigation at the present time.

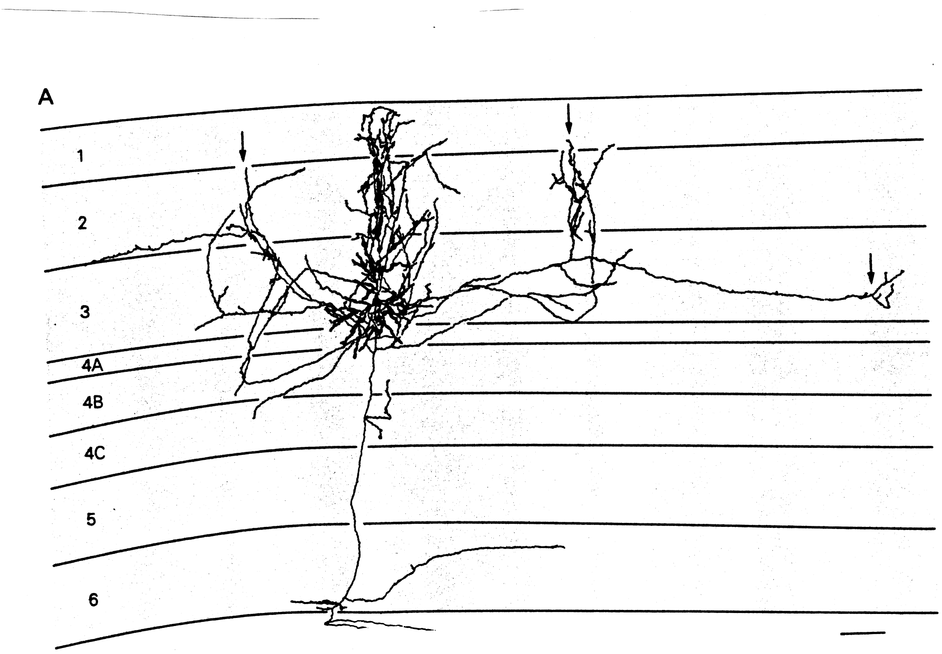


Figure 17.6: Lateral connections within area V1. This layer 3 pyramidal cell from a macaque monkey sends axonal branches across V1. The patchy terminations in layers 2 and 3 suggest that the cell is seeking out specific targets at regularly spaced intervals. [Adapted from KSJ 2000, p. 542, Fig. 27-18; after McGuire, Gilbert, Rivlin and Wiesel, 1991.]

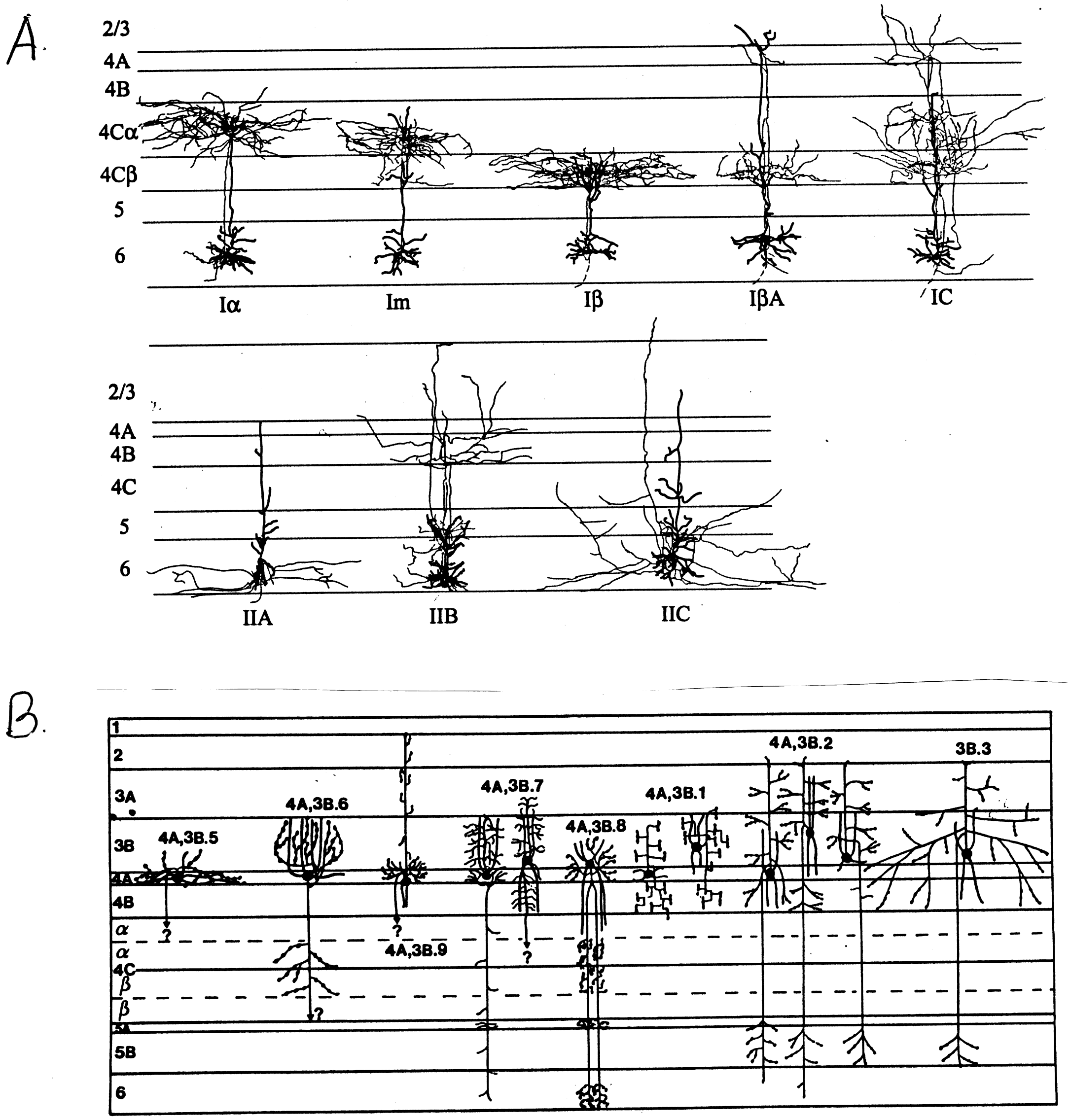


Figure 17.7: Some representative details of V1 anatomy. A: Eight varieties of pyramidal cells from layer 6. B: Twelve varieties of inhibitory neurons from layers 3B, 4A, and 4B. [A from Briggs and Calloway 2001; B from Lund and Yoshioka 1991. Both pictures via Callaway, 2004. A: Fig. 42.3, p. 687; B: Fig. 42.4, p. 690.]

17.3 Physiological properties of V1 neurons

What are the physiological properties of neurons in primary visual cortex? To what visual stimuli do they respond? Do they have definable receptive fields? Prior to the 1950s,these questions lay unanswered. In the late 1950s two students in Stephen Kuffler's laboratory, David Hubel and Torsten Wiesel, took on the problem. For several months they worked at recording from single cortical cells in cats. Following Kuffler's approach to retinal ganglion cells, they used an ophthalmoscope modified to take a set of slides with opaque spots or holes of different sizes. In this way they produced dark and light spots of different sizes, and attempted to plot the receptive fields of cortical cells. They found that the receptive fields could indeed be characterized. Like ganglion cells, the receptive fields of cortical cells had ON and OFF regions. The receptive fields generally seemed to be elongated rather than round, but the neurons did not seem very responsive to small spots.

David Hubel describes what happened next in his Nobel lecture (1982). "Our first real discovery came about as a surprise.We were inserting the glass slide with its black spot into the slot of the ophthalmoscope when suddenly over the audiomonitor the cell went off like a machine gun. After some fussing and fiddling we found out what as happening. The response had nothing to do with the black dot. As the glass slide was inserted its edge was casting onto the retina a faint but sharp shadow, a straight dark line on a light background. This was what the cell wanted, and it wanted it, moreover, in just one narrow range of orientations.. (p. 516)" Hubel and Wiesel changed their paradigm, and began mapping the receptive fields of cortical neurons with stationary and moving bars and edges. They soon graduated from cats to monkeys, and the rest is history. [What is the moral of this story?]

17.3.1 Orientation tuning

Some examples of typical receptive fields of cells in cat primary visual cortex are shown in Figure 17.8A. An early discovery was that the receptive fields of many cortical cells do not have the center-surround structure typical of ganglion cells or LGN cells. Rather than disk-shaped centers and donut-shaped surrounds, the receptive fields of these cells are markedly elongated, showing parallel regions dominated by ON vs. OFF responses. As one might guess from the receptive field maps, these cortical neurons are tuned to respond best to an edge or a bar of light, presented at a particular location on the retina at a particular orientation.

As the next step, Hubel and Wiesel traced out the *orientation tuning* of these neurons more quantitatively. That is, they determined the responses of cortical neurons to bars or edges varied systematically along one stimulus dimension, that of orientation. The result from a cell in cat cortex is shown in Figure 17.8B. This particular cell responded optimally to a vertically oriented bar. The response diminished as the orientation varied from vertical, and for orientations more than about 30 degrees from vertical the cell no longer responded. Figure 17.8C shows similar data from another cell, presented in the format of an *orientation tuning curve.*

The use of the term *tuning* points to the essential importance of these discoveries. As we have seen, ganglion cells and LGN neurons respond rather indiscriminately to contrast within their receptive fields, and will respond to many different spatial patterns within a specified retinal region. But even within a specified retinal region, orientation-tuned cortical neurons respond only to bars or edges at a small range of orientations, and completely ignore most other spatial patterns, including

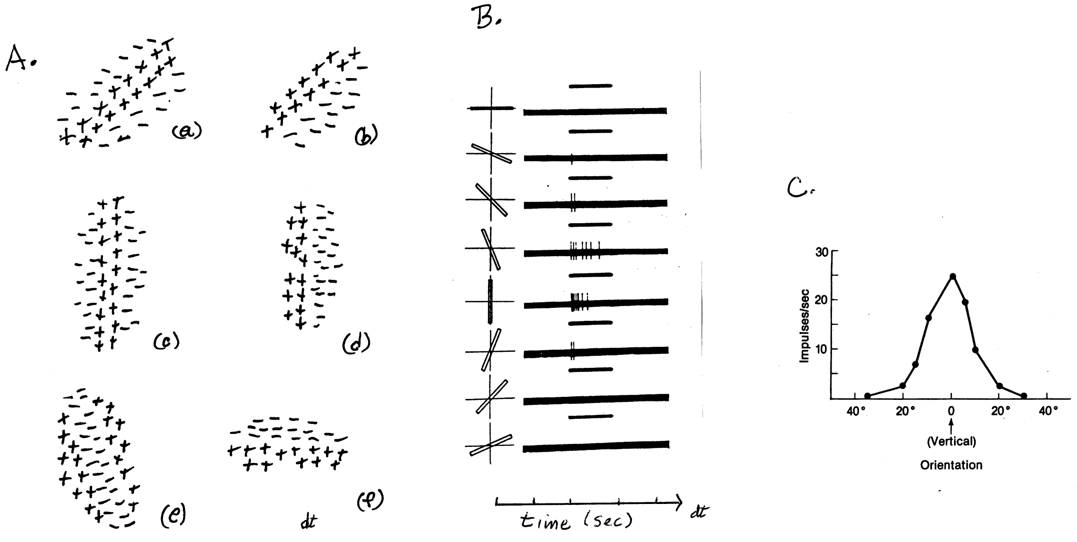


Figure 17.8: Receptive fields and orientation tuning of cortical neurons in cat primary visual cortex. A: six oriented receptive fields, with their ON regions (+) and OFF regions (-). Different neurons respond optimally to bars (left column) or edges (right column), at each of the different orientations around the clock. Some of the cells that respond to bars have ON centers and OFF surrounds (cells a, c) and others have OFF centers and ON surrounds (e). B: This cell responded optimally to a near-vertical bar, less to bars slightly off vertical, and not at all to bars tilted more than about 30 degrees from vertical. The short horizontal lines mark the presentation of the stimulus. C: data from a similar neuron, presented as an orientation tuning curve. [A: dt; B: Modified from H&W, 1959, Fig. 3, p. 579; C: from Goldstein, 1999, Fig. 3.9, p. 77.]

bars and edges at other orientations!

This degree of *stimulus selectivity* or *stimulus specificity* is a novel and remarkable feature of early cortical processing. It raises the intriguing possibility that an individual cortical neuron might be specialized to respond optimally to only a small and highly selected range of patterns of light (for example, a light/dark edge oriented at about 45^o, located 10^o directly to the right of the fovea). The flip side of specificity, of course, is that many neurons will be needed to represent all of the many possible visual stimuli. [Does this sound familiar?]

17.3.2 Simple, complex, hypercomplex, and non-oriented cells

In their early work, Hubel and Wiesel divided cortical cells into three types, which they called *simple, complex*, and *hypercomplex cells. Simple cells* are like those described above: they have oriented receptive fields with well-defined ON and OFF regions, and respond best to an edge or to a bar of a specific width and orientation, at a particular location on the retina. They also have relatively linear spatial summation properties; for example, two spots of light within the ON region (or the OFF region) summate their effects in simple ways, and spots of light presented simultaneously in the ON and OFF regions cancel each others' effects.

Complex cells also respond best to oriented lines or edges, but their receptive fields do not have distinct ON and OFF regions, and cannot be defined with spots of light. That is, complex cells respond to an oriented bar of a particular width anywhere over an extended region of retina. An example of a complex cell is given in Figure 17.9A. This cell responds to narrow horizontal bar of light in many locations within an extended receptive field, but does not respond to a wider bar. Another distinguishing feature of complex cells is their nonlinearities: their responses to complex stimuli cannot be predicted from summing their responses to simpler ones.

Hypercomplex cells are marked by a different characteristic. They respond best to bars or edges of a particular length. If the bar or edge is too long, so that its ends fall outside the initially defined receptive field, the response of the cell is diminished. An example of a hypercomplex cell is shown in Figure 17.9B. These cells are also called *end-stopped* cells.

Finally, it turns out that in primates, most cells in the input layer – layer 4 – actually have *non-oriented* center-surround receptive fields, mimicking the receptive fields seen in LGN. Thus, the elegant anatomical circuitry we saw in Figure 17.5 must be used to create the orientation tuning and other characteristics seen in simple, complex and hypercomplex cells.

These categories of neurons – non-oriented, simple, complex, and hypercomplex, cells – served the field well for many years. It has been increasingly realized, however, that rather than being discrete types of neurons, Hubel and Wiesel's categories may mark points on various continua. For example, neurons in the input layer, layer 4, show varying degrees of orientation specificity, from non-oriented to strongly oriented; and both simple and complex cells show varying degrees of end-stopping. In consequence, there is no current consensus on the number or definition of physiological types of neurons in V1 cortex. Perhaps partly for this reason, there is no agreed upon identification between physiological cell types and the anatomical cell types shown in Figures 17.5-17.7.

17.3.3 Selectivity on many stimulus dimensions

Once orientation tuning was discovered, the question became, are cortical neurons tuned on other stimulus dimensions, and if so, which ones? Are there other features of the stimulus for which V1 neurons show their remarkable specificity? The answer is yes.

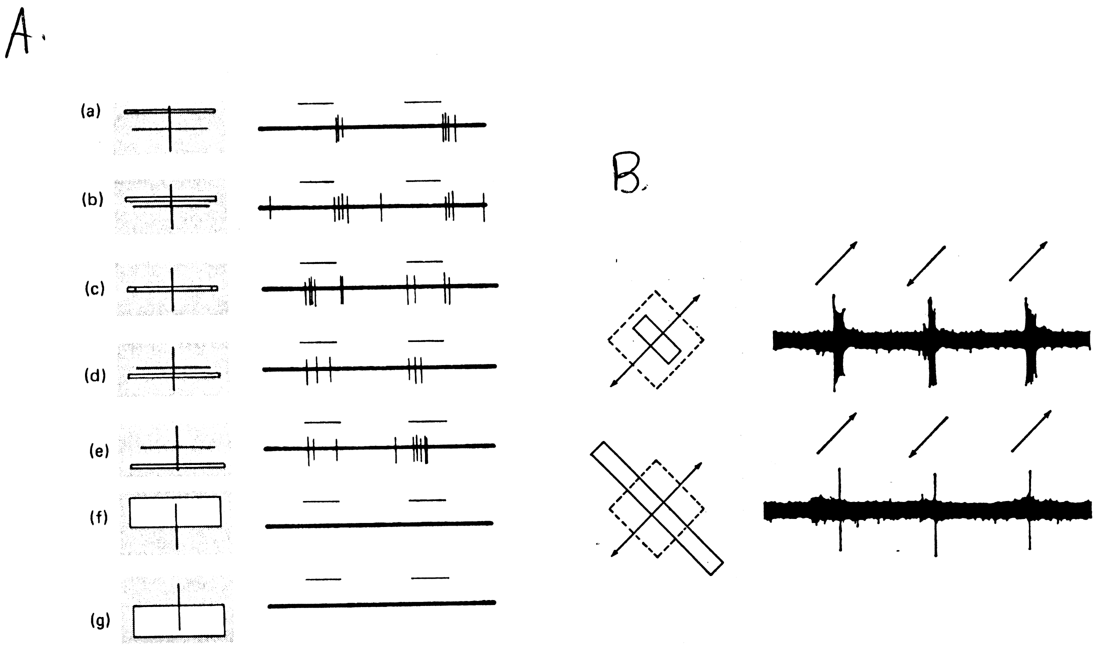


Figure 17.9: Complex and hypercomplex (end-stopped) cells. A: A complex cell from primary visual cortex in the cat. The crosses in the left column mark the location of the receptive field. The short horizontal lines in the right column mark the time of stimulus presentation (1 sec). A narrow horizontal bar presented at many locations in the receptive field (a-e) increases the activity of the neuron. A broader bar (f, g) has no effect. B: A hypercomplex cell from monkey V1. The dashed box in the left column shows the receptive field of the neuron. The arrows indicate that the stimulus was in motion. In the right column, the upper trace shows that a short slit of light moving in either direction excites the cell. The lower trace shows that a longer bar is much less effective. [A from Hubel and Wiesel, 1962, Fig. 3. B from Hubel and Wiesel, 1968, Fig. 4, p. 222.]

In their early work, Hubel and Wiesel showed that many cortical cells that are tuned for orientation are also tuned for the *direction of motion.* Figure 17.10A shows an example of a complex cell in monkey cortex that is tuned for both orientation and direction of motion. This neuron was tested with bars of light of many orientations, moving orthogonally to the orientation of the bar (left column). As we have come to expect, the cell responded only to a narrow range of orientations (rows 3 and 4). If the bar of light was not aligned with the orientation tuning of the cell, the cell did not respond to the bar for either direction of motion. But if the bar was oriented correctly, the cell responded to motion of the bar upward and to the right, but not downward and to the left. Notice that this cell is selective on two different stimulus dimensions – the stimulus must have not only the right orientation, but also the right direction of motion.

Hubel and Wiesel also showed that V1 neurons vary in their degree of *binocularity*, or *ocular dominance.* That is, whereas some V1 neurons respond to light in only one eye, many others have receptive fields in both eyes. When a neuron is binocular, the two receptive fields usually have corresponding orientations in corresponding retinal locations in both eyes. Some of these neurons respond more strongly to one eye, some to the other eye, and some equally to both. Data concerning ocular dominance are often summarized in an ocular dominance histogram such as that shown in Figure 17.10B. In this graph, the abscissa shows the degree of ocular dominance in seven categories. Cells in category 1 respond only to the contralateral eye (the left eye if one is recording from the right hemisphere, and vice versa) whereas cells in Category 7 respond only to the ipsilateral eye. Cells in the intermediate categories have receptive fields in both eyes, but with varying degrees of dominance; for example, cells in Category 4 respond equally to inputs from both eyes.[1]

It has also been shown that V1 neurons are tuned for *spatial frequency.* For example, in the early 1980s Russell De Valois and his colleagues (De Valois, Albrecht, and Thorell, 1982) studied the responses of V1 cells to sinusoidal gratings. They showed that sets of V1 neurons with overlapping receptive fields nonetheless responded optimally to different spatial frequency ranges and different orientations. Figure 17.10C shows a set of orientation-tuned V1 neurons with receptive fields that all overlapped in a single retinal region. Remarkably, these neurons are narrowly tuned in spatial frequency, with different cells serving the same retinal location being tuned to different spatial frequencies. [Does this sound familiar?]

Finally, many V1 cells are also tuned for *color.* For example, different simple cells have different null planes and optimal response axes. Many simple cells respond to isoluminant chromatic bars or gratings, and are selective for wavelength composition. An example from the work of Thorell, De Valois, and Albrecht (1984) is shown in Figure 17.10D. This cell was tested with a set of 0.2 degree bars of various colors, set to the cell's optimal orientation, embedded in an isoluminant white surround. The cell responded maximally to a green bar, and minimally to an orange one. It also responded to black and white bars with the same spatial characteristics. Moreover, the cell was tuned for bar width (or spatial frequency), and did not respond to homogeneous fields of light, either chromatic or achromatic. (The chromatic coding properties of V1 cells will be discussed in greater detail in Chapter xx (Lightness and Color)).

[1] V1 neurons also differ in their tuning for *binocular disparity* – small differences between the retinal images of the same object in the two eyes. Discussion of binocular disparity will be postponed to Chapter xx.

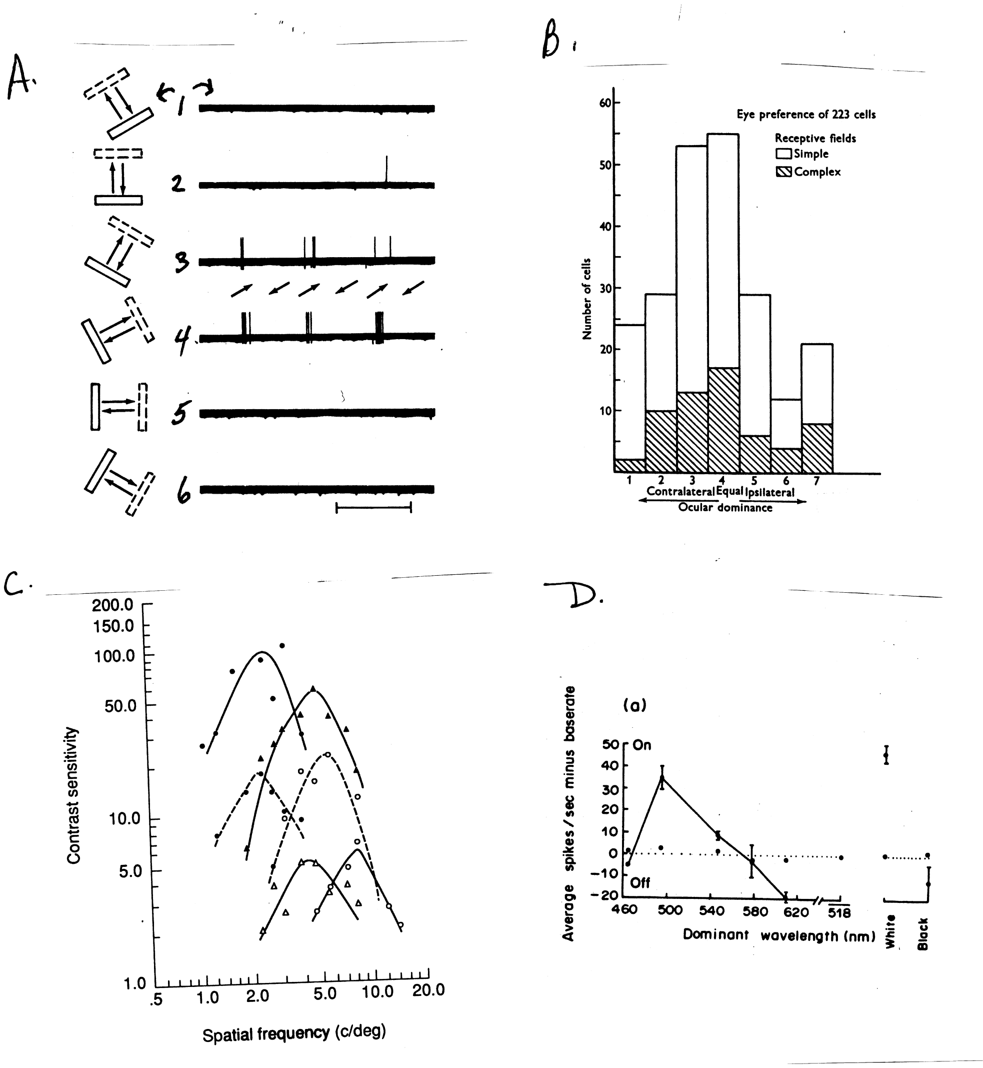


Figure 17.10: Four more varieties of tuning in V1 neurons. A: tuning for the direction of motion. The left column shows the orientation and direction of motion of the stimulus bar. This neuron responds maximally to an orientation slightly off vertical (row 4). The right column shows the responses of the cell. If the bar is aligned correctly, the cell will respond to motion of the bar – but only in one direction (arrows). B: an ocular dominance histogram for a set of about 200 cortical neurons in the cat. Most neurons respond to an oriented bar of light of the same orientation through either eye. However, the relative responsiveness to the two eyes varies. C: spatial frequency tuning for six neurons in monkey V1 cortex, all encountered in the same electrode penetration. Each neuron is tuned for spatial frequency, and the different neurons are tuned to different spatial frequencies. D. Wavelength (color) tuning. The response of the neuron varies with the wavelength of light. The cell also responds to white and black. [A from Hubel and Wiesel, 1959, Fig. 8, p. 583; B from Hubel and Wiesel, 1962, Fig. 12; C from De Valois, Albrecht, and Thorell, 1982, via Spillmann and Werner, 1982; D from Thorell, deValois, and Albrecht, 1984, Fig. 10, p. 759.]

17.3.4 The bottom line: Sparse coding for multiple stimulus features

Why did the EDC give the primary visual cortex 500 million neurons, when there are only about 1 million input axons arriving from the LGN? The most likely reason is that the multiplicity of V1 neurons allows V1 neurons to have the luxury of a much increased stimulus specificity, and thereby a *sparser code* for at least some stimulus variables.

What information is conveyed by a high activity level in a V1 neuron? When Hubel and Wiesel initially discovered orientation tuning, one could imagine that a high activity level in a V1 neuron signaled that a line of a particular orientation – say vertical – was imaged in a particular location on the retina. Other neurons in the same module, tuned to neighboring orientations, would also fire but at lower rates, yielding a small active population with the highest firing rate occurring in the neuron best tuned to vertical. The right pattern of activity in this relatively small population, then, would code the presence of the vertical line. A line at a different orientation would be represented by a similar pattern of activity in a set of neurons with the appropriate range of orientation tuning. This scenario suggests that V1 carries a sparse code for the locations and orientations of line segments.

This picture of V1 coding has to be modified, however, because we now know that each V1 neuron is tuned on many stimulus dimensions. To activate a V1 neuron optimally, the stimulus probably needs to have the right orientation, the right bar width or spatial frequency, the right direction of motion, and the right color, and be presented to the appropriate eye or to both eyes together. And other neurons within the module, tuned to slightly different combinations of stimulus features, would fire but at lower rates. This scenario suggests an even sparser code, since fewer neurons will be tuned to any given combination of features. Moreover the code is no longer sparse for the orientation of line segments, but rather for multidimensional combinations of stimulus features.

Notice that one cannot say simply that V1 (or any other brain area) carries a sparse code. One must ask, a sparse code for what features of the retinal image or the physical world? The V1 code is sparse for combinations of stimulus features, but not sparse at all for objects. Neurons in V1 would not fire similarly for all rocking chairs, or all faces, or for the same rocking chair or face presented from different viewpoints. If we are to find sparse codes for objects and classes of objects, much recoding still needs to be done.

17.4 More neuroanatomy: The modularity of cortical processing

Now that we know something of the physiological properties of V1 cells, we can ask: How are these neurons arranged within the cortex? We already know that there is some organization, in that cells with neighboring receptive fields lie close to one another, preserving the topological map. But in addition, within the local regions there is a higher degree of organization – an amazing, fine grained, *modular* pattern of organization, repeated many times in areas approximately 1 x 1 mm in size in the human cortex. There are about 1000 [xx ??] modules over the extent of V1 in each hemisphere.

How were these modules discovered? To do their classic experiments, Hubel and Wiesel inserted microelectrodes into the cortex. They started each electrode track at the cortical surface, and advanced the electrode until it encountered a cell. They then determine the orientation tuning of that cell; drove the electrode on along a straight line until it encountered the next cell; determined its orientation tuning; and so on.

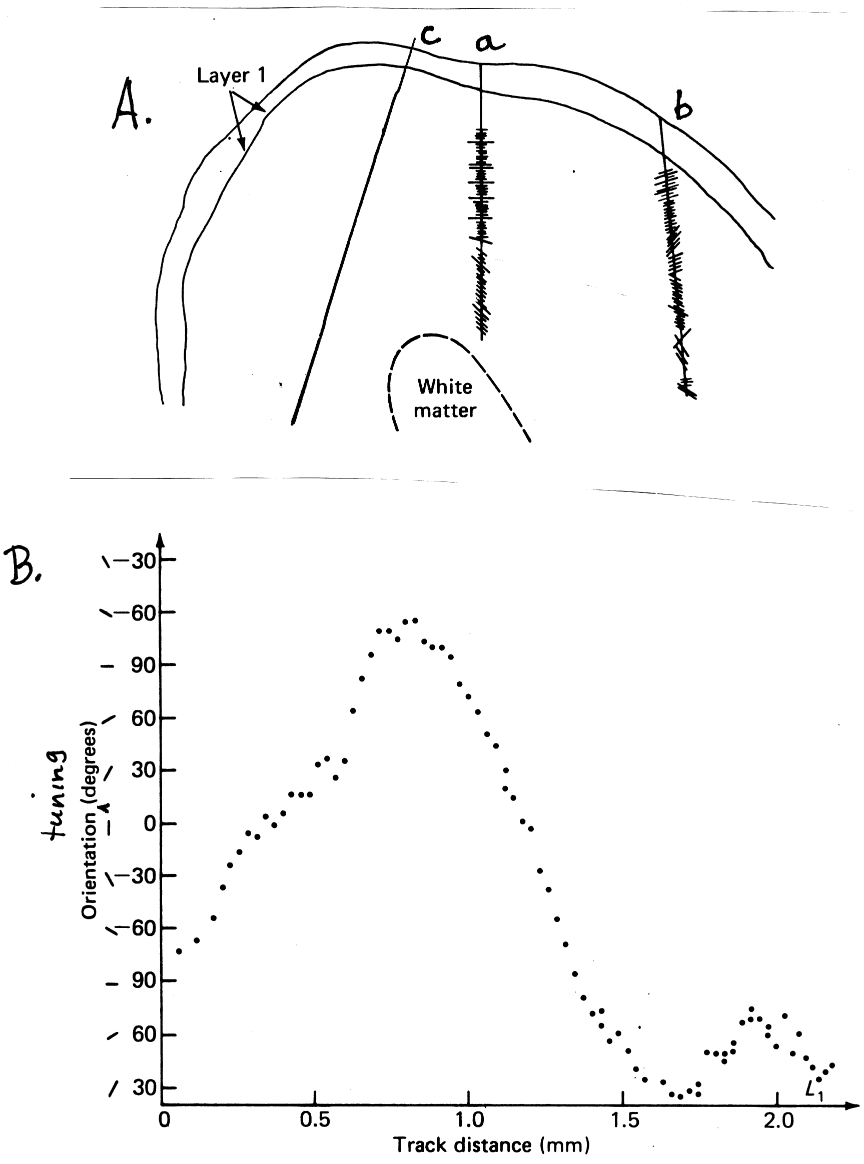


Figure 17.11: Early evidence for orientation columns and for the regular variation of orientation tuning across the primate cortex. A: three electrode tracks from V1 of macaque monkey cortex. The orientations of the short lines on tracks a and b show the orientation tuning of the cells encountered. Tracks a and b are nearly perpendicular to the cortical surface, and reveal long series of neurons with the same orientation tuning. Track c is more parallel (tangential) to the surface. B: the orientation tuning of a series of cells encountered on a tangential penetration. The orientation tuning varies regularly clockwise for about 1 mm, and then counterclockwise for another half mm. These data suggested that orientation tuning might vary systematically across V1 in some repeating (modular) pattern. [From Hubel and Wiesel, 1974, Fig. 2A.]

Figure 17.11A shows examples of two electrode tracks, a and b, that are nearly perpendicular to the cortical surface. In these tracks, long series of cells with similar or identical receptive field orientations were encountered. Hubel and Wiesel thus suggested that cells tuned for with similar orientations might be found in a vertical region, or *orientation column*, within the visual cortex.

Figure 17.11A also shows an electrode track, c, that is more nearly *tangential* (parallel) to the cortical surface. The results of such a tangential penetration through V1 cortex of a monkey are schematized in Figure 17.11B. As the electrode was advanced, the orientation tuning of the cells encountered changed regularly in a clockwise direction for almost 1 mm, then counterclockwise for almost another mm. Similarly, regular variations in ocular dominance – left eye dominant, right eye dominant, and so on – were found when electrodes were advanced in the orthogonal direction across the cortex. Such results led to the hypothesis that visual cortex might have a repeating, *modular structure*: across the cortex, repeating patterns of regular variations of orientation and ocular dominance tuning might be found.

More recently, it has been possible to use techniques that reveal the pattern of all of the V1 modules at once. *Ocular dominance* columns can be made visible by using *autoradiography.* That is, a radioactive substance that is absorbed by neurons is injected into one eye. This substance is taken up by retinal neurons, transported down the axons of ganglion cells, across the synapses in the LGN, and on to the cortical neurons activated by the injected eye. The resulting pattern of radioactivity in the cortex in fact reveals an alternation of left eye and right eye regions, the ocular dominance columns, as shown in Figure 17.12A.

To reveal the pattern of *orientation columns*, a different technique has been used. The chemical 2-deoxyglucose (2-dg), injected into the brain, accumulates in regions where cells are active. It is possible to inject an animal with 2-dg, expose the animal to (say) a pattern of vertical stripes for a few minutes, and then sacrifice the animal. The pattern in the processed brain tissue reveals the locations of V1 neurons that are tuned to vertical lines. An example of orientation columns revealed by the 2-dg technique is shown in Figure 17.12B.

17.4.1 The structure of V1 modules: Ice cubes vs. pinwheels

These data suggest that V1 cortex is made up of a repeating pattern of *modules* (or *hypercolumns*), within which orientation and ocular dominance domains are jointly arranged. The classic model of cortical modules, called the *ice cube model*, is shown in Figure 17.13A. Along one dimension of the individual module, the orientation tuning of the cells varies, with neighboring orientations occupying parallel, slab-shaped columns in the cortex. Along the other dimension, each module contains one left eye-dominated column and one right eye-dominated column.

An added complication on modules and their models arises from the fact that when V1 is stained with a stain called cytochrome oxydase, it is seen to contain a distinctive pattern of *puffs* or *blobs* – small groups of cells – that are prominent in the superficial and deep layers. The blobs have been incorporated into the ice cube model by placing a blob at the center of each half of each ice cube, as also shown in Figure 17.13. Physiological evidence suggests that the cytochrome oxydase blobs contain a relatively high concentration of cells tuned for color, adding a third dimension to the range of stimulus dimensions arranged systematically within the modular structure.

A more recent model, called the *pinwheel model*, is shown in Figure 17.13B. the pinwheel model suggests that the orientation compartments are not shaped like parallel slabs. Instead, within each ocular dominance column, orientation varies in segments shaped like the blades of a windmill. One

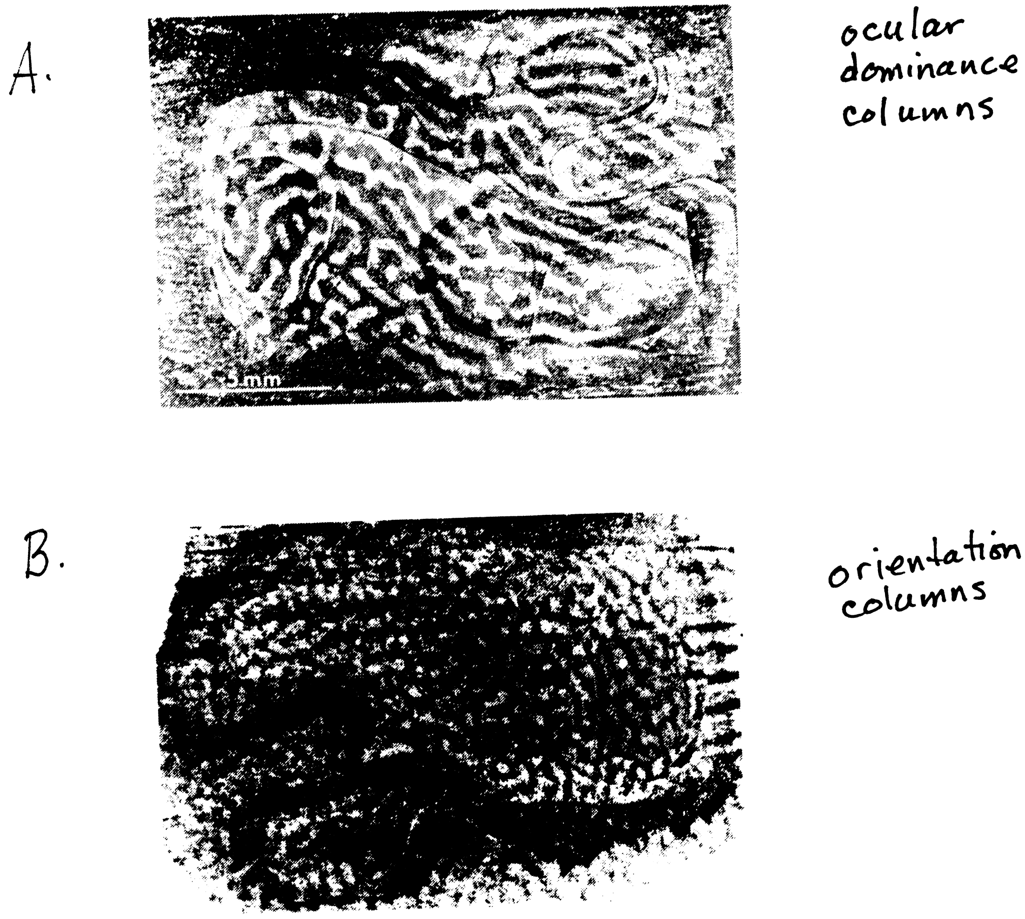


Figure 17.12: The modular structure of area V1. A: Ocular dominance columns, defined by autoradiography; B. Orientation columns, defined by uptake of 2-deoxyglucose (2-dg). Both the ocular dominance columns and the orientation columns make irregular patterns across the surface of the cortex. [From Hubel, Wiesel, and Stryker, 1978, via Levine and Shefner, 1991, Fig. 9-3, p. 187.]

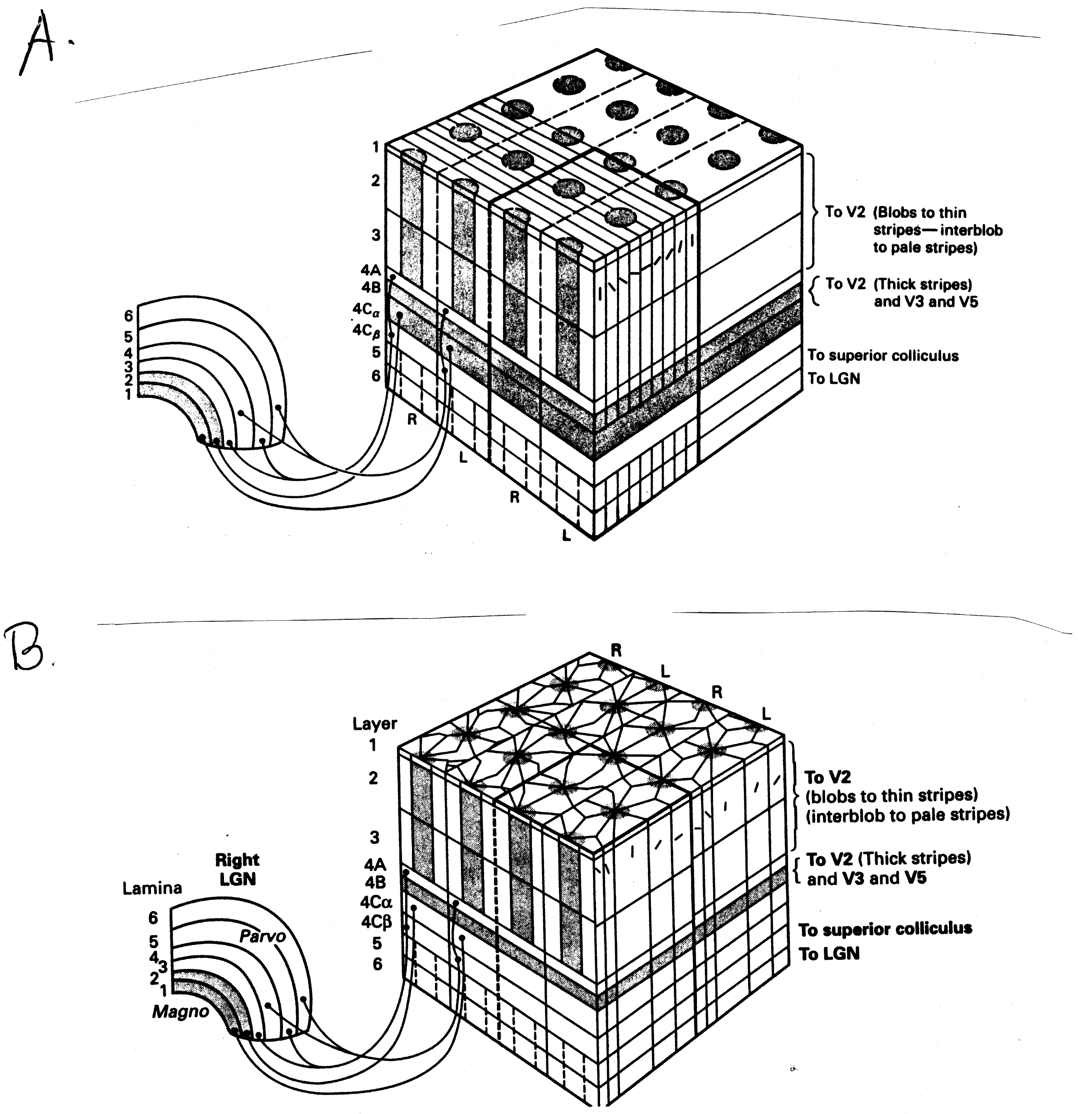

A.
1
2
3
4A
4B
4Cα
4Cβ
5
6
6
5
4
3
2
1
R
L
R
L
To V2 (Blobs to thin stripes— interblob to pale stripes)
To V2 (Thick stripes) and V3 and V5
To superior colliculus
To LGN
B.
R
L
R
L
Layer
1
2
3
4A
4B
4Cα
4Cβ
5
6
Right LGN
Lamina
6
5
4
3
2
1
Parvo
Magno
To V2 (blobs to thin stripes) (interblob to pale stripes)
To V2 (Thick stripes) and V3 and V5
To superior colliculus
To LGN

Figure 17.13: Modules in V1: Ice cube and pinwheel models. A: Ice cubes. Each cube contains one left and one right ocular dominance column (marked L and R), and a set of orientation columns (marked by lines of varying orientation). The grey regions in layers 2 and 3 show the blobs. In this model the two characteristics – ocular dominance and orientation – vary in more or less orthogonal directions within each module. B: Pinwheels. In the pinwheel model, orientation tuning varies like the blades of a windmill around the blobs. [A: Levine and Shefner, 1991, p. 190, Fig. 9-5; B. Levine, 2000, Fig. 8.5, p. 140.]

variant of the pinwheel model also incorporates systematic variations of spatial frequency – a fourth stimulus dimension – with the lowest spatial frequencies represented at the blob centers and the higher spatial frequencies at greater and greater radial positions. And more recently, it has been suggested that color is also mapped across V1 in particular spectral patterns.

Whatever the structural details turn out to be, the modular organization of V1 serves as an example of the remarkable fine-grained organization of neural tissue and neural circuits. The modular structure of V1 also reminds us again that the retinal image is not represented point for point in the primary visual cortex. Rather, because receptive fields are of non-negligible size, it is represented small region by small region. At the same time, the modules are spatially arranged so as to preserve the overall topographic organization of the visual field as laid out in Figure 17.2.

17.5 Recipes for V1 cells

The emergence of neurons tuned to orientation and other stimulus properties represents a profound change of the visual code. The obvious next question is: how do the orientation tuning and the other properties of the V1 code come about? What is known about the neural connections from LGN axons to V1 cells, and the circuits within V1, that produce the kinds of tuning seen in V1 cells?

In 1962, based on their early work with cat cortex, Hubel and Wiesel proposed two hierarchical models, one to make simple cells from non-oriented cells, and the other to make complex cells from simple cells. These two models are shown in Figure 17.14.

Hubel and Wiesel's schematic recipe for a simple cell is based on the idea that simple cells derive their inputs from a selected set of LGN cells. The model is shown in Figure 17.14A. On the left is shown a set of LGN cell receptive fields, all having ON-centers, but spread out along a straight line on the retina. At top right are shown the cell bodies of these LGN cells, with their axons projecting downward. At bottom right is shown a putative simple cell, receiving and summing excitatory inputs from this set of LGN cells. The model could be elaborated by adding inputs from OFF-center cells whose receptive field centers lie in the OFF flanks of the simple cell's receptive field. In either case, the most effective stimulus for the postsynaptic neuron would be an oriented bar of light of a width and location that exactly covers the centers of the receptive fields of the ON-center LGN cells, and minimizes the light on all of their surrounds. Voila, a simple cell. [Design some OFF-center simple cells with various orientation preferences.]

When Hubel and Wiesel described this model in 1962, they offered it tentatively, more as a feasibility argument than as a formal model. In fact, models for simple cells have been embroiled in controversy ever since. One of the features of the Hubel and Wiesel model is that the oriented receptive field of the simple cell is basically set up by its LGN inputs – a *hierarchical*, or *feedforward model*. In a competing class of *feedback models*, it is argued that cortical processing – for example, complex processing throughout the whole cortical column within which the simple cell is located – is required to create the oriented receptive field of the simple cell.

Experiments undertaken with several techniques provide some support for the importance of feedforward connections. For example, it is possible to record simultaneously from an LGN cell and a cortical simple cell, and look for a correlation between the occurrence of a spike in the LGN cell and a spike a few msec later in the cortical cell. A strong correlation shows that the simple cell receives an input from the LGN cell. Such experiments have shown that a simple cell receives input from a set of LGN cells arranged along a line on the retina, in good accord with Hubel

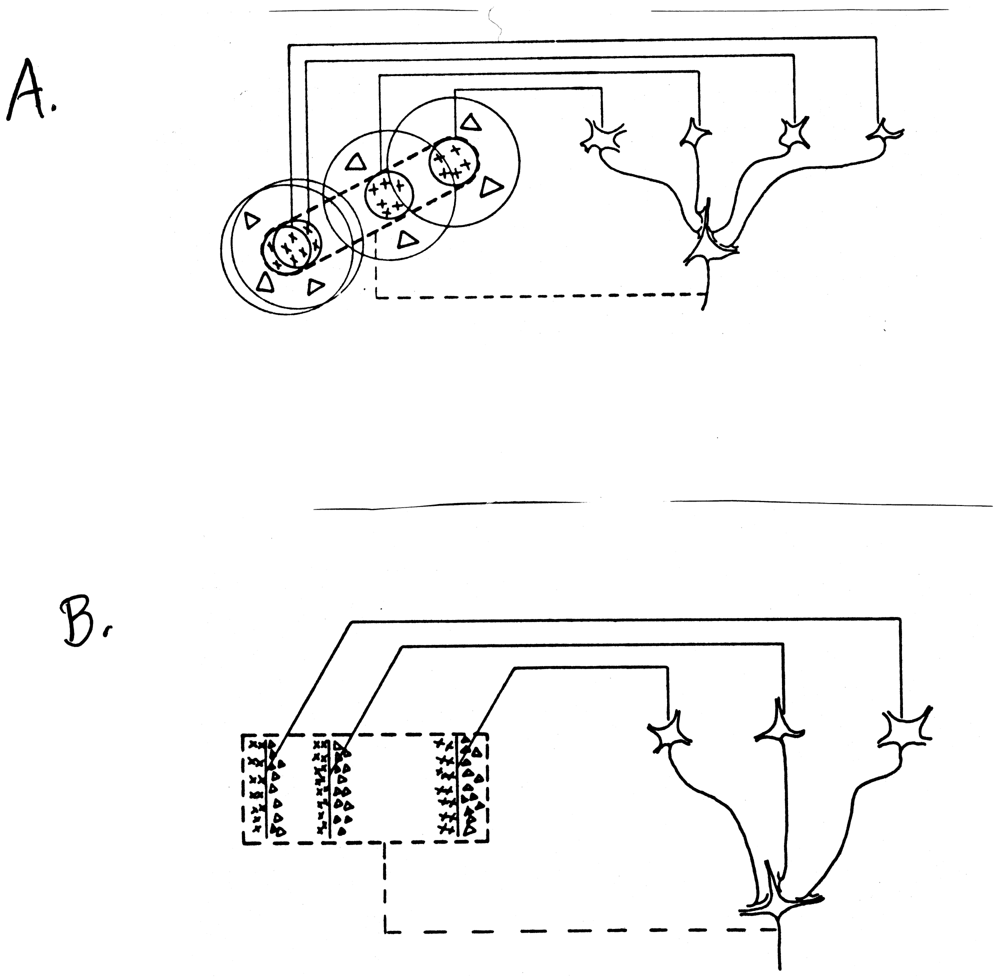


Figure 17.14: Hubel and Wiesel's recipes for simple and complex cells. A: A simple cell could be constructed from input received from a set of center-surround cells arranged along a line. B: A complex cell could be constructed from a set of co-aligned simple cells in neighboring locations. [From H&W 1962, Figs. 19 and 20.]

and Wiesel's speculation. Similarly, it is possible to record from an orientation tuned simple cell, and cool the cortex around its location. The cooling eliminates the possibility of feedback. The orientation tuning survives the cortical cooling, again lending support to a feedforward model.

On the other hand, other experiments favor a feedback model. For example, it is possible to determine the orientation tuning curve of the neuron at various time intervals after the onset of the stimulus. The result is that the orientation tuning of cortical cells is initially relatively broad, but it sharpens up over the first few hundred msec. after stimulus presentation. These data suggest that the initial orientation tuning might be controlled by feedforward circuits, whereas the later, more refined tuning might require feedback.

Hubel and Wiesel's recipe for complex cells is shown in Figure 17.14B. In parallel to the model for simple cells, complex cells are assumed to receive their inputs from a selected set of simple cells. In the cell schematized in Figure 17.14B, these are simple cells tuned to respond to vertical dark/light edges. In this case, such a dark/light edge falling anywhere in the receptive field would activate the complex cell, and a moving dark/light edge could be an especially effective stimulus, just as reported in Hubel and Wiesel's experiments.[2]

Hubel and Wiesel's model for complex cells has also been challenged. An early objection was that the latency of response of many complex cells was as short or shorter than that of many simple cells, so it made more sense to think of complex cells as having direct inputs from LGN afferents. Another argument is that there are too few simple cells and too many complex cells for the model to work easily.

In sum, the recipes for both simple and complex V1 cells are still controversial, and the subject of much work and speculation in the present day. Moreover, the modern anatomy suggests strongly that there are many different types ofV1 neurons, and therefore that many different recipes will be needed before our understanding of recipes for V1 neurons is complete.

17.6 Causal stories: What marks does V1 leave on our perception?

In earlier chapters, we have often remarked upon the ways in which neural processing at each level of the visual system leaves its marks on our perceptions. For example, we have argued that the absorption spectrum of rhodopsin creates the scotopic spectral sensitivity curve; the existence of three and only three cone types creates trichromacy; discrete sampling by the photoreceptors creates alias patterns; the convergence of rod- and cone-initiated signals in the retinal circuitry creates the psychophysically detected rod/cone interactions; and so on. We now ask this question again at the level of V1 cortex. What does V1 do? In what ways do the properties of V1 neurons leave their marks on our perception? How do V1 neurons participate in causal stories? Questions like this one have a variety of interpretations, and demand different kinds of answers.

We will discuss four examples. First, we will address the argument that the only meaningful question concerns neural computations. Second, we will examine some arguments about detection

[2] The aficionado may notice that these two models are fundamentally different in the kind of summation they posit. The simple cell model uses an *and* operation. The inputs from neurons, A, B, summate, so the combination rule is of the form A *and* B *and* . And the most effective stimulus is one that covers the right parts of the receptive fields of all of the input cells. The complex cell model uses an *or* operation. The cell is driven by inputs from A *or* B *or* C., and a bar at any one of several locations is posited to be fully effective in driving the complex cell. The former is linear; the latter is non-linear. [Why?]

thresholds that might be imposed by V1 neurons, taking examples from the evidence for multiple spatial frequency channels (Chapter xx). Third, we will entertain some tempting arguments and intuitions that arise from the sparse coding of edges and other stimulus features seen in individual V1 neurons. And fourth, we will speculate briefly on the role of V1 in the creation of conscious perception.

Before we start, a reminder about linking propositions. We have argued earlier in this book (Chapter xx, xx) that linking propositions will always be involved in causal stories. When detection thresholds are involved, we have the luxury of using Identity propositions, which as Brindley argued, cause little concern or controversy. However, many of the tempting causal stories we will encounter from now on will involve suprathreshold aspects of perception, in which the subject makes judgments about clearly visible stimuli. In these cases, more complicated linking propositions will be involved. [Watch for them as we go along.]

17.6.1 Neural computations

First, some vision scientists would argue that the most meaningful (or even the only) interpretation of the question, What is the role of V1? is, What neural computations are carried out within V1? This question we have already addressed briefly – the circuits in V1 create neurons that are multiply tuned for the locations, orientations, directions of motion, and colors of edges – and many additional characteristics of V1 circuitry are known. It is sometimes argued that the other interpretations of the question are too vague to be worth pursuing.

The question of the code transformation is, of course, fundamental. DT would argue, however, that she and many other vision scientists have another, more perceptual set of meanings in mind. To us the question means something like, In what ways does the V1 code provide new and useful raw materials for modeling perceptual states and processes? How do the computations accomplished in V1 allow us to make simpler causal stories about the system properties of vision?

17.6.2 Thresholds and weakest links

The second interpretation of the question, What does V1 do? has to do with visual thresholds. Several examples arise from the system properties introduced in Chapter xx [Spatial Frequency Channels]. There we presented psychophysical evidence to suggest that at some stage of visual processing, the incoming visual information is represented in terms of its spatial frequency components, and that different ranges of spatial frequency and orientation are processed relatively independently in separate information channels. We summarized this work with the multiple channels model of Wilson and Gelb (1984).

Might the orientation-tuned neurons seen by Hubel and Wiesel, and the spatial-frequency-tuned neurons elaborated by De Valois et al, provide the neural substrate for the spatial frequency and orientation tuned channels used to model the psychophysical data? The approach is tempting because if stimuli in different spatial frequency ranges and at different orientations are processed by different V1 neurons, then the signals generated by stimuli of different spatial frequency and orientation cannot interact. Thus, one would predict that thresholds measured in such psychophysical paradigms as masking, summation, and adaptation should all be specific to spatial frequency and orientation; and to a considerable degree, they are.

In some cases, psychophysically based models and physiological data on V1 neurons match in quantitative detail. Recall that in the Wilson and Gelb model, both the orientation tuning and the spatial frequency tuning of the psychophysically defined channels vary with spatial frequency, with both kinds of tuning becoming narrower as spatial frequency increases. Figure 17.15 shows a comparison between the spatial frequency and orientation bandwidths postulated in the Wilson and Gelb model and those estimated from the single unit recordings of De Valois et al (1982). The correspondence between the details of tuning predicted by Wilson and Gelb and the details of tuning found in V1 cells is quite remarkable. Thus, it seems very reasonable to argue that the channels derived by Wilson and Gelb from psychophysical data could be provided by this set of V1 neurons.

Now, might these neurons underlie the two other main system properties described in Chapter xx? First, Graham and Nachmias' (1971) summation-at-threshold experiments (Figure Xx) showed that component gratings F and 3F, presented as a compound grating, are detected independently, and do not summate their effects at detection threshold (Figure xx.xx). In Figure 17.10C, the tuning of V1 neurons looks narrow enough that a single neuron optimally activated by F would not respond to 3F; F should be detected by one set of V1 neurons and F3 by another. In that case, the neural signals from F and 3F could not summate, consistent with Graham and Nachmias's results.

Second, Blakemore and Campbell's (1969) adaptation-at-threshold experiments showed that adaptation to a grating of one spatial frequency left a notch in the CSF in the vicinity of that frequency (Figure xx.xx). Individual V1 neurons are known to adapt, in the sense that they become less sensitive after prolonged exposure to stimuli to which they respond strongly (e.g. Sclar, Lennie, and DePriest, 1989). Since different V1 neurons are strongly activated by stimuli of different spatial frequencies, gratings of different frequencies should desensitize different cortical neurons. In this way different gratings should create different notches in the CSF, consistent with Blakemore and Campbell's results. Thus, both summation-at-threshold and adaptation-at-threshold experiments can plausibly be modeled with frequency-tuned V1 neurons.

Three final points. First, since they are concerned with detection thresholds, these causal stories involve Identity linking propositions, leaving us on relatively safe and familiar philosophical ground. The causal stories in the next two sections depart from this simplicity.

Second, notice that all of the psychophysical data discussed above arose from detection experiments. The implicit argument in these causal stories is an elaboration of the *weakest link* argument that we saw earlier in modeling the Westheimer function. Recall that our model of the Westheimer function depends on the assumption that when a subject sets a tiny test spot to detection threshold, the bottleneck that limits detection across the whole range of sizes of the background field is a single ganglion cell with a center-surround receptive field. The parallel assumption in the present case is that when the subject sets a test grating of a particular spatial frequency to contrast threshold, the bottleneck that limits that threshold is a V1 cell tuned to that spatial frequency. But the tuning curves of V1 neurons are relatively narrow, so the argument must be elaborated to say that as we change the spatial frequency of the test stimulus, we change the neuron that does the detecting. Each V1 neuron functions in turn as the weakest link in the most sensitive channel over a narrow range of spatial frequencies, but different V1 neurons control thresholds over different spatial frequency ranges. The overall CSF would, of course, be determined by the upper envelope of the responses of these V1 neurons.

And third, obviously, the speculative parts of these arguments need to be followed up by quantitative modeling, and by further detailed experiments at physiological and/or psychophysical levels. For example, one needs to test some V1 neurons with the actual stimuli used by Graham and Nachmias – compound gratings with F and 3F components – and see whether or not they summate

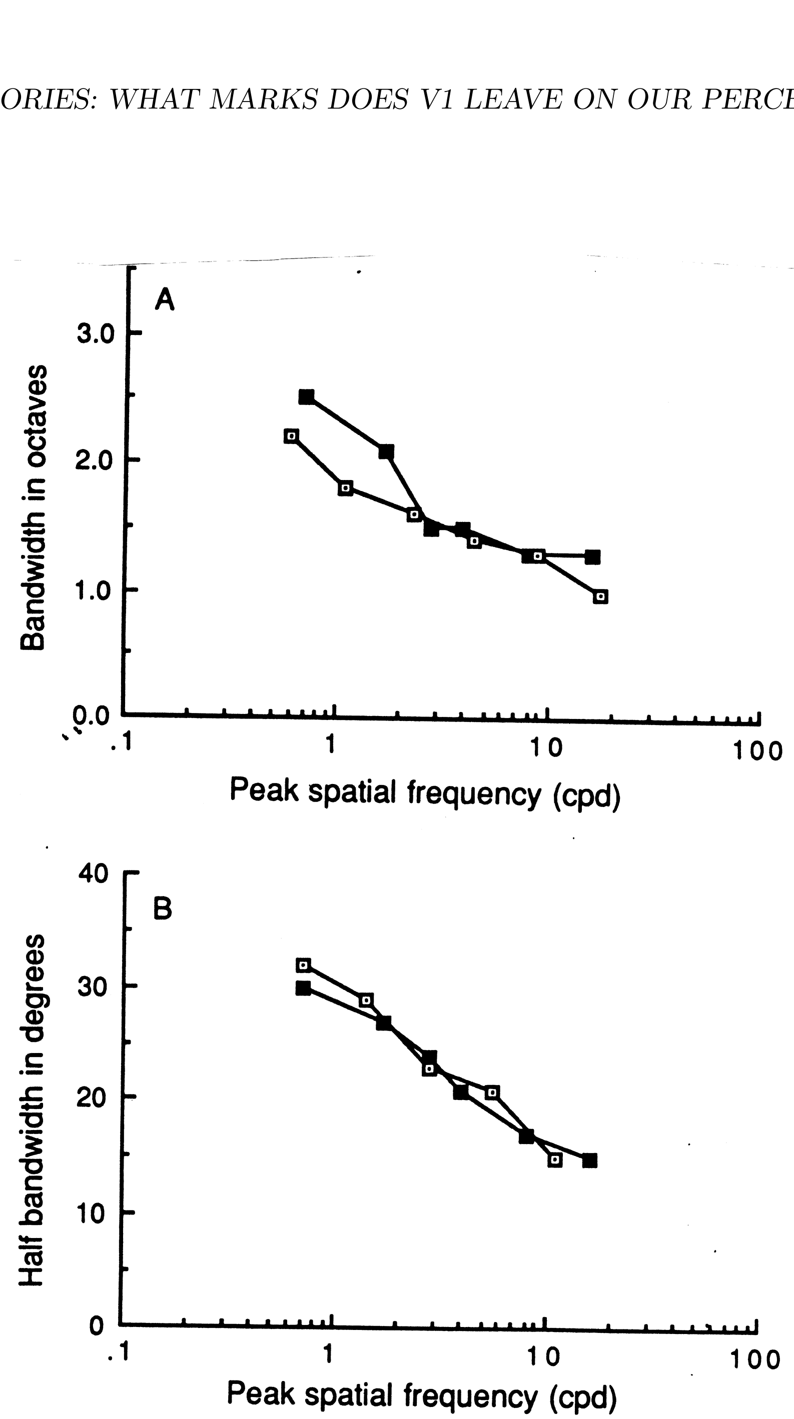


Figure 17.15: Comparison of psychophysical and physiological estimates of bandwidths for orientation and spatial frequency. Notice the striking agreement between the two data sets, particularly in the variation of bandwidths with both spatial frequency and orientation. The agreement of detail suggests that these cortical cells provide the immediate neural substrate of the spatial frequency channels defined in the Wilson and Gelb model. [From Wilson et al Ch in Spillmann and Werner, 1990; p. 248, Fig. 13. Note – figure does not specify which data set is which!]

their effects at detection threshold. Similarly, one needs to show that adaptation to a grating of a particular spatial frequency changes the responsiveness of V1 neurons that lie within its spatial frequency tuning curve, and not to gratings outside it. Some of these kinds of experiments have been done (e.g. Sclar, Lennie, and DePriest, 1989); others await.

In summary, how does V1 leave its marks on perception? Quantitative comparisons and qualitative models suggest that for sinusoidal gratings, V1 neurons might be the weakest links in a set of most sensitive physiological channels. In this way they might provide the neural substrate for at least some of the system properties concerning multiple spatial channels, as introduced in Chapter xx.

17.6.3 Sparse coding of stimulus features

As a third example, questions about the role of V1 invite us to think harder about the importance of the tuning properties of V1 neurons, and especially about the meaning of sparse coding.

Think for a minute about the coding of edges at various levels of the visual system. At early processing levels, information about the presence, orientation and location of an edge is carried by means of population codes. At the level of the photoreceptors, the edge is represented by broad regions of low and high quantum catches; at the ganglion cell level, it is represented by a dog-ear pattern across a population of ganglion cells (Figure Xx). A recording from a single photoreceptor or a single ganglion cell would yield little information about the presence or location of the edge.

However, remarkably, in V1 an edge at a particular location creates activity in only a relatively few neurons – the neurons tuned to the appropriate orientation at the appropriate retinal location – and recording from the right individual neurons would yield considerable information about the presence of the edge. In other words, as we have said, V1 neurons create the sparse coding of edges. (More broadly, they provide a sparse code for the many different aspects of the visual stimulus to which they are tuned.)

Now, when sparse coding occurs, vision scientists feel a huge temptation to assign a special role in the perception of whatever is coded sparsely, to the neurons that provide the sparse coding. On this interpretation, if the question is, what does V1 do? then one of the answers is, neurons in V1 create and provide a sparse code for the locations and orientations of edges and other stimulus features.

So we have isolated a third kind of meaning to the question, What does a given level of visual processing do? When we ask about the function of a higher cortical level we will often be asking a question about what stimuli are sparsely coded by individual neurons at that level. Interestingly, as we will see, at higher levels the sparse coding for edges is lost again, and what emerges may be sparse coding for other, more sophisticated features of shapes and/or physical objects.

A final caveat in regard to the sparse coding argument. In order to be assigned a role in the perception of a perceptual variable, a given processing level must not only provide sparse coding of that variable. In addition, it needs to provide the *first* level of sparse coding of that variable. That is, if we did the appropriate experiments in the appropriate V1 neurons, we would doubtless still see scotopic spectral sensitivity, trichromacy, alias patterns, and the other consequences of early visual processing; yet we would not want to say that V1 neurons cause these perceptual characteristics. Instead, most vision scientists would probably want to attribute the cause of a perceptual phenomenon to the *earliest* level at which the activity of individual neurons provides a sparse code for a particular set of stimulus features. Fortunately, neurons at levels earlier than

V1 have never been shown to exhibit marked orientation selectivity, so our argument about V1 creating sparse coding of lines and edges remains secure.

17.6.4 Neural correlates of consciousness

Fourth and finally, some vision scientists (like DT) are incurably interested in the question of the neural correlates of conscious visual perception. From this perspective, the question, what is the role of V1? has a fourth interpretation. Since individual V1 neurons respond strongly to lines and edges the question becomes, might activity in these neurons be a sufficient condition for our conscious perception of lines and edges at particular orientations and locations?

Among vision scientists who entertain questions about consciousness, there are two views. The first view (Crick and Koch, 19xx) is that the neurons that give rise to consciousness will be concentrated in one high-level location within the brain, at a much higher level than V1. But the second view, represented by Daniel Dennett (19xx), xx suggests that the neural correlates of consciousness are distributed. When a stimulus becomes sparsely coded at a given level of visual processing, it is argued, that level provides the neural correlate of being conscious of that stimulus. On this view, activity in V1 neurons would map to the perception of lines and edges. Later recodings could lose the sparse coding of edges, but that's OK, because the conscious perception of edges can always be generated by calling on V1. Later recodings would yield the neural correlates of other, more complex stimulus characteristics at higher levels. We return to these questions in Chapter xx (Consciousness).

In sum, we have explored four different interpretations of the question, What marks does V1 leave on our perception? Each of these interpretations stretches the scientist's imagination a little further, and different vision scientists differ in how many of these interpretations they are willing to entertain. Some stop at the first interpretation: What computations are done in V1? Some will go as far as the second: What thresholds are limited by V1 neurons? Some will even tolerate the third: What stimulus features are sparsely coded by V1 neurons? And a brave few will be delighted by the fourth: In what way does V1 contribute to our conscious perceptions? [Which of these questions do you think are meaningful? Which are you most interested in? Why?]

17.7 The object recognition problem: Bug detectors and grandmother cells?

New that we have some acquaintance with the V1 code, let's look ahead to the ultimate problem of visual perception: the problem of object recognition. We know that we are capable of recognizing objects. On the Universal Linking Proposition, there must be neural processes that allow us to do so. Moreover, on the Isomorphism proposition, we would like to think that these processes bear some resemblance to the system properties of object recognition at the perceptual level. So the question becomes, what form might the neural code take, to provide the neural basis of object recognition?

Hubel and Wiesel's work showed that there are neurons that respond only to surprisingly specific stimuli, and gave us hierarchical models to suggest how these neurons might come about. Their work freed our imaginations, and allowed us to speculate, at least momentarily, that maybe there are neurons and higher and higher levels of the hierarchy that respond to ever more specific visual patterns. Perhaps some highly specific, high level neurons form the basis of object recognition.

The excitement was increased by the work of Jerome Lettvin and his colleagues on receptive fields in the frog retina. The most famous of Lettvin et al's papers was entitled, "What the frog's eye tells the frog's brain" (ref xx). Lettvin et al made the case for five different types of ganglion cells in the frog's retina, which they called various kinds of *detectors.* Of the five types, the last is the most interesting, as it was called a *convex moving contour detector.* These neurons were reported to respond most strongly to dark blobs looking very much like bugs. Lettvin et all suggested that they could be *bug detectors.*

The notion of a bug detector incorporates three overlapping ideas. First, it suggests that a single neuron within the visual system has a high degree of stimulus selectivity, and responds, for example, only to a bug at a particular location in visual space. Second, activity in the bug detector could represent to the frog the presence of the bug at its particular location. And third, it might be relatively simple to connect activity in the bug detector to a motor response. Activity in the right bug detector, a tongue flip in the appropriate direction, and the frog would have a meal, as shown fancifully in Figure 17.16.

To many vision scientists, the analogy between V1 simple cells and the frog's convex moving edge detectors was immensely appealing. V1 simple cells certainly seemed designed to respond to lines at particular orientations and locations. It was tempting to call them *line detectors.* Moreover, hierarchical processing of the kinds suggested by Hubel and Wiesel (Figure 17.13) could in principle be used to make detectors with increasingly complex and novel features at later stages of the visual system. Take the case of a triangle, as shown in Figure 17.16B. A triangle on the retina would activate a small ensemble of V1 neurons with their receptive fields in just the right locations, aligned with the edges of the triangle. That is, the triangle would be represented diffusely in V1, by the activity of a population of localized line detectors, with maybe some endstopped cells at the corners. At the next level, this ensemble of neurons could all summate their inputs onto a single neuron – a triangle detector – that would respond optimally when this particular triangle was present on the retina.

Thus far there is an obvious problem with the triangle detector, because it responds only to a particular triangle of a particular size and orientation at a particular retinal location. How can we make a triangle detector that generalizes across these extraneous variables? The answer is conceptually simple – just create the next level of the hierarchy. Summate the inputs from triangle detectors of many sizes, shapes, and orientations at many locations, as shown in Figure 17.16B, and you have a triangle detector that generalizes across size, shape, orientation and spatial location. And perhaps later levels of hierarchical processing would create detectors that respond to more and more specific visual patterns and objects, over broader and broader ranges of location. For example, there could be face detectors, tree detectors, and fire hydrant detectors (these are said to be especially important for dogs).

In the end, the wittier visual scientists invented two fanciful kinds of cells to represent the extreme of the detector approach to object recognition. The first example was the *grandmother cell.* The idea is that maybe there is a neuron that increases its activity if and only if you are looking at your grandmother. The second example was the *yellow Volkswagen detector* – a cell that responded optimally to the presence of a yellow VW in the retinal image[3].

Although the examples are facetious, the change of perspective was immensely exciting. Before the discovery of orientation-tuned neurons and the invention of hierarchical models, many of us

[3] It took DT 15 years to finally get the joke – in the 1960's, the popular model of VW detector was the "beetle, or "bug", so the yellow VW detector was actually a spoof on a "bug detector" for human beings.

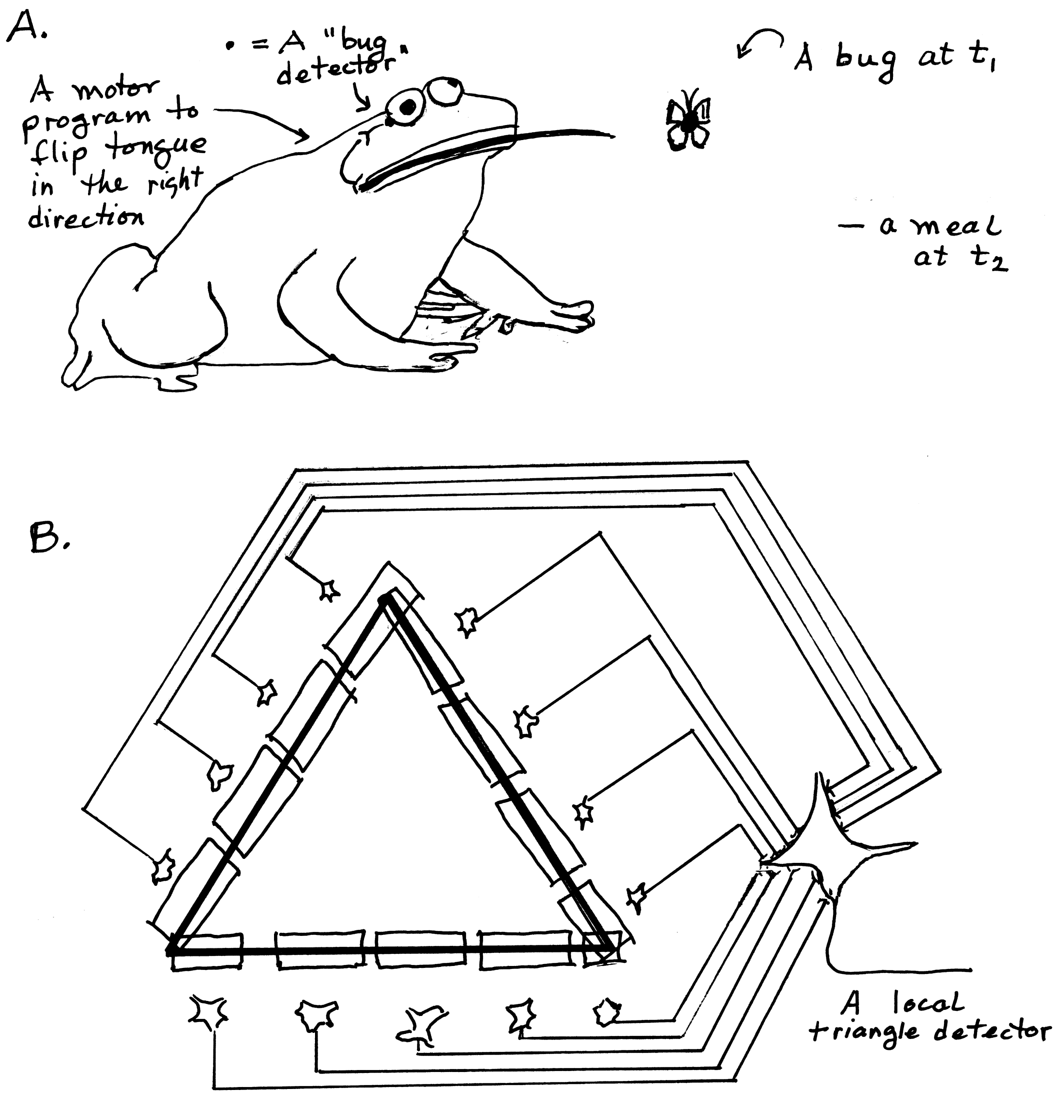

A.
• = A "bug detector"
A motor program to flip tongue in the right direction
A bug at t1
— a meal at t2
B.
A local triangle detector

Figure 17.16: Bug detectors and triangle detectors. A: A frog with a bug detector – a retinal ganglion cell that responds best to a convex moving contour in a particular location in space. In principle, an increase in the firing rate of this neuron could signal the presence of a bug in that location. A motor circuit that flips the tongue to that location provides the frog with a meal. B. A set of line detectors in carefully chosen locations provides the input to a local triangle detector – a cell that would respond optimally to this particular triangle at this particular location. Inputs from many local triangle detectors could create a global triangle detector that responded to any triangle over a wide range of retinal locations.

vision scientists had no clue how object perception was done. After these discoveries, we had a simple and attractive basis for intuition and speculation. It is immensely appealing to believe that perception of a single, unified object is represented by a single, unified neural signal – for example, a high level of activity in a single high level neuron. (Note the isomorphism in the linking proposition.)

There are, however, obvious problems with detector models. One problem is that you would need a different neuron for each object you recognize, and there are simply not enough cortical neurons available to dedicate this large number to the object recognition task. A second problem is that, if you literally had only a single neuron to represent each object, object recognition would be highly vulnerable to cell death. If you couldn't find your yellow VW one morning, you wouldn't know whether you had lost your yellow VW, or just your yellow VW detector! Of course we could have two or three of each, but that worsens the problem of numbers of available cells.

The alternative, of course, is to assume that at all levels of the hierarchy, objects will always be represented by population codes. The code for a complex object will always be a pattern of activity in an ensemble of neurons – the grandmother ensemble for grandma, and the yellow VW ensemble for a yellow VW, but always an ensemble. Since different combinations of activity in the population stands for different objects, the population code view overcomes the not-enough-neurons problem.

However, the population code view also has problems. The first is that as we have sketched it, it's empty of specifics – it seems like a handwave designed to damage the detector model, rather than a serious alternative. To endorse it, one would have to begin to develop a specific population code (how sparse are the representations of objects?). And the second problem is one that DT calls the *problem of joint action*: What entity registers the presence of the particular pattern, to "know" that it is the pattern for a yellow VW? If all of the neurons in the pattern code summate to activate a single neuron at a yet later level of the hierarchy, we are back to detectors. If not, how does the set of neurons act together, to allow us to perceive and respond to the object represented?

This discussion of detectors is intended only to seduce you – to get you started thinking on the problem of object recognition. We return to the (unsolved) problem of object recognition in Chapter xx.

17.8 Recent trends: Context effects

Up to this point in the chapter it has been possible to sustain the perspective that V1 neurons code for the presence of highly specific features or combinations of features at specific retinal locations, and that V1 still retains an approximate topographic map. However, the lateral connections seen anatomically suggest that V1 may carry out computations across broader reaches of visual space. More recently, many physiological studies have supported the more complicated view.

As we have said before, the *receptive field* of a neuron is that region of the retina within which visual stimuli affect the activity of the neuron. Empirically, the receptive field of a neuron is established by stimulating the retina with small bright or dark spots or bars of light, and finding the retinal regions upon which these stimuli lead to increases or decreases in the firing rate of the neuron. By the original definition of a receptive field, stimuli outside the receptive field should never affect the activity of the neuron.

However, sometimes light outside the receptive field does affects the neuron's activity. It does so indirectly, by changing the response of the neuron to stimuli that lie within the receptive field. These phemonena are called *context effects*, in the sense that the stimulus context outside the receptive field influences the neuron's response to a stimulus within it. The occurrence of context effects has led vision scientists to replace the term *receptive field* with the term *classical receptive field (CRF)*, to distinguish it from responses brought about by context effects.

An interesting example of context effects in monkey V1 neurons was provided by James Knierim and David Van Essen (1992). Knierim and Van Essen trained monkey subjects to maintain fixation on a fixation point for a few seconds at a time. During the fixation intervals, the CRF of a V1 neuron was defined using oriented bars of light. Next, the orientation and width of the optimal stimulus – the bar that provided the maximal response from the neuron – was determined, and this stimulus was placed in the CRF, as shown in Figure 17.17A and B. In these figures, the CRF of the neuron is indicated by the small grey box at the center of the stimulus pattern, and the optimal stimulus – a short horizontal line – is indicated within it. Finally, other stimuli (i.e. stimulus *contexts*) were added outside the CRF, as also shown in Figure 17.17A and B. These stimuli were large fields of line segments, either aligned parallel to the optimal stimulus (A) or orthogonal to it (B). The question is, will the stimuli outside the CRF change the response of the neuron to stimuli inside the CRF?

Some examples of the responses of orientation-tuned V1 neurons to these stimuli are shown in Figure 17.17C and D. In these figures, the numbers in the middle of the figure and the icons at the bottom indicate the stimuli. Stimulus 1 was the optimal stimulus for the the particular neuron reported in Figure 17.17C: a bar tilted to the left of vertical; and as expected, the neuron responded strongly. Stimulus 4 was the field of line segments presented alone outside the CRF, and as expected, the neuron did not respond. Stimulus 2 was the optimal stimulus plus the field of bars, with the bars oriented parallel to the optimal stimulus, and stimulus 3 was the optimal stimulus plus the field of bars oriented orthogonal to it. In both cases, even though it was outside the CRF, the field of bars dramatically reduced the response of the neuron! Stimuli 5-8 show a similar pattern for a non-optimally oriented bar within the CRF.

Figure 17.17D shows a second interesting result. In this neuron too, by design, the optimal stimulus – a right diagonal bar – produced the largest response (stimulus 1), and a field of bars outside the CRF produced no response (stimulus 4). But the response of the neuron to the optimal stimulus varied with the orientation of the bars outside the CRF. Bars parallel to the optimal stimulus (stimulus 2) preserved only a minimal response from the neuron, whereas orthogonal bars (stimulus 3) preserved a larger response. Stimuli 5-8 show that this neuron, unlike the neuron of Figure 17.17C, had narrow orientation tuning for stimuli within the CRF.

Context effects like these are interesting for two reasons. First, at the level of anatomy and physiology, context effects provide a possible use for the horizontal connections made by pyramidal cells in V1 (Figure 17.6). These horizontal connections may well allow interactions across cortical modules. They might, for example, connect orientation columns all tuned to the same orientation, allowing neurons tuned to the same orientation to enhance or inhibit each others' responses across the cortex. Perhaps some coordination of signals across a set of neurons tuned to horizontal lines, but located in separate modules, could provide the neural basis for seeing a longer line. [What linking proposition is involved in this argument?]

And second, context effects force us to rethink the functions of V1 neurons. In the earlier view, V1 neurons could be imagined to respond to local features or combinations of local features in the incoming signal, an carry a local-region-by-local-region picture of the world. But if we acknowledge the existence of context effects, we can no longer consider the responses of V1 neurons to be purely spatially local. We will need to rethink the computations accomplished in V1, and the role of V1

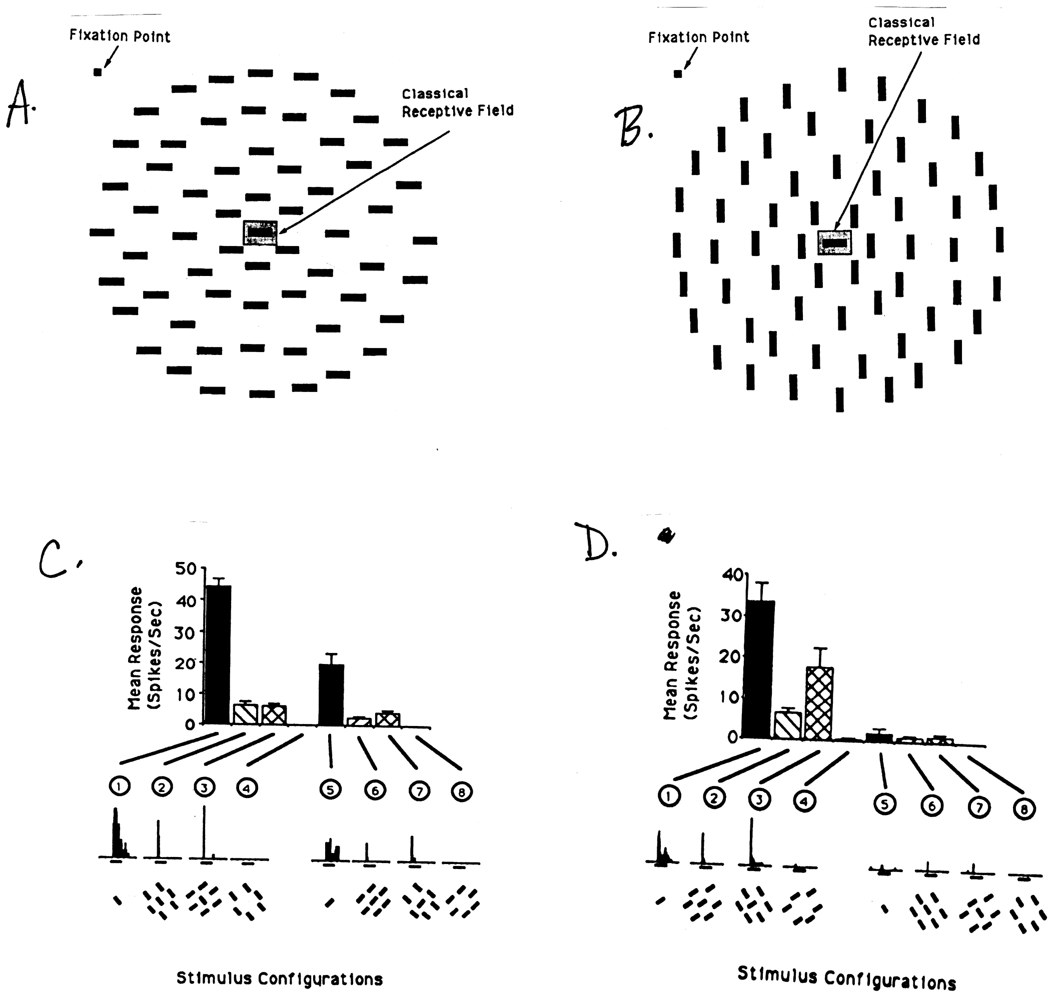


Figure 17.17: Context effects in V1 neurons. A and B: The stimuli. The small grey square marks the classical receptive field (CRF) of the neuron. A single horizontal bar falls within the CRF, whereas the "context" of short vertical or horizontal bars fall outside it. C and D: The reponses of V1 neurons are strongly influenced by the stimulus context. In C, both orientations of context bars reduce the activity of the V1 neuron about equally. In D, the co-aligned bars reduce the response more than do the orthogonal bars. From Knierim and van Essen, 1992. [A: dt modified from Fig. 1; B: Fig. 1, p. 963; C: Fig 2A, p. 964; D: Fig. 4B, p. 965.]

in visual processing.

One specific example involves a possible neural substrate for a psychophysical phenomenon called *popout*. Look back at Figure 17.17A and B, in which a small horizontal like is presented within the receptive field of a V1 neuron, and two different stimulus contexts are presented outside the receptive field. You will probably agree with the common perceptual observation that when the flanking bars are oriented parallel to the center bar (Figure 17.17A), the center bar tends to lose its identity in the field of bars – it becomes less perceptually salient. In contrast, with orthogonal alignment (Figure 17.17B), the center bar ”pops out” – it becomes more perceptually salient. In their study of context effects in V1 neurons, one of Knierim and van Essen's motivations was to ask, will V1 neurons show properties analogous to the perceptual phenomenon of popout?

Clearly, judgments of perceptual salience are Class B experiments, and linking propositions must be adopted if one is to model them. Suppose one adopts the linking propositions that activity in V1 neurons determines (or influences) our perception of lines, and that a faster firing rate in V1 neurons corresponds to greater perceptual salience. Under these assumptions, the context effects illustrated in Figure 17.17 provide a qualitative account of the perceptual phenomenon of popout, and suggest a role for V1 neurons in creating it.

17.9 Summary: The fifth transformation

In the scheme we have been developing, the fifth transformation of the visual code occurs within area V1 of the visual cortex. Unlike the case of the fourth transformation, from retinal ganglion cells to LGN cells, there is a profound recoding of visual information in V1. In the input layer (layer 4) in primates, the receptive fields of many cortical neurons still show the center-surround organization typical of retinal ganglion cells and LGN cells. But after processing within V1 itself, neurons are tuned on a variety of stimulus dimensions. Most cells have elongated receptive fields, and respond selectively to bars, edges and grating patches at particular orientations. Many are tuned to particular, relatively narrow ranges of orientation and spatial frequency, directions of motion, colors, and other stimulus features. The 500-fold increase in the number of neurons in V1, above the number of incoming LGN axons, allows different V1 neurons at the same topographic location to respond to different combinations of orientation, spatial frequency, direction of motion, ocular dominance, and other variables, creating a sparse code for specific combinations of these stimulus features.

In keeping with the formulations in earlier chapters, we now ask: In 25 words or less, how can we characterize the form of the visual code after the 5th transformation, at the level of V1? What information is carried in the activity of individual V1 neurons? A reasonable consensus answer, entertained for the past 40 years, is: V1 neurons code the *elements of form perception.* Regardless of whether we are feature creatures or frequency freaks, it is intuitively appealing to see V1 neurons as creating a code that is still largely tied to simple stimulus dimensions, and whose elements will be useful as building blocks for later levels of hierarchical processing.

All of these views are challenged by the recent findings of context effects in V1 neurons. If stimuli outside the classical receptive field can influence the firing rate of a neuron, no simple descriptions of the properties of V1 neurons can be given, and who knows what complexities will ensue. Moreover, if context effects create neural firing patterns that seem isomorphic to some of the characteristics of higher-level perception, as suggested by Knierim and van Essen and many others, we may need to revise completely our perspective on V1. We used to think of V1 as laying out the elements of form perception. But instead, maybe V1 is an integral part of the process of form perception. We will return to this theme in Chapters Xx and xx.

In the next chapter, we step back to paint a broad overview of the anatomy and physiology of higher levels of cortical processing.

Chapter 18

Cortical Cartograph

In earlier chapters of this book, we have pursued the central set of questions schematized in Figure 1.xx. First, we have traced incoming visual information through a series of five code transformations, ending up in cortical area V1. Second, we have introduced a (carefully selected) set of system properties, mostly centering around questions of detection and discrimination thresholds. And third, we have used analogies and isomorphisms between particular visual codes and particular system properties to argue that many of the (carefully selected) system properties are well explained by the early sensory processes.

However, you have probably noticed that many of the most interesting system properties of vision – questions of higher-level perception, such as the perceptual constancies and the perception of objects – have not been addressed. The obvious reason is that most of the processes and representations that provide the substrate for higher-level perception probably lie at higher cortical levels. But unfortunately, our knowledge of higher-level cortical anatomy and physiology are relatively primitive compared to our understanding of retinal processing.

The vision scientist David Van Essen (2004) describes the next task as *cortical cartography*: the making of cortical maps. He remarks, "It is of fundamental neurobiological importance to know how the visual cortex is subdivided into anatomically and functionally distinct areas, just as knowledge about the earth's geographical and political subdivisions is critical to fields such as political science and history.the objective of accurately and completely charting all of visual cortex remains far from complete. Altogether, our overall knowledge about visual-cortical organization arguably corresponds to that which cartographers of the earth's surface had attained around the end of the 16th century." (Van Essen et al, 2001, p. 1360).

The task is daunting. The cortex doesn't come with a paint-by-numbers kit – there are no black lines on its surface separating one region from another, and no recommended color palette. And yet the questions are clear. How much of the cortex is concerned with processing visual stimuli? How many subregions are there, and where are they? How are they interconnected? What are their functional characteristics, and what computations does each one carry out?

The recent arrival of functional brain imaging adds immense excitement to the enterprise. With functional imaging, almost miraculously, it is now possible to study the responses of the living human brain to selected visual stimuli, or to changes from one visual stimulus to another. Thanks to functional imaging, a direct cartography of the human brain is now being undertaken, with all of the same questions in mind.

In the present chapter, we provide an overview of visual cortical processing beyond V1. The chapter is divided into two main parts. In the first, we summarize what has been learned about the anatomy and physiology of post-V1 processing in the macaque monkey. In the second, we sample some of the relatively little that is known so far about human visual processing, based on functional brain imaging studies. In both cases, the reader is forewarned – knowledge of cortical processing is much more primitive than knowledge of retinal processing. There are many questions on which the experts disagree, and the whole field must be seen as works in progress.

18.1 An overview of visual areas in macaque monkey cortex

18.1.1 Four criteria for defining cortical areas

What criteria shall we use for carving up the cortex in a meaningful, functional way? On what basis shall we call one region of cortex area A, and a neighboring region area B? There is a major consensus that four criteria are important, although not every candidate area lives up to every criterion. The first two criteria are mainly anatomical, the third depends on both anatomical and physiological evidence, and the fourth is largely physiological.

The first criterion is that a region of cortex that is a candidate for being called a visual area should have a *distinctive cellular structure.* For example, historically, the distinctive stripe in layer 4 of area V1 was visible to the naked eye, and set V1 off from surrounding regions as a distinct area – the so-called striate cortex. More recently, more and more subtle methods have been developed for probing the cellular structures, molecular structures, anatomical transmitters, or other properties of neurons in different brain regions. All of these properties are used to seek out homogeneous regions that then become candidate areas.

The second criterion is that the candidate area should have a distinctive and coherent *pattern of inputs and outputs*; for example, it might have inputs from an established visual area, and project to two others. Evidence about patterns of connections is obtained by tracing the neural pathways that connect one region to another. Since the axon is the part of the neuron that makes the long connection from one brain site to the next, anatomical pathway tracing techniques depend on following axons.

The earliest technique for following axons was to make a lesion, large or small, to destroy cell bodies in one region of the brain. When the cell body is destroyed, the axon dies. Thus, after an appropriate period of time, the pattern of degenerated axons reveals the connections from the site of the lesion to the other regions of the brain to which the lesioned site projects. More recently, techniques have been developed that depend on the transportation of dyes and other chemical materials both forward and backward along the axons, often by the metabolic processes of the axons themselves. Some chemical or material that the axon will transport – the *tracer* – is injected into one location, called the injection site. The tracer can be selected so that the transportation is *retrograde* – backward to the sites that project to the injection site; or *anterograde* – forward to the sites to which the injection site projects; or both retrograde and anterograde tracers can be used at the same time. These studies can be used to map the interconnections among areas, and look for the coherent patterns that would help to define distinct visual areas.

The third criterion is that the candidate area should contain a *topographic map.* Topographic maps can be searched out by injecting two or more tracers with different properties – for example, two dyes of different colors – into neighboring locations within the injection site, and observing the pattern of projections in the target structures. Topographic maps can also be sought by physiological means, by mapping the retinal locations of the receptive fields of neurons across the candidate area. This criterion, however, becomes more difficult to apply at higher levels of the visual system. Receptive field sizes in general increase more and more at higher and higher levels, and the larger the receptive fields, the harder it is to define a retinotopic map.

The fourth criterion is that the candidate area should be (to some degree at least) *functionally homogeneous.* This criterion is exclusively physiological, and depends on recording the receptive fields and other response properties of single neurons. Within a region of cortex, neurons may all have a common set of properties, such as responding best to oriented lines and edges, or responding to motion and not to color, or vice versa. Such functional homogeneity argues in favor of calling the region a distinct visual area.

18.1.2 Areas V1, V2, V3, V4, and MT

All of these techniques have been applied intensively to the cortex of the macaque monkey over the past few decades. As of the present time, there is a good consensus among vision scientists that several regions immediately forward of area V1 in macaque cortex meet most or all of these criteria. These areas have by and large consistent locations and a consistent nomenclature: visual areas V2, V3[1], V4, and V5 (also called *MT*, which stands for the *middle temporal* area). In contrast, beyond V4 and MT there is less agreement on the numbers and locations of visual cortical areas; both the divisions among areas and the nomenclature are inconsistent across laboratories. For purposes of simplicity, we will lump all of the candidate areas beyond V4 and MT into two groups, named for the brain locations that they occupy: the *posterior parietal* (PP) and *inferotemporal* (IT) areas.

But where are they? In trying to sketch the locations of these areas on the cortex, we immediately run into problems. The primate cortex can be thought of largely as a two-dimensional sheet, but with a highly convoluted three dimensional structure. As previously noted (Figure 17.xx), area V1 lies deep within the calcarine sulcus, which itself lies on the inside (medial) surface of the cortex. Other important visual areas also lie within sulci, beneath the surface of the cortex. In addition, there are major individual differences in the locations of the various visual areas. Thus, no sketch of visual areas on the visible surface of the brain can convey the locations of the various visual areas very well. Figure 18.1A shows one approach to getting around this problem. In this figure, the cortex has been conceptually stretched a little, so that some of the sulci open up, revealing the locations of previously buried visual areas.

The major interconnections between these early areas are shown in Figure 18.1B. Most of the output from V1 goes to V2, with some going to V3 and to MT. The output from V2 goes to V3, V4, and MT. The output from V3 goes to V4 and MT. The major output from MT projects forward to other centers in the posterior parietal cortex. The major output from V4 projects downward to other centers in inferotemporal cortex. There are major cross connections between MT and V4, as shown; and also between various subparts of posterior parietal and inferotemporal cortex (not shown, but see Figure 18.4)[2].

In sum, Figure 18.1B provides a skeleton upon which to hang your developing understanding of visual processing in macaque monkey cortex. Aside from the V1-V2-V3 sequence, the most

[1] Different authors divide area V3 into different subparts (see, for example, Figure 18.4). We will ignore the subparts for simplicity.

[2] Notice that neither Figure 18.1A nor 18.1B captures the topography within the individual visual areas, nor the detailed pattern of connections that will be needed to preserve topography from one area to the next. We will return to this topic below.

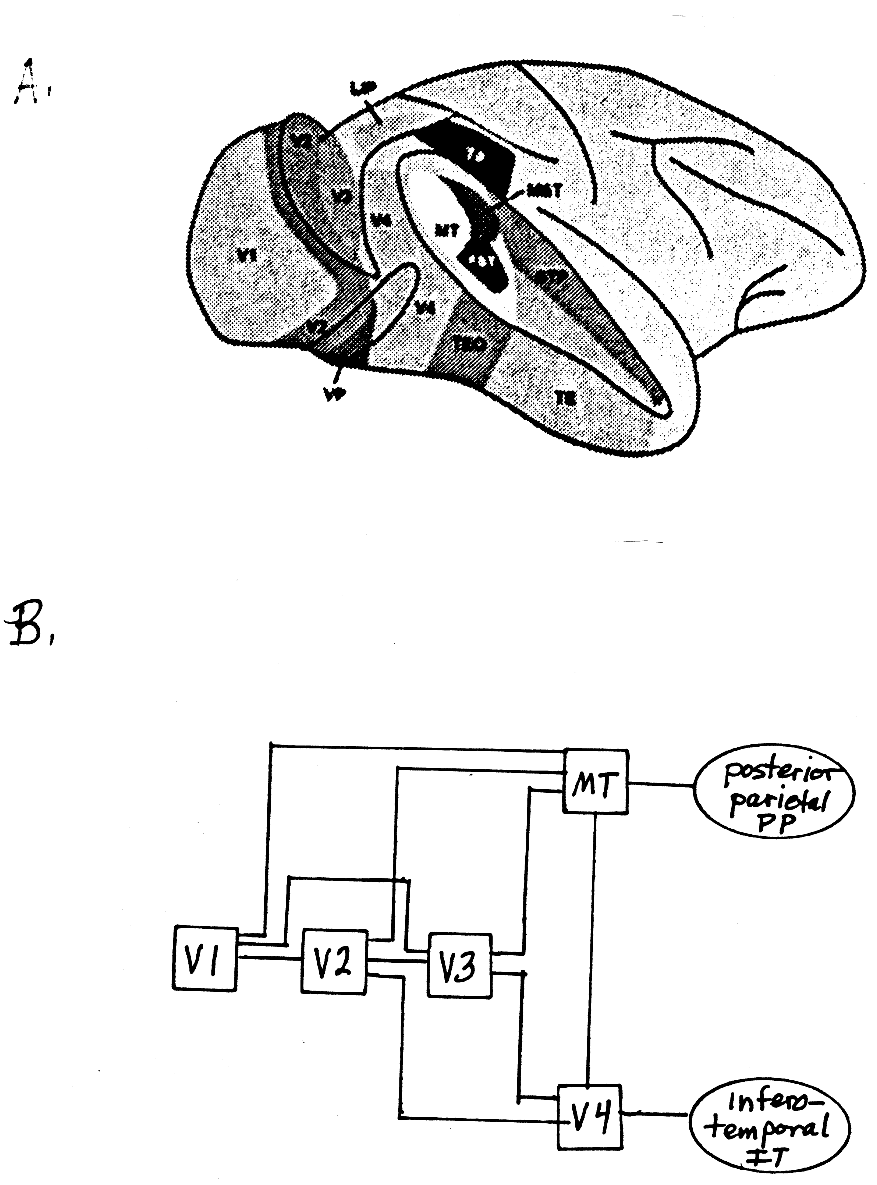


Figure 18.1: An overview of visual areas in macaque monkey cortex. A: A slightly exploded view of the right hemisphere, with area V1 unfolded from the calcarine sulcus. Five major visual areas are marked: V1, V2, V3, V4, and MT (also called V5). Inferotemporal (IT) cortex lies downward from V4, and includes the areas marked TEO and TE. Posterior parietal cortex lies forward of MT, and includes the areas marked LIP, 7a, MST, FST, and STP. B: The major connections among areas in early visual processing. [A: Adapted from Mishkin et al, 1983, via Ungerleider and Pasternak 2004, Fig. 34.1, p. 542; B drawn by DT.]

important thing to remember is the names of the two separate areas that come next in the sequence of early processing: V4 and MT (V5). These two areas have been studied extensively, and we will refer back to them often, comparing and contrasting their functional properties.

18.2 A radical view from the 1980s: Parallel processing streams

Prior to the 1980s, most visual scientists probably implicitly thought about post-V1 processing in much the same terms as pre-V1 processing. We probably expected to find a continuing, unified hierarchy of code transformations from one anatomical stage to the next, with more and more complex aspects of perception and cognition being made explicit at each level, and leading to some final unified high level representation of the physical world and the locations of objects within it.

Accumulating evidence in the 1980s led several research groups to propose a radical theoretical revision: the existence of *parallel processing streams.* The idea was that after signals arrive in cortex,, the visual system diverges into at least two separate anatomical pathways, or *streams*, with different aspects of visual processing being carried out more or less independently in the different streams. One stream, called the *dorsal* (or *parietal*) stream, runs through area MT, and occupies a series of cortical areas in the posterior parietal lobe. The other, called the *ventral* (or *temporal*) stream, runs through area V4, and occupies a series of areas in the inferior temporal lobe. The sketch of interconnections among visual processing areas already shown in Figure 18.1B incorporates these two processing streams, with MT and the posterior parietal areas making up the dorsal stream, and V4 and the inferotemporal areas making up the ventral stream.

As will be seen below, the radical argument is not just that the dorsal and ventral streams diverge anatomically. The radical argument is that different cortical streams provide the computations and representations needed for different aspects of visual perception and vision-based motor action. Taken to its logical extreme, the parallel processing view can be taken to suggest that there is no single, final, unified representation of the physical world within the visual system – the dorsal and ventral streams just diverge, and serve different purposes. This radical claim made many visual scientists sit up and take notice.

18.2.1 A tangent: Do M and P pathways map 1:1 to dorsal and ventral streams?

At the level of the retina and LGN, we already know that there are at least two parallel channels: the M and P pathways (the K pathway is ignored for simplicity here). Might the M- and P-initiated signals be kept separate throughout the complexities of processing in V1? If so, then a 1:1 correspondence could occur between the precortical M and P pathways and the two cortical streams. In keeping with this simple idea, Margaret Livingstone and David Hubel (1988) suggested that M-initiated signals could provide the input to the dorsal stream, and P-initiated signals to the ventral stream.

The simplicity of this scheme caused it to be widely accepted, and popularized in many textbooks. But the truth turns out to be more complicated. Many more recent anatomical studies have shown interconnections between M- and P-driven cells within V1. More conclusively, many physiological studies have shown that M and P inputs mix together within individual neurons in V1, particularly in the superficial layers – layers 2 and 3. Some of the neurons in these layers receive inputs initiated solely or predominantly by M cells, and others receive both M-initiated and

P-initiated inputs. But few if any neurons in the superficial layers receive exclusively P-initiated signals. In consequence, there can be no stream beyond V1 that carries a purely P-initiated signal.

These data point to an interesting asymmetry between the dorsal and ventral streams. As Livingstone and Hubel suggested, the dorsal stream is largely dominated by inputs that originate in the M pathway, albeit with small contributions originating from the P pathway. In contrast, the ventral stream receives major contributions originating from both M and P pathways.

Beyond anatomy, several other lines of evidence are consistent with the idea of parallel processing streams in monkey visual cortex. This evidence is reviewed in the next few sections.

18.2.2 Physiology: Neurons with different specialties in different cortical areas

If the dorsal and ventral pathways carry out separate functional tasks, then neurons located in these two streams should show selectivities for different stimulus dimensions, and/or respond optimally to different categories of stimuli, consistent with their functional tasks. In fact, although there is some overlap, neurons in the dorsal vs. ventral streams do show differences in their patterns of stimulus selectivity. These selectivities will be reviewed and defined more thoroughly in greater detail in subsequent chapters, but the general differences will be summarized briefly here.

We begin with the dorsal stream. As discussed previously, the inputs to the dorsal stream are largely of M-cell origin. Hence we might expect that neurons in the dorsal stream would make use of the characteristics of M cell signals, including sensitivity to high temporal frequencies. And indeed, neurons in MT are particularly responsive to motion, which requires the analysis of high temporal frequencies, and show pronounced selectivities for the direction and speed of motion.

Neurons at higher levels of the dorsal stream retain these selectivities. But in addition, new kinds of tuning emerge. Many neurons are selective for larger and more complex patterns of motion, such as optic flow patterns (e.g. rotation, expansion, or contraction). We return to the properties of dorsal stream neurons in Chapter 22.

What about the ventral stream? As discussed previously, the ventral stream receives both M- and P-cell-initiated inputs, after processing in V1. Not surprisingly, then, neurons in area V4 – the gateway to the ventral stream – show pronounced selectivities for spatial frequency and orientation, but are also strongly selective for color. Some neurons in V4 have even been reported to reveal an apparent color constancy, responding to the surface characteristics of objects rather than being controlled solely by the spectral composition of the illuminant; however, this report is controversial.

Later in the ventral stream, in inferotemporal (IT) cortex, receptive fields become increasingly large, and the combinations of features to which individual neurons respond optimally become increasingly complex. Many neurons respond optimally only to particular complex shapes, or even to ecologically important forms such as hands and faces. In addition, some IT neurons respond to the same shapes across variations of retinal location and retinal size – useful features to serve as correlates of the perceptual phenomena of size constancy and object perception. Some IT neurons also show form cue invariance – for example, if they respond best to a square, they respond to the square across variations of the physical cue that defines it. We return to the properties of ventral stream neurons in Chapter xx [Forms and objects xx].

In sum, how different are the responses of neurons in dorsal vs. ventral streams? The answer depends on your point of view. Figure 18.2 shows an early summary of the stimulus dimensions to which neurons in areas V4 vs. MT respond (Felleman and Van Essen, 1987). Many neurons in both areas are tuned for orientation, and the same is true for binocular disparity. But in V4, a majority of neurons are selective for color, and almost none are selective for the direction of motion. In MT, the reverse is true: almost none of the neurons are selective for color, but virtually all are selective for the direction of motion. Clearly, given these different responses to color and motion, there must be processing specializations between dorsal and ventral streams, consistent with the notion of different functional roles. And as we will see, the more sophisticated our experiments have become, the more distinctive the response patterns found in neurons at different levels and in different streams.

18.2.3 Lesion studies: Effects of cortical damage

If neurons in dorsal and ventral pathways are tuned to different stimulus dimensions, then selective lesions in specific brain areas should yield selective visual deficits. Moreover, if this tuning changes and becomes more sophisticated as we progress hierarchically up each stream, these deficits should vary with the hierarchical level of the lesion. Evidence of this kind has been provided by studying the specific deficits shown by both human patients and monkeys with lesions of specific cortical areas.

In human patients, of course, lesions are accidental, and their sizes and locations cannot be controlled. However, in general, lesions associated with the dorsal stream can lead to selective losses of the perception of motion, while form and color vision are retained. Lesions associated with the ventral stream can lead to selective losses of color vision, form vision, and/or object recognition. For example, some patients with lesions in IT cortex lose the capacity to recognize specific classes of forms, such as animate or inanimate objects, or even to distinguish among different faces. Others can lose the capacity to recognize or name colors. Often color and form deficits occur together. Many of these patients retain excellent visuomotor capabilities.

A particularly nice comparison of the different losses that can occur with damage to the dorsal vs. ventral stream has been provided by Milner and Goodale (1995). In this set of laboratory tests, both normal subjects and patients with brain lesions were tested. The subjects were asked to discriminate between objects, a task that should be compromised by lesions in the ventral stream. They were also tested with a grasp point task: they were asked to pick up a set of asymmetrical objects of different shapes, chosen so that unless the correct grasp points are used, the object will slip out of the hand. The grasp point test depends on visuo-motor coordination, and so should be compromised by lesions in the dorsal stream.

The results on the grasp point task for one normal control subject and two patients with different lesions are shown in Figure 18.3. The control subject discriminated well among objects, and chose effective grasp points. Patient RV, with dorsal stream lesions, discriminated well among objects, but had difficulty with the grasp point test; and patient DF, with ventral stream lesions, did well on the grasp point test but had difficulty with object discrimination.

Patient DF provides a particularly well documented case of the effects of ventral stream lesions. Many of DF's low-level visual functions, such as acuity, flicker resolution, and color discrimination, were tested and found to be relatively normal. But DF was unable to discriminate among simple forms or among letters. The contrast of lost and spared visual abilities was strikingly revealed by testing DF with a ”posting” task. This task involved a slot that could be rotated to different angular positions. DF could not consciously perceive or report the orientation of the slot, but could orient her hand correctly to insert it into the slot. Moreover, she could create appropriate hand postures to reach for and pick up objects, while being unable to name the objects verbally.

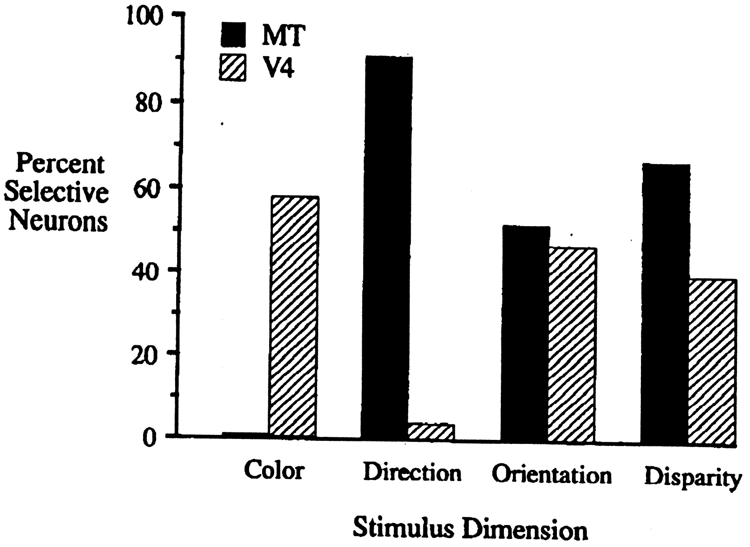


Figure 18.2: Comparison of the stimulus selectivities of single neurons in areas V4 and MT. The abscissa shows four stimulus dimensions – color, direction of motion, orientation, and binocular disparity – and the ordinate shows the percentage of neurons in each of the two areas that is selective for that stimulus dimension. The data are combined across many early studies. Note that neurons in V4 are selective for color but not for direction of motion, whereas MT neurons are selective for direction of motion but not for color. Also note the similarity of responsiveness to orientation and to binocular disparity between the two areas. [Adapted from Felleman and Van Essen, 1987, via Merigan and Maunsell, 1993, Fig. 3, p. 382.]

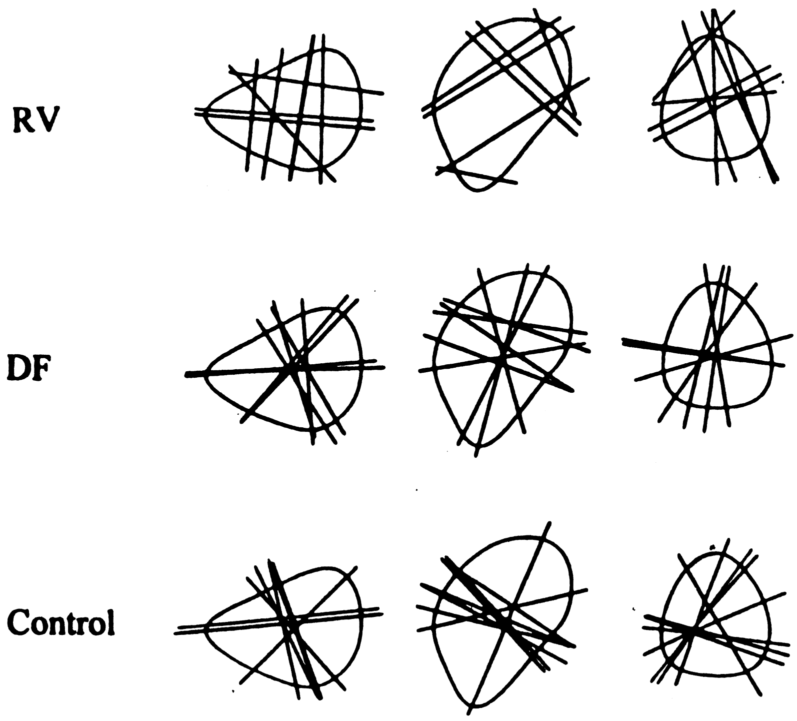


Figure 18.3: The grasp point task. Each patient or subject is shown several irregularly shaped objects and asked to pick them up. The ends of the lines indicate the places at which the subject placed his or her fingers. Patient RV, with a dorsal stream lesion, could discriminate among the objects verbally, but could not grasp them effectively. Patient DF, with a ventral stream lesion, could not discriminate among the objects verbally, but her grasping movements were effective. A normal control subject could both identify the objects and grasp them effectively. [Milner and Goodale, 1995, Fig. 5.7, p. 132]

In monkeys, lesions are experimental rather than accidental, and can be aimed at specific cortical areas. The losses of function seen after dorsal stream lesions are generally consistent with the selectivity of dorsal stream neurons for motion. Lesions in area MT can lead to deficits in motion perception and eye movements. However, monkeys often recover from these lesions, suggesting that the remaining brain regions can take over the lost functions. More extensive lesions that include both MT and parts of the posterior parietal cortex can lead to more permanent losses. Lesions in posterior parietal cortex can lead to losses or deficits in specific motor functions, including eye movements, and/or reaching for and grasping objects. These lesions can also lead to more diffuse deficits, such as a relative neglect of objects located in the visual field affected by the lesion, and/or a general spatial disorientation.

In the ventral stream, the losses of function seen after lesions are also broadly consistent with the selectivities of ventral stream neurons for form and color. Lesions of V4 cause small deficits in color discrimination. Since V4 has sometimes been thought to be a major processing region for color, the small magnitude of the losses is surprising. Lesions in later inferotemporal locations tend to lead to larger and more permanent color deficits. For form discrimination, the trend is the other way: V4 lesions tend to yield larger and more permanent deficits, with IT lesions having smaller and more transitory effects.

Finally, there are two lessons to be learned by comparing the devastating and permanent effects of LGN and V1 lesions to the generally smaller and more transitory effects of lesions in higher cortical areas. First, as Merigan and Maunsell (1993) point out, if there were a 1:1 mapping between the M pathway and the dorsal stream, or the P pathway and the ventral stream, the effects of lesions at LGN and in the corresponding cortical stream should be highly similar. The dissimilarities thus support the earlier argument that there is no 1:1 mapping between LGN pathways and cortical streams. And second, the smaller impact of cortical lesions suggests that higher cortical processing is diffuse and to some extent redundant. It is as though cortical lesions confined to a single visual area can leave behind a brain that is still sufficient to support most aspects of visual function. In addition, the transitory nature of the deficits suggests that cortical processing retains some plasticity – other brain regions can take over the tasks of the lesioned areas.

18.2.4 Psychophysics: A red herring?

It is sometimes argued that psychophysical and perceptual data can be used to support the notion of parallel processing streams in visual cortex. These arguments were based on the apparent functional separation of different aspects of visual perception, such as color and motion. That is, as discussed in Chapter xx, it is perceptually remarkably difficult to judge the direction of motion of a moving isoluminant chromatic stimulus. It can be argued that this failure of motion perception at isoluminance comes about because of the separate analysis of motion in the dorsal stream and color in the ventral stream. The argument is that if the dorsal stream has no information about color variations, and the ventral stream can't analyse motion, there is no way the subject can analyse the motion of a spatial pattern defined solely by color variations.

DT would argue, however, that these psychophysical findings provide no independent evidence for the existence of anatomically distinct parallel processing streams. Why not? As we discussed in Chapter 13xx, the loss of a visual function at isoluminance suggest that all of the neurons that perform that function have null planes or response minima at isoluminance. But functional isolation can come about from parallel processing by different neurons, such as the L, M and S cones, the M, P, and K ganglion cells, or their recoded cortical counterparts, whether these neurons are interleaved in a single cortical area or separated into anatomically distinct streams. Thus, in DT's view, psychophysical evidence of functional separation has little bearing on the existence of anatomically defined parallel processing streams.

18.3 How shall we characterize the functional roles of the different streams?

If we accept the parallel processing view, there is one final question: how shall we characterize the functional roles of the dorsal and ventral streams? To illustrate the interest and controversiality of the question, we will discuss four views of the answer.

First, in the early 1980s, based on studies of the effects of lesions on various discrimination tasks in monkeys, Leslie Ungerleider and Mortimer Mishkin (1982) introduced the idea that cortical visual processing is divided between dorsal and ventral streams. They proposed that the dorsal stream processes information concerning the *locations and motions* of objects, whereas the ventral stream processes information concerning the shapes, colors, and surface characteristics of objects, and has *object recognition* as its main function. They characterized the functions of the two streams in a single word each: *Where?* for the dorsal stream, and *What?* for the ventral stream.

A second view was proposed in the late 1980s by Livingstone and Hubel (1988). In keeping with their idea of continuity between precortical pathways and cortical streams, they suggested that the dorsal and ventral streams carry information about different *stimulus dimensions.* The dorsal stream was seen as analyzing *motion and stereopsis*, and the ventral stream as analyzing *color and form.*

A third perspective was introduced by David Van Essen and his coworkers (e.g. Van Essen and De Yoe, 1995). These authors argue that higher levels of the visual system are not designed to process stimulus dimensions; rather, they are designed to extract useful information about features and objects in the environment. But cues to the features of an object are coded in many different stimulus dimensions. For example, the pattern of motion in the stimulus can be a cue for depth and form as well as for motion per se; spectral composition can be a cue for form as well as for color; and so on. Therefore, the processing in each cortical stream must combine information from all of the stimulus dimensions that are useful for the particular visual functions processed by that stream. In terms of visual functions, Van Essen et al endorse the view that the dorsal stream carries out an analysis of *three-dimensional spatial relationships*, and the ventral stream is concerned with *pattern and shape recognition.*

A fourth view of the functions of the dorsal and ventral streams was proposed by David Milner and Melvin Goodale (1995). These authors suggest that all of the earlier proposals are flawed in focussing too narrowly on the analysis of stimuli at the expense of motor function. Milner and Goodale propose that the function of the dorsal stream is to interface with motor systems, and provide the basis for *action*; whereas the function of the ventral stream is to serve *object recognition and conscious perception.*

In summary, can we formulate a combined view of the functional roles of the dorsal and ventral streams, in 25 words or less? Yes, if we allow 25 words for each stream. Although there are important differences among the different theoretical views, a reasonable melding might be that the dorsal stream computes the physical locations and motions of objects, and interfaces with control systems needed for motor actions. In contrast, the ventral stream carries out the computations needed for object recognition, and provides the basis for conscious perception.

18.4 The Binding Problem

One more interesting puzzle arises out of the parallel processing view. This puzzle has come to be called the *binding problem.* If it is true that different aspects of perception are carried out in different, anatomically distinct streams, then is the information ever reunited? If not, then how do we achieve the unified perception of an object, and the perception of a single unified world? If shape were represented in one area, color in another, and motion in another, then how would we unite this information to see a cardinal flying? Is it necessary for all aspects of the percept to be reunited in a single anatomical location later, or not? Vision scientists disagree on the logic of whether one needs a *binding site*, or a *binding process*, or neither of the above. We will return to these questions in Chapter 27 when we wrestle with the problem of consciousness. [There are linking propositions lurking here. What are they? What should the perception of a unified object be taken to suggest or imply about the form of the neural signal that underlies the perception?]

In the meantime, DT would like to argue that if there's any such thing as a binding problem, then it's actually been with us all along, and does not arise from the presence of parallel processing streams. The argument is that any time you have a population code you have a binding problem. That is, the distributed representation of the spectral properties of a surface by activity in three cone types, or in M, P and K cells, poses as much of a binding problem as is posed by the existence of parallel processing streams in the visual cortex. If distributed coding is a problem, it's as much a problem when the neurons carrying the different elements of the code lie side by side in the retina, or separated by a tenth of a millimeter in the LGN, as it is is they are separated by several centimeters in the visual cortex. [Do you agree? Why or why not?]

18.5 An update on macaques: 32 visual areas, 305 connections, and still counting

For the anatomical afficionados, of course, there are many more areas and interconnections than are shown in Figure 18.1. The final count and nomenclature have not yet been achieved, and knowledge is still in considerable flux. As of the 1990s, for example, David Van Essen and his associates recognized 32 visual areas in macaque visual cortex, as shown in Figure 18.4 (Van Essen and De Yoe, 1995). In this figure, the cortex has been flattened out so that all the areas (including those buried in the cortical folds) can be seen. Area V1 is pulled out of the calcarine sulcus, cut at its margins, and drawn as an oval at the left of the figure. The true relative sizes of the different areas are preserved on the flattened map.

In Figure 18.4, color similarity is used to capture the major patterns of information flow among these visual areas. Areas V1, V2, and V3 are represented in shades of violet. Area MT, and the many divisions of posterior parietal cortex that form the dorsal stream, are shown in reds, oranges and yellows. Area V4, and the many divisions of inferotemporal cortex that form the ventral stream, are shown in blues and greens. There is good correspondence between the simplified scheme from Figure 18.1 as far as it goes, and the more complex one in Figure 18.4. However, there are many

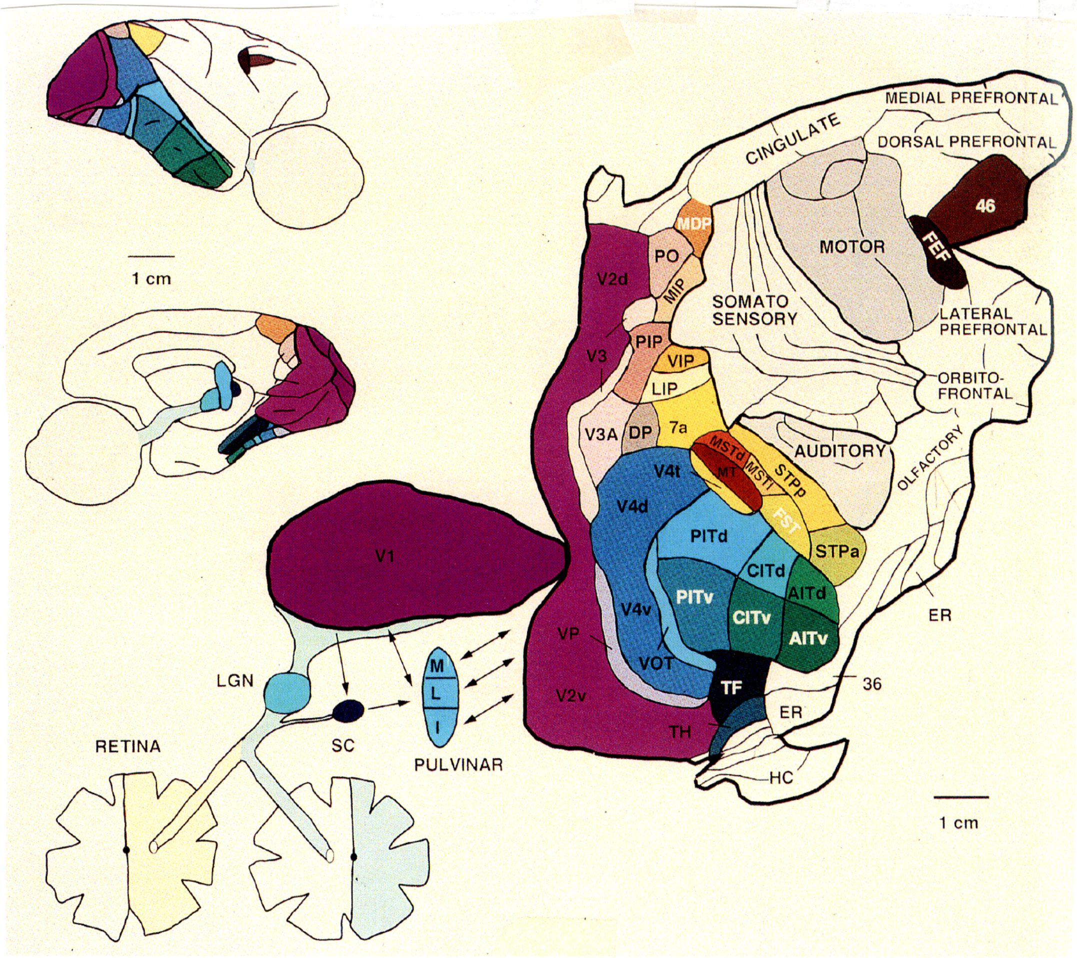


Figure 18.4: (COLOR PLATE). A more complex view: 32 proposed visual areas in macaque monkey cortex. A: a lateral view. B: a medial view. C: an unfolded (flattened) map of the whole cortex. The projections from the retina to the cortex are shown at the lower left. Virtually all of the visual input comes from the LGN via area V1, but a small projection to V2 via the superior colliculus (SC) and the pulvinar is also shown. Areas V1, V2, and V3, which are common to both dorsal and ventral streams, are shown in shades of purple and pink. The ventral stream is shown in blues and greens, and the dorsal stream is shown in reds, oranges and yellows. [Van Essen, Anderson, and Felleman, 1992, via Van Essen and De Yoe, 1995, color plate 8.]

more visual areas in Figure 18.4, and some of the areas that are already familiar have been broken up into subparts.

Similarly, the complete pattern of 305 interconnections reported by Van Essen and De Yoe (1995) is shown in Figure 18.5. The color coding of the visual areas is repeated from Fig. 18.4. With effort, the dorsal and ventral streams shown in Figure 18.1 can be traced out in Figure 18.5. However, there is much here that violates any idea of a strict parallel processing system. Aside from its complexity, the striking thing about Figure 18.5 is the many levels of cross-talk between the dorsal and ventral streams. A purely parallel system would have no such connections! Since there is so much cross talk, Van Essen and De Yoe prefer the term *concurrent* as opposed to parallel processing.

Several final issues deserve mention. First, remember that there is still disagreement among research laboratories over both the substance – the numbers and locations of visual areas – and the names that should be applied to them (summarized by Van Essen, 2004). And second, notice that this schematic says nothing about the relative strengths of connections between various areas. In fact, some of these pathways are composed of millions of axons; others, only a few. So when it comes to processing schemes, some of these pathways and connections will doubtless carry more information processing weight than others.

Third, is there order in the complexity? Van Essen and De Yoe argue that starting with V1, the cortical areas illustrated in Figures 18.4 and 18.5 are arranged in a 10 stage hierarchy, as shown by the 10 discernable horizontal rows of boxes in Figure 18.5. The justification for the hierarchical ordering stems from a detailed and consistent pattern of connections found among the many visual areas. As discussed in Chapter 17, the general rule is that when two cortical areas A and B are interconnected, feedforward connections originated in the earlier cortical area (say area A) in layers 2 and 3, and enter the later cortical area (say area B) in layer 4. In contrast, feedback connections usually originate in both superficial and deep layers of the later area, and terminate in these same layers in the earlier area, avoiding layer 4.

When Van Essen and his coworkers followed this pattern of interconnections through the cortical maze, they found that the pattern was consistent with a unidirectional overall pattern of information flow. If area A sends projections from its layers 2 and 3 to layer 4 of area B, area B never sends connections from its own layers 2 and 3 back to area A. Thus, in some ways it will be useful to consider the visual system as a largely hierarchical series of perhaps 10 more transformations, beyond the first five transformations we considered in earlier chapters.

The final issue, however, is the role of feedback. Historically it has been tempting to think of the visual system as a single processing hierarchy, or two hierarchies if we separate dorsal and ventral streams. Yet there is strong evidence for feedback from higher to lower levels of the hierarchy. In Figure xx, only feedforward connections are shown, but in fact, most of the connections are reciprocal: for almost all of the feedforward connections, there is at least a small feedback connection between the same two structures. The question is, how important is feedback, and what functions does it serve? At the present time opinions differ on this issue. All researchers acknowledge the presence of feedback pathways, but many still consider that the major processing pathways are largely feedforward. Others stress the critical influence of feedback, particularly in mediating the influence of attention and other top-down processes.

A skeptic might even argue that with so much cross-talk and feedback, the ideas of hierarchy and/or parallel or concurrent processing streams should simply be discarded. Perhaps we should start over, and conceive of visual cortex as a single humongous processing network, with neu-

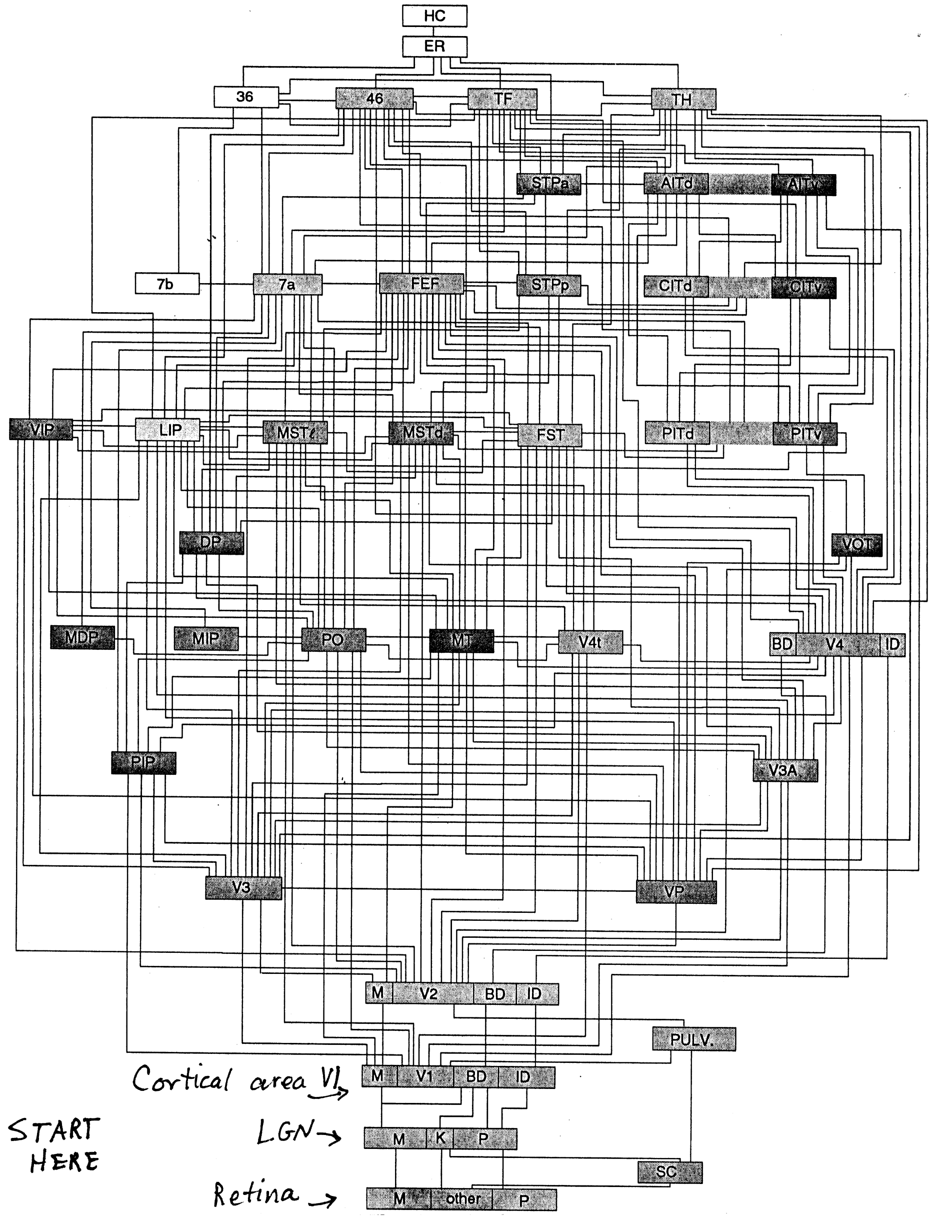


Figure 18.5: (COLOR PLATE). The 305 proposed feedforward connections among the 32 visual areas shown in Figure 18.4. The small squares show the different visual areas, with the color code repeated from Figure 18.4. V4 and MT are near the center of the diagram. The input from the retina comes in at the bottom. The dorsal stream is generally to the left, and the ventral stream is generally to the right, but the two streams are heavily interconnected. The connections shown are all feedforward; most of these pathways have feedback projections as well. [Van Essen, Anderson, and Felleman, 1992, modified by Van Essen and de Yoe, 1995, Fig. 24.4, p. 388. NEED TO GET COLOR PLATE xx]

rons with different properties contained in different cortical areas, but little hierarchical structure. [Think about this. Is it time for a change of paradigm?]

18.6 Miracle of miracles: Images of the human brain

Weturn now to our second major topic: brain imaging. One of the most remarkable scientific breakthroughs of the last 20 years has been the development of brain imaging techniques. The earliest techniques, such as *computer tomography (CT)*, made it possible to create static images of structures within the brain. More recently, functional imaging techniques such as *positron emission tomography (PET)* and especially *functional magnetic resonance imaging (fMRI)* have been developed.

The importance of these techniques cannot be overstated. With functional imaging, it is possible to trace changes in the activity of individual brain structures as a human subject looks at different visual stimuli or carries out different perceptual tasks. Almost all of the information we have accumulated on the visual brain in the past 50 years has been gleaned from monkey subjects, and we have been guessing that the human visual system would be similar. But with the use of imaging techniques, humans have returned to center stage, and an independent cartography of the human brain is now in the early process of construction.

18.6.1 fMRI and the BOLD signal

Currently the most advanced functional brain imaging technique is fMRI, and our attention will center on this technique. In non-technical terms, fMRI depends on the fact that activity in a particular brain region causes an increase in the flow of oxygenated blood to that region. For reasons that are not completely understood, not all of the extra oxygen is used, with the consequence that for a few seconds, there is a slight increase in the fraction of oxygenated to deoxygenated blood in the region. The change in blood oxygenation changes the local magnetic field within the activated brain region. This change can be picked up and localized by the fMRI equipment. The resulting signal is called the *blood oxygen level-dependent (BOLD) signal.*

With time, fMRI techniques are becoming more and more sensitive and more and more sophisticated. In particular, signal to noise characteristics have improved many fold since the earliest studies. Moreover, the spatial and temporal resolution of the BOLD signal have improved consistently over time. With current instrumentation, it is possible to resolve BOLD signals from cortical regions separated by as little as about 1.0 mm. The temporal resolution has been limited by the relatively slow time course of the changes in blood flow that underlie the BOLD signal. However, sophisticated data analysis techniques are reducing temporal integration periods from several seconds to one or two. And finally, it has become possible to use fMRI and psychophysical techniques together, while the subject is in the magnet performing the psychophysical task.

In the present chapter we will place our emphasis on three aspects of the knowledge that is being derived from fMRI work. The first is the early search for the human homologs of monkey areas MT and V4. The second is the discovery of precisely predicted topographic maps in some early human visual areas. And the third is the development of an fMRI-adaptation paradigm.

18.6.2 Early studies: Discovery of human areas hMT+ and hV4

By the early 1990s, fMRI techniques had become sufficiently well developed to be used fruitfully with human subjects. But even with fMRI techniques and monkey maps in hand, there is still no paint-by-numbers kit for the human brain. How shall we start? One possibility is to assume that areas similar to those in macaque, with similar functional characteristics, will exist in similar locations in human visual cortex. In the 1990s, areas MT and V4 played a central theoretical role in views of the cortex, as they marked the anatomical divergence between the dorsal and ventral streams. Thus, as soon as fMRI techniques became available, the race was on to locate believable human homologs of monkey areas MT and V4.

In terms of experimental approaches, the early fMRI studies made use of a *differencing paradigm.* That is, a pair of stimuli are used: one that is expected to lead to a high level of activity in a particular cortical area, and another that is as much like it as possible, but is expected to be less effective. The two stimuli are presented repeatedly, in alternation, for a few seconds each, and the difference in the BOLD signal is computed. A reliable difference between the two signals suggests that the brain region under study is well activated by the stimulus characteristics that differ between the two members of the stimulus pair.

In 1995 Roger Tootell and his associates (Tootell et al, 1995) used a differencing paradigm to look for a motion-selective region that might correspond to macaque area MT in human subjects. Since moving stimuli are highly effective in activating individual neurons in macaque MT while stationary stimuli are not, Tootell and his colleagues recorded BOLD signals arising from random dot patterns that were either moving or stationary.

The results from two different cortical areas, tentatively identified as human V1 and human MT, are shown in Figure 18.6. In putative human V1, both moving and stationary stimuli give rise to BOLD signals of similar magnitude; but in putative human MT, the BOLD signal from moving stimuli is much larger than that from stationary stimuli. Moreover, this motion-sensitive cortical region lies near the intersection of the human occipital, parietal and temporal lobes, arguably not too far from the corresponding location of area MT in monkeys. Similar results have been seen in several laboratories. This region has been named *human MT+*, or *hMT+*, with the + sign indicating that it may actually be homologous to monkey MT *plus* other neighboring motion-selective regions, rather than just to monkey MT. There is reasonable consensus, then, that the human homolog of macaque MT has been identified with fMRI. Remarkable!

Similarly, to search for the human homolog of V4, one could again use a differencing paradigm, comparing BOLD signals from spatial patterns that vary in color (or color and luminance) vs. the same spatial patterns that vary only in luminance. With paradigms of this kind, several research groups have reported the presence of a color-selective region or regions on the ventral surface of human occipital cortex. Some of these regions have come to be called *human V4*, or *hV4*. However, color signals have been seen in several locations on ventral occipital cortex, leading to controversies over exactly which region or regions really correspond to macaque V4.

A problem that arose immediately, of course, is that as fMRI techniques become more and more sensitive, and the choices of stimulus pairs more and more sophisticated, more and more motion-sensitive and/or color-sensitive cortical regions will inevitably be found. And in fact, as common sense would dictate, in addition to the specially targeted areas that we all hope are homologous to V4 and MT, most fMRI studies reveal both color and motion signals in early cortical areas probably homologous to V1 and V2, as well as in many higher-level areas. With fMRI many parts

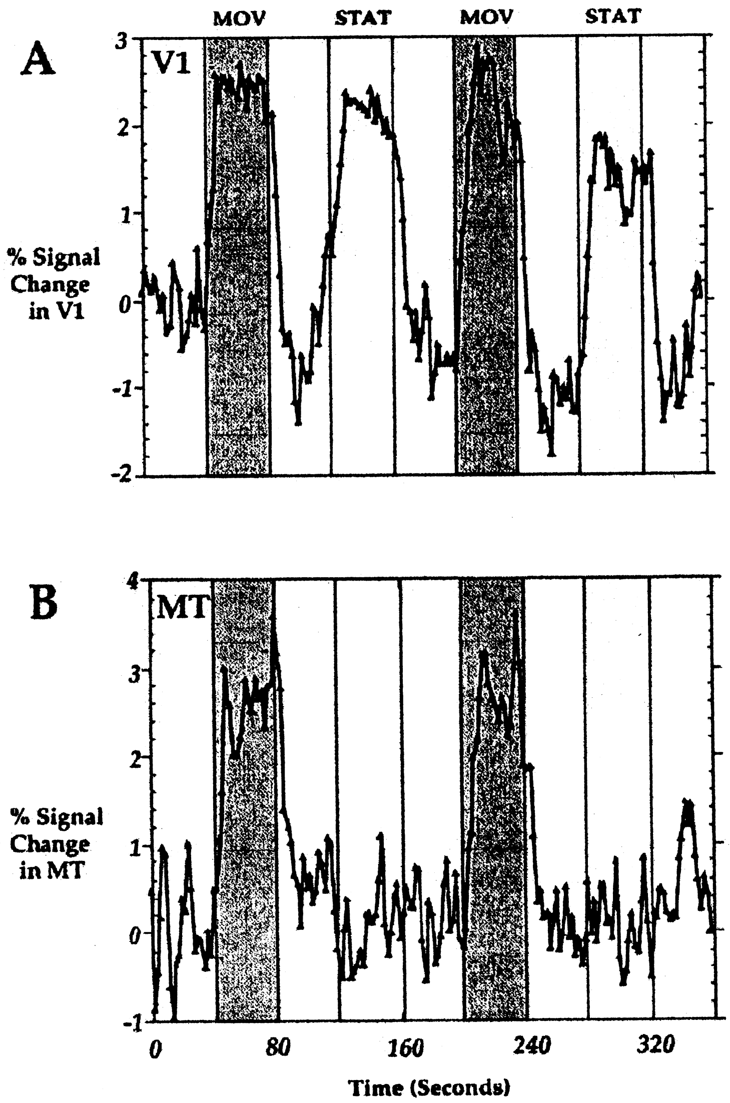


Figure 18.6: BOLD responses to moving vs. stationary stimuli. BOLD signals were recorded from putative areas V1 and MT in human cortex. The abscissae show time in seconds; the ordinates show the percent change in the BOLD signal. In the experiment, random dot patterns moved radially during two 40 sec periods (labeled MOV), and were stationary in two other periods (STAT). These periods were separated by 40 sec of baseline stimulation without random dots. A. Signals from putative V1. Moving and stationary targets produce similar BOLD signals. B. Signals from putative MT. Moving targets produce much larger BOLD signals. [Tootell et al, 1995, Fig. 5, p. 3221.]

of the cortex can be looked at at once, and the problem of motion- or color-selective areas quickly became one of abundance rather than scarcity.

In fact, to many vision scientists, it seemed odd from the beginning to search for "the" motion-sensitive or "the" color sensitive area. Rather, it seems logically necessary that information about motion and about color must be carried through all of the levels of at least one visual processing pathway, and that signals arising from the various recodings of motion or color should be demonstrable at the various levels of the hierarchy. From this perspective, the challenge becomes how to choose stimulus pairs that can sort out the codings and recodings of motion and color signals on their way up the visual processing hierarchy. [How could you find out whether area hMT+ is truly selective for motion, or just for temporal change? How about selective for the *direction* of motion, or just for motion per se? And what about changes of color code?]

18.6.3 Topographic maps

A second major thrust of fMRI work to date concerns the identification of specific topographic maps in human visual cortex. The work is based on adopting the working hypothesis that a detailed homology exists between human and monkey cortex. The topographic maps seen in monkey cortex are then used to make detailed predictions about the features of topographic maps in human visual cortex.

To illustrate the approach, we need to go into detail about the topographic maps in macaque V1, V2, and V3. The layout of these maps is shown schematically in Figure 18.7. As previously shown in Figure Xx [in V1 chapter], area V1 in the left hemisphere contains a full topographic map of the right half of the visual field (the left half of each retina), with the fovea being represented at the back of the occipital lobe, and increasingly eccentric retinal regions being represented at more forward locations within the calcarine sulcus. Moreover, the lower visual field projects to the upper bank of the calcarine sulcus, and the upper visual field to the lower bank, to provide a full topographic map of the right visual field. These maps are repeated, with variants, in areas V2 and V3. In each case, retinal eccentricity is laid out from back to front on the medial surface of the occipital cortex.

The relations among the maps in V1, V2, and V3 are also distinctive. Since area V2 surrounds V1, in the schematic of Figure xx V2 is split into dorsal and ventral halves, V2d and V2v respectively. Similarly, area V3 has borders contiguous with both the dorsal and the ventral borders of V2, and is drawn as split in the same way, forming halves V3d and V3v. In the left hemisphere, V2d and V3d contain representations of the lower right quadrant of the visual field, and V2v and V3v contain representations of the upper right quadrant.

Finally, the topographic map in V1 is projected point by point to V2 and V3, but with some twists. As emphasized by the labels at the right of Figure 18.7, the maps reverse in orientation at the boundaries between areas. That is, the boundary between the dorsal part of V1 and V2d is composed of two adjacent representations of the lower vertical meridian (LVM). Similarly, the boundary between V2d and V3d is composed of two adjacent representations of the horizontal meridian (HM). Similar reversals can also be seen at the boundaries between V1, V2v and V3v.

Why are these details interesting? If areas homologous to V1, V2, and V3 in macaques exist in humans, one would predict that the same complex set of topographic maps, with the same interrelationships, would be found. Stimuli that fall on progressively more peripheral retinal regions should yield BOLD signals that arise from progressively more forward parts of the calcarine sulcus

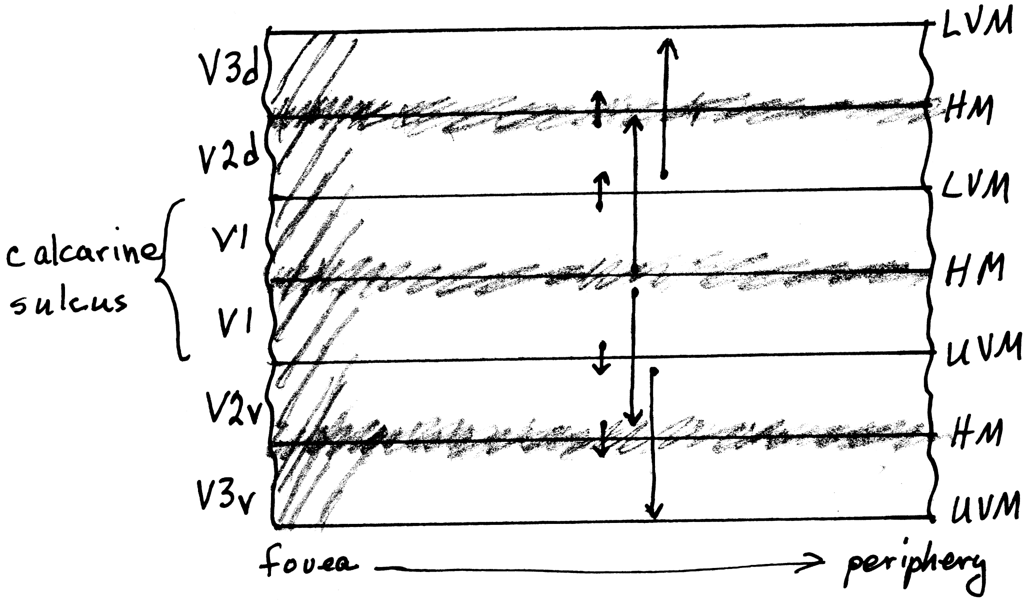


Figure 18.7: Anatomy and topography of macaque areas V1, V2, and V3 (highly schematic). Area V1 lies in the calcarine sulcus, with the horizontal meridian (HM) lying at its bottom, the lower visual field on its upper (dorsal) bank, and the upper visual field on its lower (ventral) bank. The dorsal half of V1 projects to the dorsal half of V2 (V2d), which then projects to the dorsal half of V3 (V3d); similar arrangements hold for the ventral areas. The topographic map of the lower visual field in the dorsal half of V1 is preserved in V2d and V3d, but the map reverses at each boundary between areas. The small arrows show the point to point projections between areas. For example, the representations of the lower vertical meridian (LVM) in V1 and V2d adjoin each other at the V1/V2d boundary, as do the representations of the horizontal meridian (HM) at the V2d/V3d boundary. Similar patterns hold for the corresponding ventral areas. The hatched region at the left predicts the pattern of response that should be seen for a foveal stimulus, and the shaded horizontal regions show the predictions for stimuli on the horizontal meridian.

and the surrounding medial surface of the occipital cortex; and stimuli that fall on the various meridia should yield BOLD signals that reveal the boundaries between particular pairs of areas.

An illustration of fMRI data from such a cortical mapping study is shown in Figure 18.8. Figure 18.8A shows that the location of the BOLD signal varies as it should with the eccentricity of the stimulus. Figure 18.8B shows the BOLD signals that reveal the boundaries between the visual areas. Given the complexities of the predicted maps, and the correspondence of the fMRI maps with the predictions, there can be little doubt that the human homologs of areas V1, V2, and V3, and the specifics of their topographic maps, have been established. Figure 18.8B also shows three subjects, to convey an idea of the extent of individual differences in human cortical cartography.

The importance of these fMRI mapping techniques cannot be overstated. Using these techniques, a *locator scan* – a mapping of the locations of specific cortical areas – can now be carried out in an individual subject in about half an hour. With the subject's visual areas and their topography now defined objectively, one can choose a *region of interest* (ROI) – say, foveal V1 – and study its responses to specific custom-designed stimuli. The results can then be compared across retinal areas and across subjects with much more certainty than before.

Beyond V1, V2 and V3, at least a dozen additional visual cortical regions are turning out to have topographic maps definable with fMRI mapping techniques. Any of these areas could potentially be located in individual subjects with a locator scan, and defined as a ROI. Specific cortical areas can then be tested with specific custom-designed stimuli, and codes and code changes potentially identified.

18.6.4 The fMRI-adaptation paradigm

A third major advance in fMRI work is the use of *selective adaptation paradigms* (cf. Chapter xx). Imagine a cortical region such as V1, containing different subsets of neurons selective for different stimulus characteristics, such as the orientations of line segments. With the differencing paradigm it would probably not be possible to demonstrate this orientation tuning, because lines or gratings at different orientations would probably all produce similar BOLD signals.

However, a selective adaptation paradigm could be just the ticket. We could begin by selecting a region of interest (ROI) and measuring the BOLD signal to lines or gratings at each of many different orientations. We could have the subject adapt to stimuli at a particular orientation, and then repeat the initial measurements. If orientation tuned, selectively adaptable neurons exist in the ROI, then the magnitude of the BOLD signal would be decreased by adaptation, with the largest signal loss occurring at the orientation of the adapting stimulus.

An experiment of this kind was carried out recently by Fang Fang, Scott Murray, Daniel Kersten, and Sheng He (subm). fMRI locator scans were first carried out on each subject, so that ROI corresponding to V1 and several higher visual areas could be defined. The experiment proper then began. The stimuli in the main experiments were arranged as shown in Figure 18.9. The subject's attention was controlled, and confined primarily to the center of the display, by having the subject fixate a fixation target and report the occurrence of luminance changes. While she was performing this task, the subject was first adapted for 20 sec to a set of 16 Gabor patches of varying orientations. FMRI responses were then obtained in response to each of four sets of test targets, consisting of the same pattern of Gabor patches, but with each patch rotated away from the orientation of its local adapting patch, by either 0, 7.5, 30, or 90 degrees.

The results are shown in Figure 18.10. The fMRI signal was indeed selectively reduced by

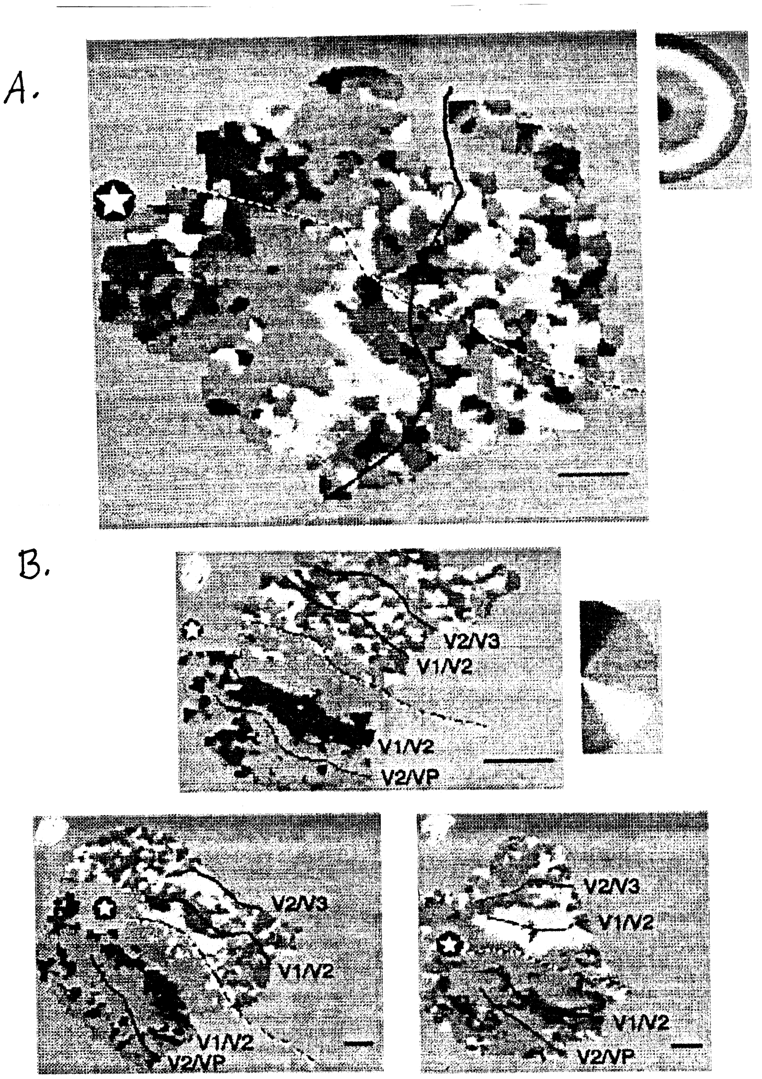


Figure 18.8: [COLOR]. Cortical topography revealed by fMRI. Each panel shows a flattened map of the medial surface of the left hemisphere. The dashed lines show the bottom of the calcarine sulcus, and the stars show the foveal representation. A: Effects of stimulus eccentricity. The solid line shows the limit of the meaningful data in the experiment. The color code shows the eccentricity of the most effective stimulus for each region of cortex – dark blue for the fovea, red, orange and yellow for increasing eccentricities, and green for the most peripheral stimulus locations used (12 degrees). B: Borders between areas V1, V2, and V3. The solid lines show the inferred borders. The color code shows the radial position of the most effective stimulus – dark blue for the upper vertical meridian, orange for the horizontal meridian, and green for the lower vertical meridian. The three maps in B show results from three individual subjects. (The labels differ slightly from those in Figure 18.7. In particular, VP in Figure 18.8 corresponds to V3v in Figure 18.7.) [From Engel et al, 1997, Figs. 7 and 8, p. 187.]

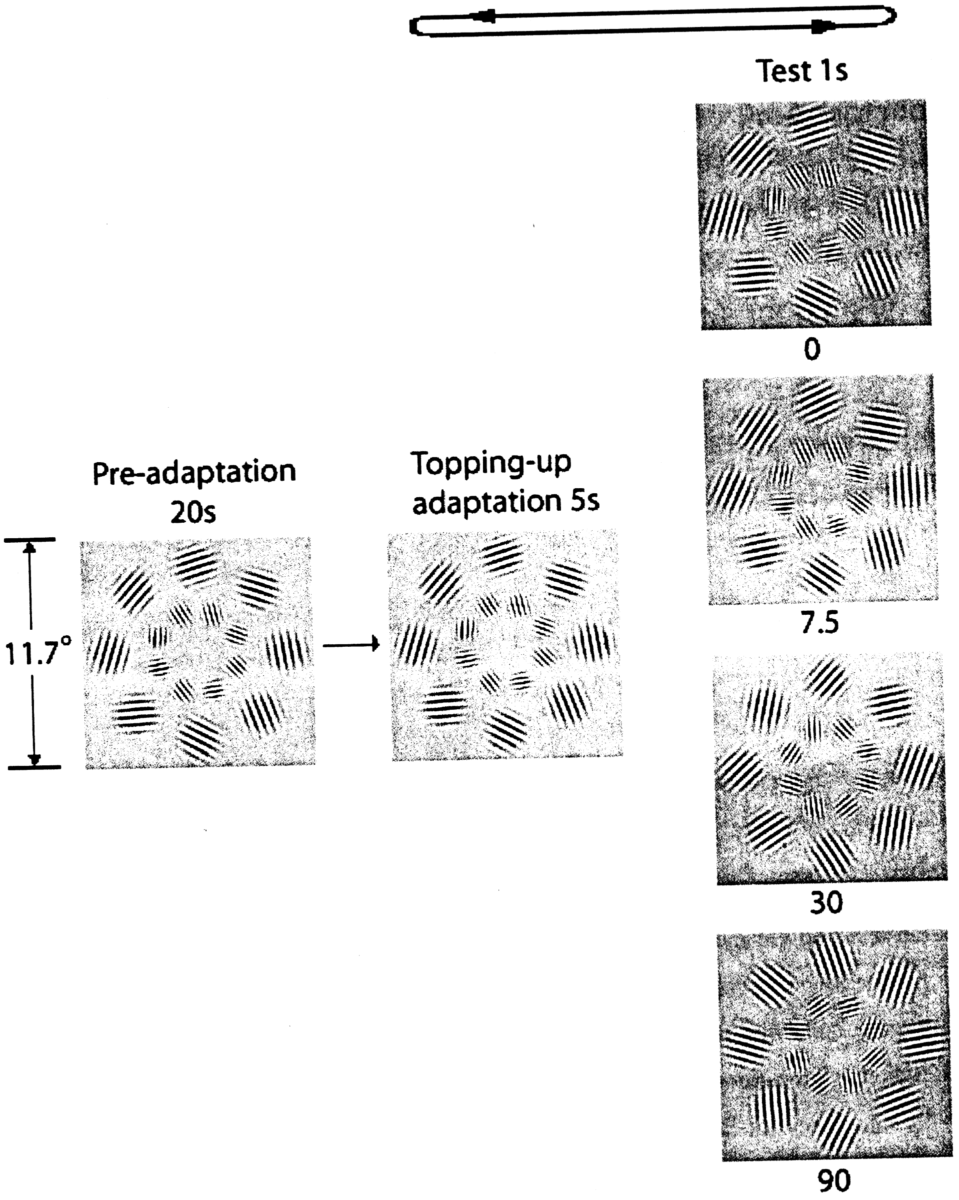


Figure 18.9: The fMRI adaptation paradigm. The figure shows the stimuli used by Feng et al (subm). The subject fixated at the white dot. Two rings of grating patches, randomly oriented, were presented for a 20 sec pre-adaptation period. Between test trials, a 5 sec "topping up" adaptation period was also used. Test stimuli were presented for 1 sec, and differed in the angular deviation between the grating in each test patch and the grating in the corresponding adaptation patch. These deviations were 0, 7.5, 30, 04 90 degrees (top to bottom panels).

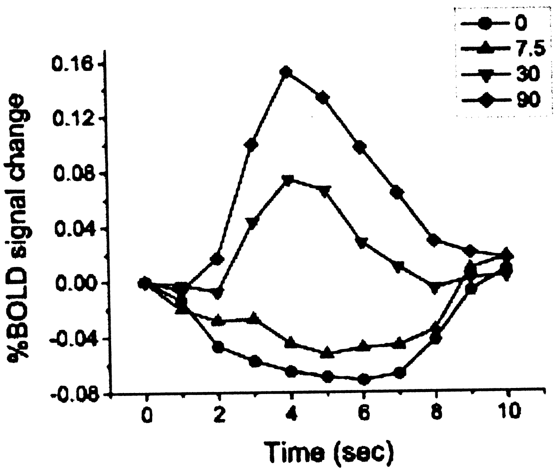


Figure 18.10: Results of Feng et al's experiment. The BOLD signal increases with the angular deviation of the test stimuli from the adapting stimuli, being smallest (actually negative) for 0 degree deviations, and largest for 90 degree deviations.

adaptation. The smaller the difference between the orientations of the local adaptation and test patches, the smaller the fMRI signal generated by the test patch, as shown in Figure xx. This result was seen most particularly in V1, but also in several higher visual areas.

Fang et al's results support the conclusion that human V1 cortex contains populations of neurons that are both orientation tuned and selectively adaptable. Of course, this is not surprising – we doubtless believed it all along on the basis of single unit studies in monkey cortex (Chapter 17xx). More importantly, these results show that the fMRI adaptation paradigm works – selective changes in the fMRI signal, generated by the adaptation paradigm, are big enough to be measured from the heads of human subjects. The Fang et al study thus validates the feasibility and potential usefulness of fMRI adaptation paradigms for studying features of the neural code. What remains is to custom design the stimuli.

In summary, fMRI work appears to be entering a new phase. As topographic maps are discovered and visual areas are objectively defined in individual subjects, fMRI experiments seem poised for a major leap forward. Rather than using functional properties to try to guess the locations of human visual areas, we can now define human ROI objectively and go on to test their functional properties. Moreover, fMRI adaptation paradigms should allow us to explore the selective tuning of neurons in individual cortical areas. Such paradigms bode well for the future of a vision science that incorporates fMRI as a major tool, and puts the human subject back at center stage.

The icing on the cake comes from the fact that fMRI, like psychophysics, can be done on both human and macaque subjects. Thus, both species can in principle be tested with exactly the same fMRI paradigms, psychophysical paradigms, and visual stimuli. Once areas are defined and homologies established, then an identity of fMRI signals between homologous cortical regions in the two species would provide strong evidence for highly similar visual codes in the homologous regions. It would then be almost safe to conclude that the properties of single units, measured invasively in macaques, are good descriptors of the single unit activity in human visual cortex.

18.7 Human vs. monkey maps: Are there any major differences?

From what we have learned so far, are there major differences between the visual areas of humans and macaques? The homology of V1, V2, and V3 seems well established, and homologies in the early dorsal and ventral streams seem within our grasp. Early results suggest that in terms of fractions of cortical surface area, the areas involved in early cortical processing – V1, V2, and V3, for example – are relatively smaller in humans than in macaques. But the areas involved in higher level visual processing are relatively larger in humans, in keeping with the greater complexity of perceptual/cognitive processing in the human species.

18.8 Summary: Cortical cartography

In the 1970s and early 1980s, visual anatomists began the task of sorting out the regions of the visual cortex engaged in visual processing, and tracing the anatomical connections among them. The task became complex very quickly. It is now agreed that many distinct cortical areas are involved in vision, but there is still incomplete agreement about how the higher visual areas should be split up, or characterized in terms of visual function.

Until about 20 years ago, most vision scientists probably thought of vision as a largely serial processing system. Then, importantly influenced by Ungerleider and Mishkin's work, we began to think of the possibility that different pathways in different areas of the brain might pursue different specializations in processing or representing visual information. At present, most (but not all) vision scientists would espose a parallel or concurrent processing view of one kind or another. But there is no consensus yet on exactly how many processing streams we should distinguish, what anatomical structures they pass through, how much cross talk goes on, or what the precise functional descriptions should be.

An exciting new tool that arrived in the 1990s is that of functional neuroimaging, especially fMRI. Modern imaging techniques have the potential of restoring studies of the human species to a central role in vision science. When used with appropriately custom-designed stimuli, fMRI can be used to define topographic maps, and thereby visual areas, in human visual cortex. We are on the brink of being able to use fMRI adaptation and other related paradigms to study the functional characteristics of neurons in these newly defined visual areas, and compare the findings in detail to the earlier knowledge gleaned from single unit studies in macaque visual cortex.

At the same time, the study of human cortical visual processing is still in its infancy. So far we know that the topic is highly complex, and we expect the detailed picture to keep changing rapidly in the immediate future. The field is marked by a certain amount of disagreement – sometimes bitter disagreement – among laboratories, as is sometimes the case at scientific frontiers. In fact, the field is worth watching solely as an exercise in scientific argumentation, let alone as a fascinating voyage of discovery into the human brain.

18.9 Strategy for future chapters

We have come to a turning point in this book. The early chapters of the book have been concerned largely with retinal processing and the marks it leaves on our perception. We were in a realm of visual science in which many system properties and many substrate properties are relatively well established, so we could stand on relatively firm ground while pursuing our philosophical interest in causal stories. In particular, we could get away with the assumption that visual processing proceeds through a comparatively simple hierarchical processing system, in which information in the incoming visual stimuli is systematically coded and recoded into ever more useful forms.

But central visual processing is much less well understood, and more complex; and there are still many disagreements. Moreover, as we will see, it will be necessary to introduce the idea of *vision as an active process*: the idea that some of the information that enters into visual perception is supplied by the central visual system, rather than by the incoming stimulus alone. With all of these complexities, it is sometimes difficult to gain a foothold on the topics of perception and central visual processing, let alone the legitimacy of particular causal stories. The satisfying precision of some of our earlier arguments is hard to find here. At the same time, it seems possible that some of the paradigms and logic we found useful when the footing was more certain, might still be useful in cases in which more uncertainty abounds.

The outline of the remainder of this book is as follows. First, we return to a philosophical chapter, considering the nature of conscious perception, the knowledge of the physical world, and the nature of neural processing in more detail. Then we proceed to chapters on several different visual functions, including image segmentation (Chapter 20), the perception of motion (Chapter 21xx), brightness and color (Chapter 23), distance and size (Chapter 24), depth and three dimensional shape (Chapter 25) and forms and objects (Chapter 26). In these chapters, our didactic strategy is reversed, and the perceptual phenomena drive the choice of anatomical and physiological topics. Finally, we end with a chapter (Chapter 27) on the nature of consciousness, and what it would mean to search for its neural substrates.

Chapter 19

Perception and Its Neural Substrates

So far, we have attended only to the retinal image and the early sensory processing of the incoming visual signals. We have traced the visual signals through several code transformations , all the way to area V1, and provided a simplified sketch of what is known about processing beyond V1. In conjunction with our serial tracing of visual signals, we have explored a variety of system properties that seem most closely tied to early sensory processing. It's now time to turn to a broader set of system properties – those deriving less closely from studies of sensory processing and more closely from studies of visual perception.

Including perception inthe mix occasions a return to the beginning – to the distinctions and relationships between three entities: the perceived world, the physical world, and the neural states that provide the physical basis for the perceived world.

First, we define perception and the perceived world, distinguish it from the physical world, and spell out some of its unique properties. In contrast to our earlier sensory view, the perceptual view centers around the perception of physical objects and their three-dimensional locations.

Second, we define the physical world and contrast it with the perceived world. We introduce the idea that perception doesn't depend on the incoming sensory information alone – it is both impoverished by sensory processing, and enriched by stored information. We argue, however, that given the confounding of physical variables in the retinal image, the properties of the perceptual world bear a remarkably close correspondence to the properties of the physical world. These correpondences provide an important new set of system properties in search of explanations.

Third, we return to the question of neural codes, and inquire into the relations between high-level neural codes and the perceived world. In particular, we will ask, what are the implications of the new system properties revealed by perceptual phenomena, for deducing or guessing the forms that neural codes might take? And (naturally!) what linking propositions are involved in such deductions and guesses?

And finally, we address the question of neural coding principles. A question that has recently been at the center of some of the theorizing in neuroscience is, what makes a good neural code?

In taking high level visual processing and its relation to perception,, we are leaving the relative safety of sensory processing and embarking on a journey in a land in which much less is known. There are relatively few studies of the neural basis of perceptual effects, and they form much of the vanguard of visual neuroscience. How can the lessons and successes of our studies of sensory aspects of vision, guide the journey? How can we use the information and principles we have learned at the sensory processing level, to generate fruitful hypotheses about what the codes will be like at the cortical level? the goal of this chapter is to try to capture and make explicit the common assumption structure within which work on central visual neurophysiology isbeing studied. We will attempt to bring together themes from philosophy, sensory studies, perception, and theories of neural processing, and weave them into a hopefully coherent story.

19.1 The perceived world

19.1.1 Perception is a part of conscious awareness

To discuss and define perception, I begin (as Descartes did) by acknowledging the existence of my own conscious awareness. Given that premise, I define perception as the part of my conscious awareness that is concerned with the status of the world around me, or (in the case of vision) in front of me. The world as represented in my perception can be called the *perceived world* (or the *phenonenal world*, or the *world as perceived*). Moreover, I extend my basic premises to acknowledge that other people – including you, the reader – also have conscious awareness, and perceptions of the world.

Another common way of saying that perception is part of conscious awareness is to say that our perceived worlds are known to each of us *from the inside*, or from a *first person perspective.* These phrases are intended to capture the difference between the kind of knowledge we have of our own conscious states, and the kind of knowledge we acquire as scientists, *from the outside* or from a *third person perspective.*

In defining perception as a part of conscious awareness, we are choosing a position on one aspect of the mind/body problem. From a perspective innocent of either science or philosophy, I know that I am conscious, and have perceptions that are somehow related to the physical world and the objects around me. I endorse the existence and legitimacy of my own conscious perceptual states, and will not concede an inch to any philosophical position that tries to abolish them, or claim that they are really something else (such as brain states). I can readily agree that perceptual states are intimately related both to the physical world and to brain states, and be intensely curious as to the nature of these relationships – as indeed we are in this chapter. But I insist that perceptual states are neither physical states nor brain states. They are conscious states, known to me from a first person perspective, and can never be legislated or defined away*.

19.1.2 Properties of the perceived world

What more can we say about the properties of the perceived world? As it happens, the perceived world is in some ways very easy to describe. This is because until we begin to study sensory systems or think about perception, we don't typically make much of a distinction between the perceived world and the physical world. So to describe the perceived world, we can more or less describe what we have always thought of as the physical world, noting the differences as we go along. What are some of the characteristics of the perceived world?

For some people, one of the most striking yet puzzling properties of the perceived world is its location – out in front of us. That is, some philosophers of perception presume (note the linking proposition) that if perceptual states arise from brain states, they should be located in the same place as brain states; namely, inside the head. But obviously, this is not the case – my perceived world is located outside my head, in the space in front of me. This property of the perceived world, mysterious to those who assumed otherwise, is sometimes called *perceptual projection*.

In any case, my perceived world lies in front of me. It is laid out in a three dimensional space that extends over my whole visual field in two dimensions, and indefinitely far in distance. It is populated with perceived entities called objects. The selection of objects changes as I move about in the world – it can be streets, buildings, people and cars in the morning, books and computer monitors at noon, and trees, rocks and rattlesnakes on Saturday. Importantly, perceived objects have a set of characteristics – perceived shape, size, color, and so on – that usually remain constant, and three-dimensional locations that are typically continuous but more susceptible to change. Other properties of the perceived world include such characteristics as the general light level, the state of the weather, and so on. [Work out a description of your currently perceived world.]

Another interesting property of the perceived world is that there is no strict dividing line between sensory processes, perception and cognition. For example, I may see a dog on the lawn. When the dog moves into my peripheral visual field, the sensory inflow loses spatial detail, but my perception continues. When the dog runs behind a tree, the incoming stimulus information required for perception of the dog is completely removed. But perception is augmented by stored information. Cognition and memory take over smoothly and automatically from perception, and I still sense, or perceive, or ”see” that there's a dog behind the tree.

Finally, a striking characteristic of the perceived world is its fragility. The perceived world loses many of its characteristics when the light level gets too low, and it goes away entirely when I close my eyes. All of these properties – especially the division of the perceived world into objects with relatively constant characteristics, are system properties in search of explanation.

19.2 The second entity: The physical world

The second of our three entities is the physical world; the world as described by physicists. It consists of matter and energy in wondrously varying configurations. It can be described at several different levels, from subatomic structures through objects as wholes to cosmological descriptions of the universe. At the scale most relevant to the present discussion, the physical world contains physical objects with physical shapes, sizes, and surface properties, at particular physical locations in three-dimensional physical space. It also contains such features as sources of radiant energy (some of which are perceived as light), and physical configurations such as concentrations of tiny water droplets that condense in the atmosphere and fall toward the earth (causing the perceptual phenomenon of rain).

But how are we to know about the physical world and its properties? Knowledge of the physical world comes basically from three sources. The first of these is our perceptions. As we said before, from a naive perspective the physical world is pretty much the same thing as the perceived world, and our moment-to-moment knowledge of the physical world is supplied by the perceived world.

The second source of information is measurements made with physical instruments – rulers, thermometers, spectrophotometers, telescopes, microscopes, thermocouples, and so on. It is the invention and use of such instruments that allows us to make a firm separation between the perceived and physical worlds (for example, the separation between perceived and physical size, and between color and wavelength composition), even as we note the striking correspondences between them.

The third and most abstract source of information about the physical world is physicists' models of its properties. Gravitational theory tells us why apples fall from trees, and quantum mechanics tells us about the nature of light. The physicist's understanding of the physical world is acknowledged to be incomplete, and is open to change as new measuring instruments are invented and new models are devised.

The physical world differs from the perceived world in that it is ontinuous, even when the lights go out or we close our eyes. We can find this out with physical instruments, and we know it from physical theory.

19.3 The relationship betweeen perceived and physical worlds

Now that we have carefully distinguished perceptual and perceived worlds, we can ask again how they relate to each other. N points need to be made.

19.3.1 The perceived world is impoverished by sensory limitations

First, the perceived world is *impoverished* by the limits of sensory processing. Any physical energy that is not transduced by our senses, or any physical stimuli that are below our thesholds, cannot affect the properties of the perceived world. You have seen many examples of information loss in Chapters 1-18. So for example, the electromagnetic spectrum is continuous over many orders of magnitude of wavelength (or energy), whereas the visible spectrum spans only a factor of less than two, between 400 and 700 nm. Lights that deliver too few quanta are undetectable; gratings of spatial frequencies above 60 cy/deg cannot be discriminated from homogeneous fields of light; and so on and so on. Much of the lesson of the earlier chapters of this book deals with the limitations of our senses with respect to the properties of the physical world.

19.3.2 The perceived world is enriched by qualia

Second, most philosophers and most vision scientists would argue that with respect to the physical world, the perceptual world is also enriched by – how to say it? – the unique properties of perception. We think of the physical property of wavelength composition as colorless, whereas the perceived color of an object has something important added. Philosophers refer to these new properties as perceptual qualities, or *qualia*. The concept of qualia as perceptual entities is elegantly (if poetically) captured in the following quotation:

...The mind, in apprehending, also experiences sensations which, properly speaking, are qualities of the mind alone. These sensations are projected by the mind so as to clothe appropriate bodies in external nature. Thus the bodies are perceived as with the qualities which in reality do not belong to them, qualities which in fact are purely offsprings of the mind. Thus nature gets credit which should in truth be reserved for ourselves; the rose for its scent; the nightingale for its song; and the sun for its radiance. The poets are entirely mistaken. They should address their lyrics to themselves, and should turn them into odes to self-congratulation on the excellency of the human mind. Nature is a dull affair, soundless, scentless, colorless, merely the hurying of material, endlessly, meaninglessly. (Whitehead (1925, p.54); quoted in Velmans, 2000, p. 112.)

A particularly nice metaphor for the relationship between the perceived and physical worlds has been presented recently by the philosopher Max Velmans. A modification of Velmans' diagram of his view, which he calls a *reflexive* model of perception, is shown in Figure xx. The idea is that the external object, here a cat, sends light to an observer's eyes. The observer's visual system captures the light, processes the incoming signal, and creates a series of neural representations of the cat.

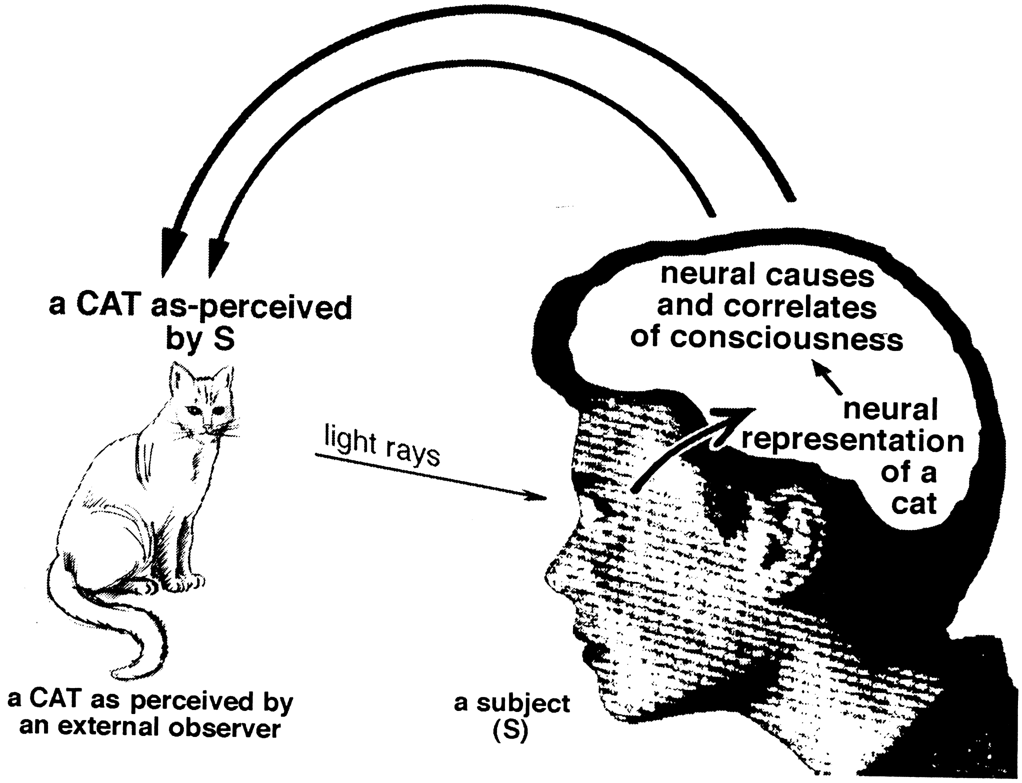

a CAT as-perceived by S
neural causes and correlates of consciousness
neural representation of a cat
light rays
a CAT as perceived by an external observer
a subject (S)

Figure 19.1: Velmans' reflexive model of perception.

At some level of neural processing, the neural representation gives rise to a perceived cat. The perceived cat, however, is not perceived to be located within the head of the observer, however, but rather at or near the location of the physical cat.

In summary, to use Whitehead's phrase, the perceived cat is "projected by the mind so as to clothe" the physical cat in perceptual properties. To unclothe and describe the physical cat, we use physical instruments and physical theory. To study the relation between the clothed and unclothed cats, we call on psychophysics.

19.3.3 The surprising veridicality of perception: Perceptual constancies

Third, once sensory processing and qualia are set aside, the striking thing about the perceived and physical worlds is how similar they are. That is, all things considered, perception is remarkably veridical: the properties of perception usually correspond closely to the properties of the physical world. Remarkably, perception carves nature at the joints, and informs us accurately about the physical properties of objects.

Why is this surprising? To capture the difficulty of the problem, perceptual psychologists classically make a distinction between *distal* and *proximal stimuli. Distal stimuli* are physical objects: states of the world and the objects in it. Houses, dogs, trees, and other real three dimensional objects are distal stimuli; so are parts or aspects of them, such as their surface properties, distances, and directions and speeds of motion. *Proximal stimuli* are the representations of stimuli as they occur in the retinal image.

The problem of creating veridical perception stems from the fact that many of the properties of objects in the the physical world are confounded in the retinal image. For example, the retinal image consists of a two-dimensional optical image, with no clear representation of the dimension of distance. The sizes of the images of objects vary with the distance of the physical object. Yet we perceive objects in a three-dimensional space, with sizes that remain constant over variations in distance. The remarkable fact of perception is that we are able to back calculate from the proximal stimulus to create perceptions that correlate closely with the properties of the distal stimulus,and that are sufficient to guide our actions and allow our survival in the physical world. In short, somehow, we have *size constancy.*

These ideas bring us to the important concept of *perceptual constancies.* A perceptual constancy is the ability to perceive the attributes of objects (distal stimuli) more or less accurately, despite the fact that the physical variables that carry information about these attributes are confounded with other physical variables in the retinal image (the proximal stimuli). If the function of perception is to tell us the characteritics of objects in our immediate environment, and allow us to respond to them wisely, then perceptual constancies are of fundamental importance.

Beyond size constancy, a second and less obvious example is *lightness constancy.* The lightness of an object is its perceived shade along the perceptual white-grey-black continuum. One of the physical characteristics of the surface of an object is its *reflectance* – the fraction of the incident light that it reflects. Lightness constancy is the tendency for objects with a constant reflectance to be perceived as a constant shade of white, grey, or black. Perceptual studies show that our lightness constancy is good. Objects that are perceived as white all turn out to reflect about 80% of the incident light; those perceived as black treflect about 4%; and those with reflectances in between are perceived as various shades of grey. That is, the reflectances of natural objects span a range of perhaps 20:1. So far, so good. But the problem is that the light incident on our retinas from white, grey and black objects is also influenced by the intensity of the light falling on them; and the incident light may vary by a factor of 1010 or more, totally swamping the relatively small difference caused by the diference in reflectances. How can we back-calculate from the intensity in the retinal image to the reflectances of objects? How can lightness constancy happen?

These and several other possible perceptual constancies – constancies of color, shape, and speed – are listed in Table 19.1. In each case, we list the constancy together with the property to which it corresponds in the distal stimulus, and the factors that are confounded in the proximal stimulus. All of the constancies illuminate the relationship beetween physical and perceptual worlds – they provide examples of the remarkable correspondence between the properties of the two Amazingly, when perception clothes physical objects with perceptual properties, the clothing corresponds more closely to the physical properties of the object than it does to the incoming sensory signal.

Th final constancy in Table 19.1 is *object constancy.* Object constancy is the capacity to perceive an object as retaining its characteristics – size, shape, color, and so on – despite variations in distance, angle of regard, spectrum of illumination, and so on. Object constancy is a combination of all of the other constancies, and is in the ultimate service of veridical perception as a whole. In later chapters, we will return to each of these constancies, and ask how they come about.

Kind of constancy	Object property	Distal (retinal) stimulus	Need to factor out	How?
brightness	reflectance	refl x illum	illum	?
color	spectral refl	s.r. x s. illum	s. illum	?
size	physical size	size x distance	distance	?
shape	physical shape	shape x angle	angle	?
speed	physical speed	speed x distance	distance	?
object	all of the above	all of the above	all of the above	?

Table 19.1: Column 1 denotes the constancy under discussion. Column 2 specifies the distal property to which the perceptual property corresponds if perception is veridical (e.g. perceived size corresponds to physical size; brightness corresponds to reflectance; etc.). Column 3 poses the problem: The object property is not represented directly in the retinal image, but is confounded with some other stimulus property (e.g. retinal image size = physical size/distance; retinal image intensity = object reflectance x illumination). Column 4 lists the confounding factor that needs to be factored out (e.g. distance or illumination). The question marks in column 5 symbolize the fact that the needed calculations are not straightforward, and require scientific discovery, both at the logical and calculational level and at the physiological level.

19.3.4 The sensory input is augmented by stored information

At this point, we add a major new concept to our treatment of vision science: the incoming sensory signal is not sufficient to specify the identities of the objects in the world in front of us. Rather, our perceptions arise from the combination of the incoming sensory signal with information stored in our brains. This added information has its origins in a combination of our evolutionary history and our developmental experiences. There is more to vision than meets the eye!

For example, who or what is depicted in the cartoon in Figure 19.2? The cartoon consists of only a few curved lines. Surely the incoming stimulus information is insufficient to specify the object. Yet many of us recognize it instantly as the profile of a face, and supply his or her image and identity from memory. This demonstration by itself should convince you that perception involves the addition of stored information. [Do you see any way around this conclusion?]

When we are operating in the real physical world (not in cartoon land), the incoming sensory signal usually carries much more information about the size, shape, color and location of an approaching object, but the *recognition* or *identification* of the object (Hi, Grandma!) and the awareness of many of its hidden properties (Where's my present?) depend on information retrieved from memory. Similarly, our actions toward physical objects are informed by our knowledge of the properties of the physical world, even when these properties are not included in the incoming sensory signal. We pet the kitten, but avoid the rattlesnake even if we have never been bitten.

Anothe dramatic case of the use of stored information arises from considering our perception in the peripheral visual field. As we have discussed in earlier chapters, our discrimination on many sensory dimensions – especially acuity and color – falls off dramatically with retinal eccentricity. If our perception of objects in the periphery were based solely on the incoming sensory information, our peripheral vision would of necessity be blurry and desaturated, and the perceived properties of an object would change as we shift our fixation and move the image of the object from fovea

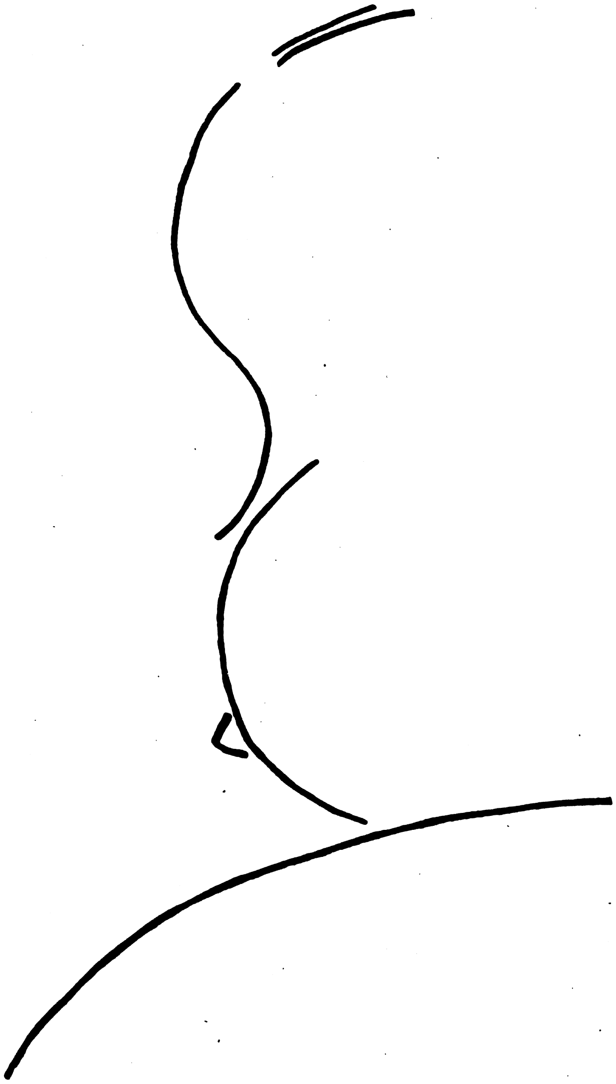


Figure 19.2: Cartoon of a famous person.

to periphery. But most people report that objects retain their properties across shifts in fixation. [Try it. Do you agree?]

A dramatic example arose while DT was writing the first draft of this chapter. DT and her friend Maureen Powers were writing at opposite ends of MP's dining room table. MP's parrot, Phred, was out of his cage, and DT noticed that he was walking around over MP's head and shoulders with a classic parrot gate. As DT concentrated her attention on the chapter, she was vaguely aware that Phred flew from MP's head to the screen door in DT's peripheral vision. Later, she became aware that Phred was walking up and down the screen. Without looking up, she could ”see” Phred's red tail and green body, and the little beady eye that had been eyeing her all week.

Now, had Phred been miraculously replaced by a dove, DT's perception would probably not have changed. That is, the perception of peripheral Phred was created out of nearly whole cloth by DT's perceptual systems! Those processes must contain some rules about the persistence of object properties and the temporal continuity of objects. In DT's perception, when Phred flew to the screen door he took his colors and identity with him, just as he does in reality. In sum, the details of our peripheral vision are created from memories of our foveal vision, plus some reduced peripheral version of the incoming sensory signals, plus some assumptions about temporal continuity.

Here's a second example: DT and her husband cohabit with a stuffed rabbit named Flopsy (it's a long story). One day as DT was working on this book, she became too hot, yanked off her grey sweatshirt, and dumped it on a table by her right hand. Soon she began to find thoughts of Flopsy intruding into her mind and distracting her attention. At first she could find no reason for this, but eventually tracked the problem. The grey sweatshirt had fallen into a pose that, in peripheral vision, looked exactly like Flopsy! Now, had DT's peripheral vision been as good as her foveal vision, she would never have confused Flopsy with the sweatshirt. But her peripheral vision is degraded, and the peripherally processed sweatshirt image was nearly identical to the peripherally processed Flopsy image – a peripheral Flopsy metamer. In fact, it had just the right properties to access memories of Flopsie, and top-down processing filled in her detailed appearance (not to mention her personality). A picture of Flopsy and her metamer are shown in Figure 19.3. By the way, the cartoon in Figure 19.2 is of Alfed Hitchcock.

To reiterate: the perceptual world contains more information than is supplied by the incoming sensory signals. Perception results from the combination of sensory signals with stored information.

19.4 Perception as a guessing game

The physical world is underdetermined in the retinal image. That is, as shown in our discussion of constancies above, many specific combinations of values physical variables are confounded to create equivalence classes in the retinal image, and at least at first glance there would seem to be no source of information that would allow us to sort them out. Retinal image size confounds physical size and physical distance, and many combinations of physical size and physical distance can yield the same retinal image size. Yet perceptually we sort them out, and see both size and distance veridically most of the time. Bumblebees can fly.

One conclusion we can draw immediately from these system properties is that there must be additional, non-obvious sources of information available, and used in generating our perceptions. These sources of information may be many, and arise from many origins, both from subtle patterns contained in the retinal image and from information stored within our brains via our evolutionary and developmental histories. Pragmatically, the EDC has doubtless been willing to glean the needed

Figure 19.3: Flopsy and her metamer. A: In focus. B: Blurred to simulate peripheral vision.

information any way it can – from any source or trick that works – preferably with precision, but with imprecision and guesswork when necessary.

What might these additional sources of information be? This inquiry leads us to a discussion that will introduce some important concepts: cues, heuristics, and the inevitably probabilistic nature of perception.

19.4.1 Cues

Some of the extra information required for veridical perception is contained in the incoming sensory signals in the form of *cues*. Cues are lawful or probabilistic regularities that occur in the mapping between aspects of the physical world and aspects of the incoming sensory signal. As such, they provide potential sources of information about that aspect of the physical world.

To continue with size and distance as examples, *distance cues* are sources of information regarding the physical distances of objects, and *size cues* are sources of information concerning the physical sizes of objects. What sources of information about distance and size, other than retinal image size, are available for use as cues?

One potential source of information is *accommodation* (see Figure xx) – the thickness of the lens needed to produce an optimally focussed retinal image. Another is *convergence* (see Figure xx) – the angle between the directions of gaze of the two eyes required to place the two images of a single object on the two foveas. Both of these quantities vary regularly with distance, and thus provide potential distance cues[1]. These and many other distance cues will be discussed more fully in Chapter 24.

[1] The size/distance example is chosen here because it is easy to grasp intuitively. However, it suffers from the complication that not all of the sources of information about distance come from the retinal image per se – accommodation and convergence cues come from the ocular musculature. In most of the cases we will encounter later, all of the cues will come from analysis of subtle patterns within the retinal image.

Cues are interesting, for several reasons. First, not all cues are obvious. Most potential cues are more or less subtle, like accommodation and convergence, and have to be discovered by intellectual analysis. Moreover, once a potential cue is discovered, one must still ask whether it is actually used by the visual system. One of the interesting aspects of the history of perception concerns the discovery of the many cues that make veridical perception possible, and the research required to find out whether each potential cue is actually used. [Try to figure out some other distance cues.]

Second, although some cues like accommodation and convergence are highly reliable, others are probabilistic and therefore more risky. Importantly, a cue can provide accurate information most but not all of the time, and still be useful. For example, consider the cue of *familiar size.* Suppose that objects have usual sizes – a house is about 30 feet high – and your perceptual system makes the assumption that most things have their usual sizes most of the time. So if I see a house on the far shore of a lake, and I see it as being the size that houses usually are, I will see it veridically most of the time. However, if it's a child's play house, the familiaar size cue will mislead me, and if I am using this cue my perception will be wrong.

Third, any given constancy is typically served not by just one cue, but by many. As we have already seen, size constancy is served by at least by accommodation, convergence, and 'usual size'; and there are many more cues to size and distance. The same is true for color constancy, lightness constancy, and so on. In studying cues, then, we will need to discover not only the various potential and actual cues, but also the rules by which they combine their effects.

And finally, many cues involve relatively sophisticated stored information and analyses. The familiar size cue is a good example. In order to use it, my visual system must identify the house as a house, and have access to stored information about the usual sizes of houses. As we will see in Chapter xx, many other distance cues rely partially on similar complex processing.

19.4.2 Heuristics: Perceptual guessing rules

When we recognize the existence of cues, the statement that the world is underdetemined in the retinal image can sometimes be seen as unduly pessimistic. For example, it is true that physical size and physical distance are confounded in retinal image size. But the signals that control accommodation and convergence contain information about distance. By using an analysis that combines these signals with retinal image size, it would seem that we actually have the information needed for the veridical perception of size, and the system property of size constancy suddenly seems less mysterious.

In other cases, however, the cues are less straightforward, and the deconfounding less than perfect. For example, the familiar size cue is not perfectly reliable – houses are bigger or smaller than the usual size on, let us say, 1% of occasions. Nonetheless, we would still be well advised to use this cue if we didn't have others available, and implement the perceptual rule, *perceive the house as being the size that houses usually are.* Using this rule would allow us to perceive the size of the house veridically on say, 99% of occasions. The EDC would unquestionably adopt this rule if it had no better cues, because using it is a lot better than nothing. It's analogous to intelligent betting in other circumstances: if you know a die contains sixes on five out of six of its faces, you will bet on sixes, even though you won't be right on every roll.

A *heuristic* is a stored interpretational rule – a perceptual *guessing rule* – that the perceiving organism applies to incomplete incoming sensory information. The function of the heuristic is to provide the best possible guess about the situation in the physical world, given imperfect or partial incoming sensory information. The rule, perceive this house as the size houses usually are, is an example of a heuristic. When the term heuristic is used, it is used to emphasize the probabilistic and therefore potentially fallible nature of the enterprise of perception.

The fact that perception is usually veridical, or nearly so, implies that the incoming sensory signal provides a set of cues that, taken together, provide considerable information about the important physical variables. It also implies that our perceptual systems contain a set of heuristics that, in combination with the incoming cues, allow our perceptual guesses to be veridical or nearly veridical most of the time. [Figure out the heuristics used with the cues of accommodation and convergence.]

Where do heuristics come from? The argument is that the cues supplied by the physical world and the perceptual heuristics of organisms that evolved in that world will be tied very closely to each other. Through some combination of genes and experiences, the EDC builds into us perceptual heuristics that are complementary to the available cues. We have particular guessing rules not because of some arbitrary rule-making on the part of the EDC, but because the guessiing rules make use of the available cues, and they evolved to work.

One of the major goals of modern perceptual research is to discover potential cues provided by the physical world through our sensory systems, to see which ones are actually in use; to discover the complementary heuristics; and to find out how the heuristics combine to yield (usually) accurate perceptions. We will develop many other examples in later chapters.

19.4.3 Perception as hypothesis

The set of ideas presented above – the ideas of cues, perceptual heurisics, and fallibility, leads to the idea that all of our perceptions can be more properly viewed as *hypotheses* than as factual knowledge about the state of the physical world. The term *perceptual hypothesis*, like the term heuristic, is chosen to emphasize the fallibility of perception.

Once again, why are the concepts of cues, heuristics and perceptual hypotheses so appealing to scientists interested in perception? Up to this point we have probably thought of sensory processing as a deterministic process (ignoring the question of noise). In this context, the terms heuristic and perceptual hypothesis are important and carry novel meanings because both carry explicit or implicit connotations of guesswork and fallibility. The perceived world is a construction that mimics the properties of the physical world as closely as it can; but it is constructed partly on guesswork, and when insufficient information is available, it can be wrong.

Where is the basic source of unreliability? Note that both the sensory system, the heuristic and the perceptual hypothesis can all be deterministic; the basic source of the unreliability is in the incoming sensory signal. Sometimes the information we need just isn't there. It's nobody's fault. Overall, it's amazing that we do as well as we do, given the apparent confounding of variables in the retinal image.

19.4.4 Perceptual errors and illusions

At other times, of course, our perceptual guesses are wrong. Momentary wrong guesses lead to momentary errors of percepton that most of us have experienced. Something that appears to be a nun in a habit turns out to be a black and white signboard; white powder spilled on a coat turns out to be a ray of sunlight, and a deer in a red hat turns out to be a fellow hunter. Perceptual

Figure 19.4: Visual illusions.

errors doubtless contribute to many traffic accidents – a road that appears to be straight actually curves away to the left, and passing can lead to tragedy.

The view that heuristics can lead us astray also provides a comfortable explanation for the existence of the classical visual illusions (Figure 19.4). The argument here is that our heuristics evolved to process scenes from the usual physical environment, and can be misapplied to line drawings. In a similar vein, it can be argued that the specifics of illusions and misperceptions can provide clues about which heuristics our perceptual systems are using.

19.5 The third entity: Neural activity and neural codes

Let's turn now to the third entity: neural activity and neural codes. In earlier chapters, we have already introduced many of the concepts related to neural coding. Just like retinal neurons, cortical neurons have firing rates, and different stimulus conditions lead to different firing rates – each cortical neuron carries close-coupled information about many features and dimensions of the retinal image at the location of its receptive field. Moreover, different cortical neurons respond to different combinations of features and dimensions, and in general information about any one specific state of the physical world is carried in an ensemble code. Moreover, recodings continue; features that are explicit in the firing rates of individual neurons at one level, can be carried by a population code at the next.

19.6 Neural states, perceptual states, and linking propositions

We have argued throughout this book that the system properties defined by psychophysics place important (although usually general) constraints on models of neural activity. We now argue that the properties of perception, such as the constancies defined above, provide a new set of system properties for visual science. What constraints do these new system properties place on our understanding of the neural codings and recodings that might or must exist within the visual cortex? Given the properties of perception, what can we deduce, or assume, or guess, or hypothesize about high level neural codes? Can perceptual phenomena give us any guidance as to what to look for when we study neural processing in the cortex?

As before, DT will argue that the answer to this question is yes, but that such arguments always involve linking propositions – assumptions about the mapping rules between neural states and perceptual states. If we are reasoning from perceptual states to neural states, we are always using linking propositions. This realization allows us to bring our earlier discussions of linking propositions forward and apply them in the new domain. What linking propositions do we use when we use the properties of perception to glean hints as to the nature of high-level neural codes?

19.6.1 A general linking proposition: Appropriate neural analyses and codes must exist

Some new general, almost tautological, linking propositions can be stated at the outset. At the most basic level, most vision scientists are committed to the belief that perceptual states come about only in conjunction with brain states; that is, that brain states are a necessary condition for the existence of the perceptual world. Moreover, computational algorithms that create the properties of perception – circuits that analyse for particular cues and combine them to create cue invariance; circuits that execute perceptual heuristics; circuits that provide access to memory, and routes for top-down influences on perception – must exist within the visual system. Similarly, high level neurons and codes that provide the immediate neural basis for conscious perception must exist within the visual system. Shades and nuances of all of these linking propositions can be found, implicit or explicit, in the modern literature on cortical processing and perception.

Can we say anything more specific? In particular, do we wish to assume that any more specific forms of isomorphism hold between perceptual states and neural states? Do we wish to assume that the neural representations on which perception is based will bear any specific form of resemblance to the properties or elements of perception?

The modern cortical visual neuroscientist needs to be aware of which variants of linking propositions enter into his assumption structure. Let's look at some examples.

19.6.2 The "same information content" and "Many:1" propositions

Of the three entities – physical states, perceptual states, and neural states – which two have the closest and most necessary resemblance to each other? It can be argued that the relationship between perceptual states and neural states must be particularly close. The perceptual world differs from the physical world – it is impoverished by sensory processing, and enriched by stored information. Neural codes differ from the physical world for the same reasons. But if the perceptual world arises solely from neural activity, the two must be in intimate correspondence. How can we capture the interdependence?

The philosopher Max Velmans (19xx) has argued that *the physiological representation and the corresponding perceptual representation must have the same information content.* Most vision scientists would probably agree with this premise. After all, unless conscious states have an existence independent of brain states, where else would the extra perceptual information come from?

A variant of this idea, however, is that the information content of the two states need not be identical, but that neural states cannot have less information than their corresponding perceptual states. If perceptual states arise solely from neural states, then it follows that all of the information included in a perceptual state must be encoded in the corresponding neural state. That is, any information represented in visual perception must also be represented within the visual system. On the other hand, it seems possible that there could be more states of the visual system than there are states of perception. That is, *the mappings from neural states to perceptual states could be 1:1 or many:1, but not 1:many.*

If either the above linking propositions is accepted, other more specific propositions follow. In particular, if stored information is added to the incoming sensory information to create the neural state that underlies a perceptual state, that stored information must be contained in the neural state that underlies the perceptual state. That is, the neural code for perception must include the information added by top-down processing. A similar argument holds for cue invariance and for the operation of heuristics. And on this view, the vividness and detail of peripheral vision implies that there will be a representation of the peripheral visual field in which more detail is expressed than one can find in the incoming sensory signal that originates from the peripheral retina.

19.6.3 Cue analysis and cue combination circuits

Some of the extra information required for veridical perception is contained in the incoming sensory signals in the form of *cues*. Cues are lawful or probabilistic regularities that occur in the mapping between aspects of the physical world and aspects of the retinal image. As such, they provide potential sources of information about that aspect of the physical world.

We have argued above that veridical perception depends heavily on cues – lawful or probabilistic regularities that occur in the mapping between aspects of the physical world and aspects of the incoming sensory signal. It is the analysis of cues that allows us as perceivers to deconfound certain variabiles that are confounded in the retinal image, and create veridical perception.

To return to our example of the perception of the distances and sizes of objects: we already know of the cues of accommodation, convergence, and familiar size. But each of these cues is subtle, and each must be analysed in turn, before the information it carries can be made explicit. A commn linking proposition in modern visual neuroscience is that *cue analysis circuits will be found within the visual system.*

19.6.4 The veridicality proposition and neural equivalence classes

Although not as tight as the mapping between perceptual states and neural states, the mapping between the physical world and the perceptual world is also remarkably close. That is, as argued above, the perceptual world bears a closer resemblance to the physical world than it does to the retinal image or the early sensory codes. Therefore, another very appealng linking proposition is that *the neural representations on which perception is based will correspond to the properties of the physical world as represented in perception.* That is, these neural codes will have a greater resemblance to the properties of the physical world than they do to the properties of the retinal image, or to the early sensory codes. There will be neural algorithms that deconfound the physical dimensions confounded in the retinal image, so that high level visual codes can bear a close correspondence to the properties of the physical world.

The perceptual phenomenon of cue invariance provides an example. The concept of cue invariance refers to the fact that many different cues can bring about the same perception. For example, the contours of an object – say, a square – can be defined by light/dark boundaries, or chromatic boundaries, or changes in texture, or in a variety of other ways; yet the square is still perceived. Therefore, the veridicality proposition suggests that the neurons that provide the neural representation of the square will also be cue-invariant; they will respond either to light-dark edges, or to chromatic edges, or to texture-defined edges, or to a variety of other edge cues. Similarly, neurons that provide the neural representation of distance should respond to many different distance cues.

The general veridicality of perception provides a similar example. Since we have size constancy and object constancy, we might hypothesize that the neurons that provide the neural basis of size and object perception will provide signals that are invariant over distance and viewing angle. Since we have veridical perception of the speed and direction of motion of an object, despite the confoundings of speed and distance in the retinal image, we might hypothesize that there should be neurons whose signals correspond better to real object motion than to retinal image motion. And so on. All of these hypotheses are instances of the use of the veridicality proposition.

A more specific veriant of the veridicality proposition can be stated in terms of the predicted equivalence classes for cortical neurons. If the veridicality proposition is correct, we might expect to find some neurons that respond equally to all distance cues, and other neurons that respond equally to an object regardless of its distance and retinal location. The same object at different locations could be coded by a fixed object signal but a different location signal.

The veridicality proposition underlies many studies in modern visual neurobiology. For example, van der Heydt and Peterhans (Chapter xx) have studied the neural basis of form perception by looking for cue-invariant cortical neurons that respond to both real and illusory contours. Albright (Chapter xx) has looked for cue invariance for motion cues ...;

van Essen's perspective on parallel processing is based on the idea that cues carried by many different aspects of the sensory input signal will be combined to yield more cue-invariant representations of perceptual variables at ligher cortical levels.

Will there be neural representations that correspond to the properties of the perceived world? How would we recognise them if we had our electrodes on them? – by their cue invariances and constancy properties.

19.6.5 Specific isomorphisms and direct coding

Can we go further? Some theorists have suggested that we can anticipate the kinds of correspondences that will be seen between individual neurons or small populations of neurons, and our perceptions. For example, Horace Barlow (see below) has argued for what he calls *direct coding*: there will be a simple and recognisable isomorphism between the properties of high level cortical neurons and our perceptions of objects.

Barlow's proposition is that "the active high-level neurons directly and simply cause the elements of our perceptions" (1972, p. 381).

Can we make this proposal more specific? Suppose that we had done a careful, exhaustive set of perceptual experiments, and could make the case that our perception of objects is based on the perception of combinations of a particular limited set of *parts*. (This idea will be developed in more detail in the chapter on form perception). Barlow's direct coding proposition would then be, the individual neurons in the set of neurons that map to form perception will each be selectively tuned to one of those parts, and the object will be represented by activity in the combination of the neurons that code its parts.

Note that this assumption is not a logical necessity. The neural code that underlies conscious perception could be complex and unrecognizable, and bear no explicit resemblance to the perceptually defined parts of objects. In that case, the mappings from neural states to perceptual states would be complex rather than simple – the neural/perceptual mappings would do a lot of the work of perception. But we neuroscientists would be infinitely gratified – and maybe not even particularly surprised – to find a simple correspondence between the elements of perception and the elements of neural codes. Maybe the EDC used common sense.

19.6.6 Parallel processing and multiple codes

Up until this point, we have been working under the implicit premise that the visual system has only a single, ”final” code to produce – the code that underlies conscious visual perception. However, our exploration of the concept of parallel processing has suggested that the visual systemmight be called onto produce two or more relatively separate ”final’ codes, to serve different purposes. The ventral stream may produce one code inthe form required to accesses memory, allows object recognition, and provide the neural substrate of conscious visual perception. But the dorsal stream may produce another code, in the form required for motor responses and coordinated action. In short, parallel processing suggests there may be no single ”final” code,but rather two or more codes specialized for different purposes.

19.6.7 The ”sufficient therefore causal” proposition

Finally, there is a very common linking proposition that sneaks into arguments in visual neuroscience, and is worth a cautionary note. The visual system contains examples of neurons that are tuned on particular stimulus dimensions, or to a particular set of stimulus properties – for example, V1 neurons that are tuned for orientation, and respond best to vertical lines. Given these neurons, it is tempting to talk as though this population of neurons codes that stimulus property: that neurons that are tuned to vertical lines are ”vertical line detectors”, and signal the presence of vertical lines. The argument is, neuron X carries the information required to do function F, therefore it ”does” function F.

Stated so starkly, it is easy to see the potential fallacies of such arguments. It would be good to be sure that function F is clearly defined. It would be good to find out whether other neurons in other places also have the right properties to ”do” function F. It would be good to make a quantitative model to make sure the information needed to ”do” function F is actually present, and that there is a feasible neural processing algorithm to go the next step. And it would be good to decide what it means to say that a particular set of neurons ”does” function F in the first place.

19.7 Principles of neural coding

In the previous section we discussed the kinds of neural codes we expect to find at central levels of the visual system, at an abstract and logical level, based on the properties of perception in combination with particular linking propositions. The final question of this chapter is, leaving perception aside for the moment, what do we know about neural codes at the level of single unit recording? What forms do we expect neural codes to take in the cortex, as they recode incoming sensory information, incorporate heuristics, and eventually form the neural basis of our perceptions and actions? Is one kind of coding better than another at a particular processing stage, or for a particular purpose? Questions of this kind are currently coming into prominence in the field of computational neuroscience. And, of course, the EDC is pleased, as they have been pondering these questions all along.

Neural codes can be classified on several different dimensions. A few of them are as follows. Some of them overlap. We begin with a principle that applies to single neurons, proceed through those that apply to populations of neurons, and end with special coding distinctions developed to treat the mapping of neural activity to conscious perception.

All-or-none vs. graded.

This distinction applies to the code carried by a single neuron. In some contexts we discuss a single neuron as though it is either firing or not firing; that is, that it has a 'trigger feature", and that its response is all-or-none. At other times, we think of single neurons as being capable of firing at many different rates, that is, as having a graded response. For a fixed population of neurons, more information can be carried by graded responses than by all-or-none responses.

[Lennie (or someone) argues that there are cases in which "the dynamic range of the neuron is not used".]

Redundant vs. independent.

A redundant code is one in which the same information is carried by more than a single neuron, so that the firing rates of a set of neurons are correlated and interpredictable. A fully non-redundant code is one in which each neuron's activity level is independent of the activity level of all of the other neurons. For a set of neurons of a given size, more information can be carried by a non-redundant code than by a redundant one.

Some of the early recodings of the visual system can be thought of as reducing the redundancy of the spatial code. For example, a typical retinal image contains spatial redundancy in the form of large regions of homogeneous illumination. The photoreceptors create a spatially redundant neural code from the spatially redundant retina image. But the center/surround properties of the retinal neurons tend to cancel out the signals from these large regions, and create a less redundant code, in which the only neurons that are active are those whose receptive fields fall in regions of high local contrast, such as edges. A redundant code in which many elements are active has been recoded into a less redundant code with fewer active neurons, each of which therefore carries more information about the image (see Figure xx).

Similarly, Bucksbaum and Gottschalk (Chapter xx) have argued that redundancy is reduced in going from a cone-based code, in which the signals in the L and M cones are highly correlated, to a postreceptoral code that contains opponent channels.

Some theorists have argued that one of thetasks of the early stages of the visual system is to reduce redundancy in the neural code.

Coarse vs. (dense ?? xx).

This distinction applies to the number of different kinds of neurons that are in use to code all of the possible values along a stimulus dimension. If relatively few neurons are used, the code is coarse; if many are used, the code is dense. For example, at the level of the photoreceptors, wavelength information is carried by the signals from only three classes of cones – a coarse code. Two-dimensional spatial location, on the other hand, is carried by the activity of a subset of 100 million photoreceptors – a dense code. At the level of V1, orientation is carried by a set of neurons tuned to different orientations, but there is as yet no agreement on the number of different orientations, and thus no agreement on the coarseness vs. denseness of the code.

Advantages and disadvantages of coarse vs. dense codes?

Sparse vs. (Compact ?? xx).

This distinction and the next apply to high-level neurons that are used to represent a particular object or scene in our perceptions. Suppose that there is a population of N neurons whose activities map to the perceptions objects or scenes. The question is, how many of these neurons, K, or what fraction of them, K/N, are involved in representing any single, particular scene? If only a small proportion of the neurons are active – K/N is small – the code is *sparse.* In the extreme of sparseness – if K were 1 for each object – we would be back to our old friends the grandmother cell and the yellow volkswagen detector.

On the other hand, if many or most of the neurons are active in representing each object or scene – K/N is large – the code is *compact.* Moderately compact codes are often referred to as *pattern, population*, or *ensemble codes.* In the extreme, the firing rates of all N neurons could participate in the coding of every object or scene. In this case the number of codes available is N times the number of distinct firing levels for each neuron. This extreme is called a *factorial code.*

Direct vs. (Arbitrary?? xx).

This distinction applies to the question, discussed above, of whether or not there is a recognisable isomorphism between the neurons that form the neural correlate of perception, and the elements of perception. If the trigger features of the neurons correspond closely to the elements that make up our perception of objects and scenes, the code is *direct.* If there is no recognisable correspondence, the code is *arbitrary.* Of course, the sparser the code – the fewer the neurons that participate in it – the more direct it is likely to be. And again, the extreme of a direct code would be a code in which the trigger features of high-level neurons correspond directly to the perception of whole objects – grandmother cells again.

Efficient vs. inefficient.

Finally, an *efficient* code is one that carries a maximum amount of information under the circumstances in which it is used. That is, in an efficient code the responses of each neuron should be graded, so that there can be many different firing rates for each neuron N in the population; and the individual neurons should be independent, so that each of the possible combinations of firing rates of the N neurons can have a distinctive meaning. Putting these two dimensions together, a factorial code is a maximally efficient code.

In addition, the modern concept of efficiency explicitly includes the idea that the code is matched to the characteristics of the environment in which it is used. That is, different codes will be maximally efficient in different environments. Trivially, if a fish lives in deep water, to which only a narrow band of wavelengths penetrates, it would be inefficient for it to have more than a single kind of photoreceptor. Less trivially, if humans were to live in an environment dominated by low spatial frequencies at vertical orientations, one would expect to find a high fraction of cortical neurons tuned to respond to these features. This notion of efficiency is closely tied to environmental and developmental concepts: the neural code evolves over generations, and develops over a lifespan, to be statistically optimal for the organism in the environment in which it lives.

Modern studies of coding efficiency are concerned with quantifying the statistical properties of natural environments, and with deriving computational models that show that the code the individual is thought to have is optimally efficient given the statistical properties of the environment. For example, some authors have argued for spatial frequency channels with particular bandwidths, on the grounds that these channels will optimize the efficiency of coding for the spatial frequency statistics of natural environments.

Advantages of sparse vs. compact codes? Olshausen and Field – let the visual system design itself in the context of natural scenes and an instruction for sparse coding, and it designs neurons with receptive field properties that look like thoseseen in V1. –an argument that the EDC uses sparse codes. [But I still don't see why xx].

In summary, vision scientists are increasingly looking to design principles – principles of neural coding such as these – to predict and/or justify the particular neural codes that are observed within the visual system. In particular, the property of efficiency is receiving much attention. Quantitative computational models increasingly incorporate design principles such as these.

19.7.1 Barlow's Neuron Doctrine

In 1972, Horace Barlow wrote a seminal paper entitled: "Single units and sensation: A neuron doctrine for perception". In this paper (pp. 380-381), Barlow laid out five dogmas which he argued underlay the then-current studies of single neurons, and the theories of neural coding implicit in them. As in the present discussion,the goal was to make these dogmas explicit, in order to exampine them. Barlow's neuron doctrine and its dogmas are still relevant today, and they provide some practice with the concepts of neural coding.

Barlow's first dogma was: "A description of that activity of a single nerve cell which is transmitted to and influences other nerve cells....is a complete enough description for functional understanding of the nervous system....." This claim is the neuron doctrine itself – understand individual neurons and their interactions, and you will understand the neural bases for perception and behavior.

Barlow's second dogma was, "At progressively higher levels in sensory pathways information about the stimulus is carried by progressively fewer active neurons. The sensory system is organized to achieve as complete a representation as possible with the minimum number of active neurons." Barlow argued that at higher and higher levels of the visual system, the number of available neurons N would get larger and larger, and the fraction K/N that participate in the code for an object would get smaller and smaller. In the extreme, of course, this line of argument leads to grandmother cells, but Barlow did not take the argument to this extreme. He argued instead for population coding at lower levels, but sparse coding in a large population of neurons at the highest levels of the visual system.

The third dogma was ”Trigger features of neurons are matched to the ... features of sensory stimulation in order to achieve greater completeness and economy of representation. This selective responsiveness is determined by the sensory stimulation to which neurons have been exposed, as well as by genetic factors operating during development.” In other words, Barlow argued for a minimally redundant and an efficient code – a code matched to the environment in which the organism evolved and developed.

The fourth dogma was ”...the active high level neurons directly and simply cause the elements of our perceptions”. In other words, Barlow argued for direct coding – a simple and recognizable isomorphism between neural activity and perception.

And finally, the fifth dogma addressed the question, what variable corresponds to impulse frequency in a high level sensory neuron? Barlow's answer was: ”The frequency of neural impulses codes subjective certainty: a high impulse frequency in a given neuron correespoonds to a high degree of confidence that the cause of the percept [i.e. the trigger feature of the neuron – DT'] is present in the external world....” Interestingly, Barlow acknowledged the presence of graded responses, but did not want to concede that the many different firing rates of a single neuron could have a wide range of different meanings in a pattern code. By using different firing rates only for the purpose of coding the relative certainty of the same object, he could make graded responses consistent with the fourth dogma – that the active high level neurons directly and simply cause the elements of our perceptions.

19.8 Summary: Bumblebees can fly

In this chapter we have explored the relationships between three entities: the perceptual world, the physical world, and the neural codes that intervene between perceptual and physical worlds.

Everyday experience tells us that our perception is almost always veridical: we can perceive the sizes, shapes, colors, and brightnesses of objects with remarkable accuracy, recognize objects, fill in the details of peripheral perception, and respond apropriately to the physical world on amoment-to-moment basis. By a bumblebees can fly argument, we know that there must be neural computations that allow these remarkable achievements to occur. In fact, we are very good at these perceptual tasks, and they are achieved automatically and effortlessly. For this reason it is sometimes hard to overcome our naive acceptance of these accomplishments, and stop taking them for granted, in order to reach the understanding that such perceptual capacities are remarkable and problematic. Such perceptual problems will be explored in the remainder of this book.

A major principle that enters into an understanding of perception is that the incoming sensory signals are combined with stored information – perceptual heuristics and information from memory – to create veridical perceptions of objects in the world. The physical world is to some degree underdetermined in the retinal image, and physical variables are often confounded. But there are cues (often subtle ones) in the image that, if you know how to intepret them, provide clues as to the characteristics of the physical world. The perceptual system must have stored knowledge of these regularities in the form of perceptual heuristics, and use them in combination with the incoming cues to create largely veridical perceptions of the physical world. These ideas are deep ones, and they will become more clear through examples in the subsequent chapters.

We also argued that the system properties of perception provide the basis of many of the visual scientist's hypotheses and speculations concerning the neural codes that underlie perception. Many of these hypotheses and speculations involve linking propositions, and analysis of them allows us to carry forward ideas we have developed in earlier parts of the book. For example, the proposition that perceptual states arise from neural states is probably regarded as certain and even tautological today, whereas Barlow's proposition that "the active high-level neurons directly and simply cause the elements of our perceptions" – that a sparse code with simple isomorphisms will provide the neural substrate of perception – remains much more speculative.

In any case, it is important to be clear about the assumptions that underlie one's beliefs about the neural correlates of perception. Why? Because the current assumption structure may work for a while, and assist us in revealing some of the properties of central visual codes. At the same time, it may limit the hypotheses we generate, and lead us to a narrower-than-optimal range of hypotheses and experiments. When the assumption structure is explicit, we can more readily ask what the alternatives are, and think outside the box.

Finally, we examined some of the dimensions – sparseness vs. compactness, directness vs. arbitrariness – on which neural codes can differ, and used Barlow's neuron doctrine as an exercise in using these new concepts of neural coding.

In the following chapters, we will examine several different aspects of perception – depth and distance, brightness and color, time and motion, and form and object perception – in more detail. In each case, we will address the psychophysical and perceptual properties of the phenomenon, together with whatever is known about the perceptual heuristics and the physiological processing that make the perceptual properties possible.

Chapter 20

Image Segmentation and Grouping

The phenomena known as *image segmentation and grouping* pose some of the most classic puzzles in the field of visual perception. The problem is as follows. The optics of the eye create only a two-dimensional retinal image, made only of quanta, with no hint of unification of its parts, nor formal boundaries between them. The photoreceptor image is similarly two-dimensional, but pointillistic, with nothing gluing together any regional subsets of photoreceptor outputs. And retinal ganglion cells make a coarser but still regionally isolated representation of the visual scene, still with no regional glue.

Yet we do not perceive a smear of light nor a set of tiny isolated points nor a set of isolated blobs. The perceptual world, in good correspondence to the physical world, is occupied by objects. Each object coheres as a perceptual unit, separated from other objects and from the spaces between objects. It follows that processes somewhere within the visual system must take a first pass at dividing up, or *segmenting*, the neural image into parts – call them candidate objects or proto-objects – that are initially processed as units. These candidate objects will be reprocessed and recoded, eventually to become the representations of perceived objects.

The complementary process to that of image segmentation is perceptual *grouping*. That is, individual objects in the visual scene do not necessarily give rise to spatially continuous regions in the retinal image. In particular, nearer objects stand in front of, and *occlude*, parts of the images of farther objects. In consequence, the retinal image of the partially occluded object is broken up into separated parts. Yet remarkably, we usually see occluded objects as complete and whole, and may not even notice that major fractions of their images are completely missing. It follows that higher levels of the visual system must group the separated parts together, and fill in guesses about the missing parts.

As usual, new system properties lead us to two kinds of questions. The perceptual question is, descriptively, how does image segmentation occur? By what rules and heuristics is the initial two-dimensional array of quantum catches divided into sensible , or candidate objects, for further processing? What defines and bounds the sub-regions? And similarly, by what rules and heuristics are the separate sub-regions that arise from multi-part or occluded objects, joined together? How does a set of separate sub-regions in the image get grouped together to become one perceived object, while another set of sub-regions is assigned to another object? And how do these processes lead so often to veridical perception?

We can similarly define the question of neurophysiological instantiations. By the Universal Linking Proposition, we know that all perceptual states and processes must be instantiated in neurophysiological states and processes. What is the nature of the code changes that take place when neural images are segmented into parts, and when subsets of parts are grouped together? Where do these code changes occur? And what are the linking propositions that are involved in the search for answers to these questions?

Image segmentation is a topic with very old roots in perceptual psychology. At the same time, it is currently a very active field of perceptual and neurophysiological investigation, and several lines of research are very suggestive as to the locations and properties of the physiological instantiations of at least some segmentation processes. We will take the occasion to review this research in some detail, because it reveals some very interesting examples of linking propositions that have been adopted by the various investigators.

20.1 Classical system properties: Image segmentation and grouping

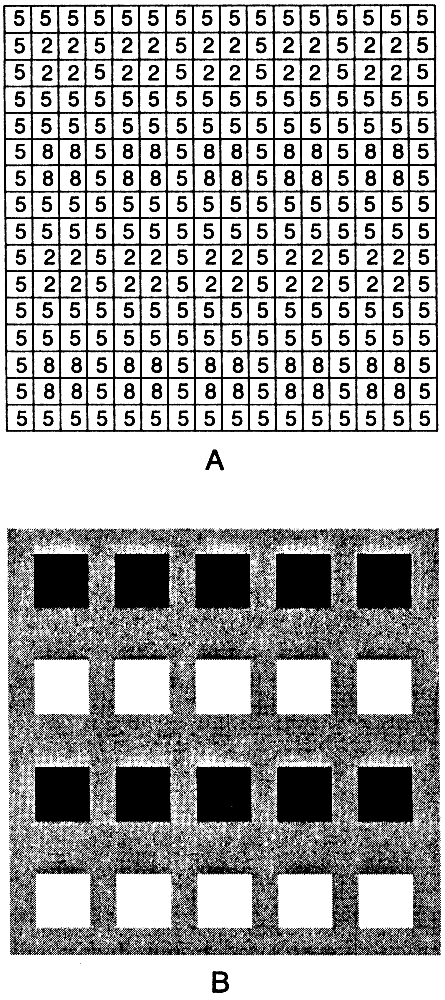
A striking illustration of the phenomenon of perceptual organization is shown in Figure 20.1 (Palmer, 1999). On the top is matrix of numbers. Your assignment is to discover the structure in the matrix. This probably takes some time and effort, if you succeed in doing it at all. On the bottom is a greyscale image that carries the same two-dimensional pattern as the matrix. You immediately perceive the greyscale image as sets of black and white squares in front of a grey background, and group the squares into four horizontal rows. Why is the discovery of structure so hard in the first case and so easy in the second? The answer must be that your visual system has special processing routines for processing greyscale images that it doesn't have for matrices of numbers. Your visual system automatically divides up, or *segments*, the incoming signals from the greyscale image into coherent regions, and *groups* subsets of these regions together for further processing. But how? And why the greyscale image but not the matrix of numbers?

Historically, the problem of image segmentation and grouping has been considered as a two-dimensional problem – the parsing of the retinal image. Also historically, the major source of information about image segmentation and grouping has been phenomenological – illustrations of segmentation and grouping phenomena on the two-dimensional pages of textbooks. We will begin with the classical approach and a set of classical demonstrations, and return later to more modern reinterpretations of these phenomena.

20.1.1 Region formation

As you look at the scene in front of you, try to imagine the retinal image that it makes. This is difficult because your visual system has already applied segmentation and grouping processes, but try to undo them and analyze your retinal image. From this retinal image you need to deduce the locations and identities of the physical objects in your neighborhood. How might you begin to divide up the image for further processing?

Probably you will notice that some regions of the scene are homogeneous or nearly homogeneous in lightness and color, and you may decide to treat these regions as units for further analysis (as in Figure 20.1). This is a good move, because many physical objects have uniform surface reflectance properties – they are uniform in overall reflectance and in spectral reflectance – and a homogeneous region of the retinal image will usually arise from a single object with such uniform



5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Figure 20.1: Image segmentation and grouping. A. A matrix of numbers does not organize itself into perceptual subunits, although if you study it you will find a spatial pattern. B. The same spatial pattern presented in shades of grey. Your visual system automatically *segments* the greyscale pattern into black and white squares against a grey background, and *groups* similar squares into rows. [After Palmer, 1999, Fig. 6.1.1, p. 256].

surface properties. Thus, the homogeneity or near homogeneity of a region of the retinal image provides a useful cuefor setting up candidate objects for further processing.

The homogeneity cue, however, quickly runs into limitations. Further perceptual analysis will show that regions that you perceived as homogeneous often are not actually homogeneous in the retinal image. For example, a three-dimensional object will be differently illuminated on its different surfaces, with the result that its retinal image will not be homogeneous. Upon analysis, the seat of DT's orange upholstered chair looks muchdifferent than the back because of its angle to the light coming in through the window, and the shadow of one of the bars of the window interrupts the homogeneity as well. A strict parsing on the basis of homogeneity would render the chair in several subparts – but it's a start. Retinal image homogeneity is probably a useful cue, but obviously other cues are needed.

Higher-order homogeneities such as *surface texture* can provide additional cues. That is, the surface of an object need not be of uniform luminance or wavelength composition to be uniform; it can be covered with a uniform pattern, or texture. In that case, the boundaries between objects are represented in the retinal image by changes in texture rather than (or in addition to) changes in luminance or wavelength composition. Some classical demonstrations of perceptual segmentation on the basis of texture are given in Figure 20.2.

But notice that the analysis of texture itself is a complex computational problem. To segment an image on the basis of texture differences, the visual system needs to analyze the statistical properties of the elements that make up the texture, and segment the image along loci at which one set of statistical properties gives way to another. In fact, the demonstrations in Figure 20.2 show that some changes in texture elements and their arrangement yield perceptual segmentation whereas others do not.

But texture segmentation gets us way ahead of our story. Let's return to the simplest case – a set of homogeneous patches of light in the retinal image. Let's decide that each such region is a unit for further perceptual processing. What's next?

20.1.2 Figure and ground, border ownership, and depth order assignment

A second classical perceptual phenomenon, related to region formation, is the perceptual separation of visual patterns into *figure* and *ground*. The claim, illustrated by demonstrations like those in Figure 20.3, is that perceptually, visual scenes spontaneously divide themselves into regions – figure and ground – with unequal perceptual properties of at least two kinds. First, it is claimed that the segmentation into figure and ground is by its very nature an ordering in depth, with the figure always appearing to be closer than the ground and occluding part of it. And second, figure and ground are separated by a common contour, but the claim is that they relate to the contour in different ways. The part of the scene that is perceived as the figure also appears to be bounded by, or to ”own” the contour – the contour bounds the figure and defines its shape. In contrast, the ground is not perceptually bounded by the contour. Its shape is indefinite, but it but continues behind the figure.

Which region will be seen as figure, and which as ground? Many stimulus factors affect figure/ground assignment. The most important is the closedness of the boundary between the two parts of the scene – the region with a closed contour will tend to be seen as the figure, as shown in Figure 20.3A. Other important factors are size, symmetry, and convexity – small, symmetrical, convex regions are more likely to be seen as figures, and large, asymmetrical, and irregular regions

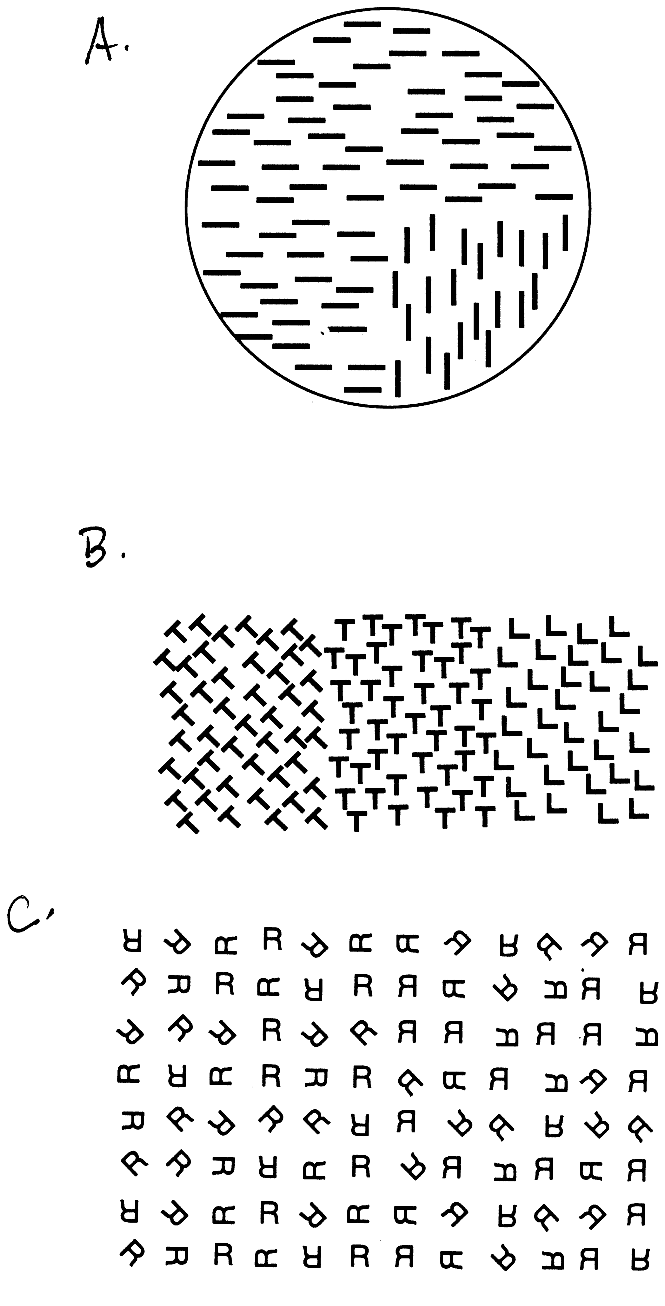


Figure 20.2: Segmentation by texture. Regions with common texture elements tend to cohere, and segment themselves from each other. A. Segmentation by common orientation. B. Segmentation by shape and orientation. Notice that the common orientation of line segments is more important than commonality of overall configuration. C. Some pattern differences, like these Rs and reversed Rs, do not support perceptual segmentation. [After Palmer, 1999, pp 276-277.]

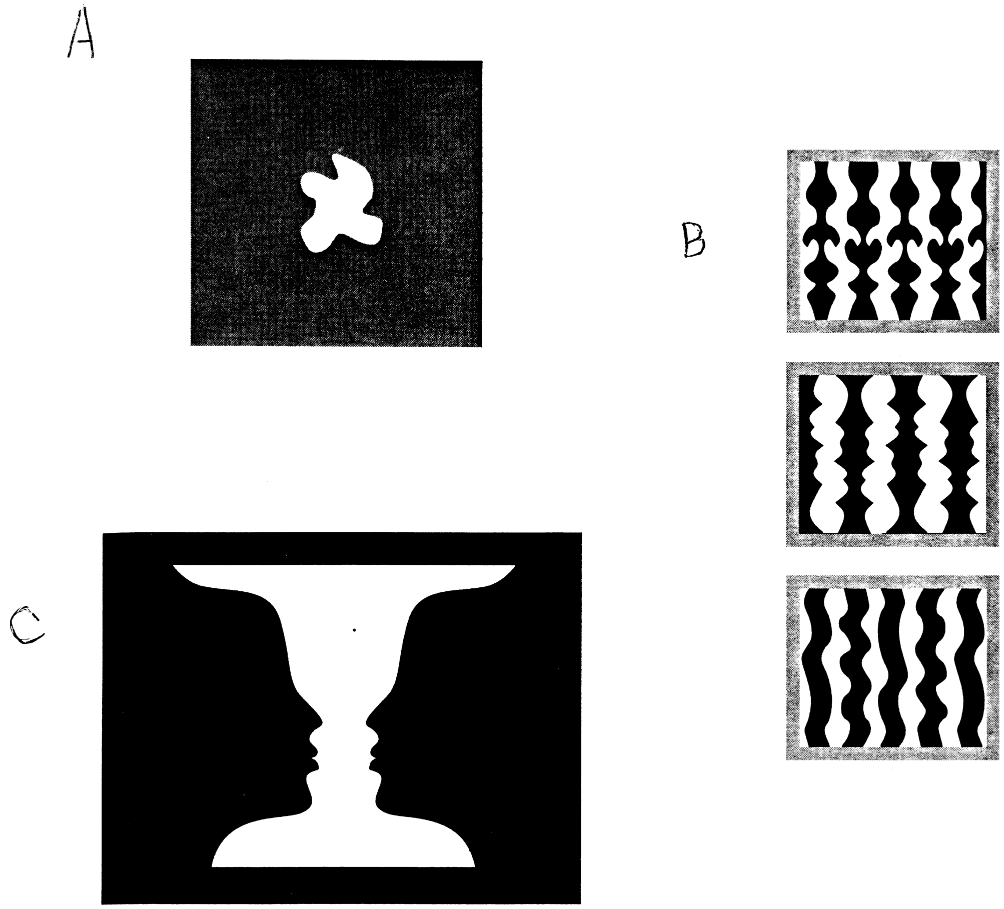


Figure 20.3: Segmentation into figure and ground. Many two-dimensional patterns divide themselves up spontaneously into regions with unequal properties: figure and ground. A. Small closed regions are usually seen as figures. B. Regions that are symmetrical, convex, and have parallel sides tend to be seen as figures. C. A reversible figure. Notice how the contour defines the shape of the region seen as the figure, whereas the shape of the ground is much less defined. [A and B from Palmer, 1999, pp. 282, 283. C from Goldstein, 1996.]

as ground, as shown in Figure 20.3B.

Finally, at least on the printed page, figure/ground assignment is not always stable. When the properties of size, convexity, and so on are nearly balanced, patterns can be seen in more than one way, and figure/ground assignments can reverse. A pattern in which figure/ground reversals occur is shown in Figure 20.3C.

20.1.3 Visual interpolation: Ambiguous contours, illusory figures, and amodal completion

Another interrelated set of phenomena closely related to segmentation and grouping are the phenomena of perceptual interpolation: ambiguous contours, illusory figures, and amodal completion. Examples of all three phenomena are provided by the classic Kanisza triangle, shown in Figure 20.4A. This figure consists of three black circular segments and three black angled lines. But most observers report seeing three boundaries that outline a bright white triangle, on a background of completed black circles and an completed outline triangle. In this figure, the boundaries of the white triangle are *ambiguous contours* – perceived contours that are not present as luminance or color differences in the retinal image. The white triangle is an *illusory figure* – a two-dimensional figure that is similarly not present in the retinal image. And the perceptual completion of the circles and outline triangle behind the illusory triangle illustrates *amodal completion* – the perceptual completion of occluded figures behind their occluders[1]. [Look around, and notice some occluded objects that you have been perceiving as complete.]

Additional examples of *ambiguous contours* are shown in Figure 20.4B-C. Figure 20.4B consists of two sets of horizontal lines with terminations along a common curved dividing locus. Most subjects report seeing an ambiguous contour along the division. Figure 20.4C is made up of six sets of nested C's, each with six or eight aligned line terminations, and it shows that sets of aligned line terminations are sufficient to generate ambiguous contours. Finally, Figure 20.4D shows that the cognitive expectation of a contour is not a sufficient condition for an ambiguous contour or an illusory figure to appear.

Some additional examples of *amodal completion* are shown in Figure 20.5. In Figure 20.5A, the black region below the square could be part of an occluded figure of many different shapes. Similarly in Figure 20.5B, your perception probably reports two rectangular solids, but the solid "in back" could have any of many shapes. At least in simple cases like these, the visual system seems to fill in the occluded part of the figure with as much regularity and simplicity as possible. No such simple statement, however, can be made for Figure 20.5B, a "mystery object" in which high-level memory processes must be involved. These phenomena are remarkable because our perception seems to go so far beyond the information given in the incoming stimulus array.

Under what stimulus conditions do these three perceptual interpolation phenomena occur? A critical stimulus feature seems to be the presence of *occlusion cues* – imagefeatures, like sets of aligned line terminations, that would most typically arise only when one object actually occludes another. When occlusion cues are present, it is as though the visual system creates the perception of an occluding object – a perceived object lying in front of the physical stimulus pattern – bounded by ambiguous contours. The occluded object is then amodally completed behind the perceptually created occluding object.

[1] There are many names for these phenomena. Ambiguous contours are also called illusory, anomalous, interpolated, or inferred contours; illusory figures are also called modal completion xx; and amodal completion is also called xx.

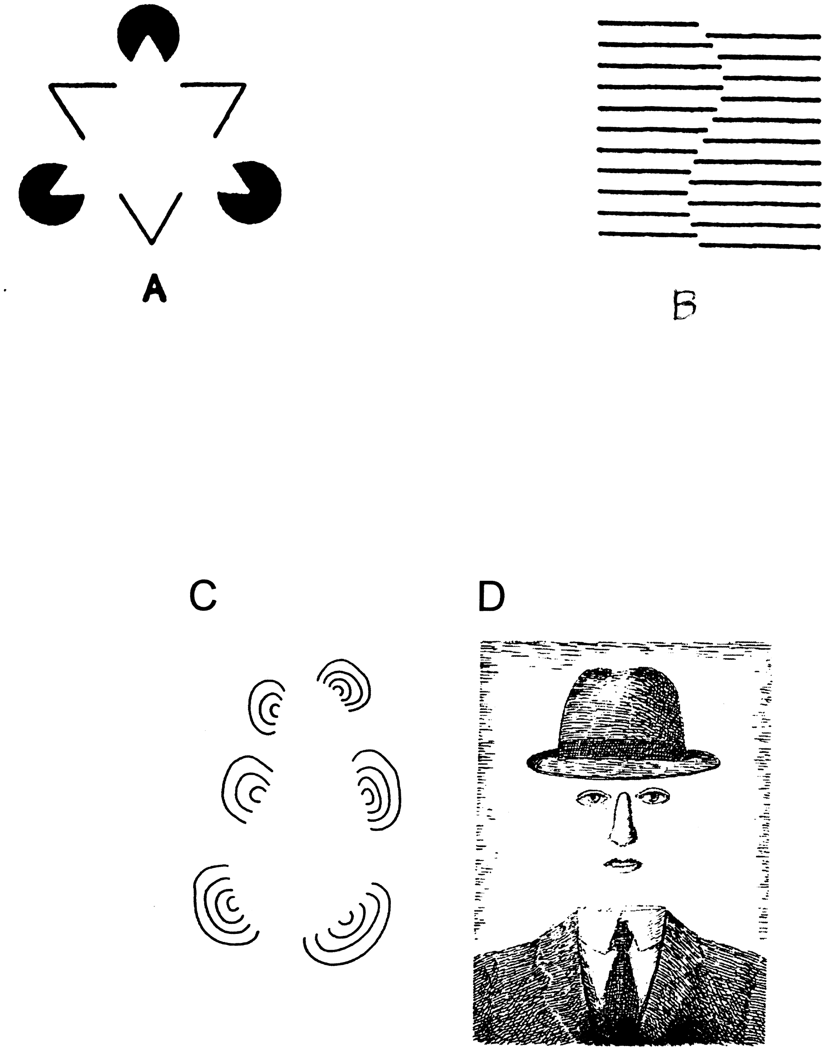


Figure 20.4: Visual interpolation. A: A classic Kanizsa triangle, showing three ambiguous contours, an illusory triangle defined by the ambiguous contours, and amodal completion of figures behind the illusory triangle. B: An anomalous contour arising from aligned line terminations. C: An illusory figure arising from aligned line terminations. D. A cognitive expectation is not sufficient to generate the perception of an illusory figure. [A. from Kanizsa? ; B from von der Heydt and Peterhans, 1989, Fig. 1, p. 1732. C,D from von der Heydt, 2004, Fig 76.1, p. 1140.]

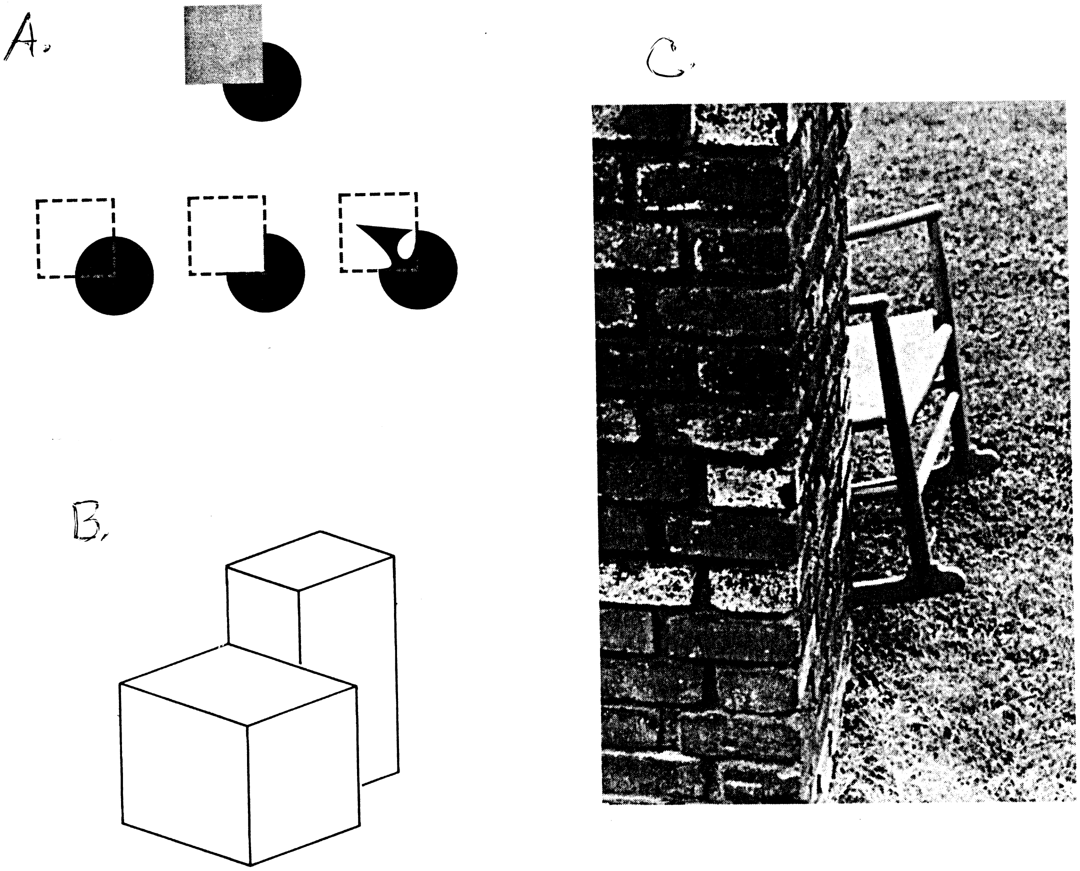


Figure 20.5: Additional examples of amodal completion. A: Imagine the many black objects that could be hidden by the square. Why did you perceive the black object as a circle? B. How do you know, or why do you perceive, that the object in back is a regular rectangular solid? C. Goldstein's "mystery object". [A from Palmer Fig. 6.4.1, p. 288; B from Goldstein, 1996, Fig. 5.2, p. 176; C from Goldstein, 1996, Fig. 5.1, p. 176.]

20.1.4 Perceptual grouping

Once the initial subregions of the image have been established, and the figure/ground assignments settled, the next problem is to organize the subregions into larger groups. For example, Figure 20.6 presents several sets of dots. How will our perceptual processes group the dots into larger wholes?

In the 1930s xx, Max Wertheimer articulated several rules that came to be called the *Gestalt laws of grouping*, and developed perceptual demonstrations of the various laws at work. Figure 20.6A shows a row of equally spaced dots. In Figure 20.6B, the spacing has changed, and you will probably see three *pairs* of dots – your perceptual processes have grouped the dots that are closer together into pairs. This is an illustration of the Gestalt principle of *proximity*: other things being equal, group together elements that are close together. In Figure 20.6C you will probably see the white dots grouped together and the black dots grouped together, so that the dots as a whole are organized into pairs. This is an illustration of the law of *similarity.* Grouping by similarity of size and of orientation are shown in Figure 20.6D and E respectively. Finally, Figure 20.6G and H illustrate the importance of continuity and of *closure* – the closedness of a form. When the form is open as in G, most subjects report two continuous lines crossing; but when the figures are closed as in H, most people report two fish kissing.

Segmentation processes operate in time as well as in space. Lights that flash in synchrony are perceived as tied together causally – the law of *synchrony.* And Figure 20.6F illustrates (in freeze frame) a grouping law that the Gestalt psychologists called *common fate* – similarity of speed and direction lead to perceptual grouping. The law of common fate can be demonstrated nicely by making two transparencies with similar figure elements on both. While they are overlapped and static on the overhead, they are perceived as a single class of objects. But if one overhead is moved with respect to the other, the figures on the two transparencies are instantly partitioned into their separate groups.

Finally, for dessert, Figure 20.7 shows some added complications. You may have to stare at this figure for a while before your perceptual organizing processes solve the puzzle it poses. Notice that in this case, your perceptual system is grouping non-contiguous regions together to form the figure. The fact that these perceptions emerge slowly suggests slow, complex, long-range computational processes at work. Interestingly, once your perceptual system has ”solved” one of these figures, it will solve the same puzzle again much more quickly. These time savings show some sort of learning process – a memory for the overall organization of a particular picture or scene. [Think about Rorschach figures. xx]

20.2 Modern themes

20.2.1 Experimental analyses of segmentation and grouping phenomena

Historically, segmentation and grouping phenomena were first introduced into the perception literature by the Gestalt psychologists in the 1920's. Interestingly, with a few exceptions, these phenomena have been passed down mostly in the form you have seen them – as demonstrations in textbooks. The fact that themost fundamental phenomena of the science can be demonstrated directly in the textbook is one of the most endearing aspects of the study of perception.

At the same time the field has often been criticized for its lack of quantitative data and its dependence on demonstrations and phenomenology. However, a little thought will show that there is no reason why a psychophysics of segmentation and grouping phenomena could not be established.

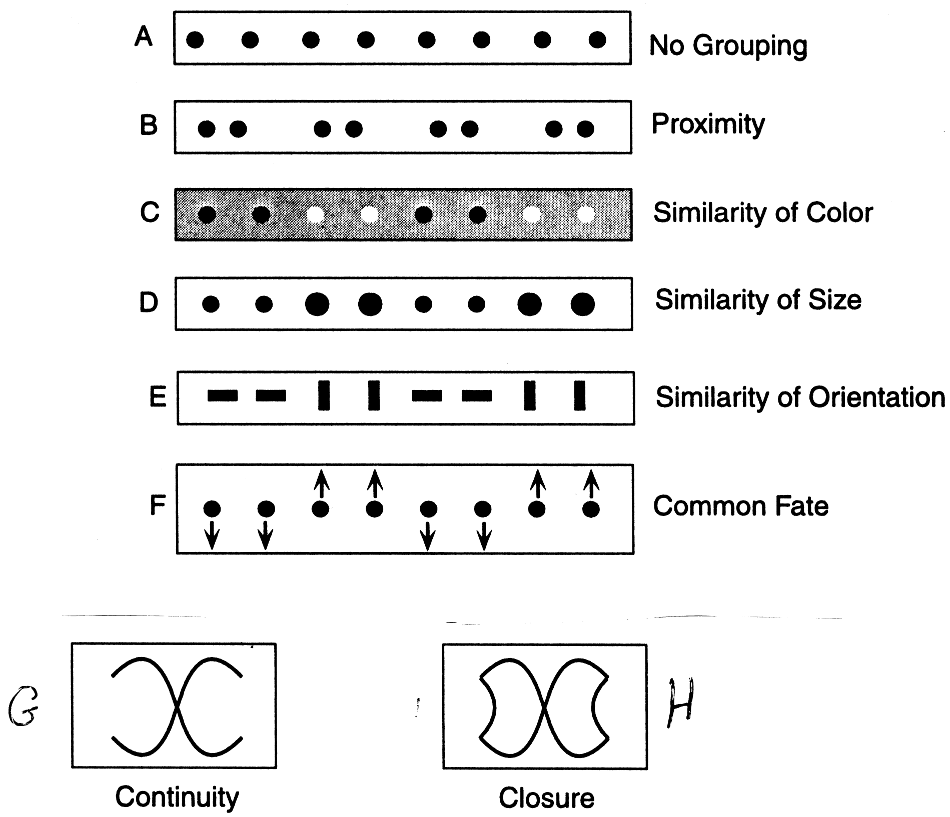


Figure 20.6: Some Gesetalt laws of grouping. Try to describe each panel in words, and notice how your words reflect the grouping laws your perceptual system imposes on what you see. [After Palmer, 1999, Fig. 6.1.2, p. 258.]

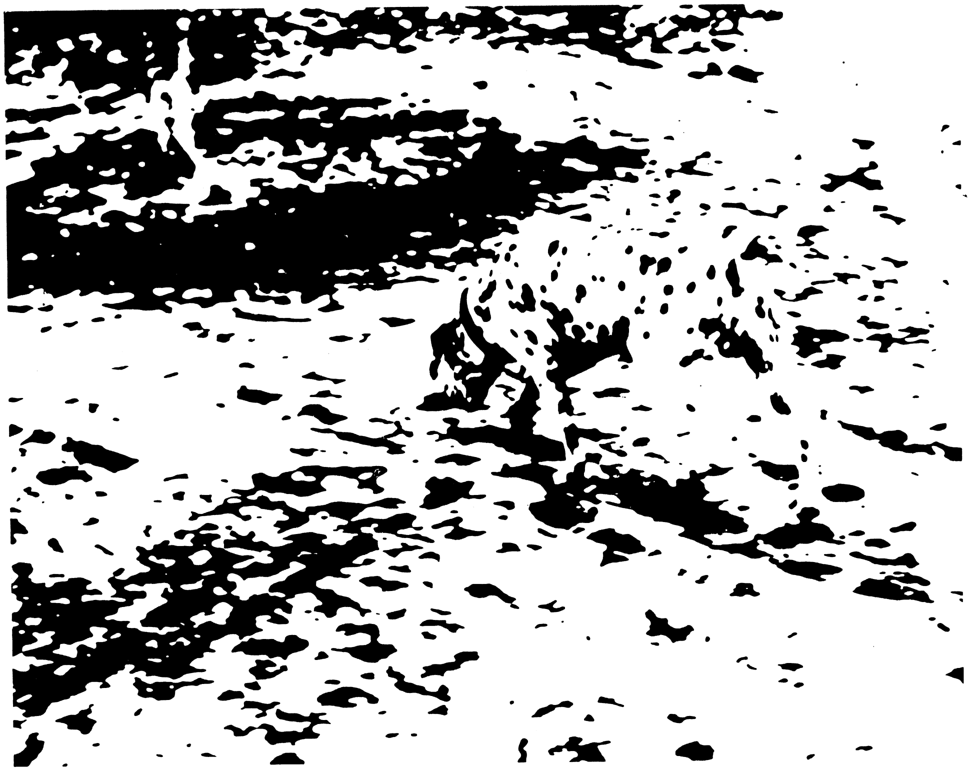


Figure 20.7: A dalmatian on a dappled street.

For example, one could readily find out how stimulus variables influence grouping, or how grouping principles such as proximity and similarity compete against each other; and it seems likely that other experimental paradigms could be imported. [Design an experiment to find out how lightness and color compete in grouping, or whether grouping depends more directly on the Boynton color code or the Hering code. Is the study of segmentation and grouping limited to Class B experiments, or can Class A experiments be used?]]

Since the 1990s, psychophysical work on segmentation and grouping has become increasingly popular. Much is now known about how stimulus variables influence segmentation and grouping processes, as well as how the various segmentation and grouping cues interact. We give more examples of quantitative studies a bit later, when we look for ordinal relationships among the various segmentation and grouping processes.

20.2.2 A new emphasis on three dimensions

Another aspect of the classic treatments of segmentation and grouping phenomena is the emphasis on only two spatial dimensions. After all, the main way of knowing about these phenomena was through the phenomenology of pictures in textbooks. By their very nature these demonstrations must be two-dimensional. However, you may have noted that hints of the third dimension did creep in, particularly in the phenomenon of figure vs. ground (the figure is perceived to lie in front of the ground) and in the interpretation of interpolation processes as ways of dealing with occlusion (which is an inherently three-dimensional problem).

Recent theorists such as Puilip Kellman, and Ken Nakayama and his colleagues (e.g. Nakayama, He, and Shimojo, 1995), have taken the next step along this path toward three dimensions, and begun to build depth cues into accounts of segmentation and grouping processes. This line of thinking is bolstered by two basic arguments. The first is that, again, the fundamental problem of vision is to reconstruct the shapes and locations of three-dimensional objects, not to process two dimensional images. But the two-dimensional shapes in the retinal image include the spaces between objects, and these do not relate meaningfully to the three-dimensional shapes of objects. Thus, it would not make sense for the visual system to undertake an initial segmentation of the two-dimensional image – many nonsensical proto-objects would be created. Rather, it makes sense for the initial segmentation to be done in three dimensions.

The second argument is that when we view a three-dimensional world containing three-dimensional physical objects, a new set of cues – depth cues – become available to use in parsing. In particular, binocular disparity – the small difference between the retinal images in the left and right eyes – varies lawfully with the depths and distances of objects. Moreover, a continuous object provides a continuous variation of binocular disparity and other depth cues. Thus, it would make sense for the visual system to rely on binocular disparity and other depth and distance cues in the initial segmentation process. (Binocular disparity and other depth cues will be discussed in detail in Chapter 25).

If we adopt this perspective, we need to completely rethink the whole question of image segmentation. The goal of segmentation and grouping processes is no longer to segment the retinal image into two-dimensional regions for further processing. Rather, it is to segment the world-to-be-perceived into three-dimensional regions defined by continuous, oriented depth planes, or *surfaces*. The argument is that such surface analysis allows better choices of proto-objects – guesses about the shapes and locations of objects in the physical world.

In particular, region analysis – the first topic of the chapter – must be completely rethought. It is no longer seen as the gluing together of two-dimensional regions on the basis of homogeneity or continuity of luminance, color, or texture. Instead, it becomes the gluing together of oriented surfaces in three dimensions, on the basis of homogeneity or continuity of depth, with color, lightness and texture relegated to a secondary role. For example, suppose that in the retinal image there is a luminance or color boundary that is not accompanied by a change in depth. Should we consider it an object boundary for further processing? The classic two-dimensional view would say yes; the three-dimensional view would say no, it's a single object that changes its illumination or its reflectance properties at that boundary.

20.2.3 An ordered sequence, or a merry-go-round?

Another question that sparks modern interest in the field of segmentation and grouping is, in what ways do (or can) the different grouping principles influence each other? For example, can interpolation processes such as amodal completion and illusory figures influence grouping? Can grouping influence figure/ground assignment? And can high-level processes such as object familiarity play a role in segmentation and grouping? The question is interesting because of the presumption embedded in the jargon of the field: if process A can influence process B, then process A is said to precede, or "come before", process B in the computational hierarchy. We will use this jargon in what follows, and analyse it more carefully at the end of the section.

A variety of experiments of this kind have been carried out by Stephen Palmer and his colleagues. For example, an experiment testing the influence of amodal completion on grouping (Palmer, Neff, and Beck, 1996) is shown in Figure 20.8A and B. The display consists of five columns of figures: two columns of whole circles on the left, two columns of half circles on the right, and a third column of half circles in the center. The subject's task is to report whether the middle column of figures groups with the circles on the left or the half-circles on the right.

The grey bar down the center of the display is a potential perceptual occluder. When the grey bar lies between two columns of half circles (Figure 20.8A), most subjects report that the center column of half circles groups with the half circles on the right. However, when the grey bar lies adjacent to the center column of half circles (Figure 20.8B), it is seen as occluding the right halves of a column of circles. Most subjects report that the circles are then amodally completed behind the bar, and group with the whole circles on the left. Amodal completion, then, can affect grouping. (In the jargon, amodal completion "comes before" grouping.)

Palmer and Nelson (2000) found similar results with illusory figures, as shown in Figure 20.8C, D. In this case, similarity of the orientations of sets of ovals (Figure 20.8C) was pitted against similarity of the shape of illusory figures created by "occluding" the ovals (Figure 20.8D) with illusory rectangles. Most subjects report that the middle row of Figure 20.8D groups with the right side of the figure, in which the illusory rectangles have similar orientations. Illusory figures, then, can affect grouping. [In the jargon, the creation of illusory figures "comes before" grouping.]

In another experiment, Palmer and Brooks (in prep xx) asked whether grouping might influence border assignment and figure/ground organization. They created displays in which regions on both sides of a contour were covered with textures of stationary or moving dots. The contour between the two sides could take on the same motion (or stationarity) as one of the textures. The result was that the contour grouped with the dots with which it had a common motion (common fate), thus claiming the contour for one side of the display; and the side of the display that owned the contour

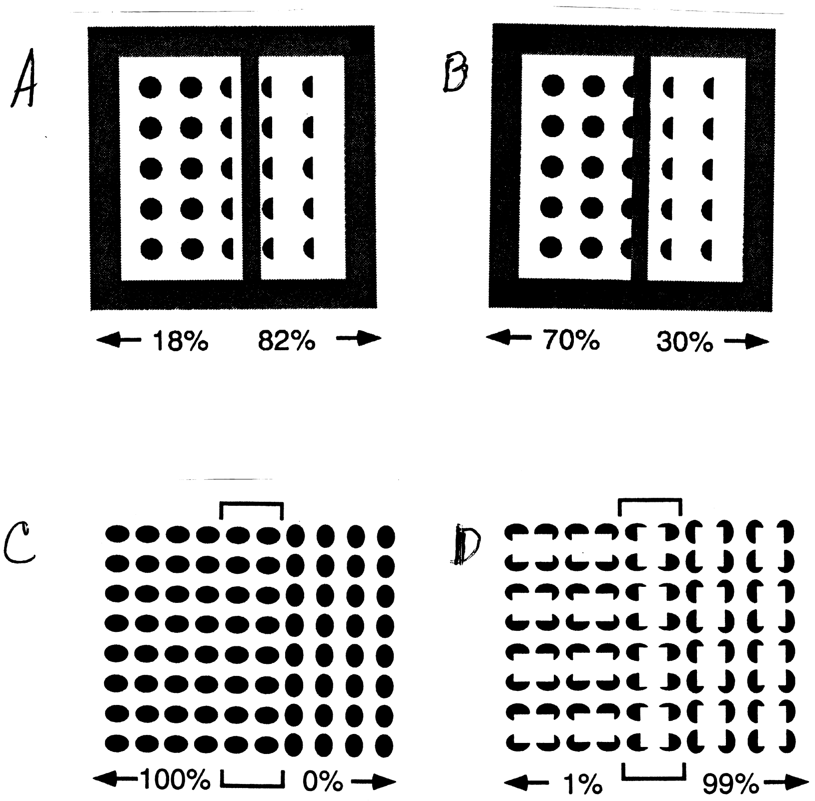


Figure 20.8: Effects of amodal completion and illusory figures on grouping. The subject's task is to judge whether the center column of figures groups with the left two columns or the right two columns. The numbers under the figures report the percentages of subjects who report each of the groupings. A,B: Amodal completion influences grouping. C,D: Illusory figures influence grouping. [A, B: Palmer, Neff, and Beck, 1996, in Palmer 2003, p. 31. C, D: Palmer and Nelson, 2000, in Palmer 2003, p. 32.]

was seen as the figure. [In the jargon, grouping by common fate ”comes before” figure/ground assignment.]

In another kind of experiment, Mary Peterson and her colleagues have shown that the familiarity of a figure – that is, information from memory – can influence figure/ground assignment. Peterson, Harvey, and Weidenbacher (1991) showed subjects ambiguous figure/ground displays, in which either side of the contour could be perceived as the figure. In each case the region on one side of the figure portrayed a familiar object (such as a seahorse) whereas the other did not. The result was that the side that portrayed the familiar object was usually seen as the figure. (In the jargon, one would have to say that object memory ”comes before” figure/ground assignment. But this conclusion seems a bit outlandish.)

[Add a Nakayama expt? 3-D surface rep influences grouping, etc.]

In summary, the various factors that influence segmentation and grouping can interact in many different ways. Amodal completion and illusory figures can influence grouping; grouping can influence border assignment and figure/ground organization; and object familiarity can influence figure/ground organization. It seems likely that many more ordered interactions of segmentation and grouping cues will be documented over the coming years.

If A can influence B, does that mean A comes before B, in information flow and/or anatomical terms? Such phrasing implicitly assumes a hierarchical model, without feedback. But it seems highly likely that cases will be found in which A can influence B, and other cases in which B can influence A. Such a constellation of results would militate against a strictly hierarchical model, and point up the faulty logic inherent in conclusions about which process comes before which.

20.2.4 Information flow models: Early or late? Hierarchical or feedback?

Where do segmentation and grouping processes fit into the overall flow diagram of visual processing, such as those shown in Figures xx (earlier Chapters)? There is general consensus that segmentation processes fall between an early feature-based processing stage on the one hand, and an object-based processing stage on the other. That is, segmentation is ”late” with respect to features, and ”early” with respect to object recognition. The term ”mid-level vision” has been coined to describe the information processing level at which segmentation and grouping processes are believed to occur.

At the more micro level, the question can be asked again: within ”mid-level” vision, in what order do the various segmentation and grouping processes occur? Which are ”early” and which are ”late” with respect to each other? There is little consensus at present. Figure 20.9 and 20.10 show a comparison of two contemporary information flow models of the ordering of segmentation and grouping processes with respect to each other. Each model has two versions.

An early information flow model proposed by Stephen Palmer and Irvin Rock (1994) is shown in Figure 20.9A. Palmer and Rock argued, mostly on logical grounds, that the various two-dimensional segmentation and grouping processes must occur in a particular order. In their model, *representations*, or *maps*, are shown as squares, and *processes* are shown by small circles between the maps.

Palmer and Rock began with the *image* – the incoming sensory signal. In their model, the process of *edge detection* applied to the image yielded an *edge map* – a representation of edges and their locations. The edge map is followed by the process of *region formation* – the forming of a set of mutually exclusive regions in a region map. Palmer and Rock argued that the region map must occur early in processing because regions must be formed before they can be further processed. The process of *figure/ground* analysis comes next, resulting in a map of what Palmer and Rock

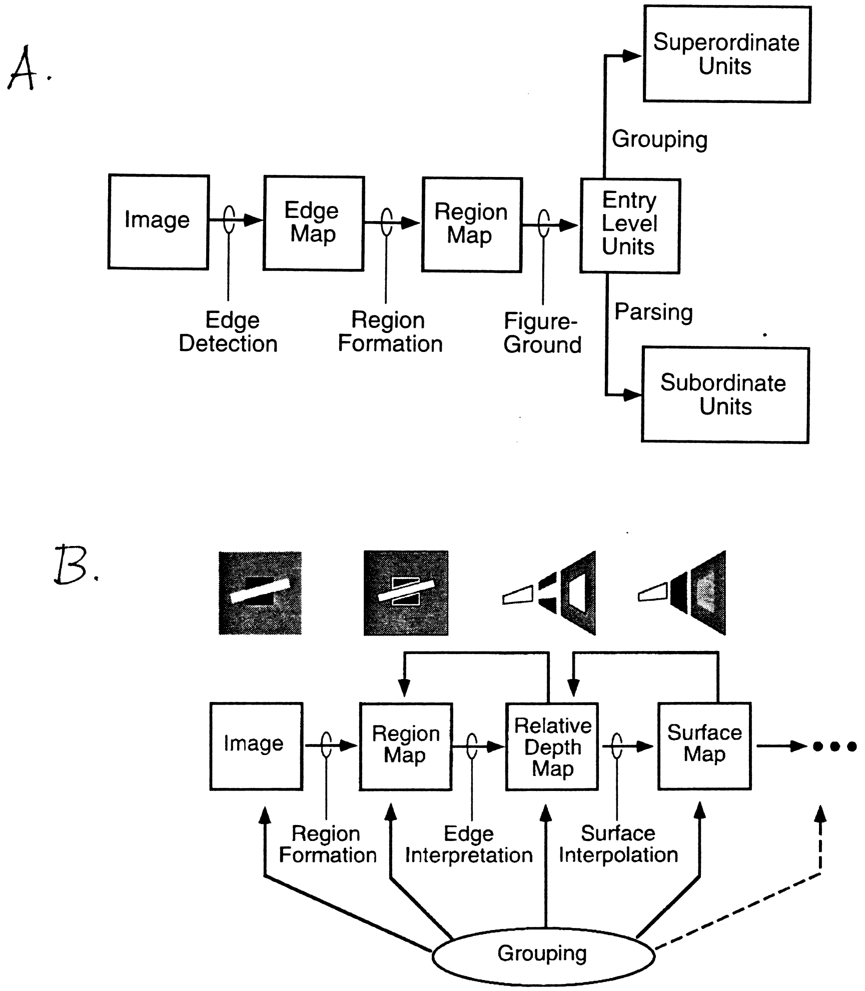


Figure 20.9: Palmer's information flow models. A: A hierarchical model showing the proposed order of classical segmentation and grouping processes. B: A more complex variant of the model, into which local feedback, multiple applications of grouping principles, and a surface map are incorporated. The icons above each stage indicates the two- vs. three-dimensional nature of the coding. Notice that a two-dimensional region map still precedes a three-dimensional surface map. [A: Palmer, 2003, p. 11; B: Palmer, 2003, p. 38.]

called *entry level units* – elements that could then be subjected to *grouping* and *parsing* processes. Grouping processes group subsets of the entry level units into *superordinate* units, whereas parsing processes divide them up into *subordinate units.* Since neither grouping nor parsing seemed to precede the other on logical grounds, Palmer and Rock speculated that these two kinds of processes operate in parallel.

Palmer and Rock's 1994 model was designed with the classical two-dimensional view of segmentation and grouping in mind. However, the complexities of interactions among segmentation and grouping processes – for example, the fact that grouping can influence figure/ground processes – led Palmer to revise this scheme (Palmer, 2003). The new model, shown in Figure 20.9B, posits that grouping processes occur repeatedly, being applied at each level to the code in the form it is in at that level. Although short feedback loops are indicated between each level and the one that immediately precedes it, the model remains largely hierarchical.

A more elaborate information flow model, less closely tied to the classical descriptions of segmentation and grouping processes, has been proposed by Philip Kellman and his colleagues (Kellman, Guttman, and Wickens, 2001). Kellman et al's initial model is shown in Figure 20.10 A. As in the previous cases, the model includes both processes, shown as squares, and representations, shown as octagons. A detailed account of this model is beyond the scope of this book, but we can briefly capture some of the flavor. [Is there an easier example?? Xx]

At the input to the model, Kellman et al posit the existence of two separate branches (”streams”) of processing, a *surface-processing stream* and a *contour stream.* The surface-processing stream detects regions of homogeneous luminance, color, texture, disparity, and/or motion, and represents these properties and their spatial locations. The contour stream detects edges from discontinuities in luminance, color, texture, disparity, and/or motion; integrates contours in three dimensions; and determines boundary assignments (border ”ownership”, and thereby figure/ground relationships). Information from these two streams is combined to form the *visible regions representation*: a representation of spatially contiguous, localized, three-dimensionally oriented regions marked by boundaries. Notice that an occluded object that occupies two separated regions in the retinal image produces two visible regions; no grouping process has yet been applied to the occlusion problem. Further boundary interpolation and surface interpolation (”spreading”) processes, including grouping processes, are then applied to produce two more representations: the *representation of units* (objects), and the *representation of shapes.* The representation of shapes forms the primary input to later processes of object recognition.

Although the details of Kellman et al's model are beyond our scope, two characteristics are important. First, with its inclusion of disparity and motion cues, the model incorporates the modern view that the initial segmentation of the incoming image is three-demensional. Second, notice that the model in Figure 20.10A, like Palmer's models, is hierarchical, or feed-forward: there is no mechanism by which later processes can influence earlier ones. However, the variant of the model in Figure 20.10B (Kellman, 2003) shows that the model is flexible in allowing feedback to be added from any of the higher processes and representations to the lower ones. In fact, Peterson et al's (1991) finding that contour familiarity could influence figure/ground relationships led Kellman to add a feedback arrow from the shape representation box back to the edge classification box, and other similar arrows could certainly be added.

What are we to make of this class of models? As discussed above, a growing number of interactions among classical segmentation and grouping phenomena are being discovered: amodal completion affects grouping; contour familiarity affects figure/ground assignment; grouping affects

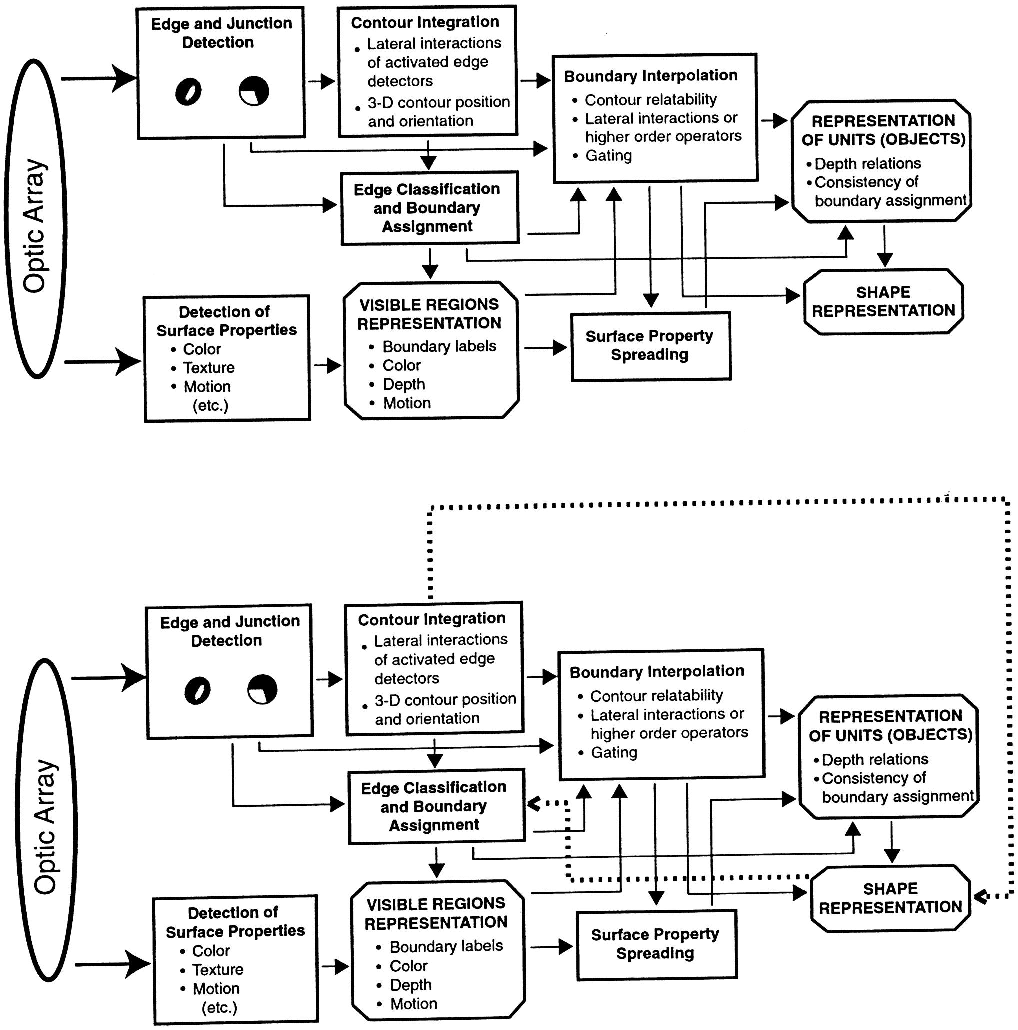

Optic Array
Edge and Junction Detection
Contour Integration
• Lateral interactions of activated edge detectors
• 3-D contour position and orientation
Boundary Interpolation
• Contour relatability
• Lateral interactions or higher order operators
• Gating
REPRESENTATION OF UNITS (OBJECTS)
• Depth relations
• Consistency of boundary assignment
Edge Classification and Boundary Assignment
SHAPE REPRESENTATION
Detection of Surface Properties
• Color
• Texture
• Motion
(etc.)
VISIBLE REGIONS REPRESENTATION
• Boundary labels
• Color
• Depth
• Motion
Surface Property Spreading
Optic Array
Edge and Junction Detection
Contour Integration
• Lateral interactions of activated edge detectors
• 3-D contour position and orientation
Boundary Interpolation
• Contour relatability
• Lateral interactions or higher order operators
• Gating
REPRESENTATION OF UNITS (OBJECTS)
• Depth relations
• Consistency of boundary assignment
Edge Classification and Boundary Assignment
SHAPE REPRESENTATION
Detection of Surface Properties
• Color
• Texture
• Motion
(etc.)
VISIBLE REGIONS REPRESENTATION
• Boundary labels
• Color
• Depth
• Motion
Surface Property Spreading

Figure 20.10: Palmer's information flow models. A: A hierarchical model showing the proposed order of classical segmentation and grouping processes. B: A more complex variant of the model, into which local feedback, multiple applications of grouping principles, and a surface map are incorporated. The icons above each stage indicates the two- vs. three-dimensional nature of the coding. Notice that a two-dimensional region map still precedes a three-dimensional surface map. [A: Palmer, 2003, p. 11; B: Palmer, 2003, p. 38.]

figure/ground assignment; and so on. It seems likely that the more experiments are conducted in the field of segmentation and grouping, the more different interactions among individual pairs of segmentation and grouping processes will be found, and the more likely it will be that feedback models will prevail over strictly hierarchical ones.

Two major questions remain open. First, if feedback is required, can one do the job with local feedback – say, from each processing level to just the next previous one, as in Figure 20.9B? Or will longer-range feedback loops, like that in Figure 20.10B, be required? And second, to what extent can perceptual experiments actually differentiate among the myriad information flow models that could presumably be generated? These are matters of debate and controversy at the present time.

Finally, both Palmer and Kellman are quick to point out that these are information flow models, based on perceptual data and speculation. There is no guarantee that anatomical regions or physiological processing stages will match up 1:1 with the processes or representations made explicit in these models.

20.3 Metaquestions

Segmentation and grouping processes are interesting from a philosophical point of view, because they provide important examples for two of the major themes of this book. How do these processes advance and enlighten our understanding of heuristics? And, do they suggest new linking propositions?

20.3.1 Segmentation and grouping suggest a new view of heuristics

As we have said before, the physical world is underdetermined in the retinal image – any given two-dimensional retinal image can in principle arise from many different states of the three-dimensional physical world. Yet our perceptual systems choose one interpretation – we perceive an object – and remarkably, our perception is usually veridical.

This remarkable veridicality is made possible by the artful combination of two factors: the cues – sources of information – in the retinal image; and the heuristics – informed guessing rules – built into the visual system by evolution and development. The argument is that for each cue, the visual system contains stored information concerning the most likely physical source of that cue; and it adopts the hypothesis, or guess, that the most likely source of the cue is present in the physical world. Moreover, psychologists have long argued that the perceptual phenomena of image segmentation and grouping provide us with clues as to the nature of these heuristics.

For example, consider the case of region segmentation. Suppose it is statistically likely (although not certain) that a continuous, homogeneous region in the retinal image arises from the continuous, homogeneous surface of a single object. The presence of a homogeneous region in the retinal image, then, provides a cue to the likely presence of a physical object with a homogeneous surface. The fact that homogeneous regions are perceptually seen as units suggests the presence of a heuristic like the following: *treat a homogeneous region of the retinal image as a candidate object for further processing.*

As a second example, consider figure/ground segmentation. Suppose that it is statistically likely that small, closed, convex, symmetrical regions in the retinal image arise from objects, and larger, irregular and asymmetrical regions from the spaces between objects. If so, then these properties provide cues as to which homogeneous areas of the retinal image are probably objects.

The corresponding heuristic would refine the choice of candidate objects: *treat small, convex, symmetrical homogeneous regions as candidate objects; do not give this status to large, asymmetrical regions, even if homogeneous.*

As a third example, consider the case of anomalous contours. Suppose it is statistically likely that particular patterns of line terminations in the retinal image arise from the occlusion of one physical object by another. If so, such patterns of line terminations provide a cue for occlusion. The corresponding heuristic would be: *treat particular patterns of line terminations as though they arise from occlusion; treat the scene as containing occluding and occluded candidate objects.* Other heuristics – maybe make the shape as simple and symmetrical as possible – would be called upon for amodal completion of the shapes of the occluded candidate objects.

And finally, the Gestalt laws of grouping can be seen as attempts to describe the heuristics used to group the parts of candidate objects into larger units. [Work out some possible heuristics based on the Gestalt laws of grouping.]

The heuristics presumed to be working in the case of two-dimensional image segmentation provide an excellent re-illustration of the fundamental concept of heuristics itself. Notice that the available cues to accurate segmentation and grouping are *less deterministic – less reliable* – than were the cues for distance (earlier Chapter xx). Homogeneous regions can arise from backgrounds (such as the sky) rather than from homogeneous objects; patterns of line terminations can be accidental, rather than arising from occlusion borders; a set of similar regions can arise from separate objects; and so on. In the extreme, all of the cues are unreliable, and the visual system must grasp at straws when it undertakes the segmentation of a two-dimensional scene. When we talk about using these heuristics, then, we are talking about the visual system having guessing rules that allow it to use two-dimensional stimuli to grasp at straws, in order guess at the nature of three-dimensional objects in the physical world. These guesses, of course, will sometimes be wrong, and misperceptions will occur.

From the modern perspective, however, one can argue that the classical two-dimensional demonstrations of segmentation and grouping phenomena do not reveal the workings of normal heuristics. Instead, they are an especially wicked class of custom designed stimuli, chosen to embarrass the visual system by forcing it to reveal its guessing strategies under very unnatural conditions. The demonstrations are presented on a two-dimensional page, with corresponding cues to the presence of a two-dimensional surface; yet they contain distance cues, triggering the implementation of heuristics whose function is to reconstruct three-dimensional scenes and objects. The perceptions we have when we look at these demonstrations may or may not reveal much about the heuristics used in more normal visual situations.

In the modern view, even the earliest processes of image segmentation are three dimensional, and the goal of early segmentation and grouping processes is to proceed directly to a three-dimensional map of the locations of surfaces. But notice that unlike the retinal image cues, the physiological cues to depth and distance are much more reliable and deterministic. An object at a given distance should virtually always elicit lawful amounts of accommodation and convergence. Thus, the amounts of accommodation and convergence are highly reliable cues to distance, and little guessing need be done. Similarly, physical depth creates retinal disparity according to fixed laws of geometry. Once the distance of an object is defined, the amount of retinal disparity it creates is a highly reliable cue to the depth of the object. The physiological cues plus a look-up table are sufficient to determine depth and distance, and structure a three-dimensional map.

On this perspective, the cues and heuristics so delightfully illustrated by the classic textbook demonstrations are secondary sources of information, used only to fine tune the already structured three-dimensional map. In fact, perhaps the unreliability of the two-dimensional cues is part of what makes the modern argument for three-dimensional surface segmentation so appealing. To create the foundation of veridical perception, build on reliable depth and distance cues and a well-behaved look-up table, rather than unreliable figural cues and probabilistic guessing rules.

Finally, there is much work on the kinds of computations that could implement segmentation and grouping processes and heuristics. This work is beyond our scope, but for references see Jacobs (2003) and Lee (2003).

20.3.2 Segmentation and grouping evoke new linking propositions

Do segmentation and grouping processes provide new insights into the concept of linking propositions, or suggest or imply new linking propositions for our consideration?

We begin with a reminder of the Universal linking proposition: Perceptual states and processes are always instantiated in neurophysiological states and processes. If region analysis, figure/ground segmentation, border ownership, visual interpolation, and grouping occur, then there are characteristics of the neural code that embody those system properties. Moreover, if we believe that these system properties are not embodied in the code at, say, the ganglion cell level, then they must be introduced by computational processes at some later stage or stages of processing, and it is of great interest to find the stage at which they are implemented.

But what shall we look for, and where shall we look? How would the neural code have to change in order to instantiate segmentation and grouping phenomena? How would we recognize the neural instantiation of an ambiguous contour? The usual approach is to use Analogy propositions. The neuroscientist begins by speculating that a neural process that is somehow analogous to the perceptual phenomenon will occur within the neural code at some particular location, and then goes to look for it. She may also look at the next earlier level, and try to show that the (speculatively) appropriate neural processes make their first appearance at a particular level of processing.

One of the important properties of segmentation and grouping phenomena concerns spatial domains. As you can see by examining Figures 20.1 to 20.8 with calibrated thumbnail in hand, the spatial domains involved in these phenomena are large – on the order of many degrees at the fovea. A commonly adopted (and perhaps even tautological) linking proposition used here is is one of common *spatial domain*: perceptual processes that spread over a given spatial domain imply neural processes that spread over the same domain. Since the classical receptive fields of ganglion cells are much less than 1o at the fovea, the argument would be that the required neural code must come into being beyond the ganglion cells. We will re-examine this argument in regard to V1 and other cortical levels below.

Beyond the spatial domain argument, what linking propositions are appealing for segmentation and grouping processes? What isomorphisms might we choose to assume and/or explore? A few specific linking propositions have been adopted, either implicitly or explicitly, in the physiological literature. For example, in regard to ambiguous contours, one might argue that an ambiguous contour looks like a contour because it sets up the same activity that a real contour does in some particular set of neurons at a particular cortical level – a cue invariance argument. In regard to figure/ground segmentation, one might argue that neural activity in the relevant cortical location will be different when a stimulus is perceived as figure than when it is perceived as ground: different perceptions imply different neural processes. One might go on to speculate as to what specific form the difference will take – for example, one might risk one's research career on the belief that firing rates of the relevant neurons will be higher when the region is perceived as figure than when it is perceived as ground. Studies based on analogies like these will be described below.

What about surface representations? Suppose one becomes convinced by modern arguments that there must exist a stage at which the locations and orientations of three-dimensional surfaces are explicitly represented. By what analogy shall we be guided, in deciding how to look for this putative representation? We could remember that V1 cells are tuned for the location and orientation of line segments, and argue (ignoring difficulties for the moment) that they provide a stage at which oriented line segments are explicitly represented. By extension, we could speculate that there should exist a stage of coding at which individual neurons would be tuned to the locations and three-dimensional orientations of surfaces. We could speculate that if the proper stimuli – surface patches, tilted in 3-D – were used, such a code might be revealed in V1, or alternatively, at a later level of processing – say, V2 or V4. To DT's knowledge, no study of exactly this kind has been done, but we will return to the question of surface representations below.

[Here's an exercise. We have said many times that neither the pointillistic representation at the photoreceptor level, nor the center-surround receptive fields at the ganglion cell level, instantiate the gluing together of the retinal image into regions. But what does? What might? Based on what linking proposition? If you wanted to find the neural basis of region segmentation, what would you look for and where would you start?]

20.4 Neural mechanisms of image segmentation

Finally, we come to the question: What is known about the neural basis of segmentation and grouping phenomena? How are these system properties instantiated within the visual system? In order to address these questions, of course, it will be necessary to use stimuli that, at least to human observers, elicit the perceptual phenomena in question. Unfortunately, none of the following studies report psychophysical data from human observers. However, all rely heavily on using stimuli that have been used extensively in demonstrations of human segmentation and grouping phenomena.

20.4.1 V1 bashing: Is it justified?

In the perceptual literature on segmentation and grouping, there is a lot of what one might call V1 bashing – claims that V1 neurons cannot provide the basis of segmentation and grouping processes. Why? These claims seem to be made on two bases. The first is an argument of spatial domain: it is claimed that the receptive fields of V1 neurons are spatially too small to mediate segmentation and grouping. The second is an argument about the capacities of codes: that the kinds of coding classically described for V1 do not have the right properties. That is, a code based on local image features or Gabor patches by its very nature contains no instantiation of segmentation and grouping processes – no spatial ”glue” to unify and separate regions.

As to the first claim, however, it is important to remember that more recent studies of V1 (Chapter 17) have shown both anatomical and physiological properties that make V1 a candidate for mediating spatial interactions over larger regions of the visual field. Anatomically, there are long range connections among neurons within V1. Physiologically, context effects are well established: stimuli many degrees outside the classical receptive field of a V1 neuron can influence its activity. And as to the second claim, it is important to remember that neurons can only reveal their responses to the stimuli with which we test them. V1 neurons have been waiting patiently since the early 1960s to be tested with stimuli custom designed to probe for segmentation and grouping processes. If the right stimuli were used, perhaps the appropriate processes would be readily revealed. Neurons can only answer the questions that neuroscientists choose to ask them.

In the past two decades, a few investigators have begun to use appropriate stimuli, and initiated the search for the neural correlates of segmentation and grouping processes. This work centers on three visual areas: V1, V2, and V4. We will review four kinds of studies: anomalous contours, figure/ground assignment, edge ownership, and amodal completion.

20.4.2 Anomalous contours

In the 1980s, Rudiger von der Heydt and his colleagues began the search for the neural correlates of image segmentation phenomena. The problem they took on was anomalous contours (Figure 20.4). They adopted the hypothesis that anomalous contours arise from heuristics designed to separate an occluded from an occluding object by creating a perceived boundary between the two. They reasoned that luminance-based contours and anomalous contours are functionally equivalent, in the sense that both can be used to code the contours of perceived objects. If so, then by an identity (or similarity) linking proposition, stimuli that provide these two sources of perceived contours should fall in an equivalence class – they should set up identical (or similar) neural signals – somewhere in cortical processing.

On this logic, von der Heydt and his colleagues decided to search for similarities of coding – cue invariance – between real and anomalous contours. Because individual neurons in V1 and V2 are well known to respond to luminance-based bars and edges, von der Heydt and his colleagues recorded in these two cortical areas. Awake, fixating rhesus monkeys were used as subjects. While the monkeys were fixating, individual cortical neurons were isolated. Their receptive fields and orientation tuning were first defined with traditional, luminance-based edge and bar stimuli. Each neuron was then tested with abutting grating stimuli that, in humans, produce anomalous contours.

The results were dramatic. Virtually all neurons in V1 responded only to the luminance-based contours. But almost half of the neurons in V2 that responded to luminance-based contours also responded to anomalous contours! Moreover, the orientation tuning curves were usually remarkably similar for the two classes of stimuli. The results from four V2 neurons are shown in Figure 20.11.

In sum, and remarkably, von der Heydt and his colleagues both found the equivalence classes they were looking for, and isolated the level at which these equivalance classes first appear in cortical processing. V1 neurons respond to luminance-based but not anomalous contours. But there exists a population of neurons in V2 that respond to both real and anomalous contours, as though they use the two interchangeably to signal the location of a contour.

Von der Heydt and Peterhans suggest that V2 neurons like these code the presence of contours – object boundaries – at particular locations and orientations in three-dimensional physical space. Notice that these V2 neurons can also be seen as implementing a heuristic a guess about the contours present in the physical stimulus. V2 neurons do not just detect retinal image contours, they signal (guesses about) the contours of objects.

20.4.3 Figure/ground assignment

In the 1990's, Victor Lamme (1995; 2004) took on the search for the neural correlates of figure/ground segregation. Lamme based his predictions on the Universal Linking Proposition: when

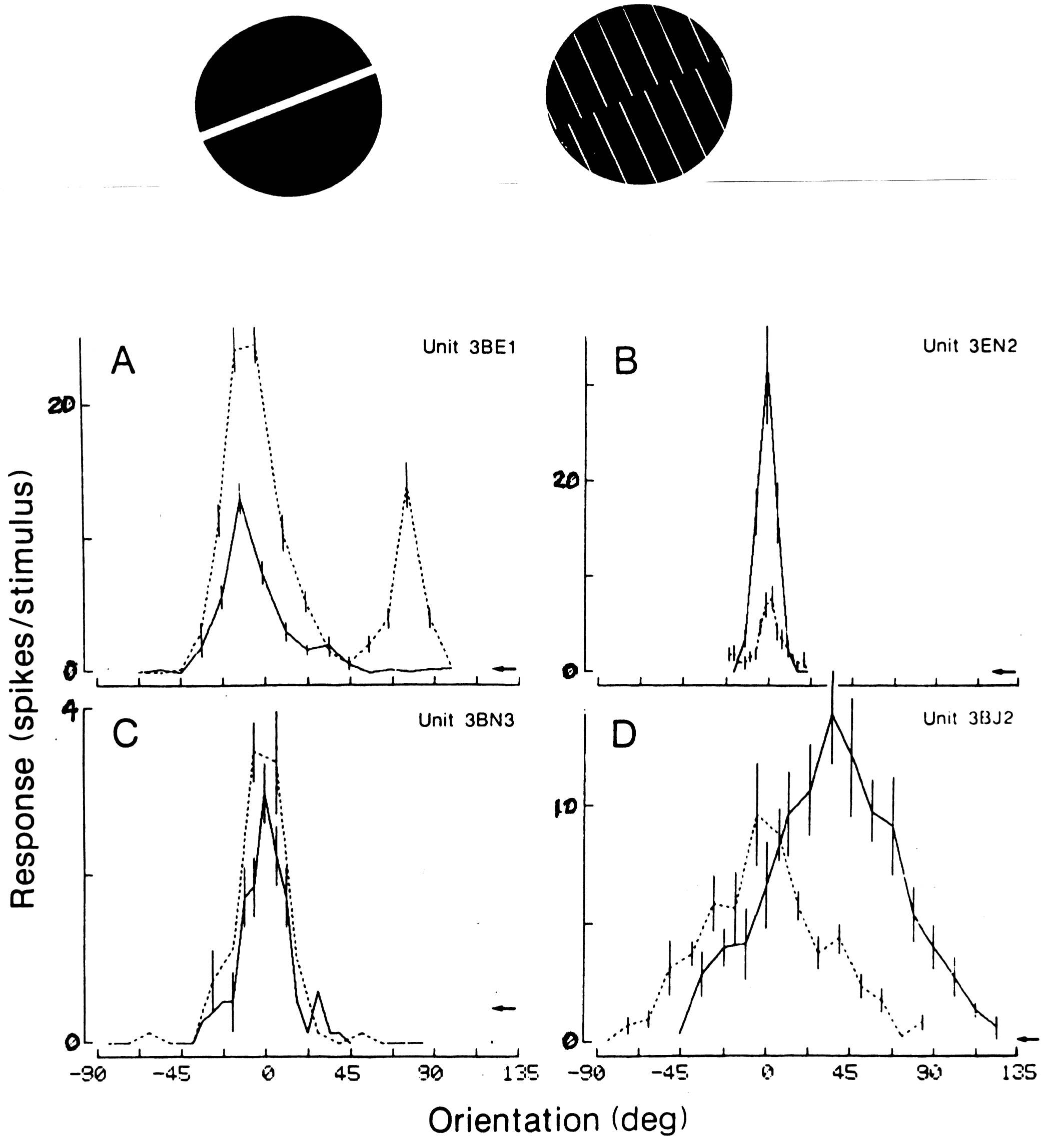


Figure 20.11: A possible neural instantiation of anomalous contours. At the top are shown the stimuli used by von der Heydt and Peterhans (1989): a real contour, and an anomalous contour aligned at the same orientation. A-D: Orientation tuning curves for four V2 neurons. The solid lines show responses to real contours; the dotted lines show responses to anomalous contours at matched orientations. In most of the neurons that responded to anomalous contours, the tuning curves for the two stimuli were quite similar (panels A-C); occasionally they were shifted (panel D). In some neurons the responses were highly similar (panel C), supporting the argument that some V2 neurons respond to real and anomalous contour stimuli in a cue-invariant fashion. [von der Heydt and Peterhans, 1989, Fig. 5, p. 1737.]

a stimulus changes from being perceived as figure to being perceived as ground, neural activity in the region corresponding to the stimulus must also change in some way. More specifically, he hypothesized that the firing rate of the neuron might be higher when it underlies a perceived figure than when it underlies the perceived ground. He chose to look for such signs of figure/ground assignment in area V1.

One of Lamme's experiments is shown in Figure 20.12. The panel at the top shows a square figure defined by left-diagonal lines, at the right, surrounded by a ground of right-diagonal lines. The icons at the left-hand sides of panels A-D show the stimulus combinations used in the experiment. In each icon, the small black rectangle shows the classical receptive field of a V1 neuron. As shown, the stimuli were presented in such a way as to cover the classical receptive field of the neuron. Notice that the stimulus within the classical receptive field – say, left diagonal lines – is the same in panels A aned B; what varies is the orientation of the lines outside the classical receptive field. But in consequence, the square of left diagonal lines is perceived (by humans) as a figure in A and as part of the ground in B. If the neuron's response varies, this would be (at least) another example of a context effect in V1 neurons, and (perhaps) a neural instantiation of the difference between perceived figure and perceived ground. Similar arguments hold for the stimuli in panels C and D.

The results of the experiment for three V1 neurons are shown in the three right hand columns of Figure 20.12A-D. In all three cases the response to the stimulus when it was (presumably) perceived as the figure was consistently greater than the response to the same stimulus when it was (presumably) perceived as the ground. The most dramatic example is given by cell 14, which responded to pattern A alone.

It is a short step to the speculation that the enhanced firing rate in V1 neurons like these produces (or begins a causal chain that produces) the perception of the stimulus as a figure, and underlies figure/ground assignment. We would, of course, want totest with many more stimuli, and look for Brindley's correspondence of many details, before we could count on this general conclusion. But it is an intriguing possibility.

20.4.4 Border ownership

More recently, von der Heydt and his colleagues (Zhou, Friedman, and von der Heydt, 2000) have carried the topic of image segmentation forward another step by exploring the question of border ownership. Zhou et al began by selecting neurons in V1, V2, or V4 that responded to lines or edges, and recording their receptive field sizes and orientation tuning curves. They then explored the responses of these neurons to a custom designed set of four stimuli, as shown in Figure 20.13.

In this figure, the small oval represents the classical receptive field of the neuron, and the dark vs. light shading represents the stimulus. In all four cases an edge of the optimal orientation lies on the receptive field of the neuron. Notice that in panels A and B the stimulus within the classical receptive field is identical in this case a slightly tilted light/dark border with the high luminance component on the left. The same is true in panels C and D, but in these two cases the high luminance component is on the right. If the neuron's response were controlled simply by the stimulus within its receptive field, it should respond equally to A and B, and equally to C and D.

But in perceptual terms, in stimulus A the edge that lies within the receptive field is 9for a human subject) part of the boundary of a bright square on the left. The square on the left is seen as the figure, and perceptually it lies in front of the background and owns the border. In contrast, in stimulus B the edge is part of the dark square on the right, which is perceived as the figure, lies in

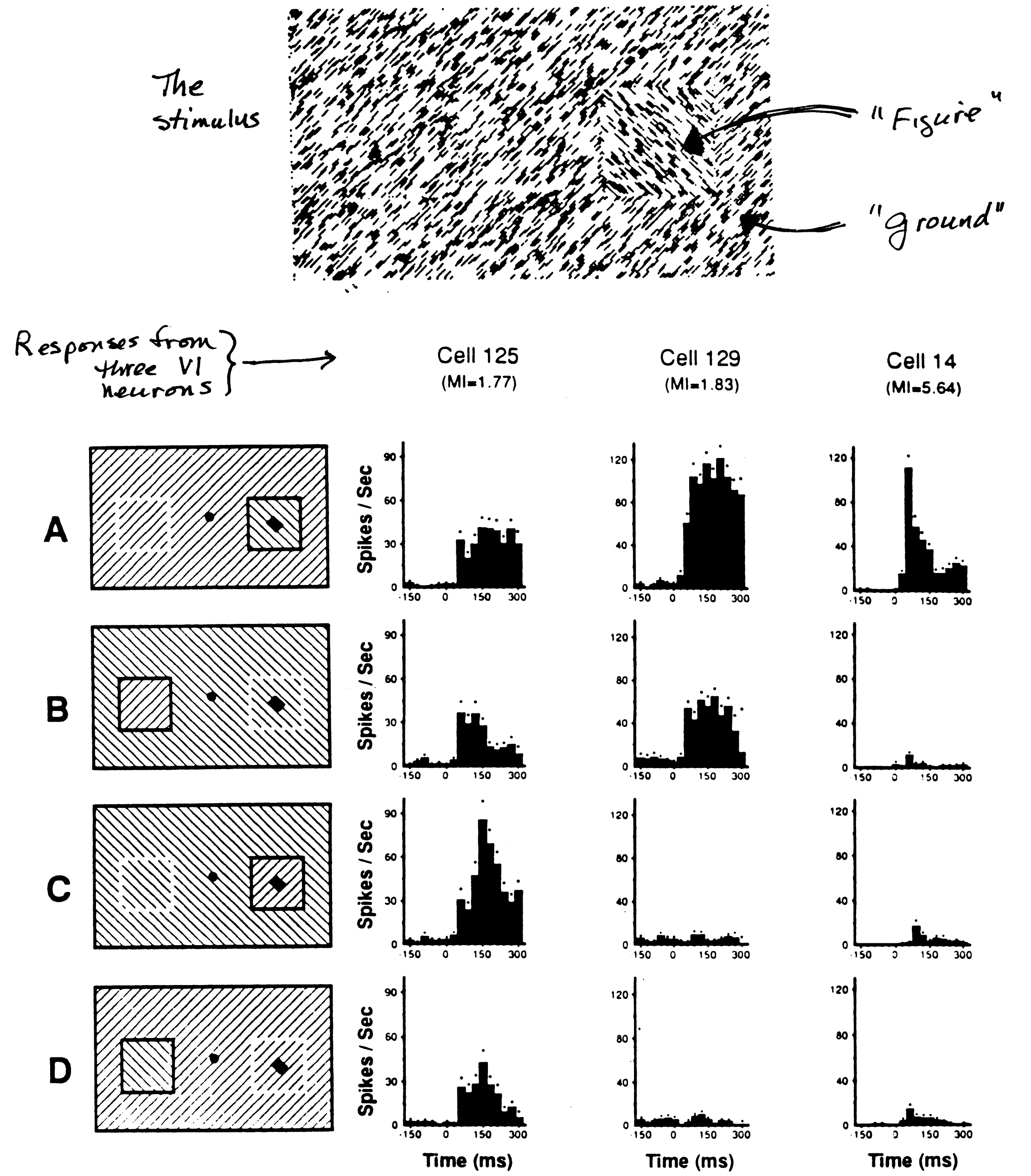


Figure 20.12: A possible neural instantiation of figure/ground assignment. The top panel shows the kind of stimuli used. In the icons at the left in panels, A-D, the small black rectangle shows the classical receptive field of a V1 neuron. The responses of three neurons are shown. For each cell, the firing rate is enhanced when the stimulus in its receptive field is perceived (by humans) as a figure, compared to when it is perceived as ground. [After Lamme 1995, Figs 2 and 5, p. 1608-1609.]

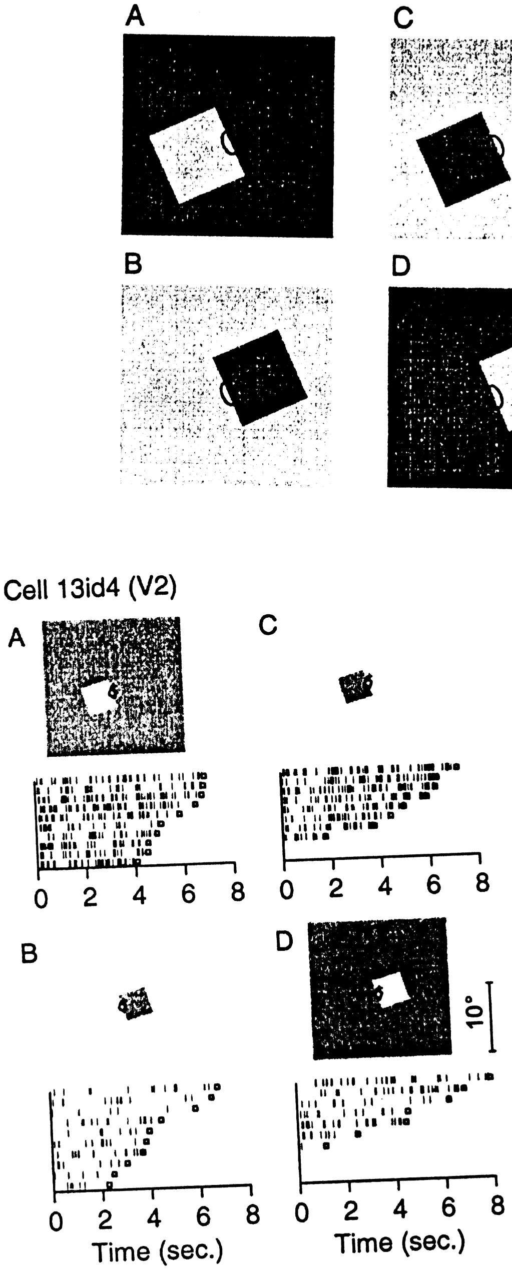


Figure 20.13: A candidate border ownership cell. This V2 neuron responded strongly and more or less equally to stimuli A and C, which human subjects perceive as squares up and to the left, but not to B and D, which human subjects perceive as squares down and to the right. Perceptually, the square "owns the border" in each case. The neuron does not just code the location or polarity of the black/white contour. Instead, it seems to code the location of the figure, and the ownership of the border, responding to squares down and to the left but not to squares up and to the right. [Zhou, Friedman, and von der Heydt, 2000, Fig. 2, p. 6596, and Fig. 4, p. 6597.]

front of the background, and owns the border. If the neuron under study signals the figure-ground relationship and the concomitant border ownership, the neural signals should differ between the two stimuli, responding to, say, A but not B. The same argument identical stimulus within the receptive field, but differing border ownership holds for stimuli C and D.

Zhou et al argued, however, that the opposit4e pairing of stimuli is of greater interest. They argued that neurons that respond strongly and equally to stimuli A and C but less to stimuli B and D (or vice versa), could be said to code *border ownership.* An example of a candidate border ownership neuron is shown in the lower half of Figure 20.13. This neuron responded strongly and similarly to stimuli A and C, as though it were signaling the presence of a figure down and to the left. It responded much less vigorously to stimuli B and D, with the figure up and to the right.

Zhou et al suggest that these neurons carry information concerning the side of the border to which the contour belongs. Concomitantly, they carry information about figure-ground definition and occlusion that is, about the depth ordering of objects. As such these neurons could play a major role in introducing some of the characteristics needed for image segmentation into the neural code.

The proportion of border ownership neurons varied among visual areas only about 3% in V1, but about 15% in V2 and V3. These statistics suggest that computations done within V2 introduce properties important to image segmentation into the neural code. Zhou et al suggest that like orientation, spatial frequency, color and binocularity disparity preferences, the property of *border ownership preference* should be added to the list of features coded by sub-populations of V2 neurons.

Since that time, there have been several studies of the cue invariance of border ownership in V2 cells. The responses of some border ownership cells are invariant over variations in the size of the figure. Some (but not all) border ownership neurons also show border ownership when tested with a variety of other border cues, such as binocular disparity, and solid vs. outline figures. Moreover, when a neuron shows border ownership for two different cues, the side of border ownership is usually consistent if the neuron showed ownership of the left border for one cue, it usually shows ownership of the left border for all cues to which it responds.

20.4.5 Amodal completion

Finally, Nakayama and his colleagues (Bakin, Nakayama, and Gilbert, 2000) have searched for neural instantiations of amodal completion (Figures 20.4 and 20.5) and related phenomena of surface representation in the activity of monkey V1 and V2 cells. As described above, amodal completion occurs under conditions of perceived occlusion – for example, when two line segments separated by a gap are seen as a single continuous line that lies behind an occluding object.

Bakin et al (2000) began with the phenomenon that a neuron's response to a line can be facilitated by a collinear flanking line, as shown by Kapadia et al, 1995 (Chapter xx). They argued that this facilitation could provide the neural instantiation of the amodal completion of a longer line that encompasses the two initial line segments.

To test for amodal completion, Bakin et al used three highly similar stimulus conditions, only one of which resulted (for human observers) in perceived occlusion and amodal completion. The question was, will facilitation of the response to the initial line occur only in the condition in which human observers see an occluding surface, accompanied by amodal completion?

As usual, monkeys were trained to fixate a fixation target. The stimuli were presented binocu-

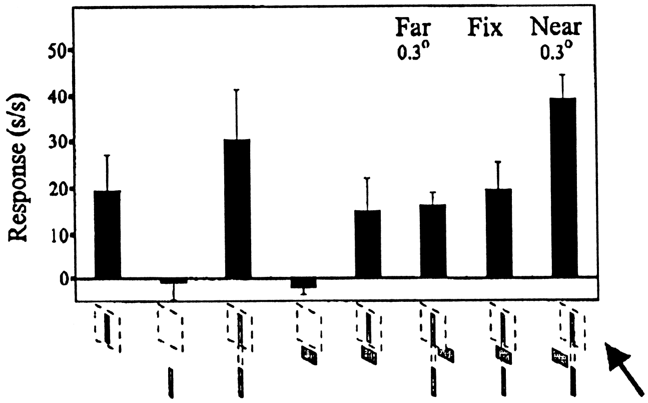


Figure 20.14: A possible neural instantiation of amodal completion. The icons underneath the graph show the eight conditions of the experiment, as described in the text. The greatest facilitation of the neuron's response is shown in the eighth condition, at the far right, in which two line segments are perceived (by human observers) to be united into a single long line behind an occluder. [Bakin, Nakayama,and Gilbert, 2000, Fig. 4, p. 8191.]

larly, and the equipment was arranged so that the binocular disparity between parts of the stimulus pattern could be varied between conditions. In consequence, for human observers, different parts of the stimulus pattern appeared in different depth planes.

The design of the experiment and the results from a single neuron are shown in Figure 20.14. Eight different stimulus conditions were used, as shown by the icons at the bottom of the graph. First, a line at the appropriate orientation was placed within the classical receptive field of a neuron, and the neuron increased its firing rate by about 20 spikes per second. This is the response against which all other responses will be compared.

Second, a colinear flanking line was presented alone, outside the classical receptive field, and as expected the neuron did not respond. Third, both lines were presented, and the response of the neuron to the first line was facilitated – increased to about 30 spikes per second (cf. Kapadia et al, 1995; Chapter xx.). Fourth, an orthogonal cross bar was presented alone, outside the classical receptive field, in the same depth plane as that used for the lines; again the neuron did not respond. Fifth, the original line and the cross bar were presented together, and the cross bar did not facilitate the response of the neuron to the line.

The sixth, seventh, and eighth conditions are the critical ones. In these conditions the two lines and the cross bar were all presented; the variable was the depth plane in which the cross bar appeared. In the sixth condition, the cross bar appeared behind the two lines – it was not an occluder – and no facilitation occurred. Similarly, in the seventh condition all stimuli were in the same depth plane – again no occlusion and no facilitation. But in the eighth condition, the cross bar appeared to lie in front of the two lines, forming (for human observers) an occluding object. In this case, the neuron's firing rate increased to about 40 spikes per second – a large facilitation effect occurred.

Perhaps facilitation of the neuron's response signals the amodal completion of a long line behind the cross bar. If so, then this neuron instantiates the neural correlate of amodal completion of the line, or introduces it into the neural code. Responses to these and other similar kinds of stimuli were common in V2 cells but rare in V1 cells. Nakayama and his colleagues argue that these data also provide suggestive evidence that V2 neurons code the locations of three-dimensional surfaces in space.

In summary, since the pioneering work of von der Heydt and his colleagues in the 1980s, several research groups have looked for neural activity that might incorporate various aspects of image segmentation and grouping into the neural code. Since no one knows what these code changes will look like, we must begin with speculations as to the nature of the code changes, and as to the anatomical locations at which they will occur. Given the necessary degree of speculation, researchers in this field seem to DT to have been remarkably lucky so far, as in each case candidate neural instantiations have been found.

20.5 Summary

In the present chapter we began by reviewing the classical textbook demonstrations of the perceptual phenomena of image segmentation and grouping. It is often argued that these phenomena reveal the workings of perceptual heuristics; for example, that the perceptual depth ordering implicit in figure/ground assignment show the use of a guessing rule with which the visual system begins to reconstruct the third dimension of the physical world. These guesses, however, would seem to be rather desperate ones. The newer perspective is to suggest that under more natural viewing conditions, scene segmentation is inherently three-dimensional. In that case, segmentation can incorporate the more reliable cues for distance and depth, and segmentation can proceed on a much surer footing.

We then reviewed some experiments that show a variety of different ordered interactions among segmentation cues. For example, amodal completion affects grouping, grouping affects figure/ground assignment, and so forth. We reviewed two examples of information flow models for the ordering of segmentation and grouping processes, but left open the possibility that no model that postulates a single serial order will be sufficient. Instead, it seems likely that image segmentation takes place as a set of highly interactive computations, and that feedback as well as feed-forward computations are involved.

In the last part of the chapter, we explored a set of studies aimed at revealing the neurophysiology of four individual segmentation processes: figure/ground segregation, formation of anomalous contours, border ownership, and amodal completion. Plausible neural instantiations of all four phenomena were found either in V1 or, more frequently, in V2. Thus, the case can be made that early cortical processing, especially in V2, contributes heavily to introducing the neural instantiations of segmentation and grouping phenomena into the neural code. These computational processes may be partially local, but may also be part of a larger neural circuit that incorporates feedback from higher cortical levels. These issues of early vs. late and feedforward vs. feedback are, however, highly controversial at the present time.

Overall, it is well to remember that this field is in its infancy. Only a few of the many possible speculations have been speculated, only a few of the possible linking propositions have been articulated, and only a very few of the many possible studies have yet been carried out, at any of the many processing levels of the visual system. Thus, the generality of the initial results is profoundly unknown.

As usual, we want the impossible. Before we count on the conclusions from these studies, we need to know more about the broader equivalance classes of all of the neurons whose properties are being studied. For example, how do the neurons that Zhou et al called border ownership neurons respond to a larger range of stimuli? Do their equivalance classes extend to many different figures, as long as the figures are, say, up and to the right of the receptive field? Do they respond to circles? Triangles? Figures with irregular borders? To be worthy of their name, these neurons would need to respond to all stimuli in which a region up and to the right is perceived as the figure. Similarly, we have no idea what the broad equivalence classes of Lamme's V1 neurons are: do they always respond more to stimuli perceived (by humans) as figures than to stimuli perceived as ground? And, even more basically, do monkeys perceive segmentation and grouping as humans do? More tightly parallel psychophysical and physiological studies will be needed if Brindley's "correspondence of very many details" is to be achieved.

Finally, the work on the neural correlates of image segmentation processes centers around two kinds of special linking propositions. The first is that the instantiation of some grouping phenomena, such as figure/ground assignment in the one case or amodal completion in the second, depend upon increases in the firing rates of particular neurons. The second is that other phenomena, such as anomalous contours in the one case or border ownership in the other, depend on cue-invariant responses from other particular neurons. These linking propositions are interesting and may be true, but their presence is inescapable, and their arbitrariness would seem to weaken the nomological network of which they are a part.

Chapter 21

Motion – Psychophysics

Imagine a blue jay flying around your campsite, looking for crumbs but keeping a minimum distance from you to avoid being caught. For the sake of the argument we will define his pattern of motion as a change in his two-dimensional location over time. If you raise a threatening hand, his motion will quickly include the third dimension – a change of distance. In addition to these translational motions, his body also rotates around his own axes in space, and it undergoes deformations – changes in shape – as he tucks in his wings and extends his feet to alight on your plate of pancakes.

Each of these different kinds of physical motion – translations, rotations, deformations – cause different patterns of motion in the image of the blue jay on your retina. Conversely, these different patterns of motion can potentially serve as cues not only to his motion, but also to his 3D shape and his identity. But clearly there is much sorting out to be done, and heuristics to be called upon, to get from the patterns of motion in the retinal image to useful representations of the locations, shapes, and identities of physical objects in the world. These multiple uses of motion cues make motion a complex perceptual topic.

As of the early 21st century, many vision scientists argue that motion is analyzed in two basic stages. The first is the analysis of *local motion* – the creation of neurons that are tuned for the direction of motion of small, fixed elements of the retinal image over short distances. It is easy to imagine that the analysis of local motion takes place in V1, where as Hubel and Wiesel and others have already shown us, neurons are created that are selective for the direction of motion of such stimuli as edges, bars, and gratings. Such local motion detectors could provide the raw materials from which higher-level neurons tuned for more complex motion patterns could be built.

In the present chapter we begin by elaborating the difficulties of motion processing that arise from the collapsing of the three-dimensional world into the two-dimensional retinal image, and from the presence of eye movements. We then turn to local motion analysis, and show how multiple channels theory has been extended to include a set of independent detectors that are selective for the local direction of motion. We then introduce two classic illusions of motion, and proceed to describe some computational schemes that might underlie the analysis of local motion and create the classical illusions.

Next, we move beyond the local motion level, and turn to the more complex aspects of motion perception, such as those illustrated by the rotations and distortions in the image provided by the blue jay. We consider the patterns of retinal image motion that arise from variations in the distances of objects, from self-locomotion, from the rotations of rigid objects, and from the deformations of shape typical of biological organisms. We end the chapter with some speculations about new heuristics that might enter into the interpretation of motion cues.

In the subsequent chapter, we turn to the physiology of motion processing.

21.1 What is motion?

In the simplest case, motion is defined as a change of location in time. The motion of an object along a straight path has two basic parameters: *speed* and *direction.* The *speed* of motion is the change of location per unit time (Δx/Δt), and the *direction* is the angular direction of the motion (leftward, rightward, left diagonal, etc.).

There are two vocabulary issues to be aware of. First, in formal physical terminology, the term *velocity* refers to both direction and speed, and to specify velocity one must specify both. But in vision science some authors use the term *velocity* interchangeably with the term *speed.* And second, the terms *motion* and the *direction of motion* are sometimes used interchangeably; for example, if not elaborated further the phrase ”neurons tuned for motion” usually means ”neurons tuned for the direction of motion”.

21.1.1 Physical motion, retinal image motion, and perceived motion

As in other areas of perception, in order to make sense of motion perception it is critical to distinguish among three concepts: *physical motion, retinal image motion*, and *perceived motion.* Physical motion is the motion of a physical object; retinal image motion is the pattern of motion created in the retinal image by the physical motion; and perceived motion is the motion as perceived and judged by a human observer.

The perceptual goal, of course, is to make perceived motion correspond to physical motion: to have veridical perception of the direction and speed of objects in the physical world. And indeed, common experience tells us that most of the time we routinely have least approximately veridical perceptions of the motions of objects. In fact, our experiences seem so simple and direct that that it takes an effort to break away from nave realism and figure out why the veridical perception of motion is a puzzle[1].

In the case of motion within the *frontal plane* – any two-dimensional plane at a fixed distance in front of the observer – the problem of veridical perception would seem to be particularly simple. Consider a moving object, and the image of that object as created by a stationary lens. For every change of location of the object, there is a change of location of the image, with speed and direction correspondences that can be readily specified through the laws of optics. Thus, the motions in the image provide reliable information – cues – to the motions of the object. Moreover, since physical objects typically have fixed shapes and exist continuously in time, the heuristics required to reconstruct the object motion should also be very simple: interpret motion of a particular speed and direction in the image as motion of a corresponding speed and direction on the part of the physical object.

[1] A vision scientist friend of DT’s was trying to explain to her father what she does. When she told him she was studying the perception of motion, he said, ”It moves, I see it move. What’s to study?”

21.1.2 Retinal image motion confounds physical motion with other variables

Would that it were that simple! The problem is, the above discussion is based on an object that is confined to the frontal plane and a lens that is fixed in space. In the real physical world, there are at least three major confounding variables. [Think of some more.]

First, physical objects move in three dimensions, not just two. As we have seen before, mapping the three-dimensional world to the two-dimensional retinal image imposes a profound loss of information. For example, many ovals appropriately tilted in distance will yield the same circular retinal image (see Chapter xx), so a circular retinal image doesn't imply a circular object. Similarly these same ovals, rotating in specific, diabolically designed patterns in three dimensions, can all yield the same pattern of retinal image motion, so a particular pattern of retinal image motion doesn't imply a particular pattern of object motion. Imagine two scenarios – a soccer ball coming toward you, and a person blowing up a soccer-ball-patterned balloon, according to a program designed to yield the same expansion pattern. In both cases your retinal image will show the same spatiotemporal pattern – a checkered image that expands over time.

In the 21st century, film, video, and virtual reality technologies provide ever more persuasive examples of this world-to-retinal-image metamerism. In addition to its possible more natural multiple sources, a given spatiotemporal pattern in the retinal image can arise either from the rotation of an object, or from a videotape of the rotation of the same object, or from a wholly artificial animation sequence designed to produce the same pattern. Analysis of the sequence of changes in the retinal image will tell us only that the physical world contains a member of a particular spatiotemporal equivalence class, but not which member. Beyond that, the shape and motion of the object must be guessed on the basis of heuristics (to which we return below).

A second major source of confounding variables arises from the fact that in the real visual system, the optics move whenever the eyes move or the person moves. Thus, a given pattern of motion in the retinal image of an object can arise either from motion of the object or from motion of the eyes. Consider the pair of cases shown in Figure 21.1A and B. In Fig21.1A an observer is looking at a an object (here an oval) centered on a fixation point F in a stationary scene. She looks at the oval for an extended period of time that includes times t_1 and t_2, and the retinal image remains the same from t_1 to t_2. In Figure 21.1B, the observer is still looking at the same scene, but decides to change her fixation from point P to point Q. In consequence, with respect to the fixation point F, the locations of all objects in the image change between t_1 and t_2 (P, Q, and the oval are all to the left of F at t_1, and to the right of F at t_2). Nonetheless, in perception the two cases remain remarkably similar and both are seen veridically: we see a stationary object against a stationary surround.

Similarly, in Figure 21.1C an object (the oval) is moving from left to right against a stationary background. The observer is fixating at F, and the observer's eyes are stationary. In this case, the moving object creates a moving retinal image. Since the retinal image is inverted with respect to the physical world, the image of the moving object moves *opposite* to the direction of motion of the object (the image of the oval begins to the right of F and ends up to the left of F), across an otherwise unchanging background.

But now suppose that the observer decides to track the moving object with her eyes. In this case, shown in Figure 21.1D, the image of the oval remains stationary on the observer's fovea, while the image of the background moves across the retina in the *same* direction as the physical object motion (the image of the oval begins to the left of F and ends up to the right of F). Remarkably,

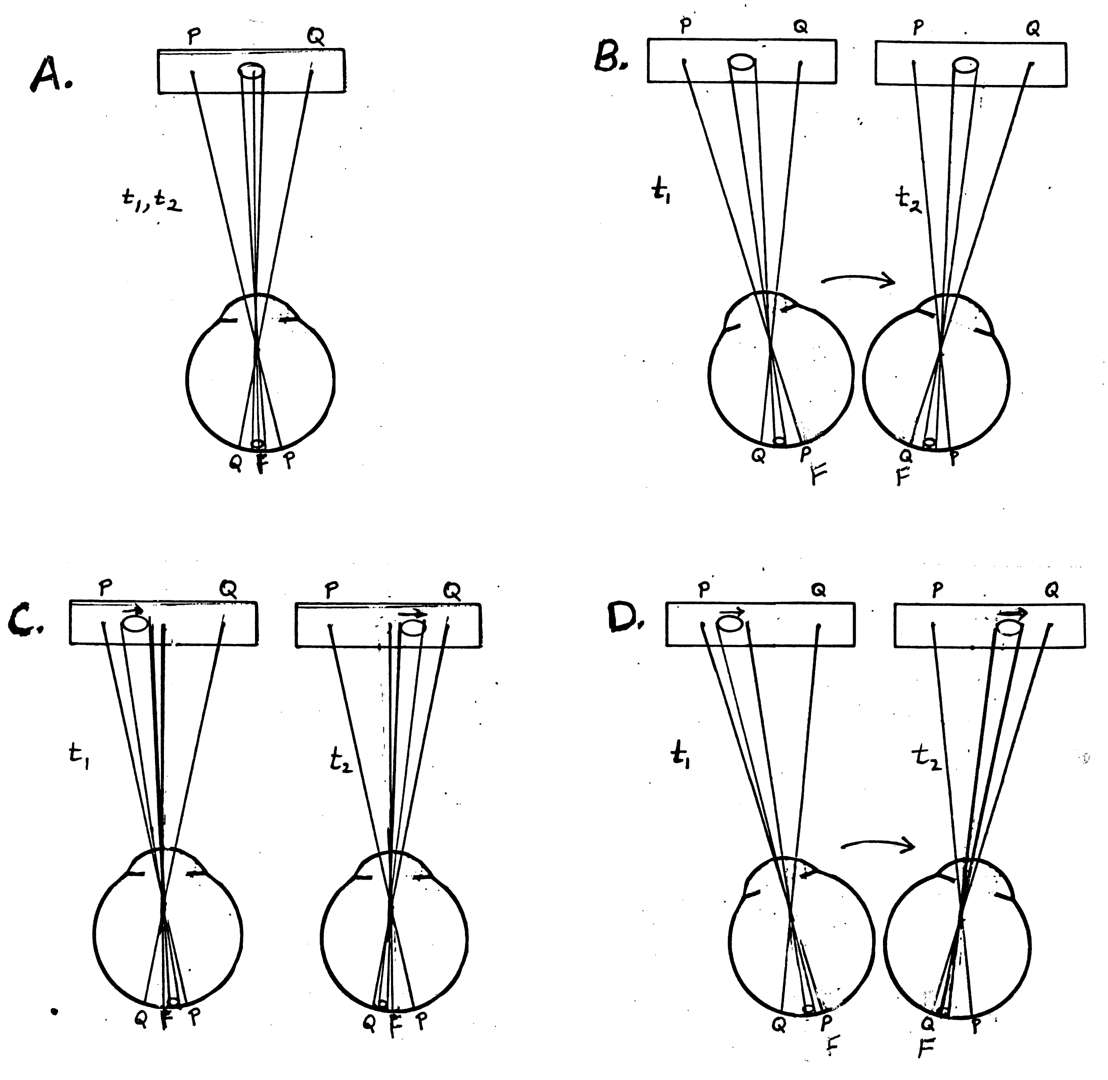


Figure 21.1: The ambiguity of retinal image motion as a cue to object motion. A: an observer viewing a stationary object (the oval) against a stationary background with a stationary eye. The oval is fixated, and imaged on the fovea at F. B: the observer changes fixation from point P to point Q. The image of the oval moves across the retina. C: the observer holds fixation while the object moves across the stationary background. The image of the oval moves across the retina. D: the same scene, but the observer tracks the motion of the object, and the image of the object is stationary on the retina. A and B lead to very different changes in the retinal image, but both lead to the perception of a stationary object; similarly, C and D lead to very different changes in the retinal image, but both lead to the perception of a moving object. In all four cases perceived motion is veridical.

again, the two spatiotemporal patterns in the retinal image in Figure 21.1C and D lead to the same (veridical) perception of the world: we see an object moving against a stationary background. [Try these conditions out yourself, by using your finger as the stationary or moving object, and either tracking it or fixating on another object.]

A third aspect of the confounding of variables in retinal image motion concerns the *speed* of retinal image motion: when distance varies, retinal speed is not well tied to object speed. That is, for a fixed physical speed, the retinal image speed decreases with the distance of the object from the observer. Yet across a wide variety of circumstances, perceived speed varies in good correspondence with physical speed, despite the confounding of retinal image speed by distance.

In summary, the veridical perception of object motion is a hard problem, not the least because many physical variables are confounded in the motion of elements in the two-dimensional retinal image. It seems likely that the analysis of motion, and the extraction of motion cues, are going to be complex, and probably involve many levels of processing.

21.2 Local motion

[I can't think of any way to get into local motion other than the pseudo- historical argument given below. I'd love to have better, or additional, suggestions. Xx]

At the same time, there is an emerging consensus that a sensible first step would be to build a set of detectors that code the direction of motion of local image elements over local regions of the retina. Later levels of processing could then combine inputs from selected subsets of these local motion detectors, to make neurons that are selective for larger and more complex motion patterns. How might one find evidence for the initial encoding of the direction of local motion across small regions of the retinal image?

21.2.1 Are there independent channels tuned for the direction of motion?

As we discussed in Chapter xx and xx, in the 1960s and 1970s, visual neurophysiologists had discovered that monkey V1 cortex contains neurons selective for the orientations of bars and edges, and for the spatial frequencies of gratings. Subsequently, visual psychophysicists had found that different spatial frequencies and orientations are detected independently, and had developed a variety of multiple channels theories of early visual processing. So the question arose: might the V1 neurons observed by Hubel and Weisel, De Valois, and others provide the physiological instantiation of the psychophysically observed (or inferred) channels? As we have discussed, comparisons of psychophysical and physiological estimates of orientation bandwidths and spatial frequency bandwidths encouraged theorists to pursue the analogy.

Now, in the same studies in which orientation and spatial frequency tuning were discovered, it was also observed that V1 neurons are selective for the direction of motion – remember Hubel's story of slipping the slide out of the projector and having an previously silent V1 neuron suddenly put out a burst of spikes. This observation leads to the question: might psychophysical paradigms like those we used to provide evidence for independent spatial frequency and orientation channels, be used to provide evidence for independent channels for the detection of different directions of motion?

In 1980, Andrew Watson, Peter Thompson, Brian Murphy, and Jacob Nachmias carried out two studies designed to test this question. In the first study they used the *summation-at-threshold* paradigm (See Chapter xx). That is, contrast thresholds were measured for leftward moving gratings, rightward moving gratings, and both component gratings superimposed. If the two directions of motion are detected by the same detector, the compound stimulus should be more detectable than either of its components. Alternatively, if the two directions of motion are detected by independent detectors, the compound stimulus should be no more detectable than the more detectable of its two components. Watson and his colleagues found that the latter was the case, and concluded that moving gratings are detected via independent detectors that are selective for the direction of motion.

In their second study, Watson and his colleagues used a paradigm we have not introduced previously: the *identification-to-detection ratio.* To understand this paradigm, we need to distinguish between two tasks: the *detection (D) of a moving stimulus*, and the *identification of the direction of motion (M) of the stimulus.* The goal of the experiment is to measure both detection thresholds and identification (direction-of-motion) thresholds, in order to determine the *M/D ratio* – the ratio between contrast thresholds for identification and for detection.

The psychophysical technique Watson et al used is called a *2 x 2 forced choice method.* In this technique, the subject views a display containing two potential locations for the stimulus. On each trial, the stimulus is presented in either the left or the right location, and it is either the leftward-moving or the rightward-moving grating. The subject's task is to report *both* the location and the direction of motion of the stimulus. The contrast of the stimulus is varied over trials until contrast thresholds for both detection and identification of the direction of motion are determined. Watson et al found that for most combinations of spatial and temporal frequency, M/D ratios were near 1.

The theoretical interpretation of this result is as follows. We assume that detection of the moving grating is carried out by the most sensitive available detector. If it takes no more contrast to identify the grating than it does to detect it – if the M/D ratio is near 1 – the same detector that detects the moving grating must also be signaling its direction of motion. In short, these detectors do not just signal the presence of a moving grating; they also provide a code for the direction of motion.[2]

In sum, there is currently a general consensus on the initial encoding of motion. Many V1 neurons are selective for the direction of motion of a stimulus within their receptive fields. Thus V1 neurons can be seen as a sensible set of local motion detectors that could provide inputs to hierarchical models of higher level motion processing. Moreover, paradigms imported from multiple channels theory readily gave evidence for the independent detection of motion in different directions. Thus, multiple channels theory can be readily extended to include channels tuned not only to the orientations and spatial frequencies of local regions of the retinal image, but also to the direction of local motion. Sounds like a winner!

[Are there similar experiments for speed? Can speed be discriminated at detection threshold? Xx]

21.3 Two classical illusions of motion perception

We now turn to the question of illusions of motion. Throughout the history of studies of motion perception, two illusions – apparent (sampled) motion, and motion after-effects – have been widely

[2] Detectors that also signal other properties of the stimulus, such as its spatial frequency, orientation, and/or direction of motion, are sometimes called *labeled lines.*

appreciated. As is often the case with illusions, these seemingly perverse system properties have been historically important because they have placed interesting constraints – or at least led to interesting speculations – on the forms that the computational and physiological analysis of motion might take.

21.3.1 Apparent motion and sampled motion

The first classical illusion of motion perception is the phenomenon of *apparent motion.* A pair of discrete flashes of a light or pattern, presented at two different locations in space with an appropriate time delay between them, creates the striking illusion that the light or pattern is moving between the two locations.

The phenomenon of apparent motion has been illustrated to a pre-caffeinated DT on many dark winter mornings. Looking out the window, she could often see a construction crane in the middle distance (Figure 21.2A). There were three strobe lights mounted on the crane in different locations, but the strobes were not synchronized, and the temporal phases of the three flashes drifted over time. So for a while lights A and B would jump back and forth and C would be stationary; then they would all flash in synchrony for a while; then they would all go on a wild three-way chase; and so on. (Amazing how little it takes to amuse a groggy vision scientist over morning coffee.)

A closely related illusion is that of *sampled motion*, shown in Figure 21.2B. In sampled motion, an ordered sequence of flashes of a stationary light or pattern, systematically spread out across space and time, creates the perception of smooth, continuous motion. For example, the lights you perceive to move on a theater marquee are actually stationary lights flashed in a well-chosen spatiotemporal sequence. For a demonstration even closer to home, go to the movies or watch television, because the motion seen in movies and on TV arises from temporal sequences of still pictures. [Consider that an assignment!]

The perceptual similarity of smooth and sampled motion raises two questions. First, what characteristics lead to the perception of equal speeds in smooth and sampled motion displays? In fact, the answer is as intuitively expected; if an element in each display covers the same distance in the same amount of time, then whether it moves continuously or jumps repeatedly along its trajectory doesn't matter – the same speed will be seen. This perceptual similarity suggests that smooth and sampled motion stimuli set up similar speed signals within the visual system.

And second, can smooth and sampled motion be indiscriminable? This question is interesting because if so, we have found a new example of metamerism (see Chapter xx on scotopic metamers and Chapter 7 on trichromacy). In 1983, Andrew Watson, Albert Ahumada and Joyce Farrell tested this question in the laboratory. They used a video system to present a stimulus (a line or grating) moving at a fixed speed across the video screen, but in either smooth or sampled motion. (The ”smooth” motion was actually also sampled, but at a rate near 2000 Hz – nearly two orders of magnitude beyond the temporal resolution of the eye – and thus was smooth enough for their purposes). They then varied the temporal sampling rate of the sampled motion stimulus.

The smooth and sampled motion were presented on randomly interleaved trials in a two-alternative forced-choice experiment, and the subject's task was to report whether the motion was smooth or sampled on each trial. Watson and his colleagues found that under the right spatiotemporal conditions, the subjects' performance dropped toward chance. That is, smooth and sampled motion can produce true metamers, implying a loss of information within the visual system.

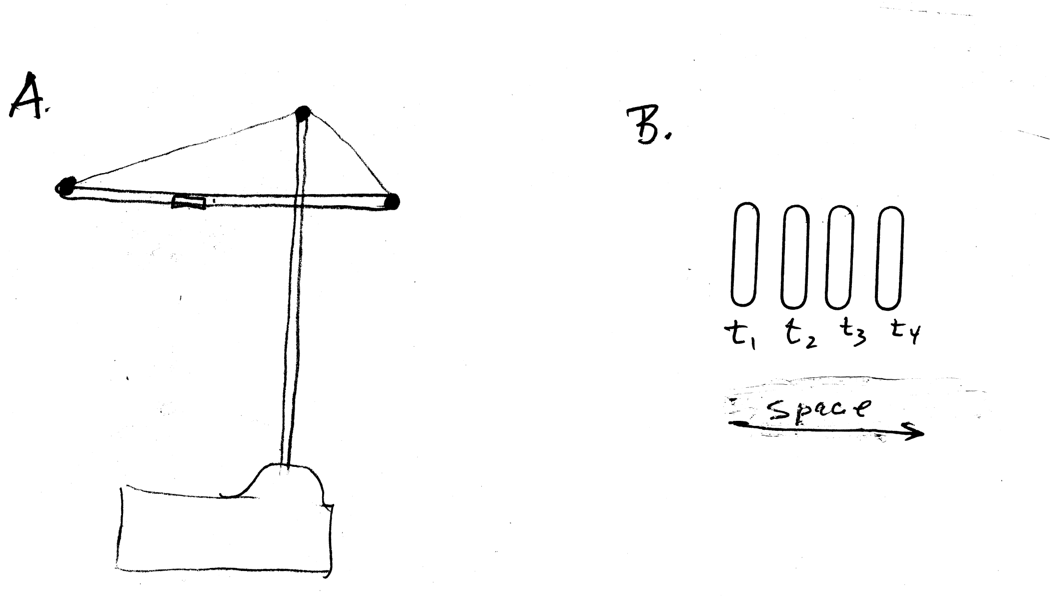


Figure 21.2: Apparent motion and sampled motion. A. Apparent motion occurs when two stimuli are flashed in different locations with a well-chosen time delay between them. A construction crane with three asynchronous strobe lights gives a striking illustration of apparent motion. As the phase relationships change among the three lights, the perceived pattern of motion changes. B. Sampled motion occurs when a set of similar stimuli are flashed in an ordered sequence of spatial positions over time. The perception is similar to that evoked by a single continuously moving object. Within the right range of spatiotemporal parameters, the perceptions evoked by continuous (smooth) and sampled motion stimuli can be indistinguishable.

21.3.2 Motion aftereffects [swb 155]

A second classical illusion of motion perception is called the *motion after-effect*, or *MAE*. It is also called the *waterfall illusion*, after the earliest reports of the phenomenon. If you stare for a minute or so at an object or scene (such as a waterfall) that is moving consistently in one direction, and then look away at a stationary scene (such as the river bank), you will perceive the stationary scene to be moving in the opposite direction.

[ADD ONE LAB EXPERIMENT on MAE HERE? – WHICH ONE?? XX]

21.4 Computational models

Up to this point, we have discussed three classical system properties of the perception of motion. The first is the evidence for the presence of local motion detectors – channels or neurons that respond differentially to different directions of local motion. The second and third are illusions of motion: the perceptual similarity (and even metamerism) of smooth and sampled motion, and the presence of motion after-effects. We now take each of these three system properties and ask, given each system property, what neural computations might produce it?

Of course, in reasoning (or speculating) from perception to computation, the goal of the computations is to generate a mathematical or neural description that can account for the system property with which we started. In her usual style, DT will argue that the account will always rest on an assumption about how the mathametical or neural description maps to the perceptual description, and thus linking propositions will always be involved in computational as well as neurophysiological models. So when computational theories are proposed, one fundamental question always is, what linking propositions are involved in the arguments from psychophysical results to putative neural computations? Or alternatively, what new linking propositions, or new characteristics of linking propositions, emerge from computational models of the properties of motion perception?

21.4.1 Encoding the direction of motion: Delay-and-compare networks

The first characteristic of motion perception is that we perceive local motion, and can distinguish its direction. What clues to computations does this provide, and via what linking propositions? Based on the Universal linking proposition, we know that sufficient information to support these perceptual accomplishments is carried, in some code or other, through every stage of a pathway that extends through all levels of the visual system. Moreover, based on the Neuron Doctrine, the fact that we see objects move in particular directions suggests that the visual system is likely to contains neurons that are selective for the direction of motion; and the fact that we see *continuous* motion suggests that *continuous* physical motion is likely to set up *continuous* activity in these neurons. In fact, classic studies of V1 neurons reveal many neurons with these properties.

In computational terms, how can we make a directionally-selective neuron? The earliest models of direction selectivity were based on modeling elements called *delay-and-compare networks* (Reichardt, 1961). Two different versions of simple delay-and-compare networks are shown in Figure 21.3. Both of these networks create neurons that respond to rightward but not leftward motion.

Figure 21.3A shows a delay-and-compare network made with only excitatory components. The inputs to the network are two hypothetical neurons or groups of neurons, A and B, with the receptive fields in A located to the left of those in B. The signals from A and B are combined

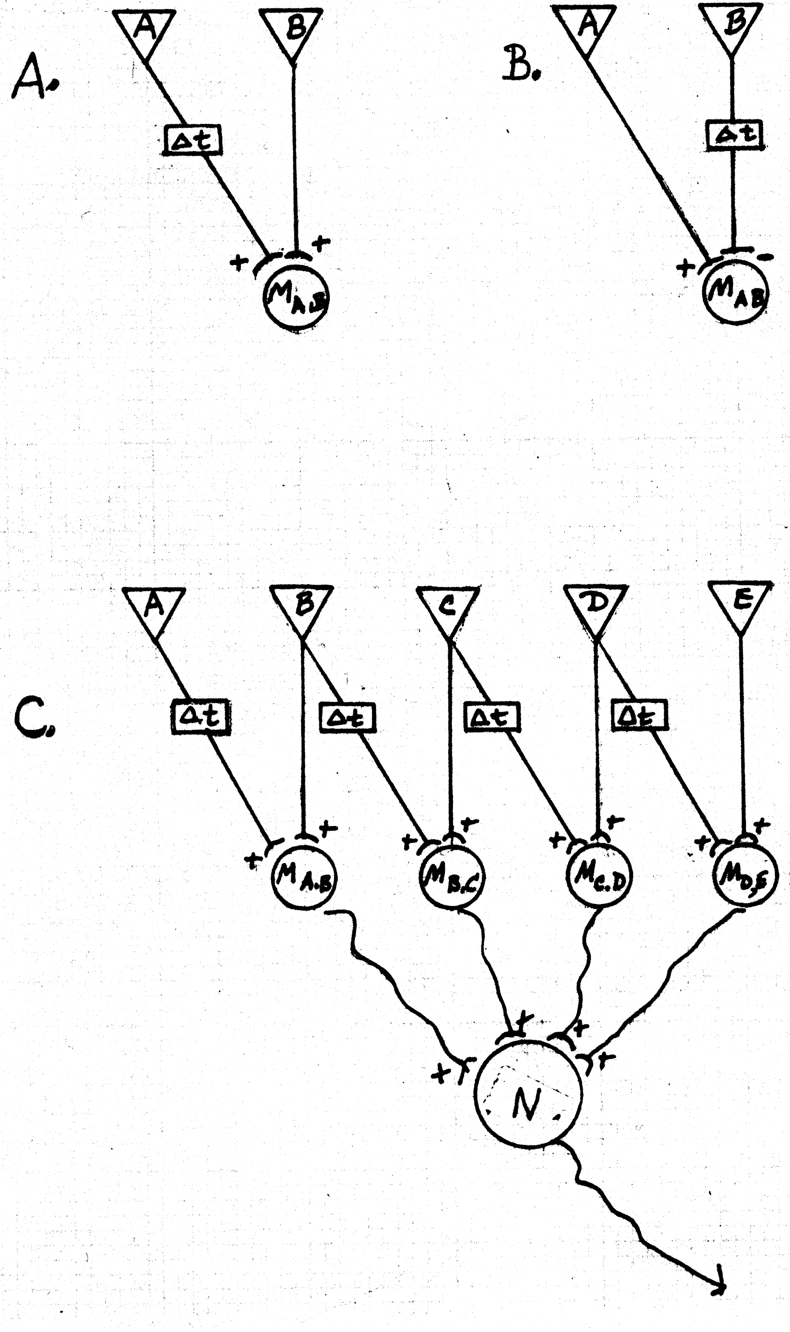


Figure 21.3: Delay-and-compare networks. A: A delay-and-compare network made from two excitatory inputs. B: A delay-and-compare network made with one excitatory and one inhibitory input. In both cases, the neuron $\mathrm{M}_{A,B}$ will respond to left-to-right but not right-to-left motion, as a light moves over the very short trajectory from A to B. C: A delay-and-compare network with a neuron N that will respond continuously while the light moves along the trajectory from A to E.

(*compared*) at an output neuron M. Two additional assumptions make this model work. The first is that the signal from B arrives at M without appreciable delay, whereas the signal from A is *delayed* by a time delay Δt before it reaches M. The second is that to activate M the signals from A and B must arrive at M simultaneously (or nearly so). The upshot is that a spot of light that moves rightward at just the right speed, passing across A and then B with just the right time delay in between, will activate neuron M. M will be activated if the neural time delay Δt is matched to the speed of the motion, such that the two signals arrive simultaneously at M. In contrast, a spot that moves leftward from B to A will generate two input signals to M that are separated in time, and M will not respond. In short, the delay-and-compare network yields an output neuron rather narrowly tuned to a particular speed and direction of motion in a local retinal region.

Figure 21.3B shows a second version of a delay-and-compare network, in which an inhibitory element is used. In this case, it is the signal from B that both undergoes the time delay and provides the inhibitory input. For this network, a point of light moving leftward across B and then A will initiate a signal from B, followed by a signal from A. If the time delay for the signal from B is matched appropriately to the speed of motion, both signals will arrive at M simultaneously, and cancel each other, so M will *not* respond to leftward motion. For rightward motion, the signal from A is initiated first, and has time to initiate a signal in M before the arrival of the inhibitory input from B. Notice that in this case a single input from A is assumed to be sufficient to initiate a signal in M, as long as it is not cancelled by a signal from B. So this network is not particularly well tuned to the direction of motion, but it does have a clear null direction – leftward. [Using delay and compare networks, design a set of neurons selective for different directions and speeds of motion. What features of these models create the direction tuning? The speed tuning?]

The circuits shown in Figure 21.3A and B are specialized to respond to motion over local retinal regions. A generalization of the delay-and-compare scheme is shown in Figure 21.3C. Here each group of input neurons contributes to one motion neuron directly, and to another with a time delay. A stimulus moving continuously across the retina would activate each first-level motion neuron (M) in turn. Summation of outputs from all of these first-level motion neurons at a second level would create a neuron N that would respond continuously as the light moved along the whole trajectory between A and E.

More sophisticated motion models, often called *motion energy models*, have also been developed. These models build upon multiple channel models of visual processing. They begin with neurons that already have elongated, spatially opponent receptive fields, and are thus already selective for particular spatial frequencies and orientations. In consequence, in these models, motion is analyzed separately for each spatial frequency and each orientation. The essential feature of these models is that signals from different subparts of the centers and surrounds of the input neurons arrive at the motion neurons with different time delays, essentially making local delay-and-compare networks. [However, it's not easy to come up with spatiotemporal schemes that actually work. Give it a shot, and then see Adelson and Bergen, 1985 or Mather, 19xx].

In summary, delay-and-compare networks are attractive because the provide an intuitively accessible scheme for creating local direction-selective neurons. The outputs of such direction-tuned neurons could then be combined hierarchically in many different ways, to create higher-level neurons selective for more complex patterns of motion.

21.4.2 Creating the metamerism of smooth vs. sampled motion

We turn next to the question of smooth vs. sampled motion. Earlier we claimed that motion metamers exist – under some conditions, physically smooth and sampled motion can be perceptually indistinguishable.

What linking proposition will be involved in an explanation of motion metamers? As before, arguments arising from the occurrence of metamers rely on Identity linking propositions. Physically smooth and sampled motion lead to different spatiotemporal patterns in the retinal image, and presumably in the photoreceptor outputs and at some of the higher levels of processing. But the perceptual metamerism of smooth and sampled motion strongly suggests (some would say logically implies) that somewhere within the visual system, the neural signals arising from smooth and sampled motion are rendered identical. The fact that smooth and sampled motion are different in the physical world, and different in the retinal image, is lost to perception, just as differences in wavelength composition are lost in the cases of rod vision and trichromatic color vision. A desirable characteristic of a computational motion model, then, would be one that renders identical the signals arising from physically smooth and sampled motion, in the spatiotemporal domains in which perceptual motion metamers occur.

How and under what conditions should smooth and sampled motion yield motion metamers? In order to be able to think about this problem, let's simplify it one more time. Let's consider a point or line of light that moves in only one dimension – say, up or down. We can represent the motion of such a point in a diagram called a *space/time*, or *x,t plot*, as shown in Figure 21.4. In this plot, time is represented on the horizontal axis, and a single dimension of space – here, up and down – is represented on the vertical axis. The x,t plot describes the location of the point of light as a function of time.

Figure 21.4A shows x,t plots of the motion of a single point, moving smoothly up or down at various speeds. A stationary point creates a horizontal line (3) in the x,t plot. A point moving slowly upward (2) yields a line with a shallow but positive slope, whereas the same point moving downward at the same speed (4) yields a line with a shallow but negative slope. Faster motion yields lines of steeper slope (1 and 5). In short, both speed and direction are represented in the slopes of lines in the x,t plot.

Figure 21.4B-E shows x,t plots for a short vertical line (rather than a point) moving downward. The motion depicted in Figure 21.4B is smooth, whereas that in Figure 21.4C-E is sampled at three different sampling rates. All three of these stimuli have the same slope as the continuous motion stimulus in B, and all are perceived to have the same speed and direction of motion. But the higher the sampling rate, the smaller will be the changes in spatial location, and the smaller the gaps in time, between one sample and the next.

We now bring back our old friends – the spatial and temporal contrast sensitivity functions (CSFs and tCSFs, Chapter xx). Recall that both the CSF and the tCSF are low pass. As the temporal sampling rate increases, the lines in the retinal image will get closer and closer together, and eventually the spatial frequency will exceed the high spatial frequency cutoff of the CSF. Similarly, the temporal frequency created by temporal sampling will exceed the high temporal frequency cut-off of the tCSF. When both these limits are exceeded, smooth and sampled motion should become indiscriminable.

Formal quantitative predictions can be made by a process analogous to that shown in Figure xx [Fourier analysis in Spatial Vision Chapter]. First, Fourier analysis is used to separate out

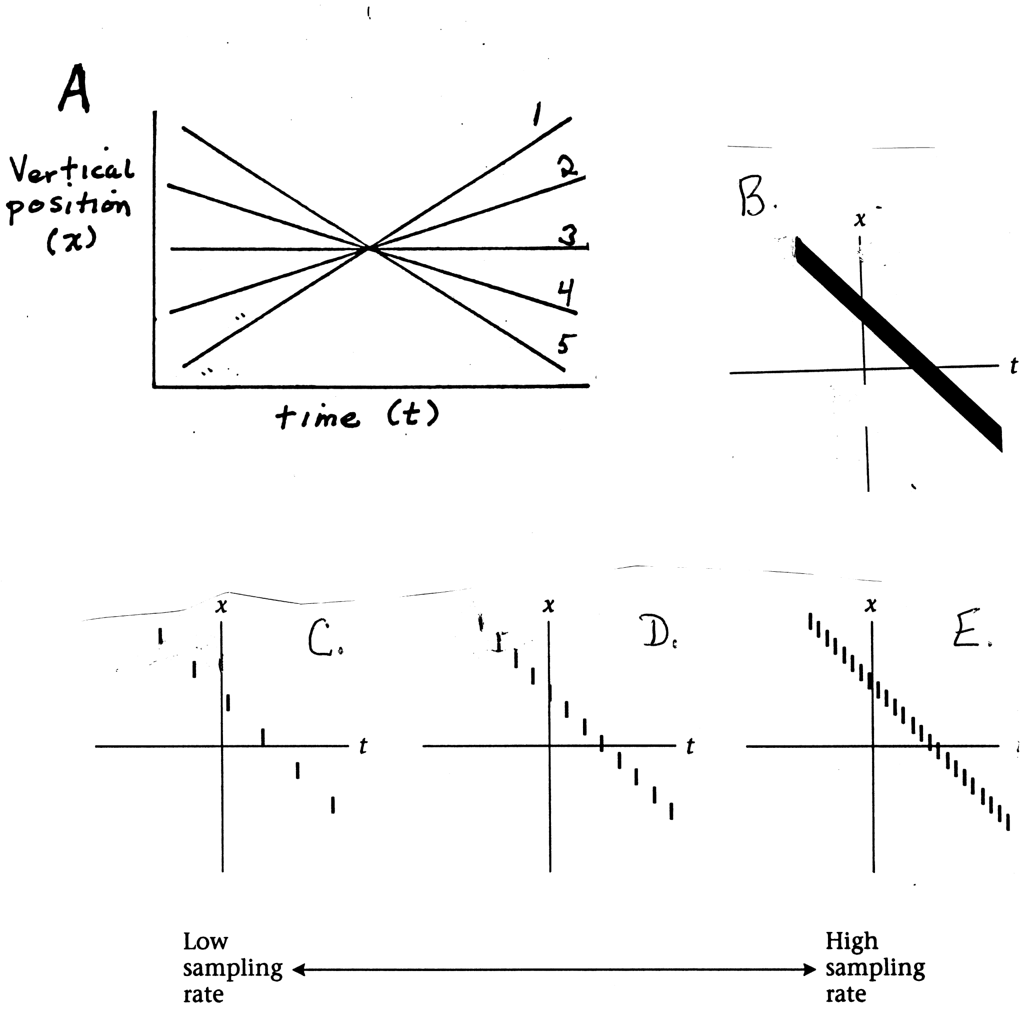


Figure 21.4: Space/time or x,t plots. In all of these plots, a single spatial dimension, x (say, up-down), is plotted against time. A: The lines in the x,t plot show the velocity – the speed and direction – of a point of light: (1) rapid upward, (2) slow upward, (3) stationary (no motion), (4) slow downward, and (5) rapid downward. B: a short vertical line moving smoothly and rapidly downward. C, D, and E: sampled motion with low, moderate, and high temporal sampling rates, all matched to the smooth motion in the change of location in the elapsed time. All four stimuli – B, C, D, and E – are perceived to move with the same velocity. At sufficiently high sampling rates, smooth and sampled motion become indistinguishable. [A drawn by DT; B-E after Wandell, 1995, p. 348.]

the spatial and temporal frequency spectra of the continuous and sampled stimuli. Second, these stimuli are passed through the spatio-temporal *window of visibility*, as represented by the spatial and temporal contrast sensitivity functions. Fourier components with spatial and/or temporal frequencies beyond the high-frequency cutoffs will be lost, and their loss produces motion metamers. In sum, smooth and sampled stimuli that differ only in spatial and temporal frequencies that lie above the high-frequency cutoffs of the CSF and tCSF should be indiscriminable.

These arguments are qualitatively highly attractive, but the jury is still out on whether or not they will prove quantitatively sufficient to account for the sampling rate at which sampled motion becomes indiscriminable from smooth motion. It may be that moving stimuli suffer further losses of spatial and/or temporal resolution, beyond those imposed by the window of visibility for stationary stimuli.

21.4.3 Creating motion aftereffects [SWB 155]

Motion after-effects (MAE) are adaptation phenomena. As we saw in Ch xx (on Adaptation), models of adaptation effects always rest on the assumption that at one or more sites, the visual system is in a different state after the adaptation than it is before, so that the same set of inputs yields a different constellation of neural activity at and beyond the site(s) of adaptation.

The existence of motion aftereffects can be explained qualitatively with a combination of three assumptions. The first is that there are separate neurons tuned for motion in different directions. The second is that these neurons are adaptable: when you stare at a stimulus moving in a single direction, the population of neurons tuned for that direction and nearby directions of motion become temporarily less sensitive.

The third assumption is that the *population code* is changed by the differential adaptation of the neurons tuned to motion in a particular direction. When you first stare at, say, rightward motion, the responsiveness of neurons tuned to rightward motion is reduced. When you subsequently stare at a stationary scene, you see it with a neural population that gives a reduced signal from the neurons tuned to rightward motion. This adaptation distorts the neural code in the population as a whole. Thus, two population codes become similar: the code that arises from (say) leftward motion in the unadapted system, and the code that arises from a stationary stimulus after the system is adapted to (say) rightward motion[3]. [What is the linking proposition in this argument? What family does it come from?]

In summary, three system properties of motion perception have been so historically compelling that they have left indelible marks on computational models of motion. These models incorporate modeling elements you have seen before. First and most basically, as perceivers we discriminate remarkably well among different directions of motion. In early motion models, delay-and-compare networks are used to create neurons selective for the direction of motion. In more recent motion energy models, multiple channels theory is extended to posit that the visual system creates individual detectors selective not only for different spatial frequencies and orientations, but also for different directions of motion.

[3] A common more specific model of the MAE is that detectors for motion in the two opposite directions – say, left and right – are coupled in a subtractive, or opponent fashion, much like the opponent coding in color vision discussed in Chapter xx. These pairs of detectors are assumed to combine their signals with a ”winner-take-all” strategy – whichever set of motion detectors has the stronger signal determines the perceived direction of motion. However, there is no particular evidence for such specific opponent coupling, and in DT's view the more general argument based on population codes is the more appealing.

Second, to account for the metamerism of smooth and sampled motion, the high-frequency cutoffs of the psychophysical CSFs and tCSFs of human vision are brought into play. The argument is that the signals that arise from smooth and sampled motion stimuli are rendered identical because they lose their high spatial and temporal frequency components.

And third, to account for motion after-effects, a population code is invoked. We assume that differential adaptation changes the population response of a set of motion-tuned neurons. The changes take the form that a leftward moving stimulus seen by the unadapted system, and a stationary stimulus seen by the right-adapted system, produce similar states of the population code, and are thus perceived as similar.

21.5 More complex motion patterns

Up to this point, for the sake of simplicity, we have discussed only the analysis of local motion and the existence of two classical motion illusions. But there is more to the story of motion. Physical objects are three-dimensional, and they undergo complex patterns of motion, such as motion in distance and rotations about their own axes. Some objects are rigid, whereas others like our bluejay also undergo changes of three-dimensional shape.

In each case, the motions of three-dimensional objects in three-dimensional space are inevitably reduced to two-dimensional patterns of motion in the retinal image. This collapsing from three to two dimensions creates a massive information loss. Nonetheless, it could still be the case that different kinds of physical object motion – changes of distance, rotations of rigid objects, changes in object shape – create different kinds of spatiotemporal patterns – motion patterns – in the retinal image. If we could sort them out, these differing motion patterns could potentially still serve as cues to the three-dimensional trajectories, shapes, and identities of objects. In combination with clever heuristics, these patterns may even enable accurate guesses about trajectories, shapes, and identities. So it is interesting to ask, more specifically, what are the commonalities in the patterns of retinal image motion caused by each of the various types of motion? And is there evidence that our visual systems are sensitive to these motion patterns?

21.5.1 Stochastic motion stimuli

We begin with a category of stimuli called *stochastic motion stimuli* (Williams and Sekuler, 1984). A set of three stochastic motion stimuli are shown in Figure 21.5. Unlike the sinusoidal gratings used to study local motion, a stochastic motion stimulus consists of a field of disconnected dots. Between frames of the display, each dot is displaced along a different trajectory. The manipulated parameter is the degree of *correlation* of motion between frames: the fraction of the dots that are displaced in the same direction. The left, middle, and right panels of Figure 21.5 depict correlations of 0%, 50%, and 100% respectively. For human subjects, as the correlation increases upward from zero, the perception of random movements yields to the impression of an overall directionality, and eventually to clear perceptions of motion flowing in the correlated direction.

Stochastic motion stimuli are theoretically attractive for at least two reasons. First, to perceive the predominant direction of motion in a stochastic motion stimulus, it is not enough to know the local motion of the individual dots. Rather, it requires the sorting out of directions, and the integration of motion information across broad retinal regions. Thus, these stimuli are designed to

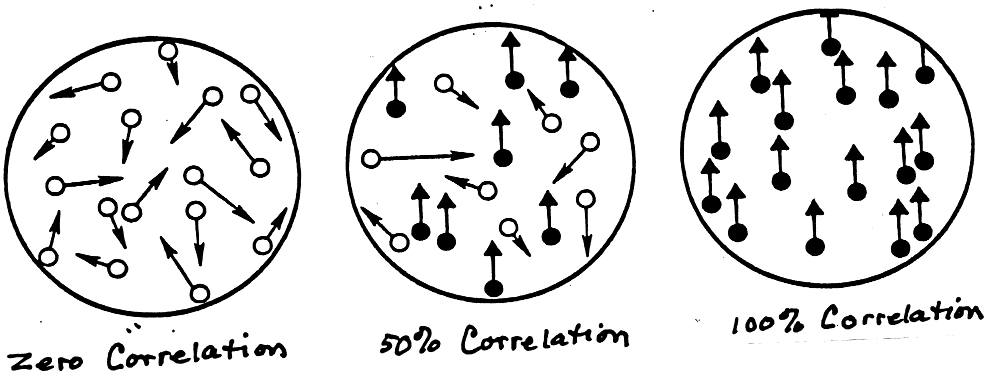


Figure 21.5: Stochastic motion stimuli. Each panel depicts two frames of a video display. In frame 1 the dots have the locations indicated by the circles, and in frame 2 they are replaced by dots at the arrow tips. The experimenter varies the degree of *correlation* between the two frames; that is, the fraction of dots whose replacement dots are all displaced in the same direction. The correlated dots are shown in black. The left, middle, and right panels show correlations of 0%, 50%, and 100% respectively. The observer's task is to report the direction of displacement of the correlated set of dots (here, upward). Stochastic motion stimuli are used to study the integration of perceived motion across spatially separated stimulus elements. (Modified from Salzman et al, 1992).

call upon a higher level of motion processing – call it *motion integration* – than are the edges and gratings used to test the properties of local motion detectors.

And second, it is possible to do forced-choice experiments (Chapter 2) with stochastic motion stimuli by measuring *motion coherence thresholds*. Suppose that the correlated dots can move in either of two opposite directions, such as leftward vs. rightward. The percentage of correlated dots can be varied from trial to trial, and the subject's task is to report the direction of motion of the correlated dots. The motion coherence threshold is the percentage of coherence required for the subject to report the direction of coherent motion correctly on (say) 75% of the trials. Motion coherence thresholds depend on a variety of parameters, but a good rule of thumb is that coherent motion of about 5% is detectable.

As you will see, stochastic motion stimuli are frequently used to study motion integration in joint psychophysical and neurophysiological experiments on monkey subjects. The fact that they can be used in forced-choice experiments leads to a relative simplicity in training the animals and in interpreting the data. We will see the data from stochastic motion experiments on monkey subjects in the next chapter.

21.5.2 Changes of distance

What patterns of retinal image motion are generated by variations in distance – translations in the third spatial dimension? Suppose you are fixating a soccer ball, and someone kicks it directly at your head. Figure 21.6A shows the *motion flow field* – the pattern of retinal image motion created by the soccer ball. The image of the moving ball will create a looming pattern – a pattern that expands symmetrically on your retina, with the center of expansion determined by your point of fixation. As the ball moves toward you, its angular subtense will increase, and thus it will get larger in the retinal image. The center point will remain stationary in your retinal image, and the farther any other point on the ball is from fixation, the faster its image will move across your retina. In addition, as the image of the ball expands, it will occlude more and more of the image of the surrounding visual scene. Similarly, Figure 21.6B shows the pattern created by an object moving away from you.

Similarly, what happens if you walk directly forward in a complex visual scene, while fixating at its center? Figure 21.6C shows the motion flow field. As in the case of the soccer ball, physical objects will make expansion patterns, and the center of expansion will still be determined by your point of fixation. But the retinal motion flow fields created by the non-fixated objects in the scene will be asymmetrical, and the expanding region of occlusion that arose from the edge of the moving ball now occurs for each object. In addition, as the distance decreases and the scene expands in the retinal image, the edges of the scene are lost as they pass beyond the edges of your visual field. Figure 21.6D shows the change in the motion flow field created by shifting fixation to the right-hand rectangle.

For the patterns produced both by approaching objects and by self-motion, complications of the viewing conditions will yield more complex patterns of retinal image motion. For example, the looming pattern created by the soccer ball change if you are not fixating the ball, or if the ball is on a trajectory that will make it hit your chest rather than your head. Similarly, as Figure 21.6D shows, the motion flow fields that arise from locomotion change if you are not looking where you are going!

Now, the functional question is, granted that these motion flow fields and their subtle variations

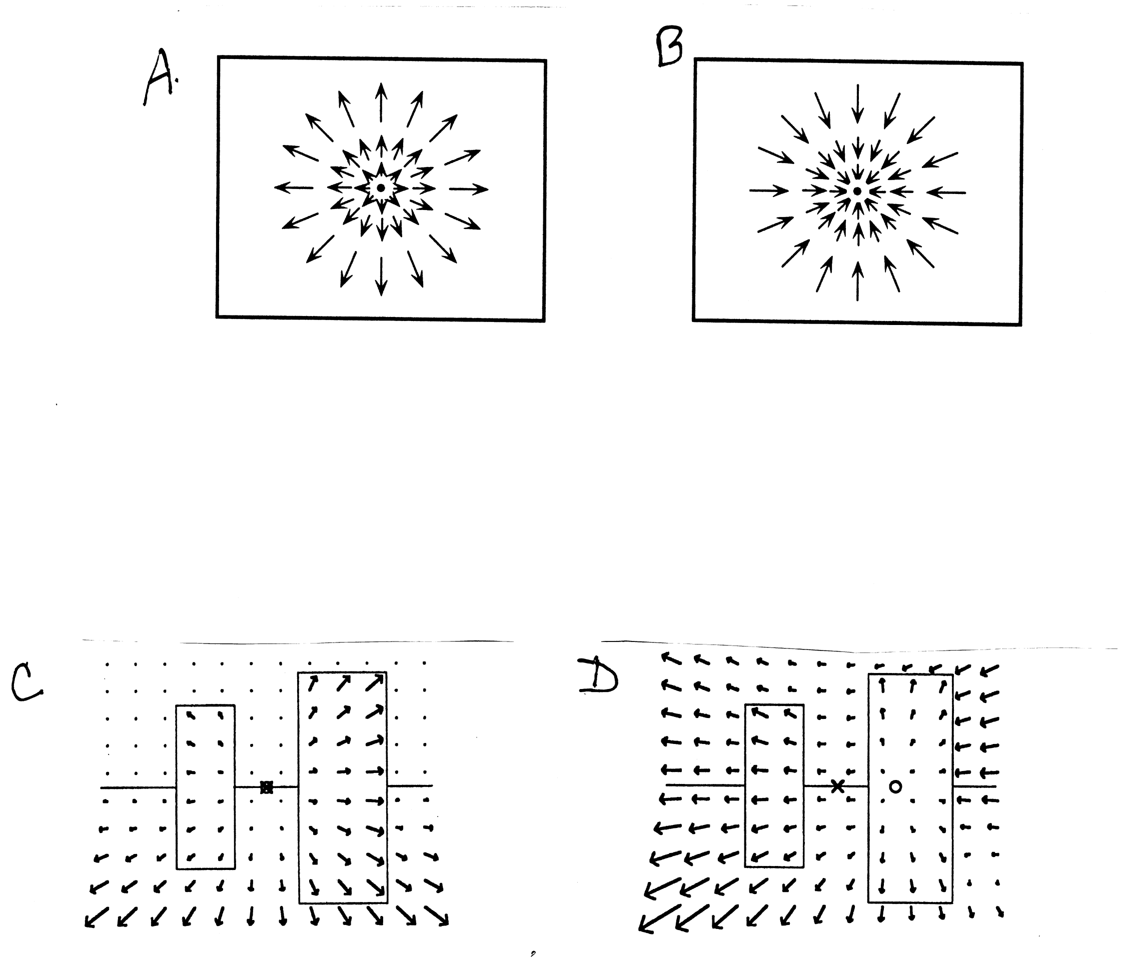


Figure 21.6: The effects of changes in distance. The arrows show motion (velocity) vectors for points of light at various locations in the retinal image. The direction of the arrow shows the direction of motion, and the length of the arrow shows the speed. A: The motion flow field produced by an approaching object fixated at its center. B: The motion flow field produced by a receding object. C: The motion flow field produced by an observer walking toward a scene containing two rectangular objects, while fixating at the X. D: The motion flow field produced when the observer continues to walk toward X but shifts fixation from X to O. [After Palmer, 1999, p. 507].

and symmetries exist, do we actually analyze them and use them as perceptual cues? Common experience suggests that we do. The soccer ball can come straight toward you or be off to the side. The fact that you catch it in both cases suggests that you process the subtleties of expansion patterns to perceive (or at least respond motorically to) the veridical trajectories of objects. Moreover, the fact that you move around successfully in the physical world suggests that the differences in motion flow fields created by locomotion are also used successfully as cues. And many laboratory studies confirm our abilities to distinguish among the different patterns of optic flow generated by changes in distance, and use them in the service of veridical perception and appropriate patterns of action.[4]

21.5.3 The rotation of rigid objects

A second important kind of motion pattern in the retinal image arises from the rotation of rigid physical objects. When a rigid object rotates, what kinds of motion patterns does it generate in the retinal image? First, all points on the object, and all points in the retinal image, will move in temporal synchrony. Beyond that, the answer is complex, and depends on the object. If the object is a homogeneous, unpattterned sphere, its retinal image will not change. But if the surface of the object is irregular in shape, the outline of the object in the retinal image will contain systematic temporal distortions of shape.

You can illustrate these effects by rotating your rigid hand in front of you and imagining the sequence of changes in the retinal image. The sequence of images will include distortions of shape; for example, your thumb will move from being seen from the side to being seen from the front. It will also include occlusions – when your thumb is on the side toward you, there may be no retinal image of your little finger.

Moreover, patterns on the surface of the object will yield patterns of motion in the retinal image that depend on the shape of the object. For example, the image will contain *velocity gradients* – systematic variations in image velocity in different parts of the image. The pattern elements near the edges of the image, since they are seen on an angled surface, will move more rapidly than those on the front face of the object.

As in the case of motion flow fields, computational theorists have made great progress in describing the general properties of retinal image motion that arises from the physical rotation of rigid objects.

[IRIS; WE NEED TO FIND A GENERAL DESCRIPTION OF THE RETINAL IMAGE PATTERNS THAT ARISE FROM RIGID ROTATION AND PUT IT HERE IN SIMPLIFIED FORM. XX.]

In sum, these patterns stand as potential cues to the three-dimensional shapes of rigid objects.

Finally, we return to the functional question: are these patterns actually used as cues to object shape? The question was first investigated by Hans Wallach and xx O'Connell in 1953. In order to eliminate other cues to three-dimensional shape, Wallach and O'Connell set up a shadow-casting arrangement, such that the subject viewed the silhouettes of rotating objects cast upon

[4] One day in the laboratory, DT inadvertently revealed the powerful effects of expansion patterns. She was controlling the size of a disk of light in a visual display by varying the diameter of an iris diaphragm. Her advisor, Tom Cornsweet, was biting a bite bar and looking into an optical system, directly at the disk. Without warning him, DT rotated the lever that opened the diaphragm, and Tom nearly tore out his teeth in his startle response to the looming pattern. (He let her have her Ph.D. anyway.)

a rear-projection screen. Observers reported that the shadow of a stationary object looked two-dimensional. But the shadows of many rotating objects looked distinctly three-dimensional, and led to strong and veridical perceptions of both the three-dimensional shape and the motion of the object. Wallach and O'Connell called this perceptual skill the kinetic depth effect (KDE). Their data showed that under many circumstances, the pattern of retinal image motion that arises from the distorting silhouette of a rotating object can be sufficient to generate a veridical perception of the shape and motion of the object.

More recently, with the advent of video displays, more complex moving stimuli can be created easily in the laboratory. The perceptual responses to such stimuli have been studied under the name *structure-from-motion*; that is, recovery of the three-dimensional structure of a simulated three-dimensional object from the motion patterns contained in the two-dimensional retinal image. In these experiments, textured objects rather than silhouettes are usually used, so that the motion of the individual texture elements can serve as cues to motion and shape. An example of a three-dimensional shape that can be reconstructed from the pattern of motion of dots on its surface is shown in Figure 21.7.

21.5.4 Non-rigid (biological) motion

A third kind of retinal image motion pattern arises from non-rigidity – the change of shape of an object. We have already seen that the rotation of a rigid object leads to particular kinds of distortion patterns in the retinal image. But non-rigid objects probably lead to similar distortion patterns. Can we tell them apart?

A very interesting class of non-rigid motion patterns, called *biological motion*, arises from the kinds of physical motion typical of animate objects, such as our blue jay flying, or a person walking. In biological motion the limbs, head and torso of the jay, or the body, arms and legs of the person, change locations with respect to each other in time. But because the parts of the animal are rigid physical objects of fixed shapes, joined together at joints, only certain patterns of non-rigid motion are possible among the parts.

These allowable physical changes will lead to highly specific patterns of motion in the retinal image. In general, the retinal image motion sequences will be characterized by synchronies constrained by the configurations of joints, having different mathematical properties than the distortions that arise from rigid rotation. [Iris – Add a more technically correct description?? Xx]

Like the motion sequences produced by variations in distance and by rotation of rigid objects, the motion sequences that arise from biological motion can be used as perceptual cues. The first evidence was provided by Gunnar Johansson (1975). Johansson attached small lights to the joints of a model, as shown in Figure 21.8. He then filmed the model walking in the dark, so that the only thing visible was the simultaneous trajectories of the various lights. When the model was stationary, subjects saw only a pattern of lights. But when the model walked, or when subjects viewed the spatiotemporal patterns created by the model walking, they immediately perceived a person walking.

More recent studies have shown that subtly different patterns can give rise to the perception of different object classes, such as male vs. female walkers, and that even individual people can often be recognized from their point light patterns. These demonstrations show that indeed, biological motion patterns carry cues that are sufficient to support veridically perceived motion, and in some cases, even specific object recognition.

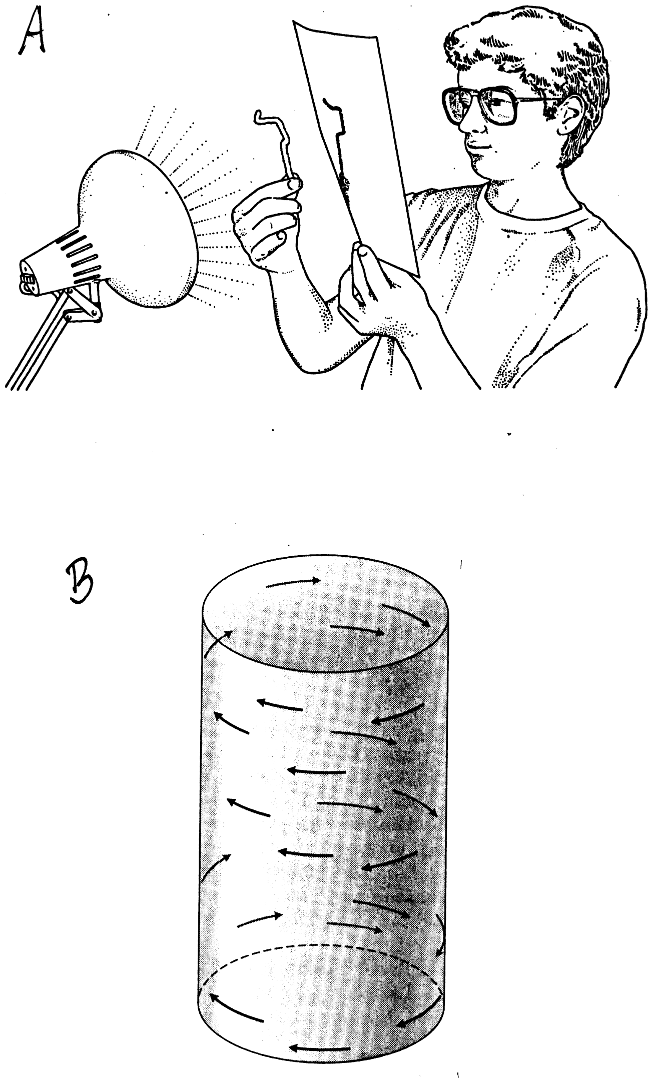


Figure 21.7: The kinetic depth effect and structure-from-motion. A: The kinetic depth effect. The shadow of a stationary object looks like a two-dimensional pattern, but rotation of the object often yields a veridical perception of booth the three-dimensional shape and the motion of the object. B: Structure-from-motion. Patterns of moving dots created by rotating a patterned object are perceived as having the three-dimensional shape implied by the motion pattern. Here a computer generates the motion flow field that would be generated by a transparent cylinder with dots on both surfaces, rotating in depth. (The subject sees only the dots, not the lines that mark the boundaries of the cylinder in the illustration.) The subject perceives a transparent cylinder rotating in depth. [A From Goldstein, 19xx, p. 392; B from Wandell, 1995, p. 362].

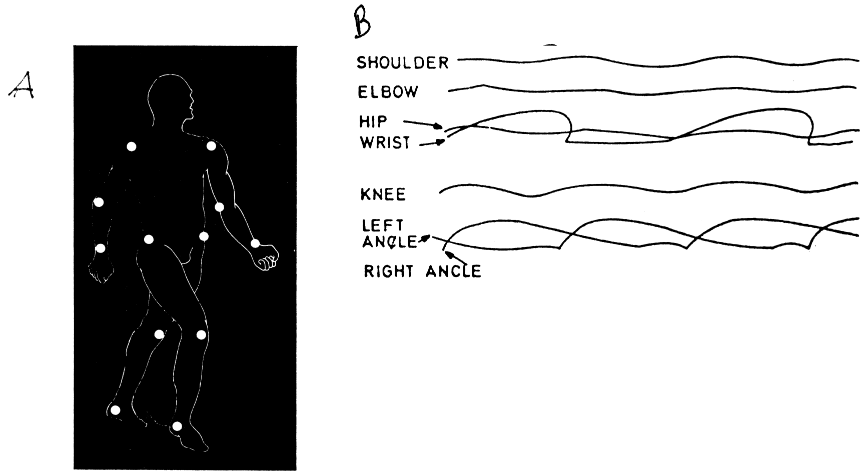


Figure 21.8: Biological motion. A: A biological motion stimulus can be created by fastening lights to the joints of a model, and having the model perform acts such as walking. B. The lights on the joints produce characteristic, phase-locked patterns of motion in the retinal image. These patterns are sufficient to create the perception of a person walking. What a set of heuristics! [A from Palmer, 1999, p. 512; B from Johansson, 1973].

In summary, different kinds of object motion lead to different kinds of motion patterns in the retinal image. Set up in isolation in the laboratory, these motion sequences are often sufficient to give rise to the veridical perception of those motions and those objects. In short, retinal image motion patterns provide powerful cues for the perception of the shapes, identities and motions of physical objects.

Of course, the motion patterns we have discussed can be combined in endless varieties. For example, imagine the pattern of image motion on the retina of a person who is walking forward and looking around, in an environment that contains other moving objects – say a group of freestyle skateboarders doing three-dimensional flips and rolls! Yet the visual system routinely parses and interprets these motion patterns, and uses them to construct veridical perceptions of rigid and non-rigid moving objects, and our own self-motion, as we walk about in the physical world.

21.6 Heuristics of motion processing

As discussed earlier, the world is underdetermined in the retinal image. Thus, strictly speaking, analysis of the sequence of changes in the retinal image will tell us only that the physical world contains a member of a particular spatiotemporal equivalence class, but not which member. Beyond that, the shape and motion of the object must be guessed on the basis of heuristics. What kinds of heuristics are used in the perception of motion?

Important constraints on interpretations come, as usual, from the nature of the physical world. Most physical objects are of fixed size and shape, and if they move, they most often move continuously along straight lines. As usual, we do not need to entertain all of the logically possible origins of a given pattern of retinal image motion. Instead, we can confine our heuristic guessing to the kinds of motions generated by physical objects in the physical world. In fact, following Helmholtz, the visual system is likely to guess that the physical object most likely to yield the particular pattern of retinal image motion, is actually the object that is present in the world.

As a first example of a motion heuristic, we have already seen that smooth and sampled motion can be indistinguishable, and explored a model for how this metamerism might come about. The question then arises, when any member of this equivalence class is present on our retinas, what shall we see? In the physical world, sampled motion is a very unlikely stimulus, whereas continuous motion is commonplace. So a sensible heuristic would be, *interpret a member of the sampled/continuous matamer set as arising from continuous motion.*

A second and third example arise from the interpretation of expansion patterns. Most physical objects have a fixed size. Thus small expansion patterns in the retinal image usually come about from changes in the distance of an object, not changes in its size. So a sensible heuristic would be: *interpret a small expansion pattern as a change of the distance of an object.* Expansion patterns for the whole retinal image usually arise from observer motion forward, so a sensible heuristic would be: *interpret a full-field expansion as self-motion.*

A fourth example arises from processing distortion patterns. As discussed above, the rotation of any object with a rigid three-dimensional shape yields a specifiable, synchronous distortion pattern in the retinal image, and the distortion patterns created by all rotating rigid objects conform to certain specifiable rules. Conversely, any distortion pattern that follows these rules probably arises from the rotation of a rigid object. Thus, the visual system probably contains a *rigidity constraint* – a heuristic something like, *whenever possible, interpret retinal image distortion patterns as due to the rotation of a rigid object.* The specifics of the distortion pattern also carry information about the specific shape of the object, and yield the shape-from-motion cue to three-dimensional shape (see Chapter 25 on depth).

A fifth example arises from biological motion. Biological motion is non-rigid motion, but it arises from the motion of rigid segments tied together flexibly at joints. The retinal image patterns that arise from biological motion also follow specific rules. xx Conversely, there is probably a heuristic to the effect that retinal image patterns that conform to these rules are perceived as moving animate objects. [Fill in xx]

A final example arises from the case of periodic occlusion. Consider an image that appears and disappears along a long trajectory. What shall we see? Since physical objects are continuous in space and time, such an image is probably most often caused by a continuous object undergoing periodic occlusion, such as a rabbit hopping through a grove of trees. The heuristic, *interpret appearance and disappearance along a trajectory as a moving object undergoing occlusion*, has probably bought many predators to a meal and prey to an untimely end.

Finally, two points. First, it is worth remembering that all of these heuristics create illusions if the retinal stimulus is not in fact created by the object/motion most likely to have created it – if a heuristic is misapplied. A nice example is the child (or adult) who perceives that the moon is following him home at night. This illusion probably arises from a heuristic to the effect that when two parts of the visual field are in motion relative to one another, the smaller part is moving and the larger part is stationary.

And second, these speculative heuristics are themselves worthy material for empirical investigation. Over what ranges of conditions does a given heuristic function? And what happens if a display is created that pits two likely heuristic against each other?

21.7 Summary

In summary, the topic of motion is complex from at least two perspectives. First, as the perceiver, we want both of two things. We want veridical perception of the motion of objects. At the same time we want to use motion cues to help us perceive the three-dimensional shapes and identities of objects, and our own patterns of locomotion in the world. Information about all of these topics is present in the motion flow fields in the retinal image, but the challenge is to sort them out.

In all probability the first step is the analysis of local motion. We have seen that across a wide range of spatiotemporal conditions, M/D ratios are near 1: similar contrast thresholds are seen for detectinga moving stimulus, and for judging its direction of motion. These data suggested the presence of independent detectors tuned to the two directions of motion. We also saw that computationally, delay-and-compare networks can be used to build neurons that are selective for the direction of local motion.

Higher levels of analysis can be built on these neurons by standard hierarchical concatenations. In judging the direction of coherent motion in stochastic motion patterns, subjects show that they can integrate motion signals across wide retinal regions; this system property suggests a hierarchical model in which higher-level neurons tuned for larger-scale motion integration are created from local detectors tuned for a similar direction of motion. Similarly, we examined the patterns of retinal image motion that arise from variations in distance, from rotation of rigid objects in three dimensions, and from biological motion.

And finally, we revisited the topic of heuristics. We saw that living in our particular physical world allows us to narrow down the interpretation of motion patterns in the retinal image. Remarkably, the cues and heuristics together are sufficient to allow us to perceive both the motion of objects and their shapes and identities veridically most of the time.

As we said earlier, there is a tendency in modern treatments of motion processing to assume that motion analysis is done in two basic steps. The first is the analysis of local motion, and the second is everything else. In the next chapter we ask, what is known about the neural substrates of motion processing? What about local motion, and what about everything else?

Ch. 21

Motion: Notes to myself: 10/19/05

SUZANNE'S EMAIL Email 1/14. Suzanne says, "..you have to begin with some type of local measurements of speed and direction (motion vectors). The average speed would not be useful.As soon as you start making separate measurements of regions of the field, it is only a matter of degree whether you have 2 or a thousand.

"Measurements of local motion per se?? M cones per se?? All is inference."

Watson and Robson, labeled lines.

Silverman, Nakayama, McKee – vel discrim depends on speed, not temporal frequency.

Welch, Nature, vel discrim in a plaid depends on the components, not the apparent vel of the plaid.

George Mather's review in Smith and Snowden, 1994, Visual Detection of Motion: "Motion detector models – Psychophysical evidence."

Email 1/18 ".. no evidence that V1 neurons encode speed per se.confound speed with spatial frequency (Movshon). " END SUZANNE'S EMAIL

MOVED 1/26 Motion integration

As discussed in the Introduction, many vision scientists accept the perspective that there are at least two stages of analysis of 2-D motion: the creation of local direction-of-motion tuned neurons, which could be created by delay-and-compare circuits; and integration across sets of such neurons to provide neurons that integrate motion signals across broader, non-contiguous retinal regions.

If there are neurons that carry out motion integration, what properties would they have? In terms of linking propositions, the spatial domain proposition, which we examined in Ch. xx on image segmentation, arises again in regard to motion. That is, perceptual processes that require integration of information over a given spatial domain imply physiological processes that integrate information over the same domain. If the domain of motion integration exceeds the receptive field sizes of V1 neurons, another processing stage is needed.

[What more can be said? Are there extant computational models of motion integration? I suppose the Newsome group has written on this. Xx] END MOVED 1/26

Chapter 22

Motion – Physiology

[Summarize L Props here]

We then turn to single unit physiology. Given the fact that different motion cues can serve different perceptual purposes, it seems l ikely that there will be a hierarchy of motion processing areas, with neurons at different levels of the visual system being selective for different aspects of retinal image motion. That is, some anatomical areas could house neurons selective for the direction of motion, others for the cue-invariant representations of the shapes of rigid objects, others for biological motions, and so on. This notion will provide a perspective from which to review the neurophysiological literature on motion.

In the final section of the chapter, we examine in detail an important paradigm for exploring the relationships between physiological activity and perception. This paradigm is simultaneous recording of both single units and psychophysical judgments in awake, behaving monkeys. It turns out that for carefully chosen neurons, in the context of carefully designed models, there are remarkably close correlations between psychophysical and neural levels of analysis. Moreover, manipulating the activity of these neurons changes the monkey's reports of the direction of motion! This finding strongly suggests that the affected neurons are on the causal chain between the retinal image and the monkey's perceptions. This paradigm, and the models developed within it, are particularly exciting because they provide new and more stringent criteria for judging Causal Stories.

TO BE CUT AS OF /11/06

Linking Propositions [This section is npw included in the "Comp models" section of Chapter 21 you have to state the L Prop before you can make sense of the Comp Model. For Chapter 22, summarize L Props in Intro. What new linking propositions, or new characteristics of linking propositions, arise from the study of motion?

First, the spatial domain argument, which we examined in Chapter 20 on image segmentation, arises again in regard to motion. That is, perceptual processes that require integration of information over a given spatial domain imply physiological processes that integrate information over the same domain.

Second, there is a common assumption that the perception of smooth motion arises from the continuous firing of neurons. Transients won't do.

Let us take each of the properties of motion discussed in Section B, and ask, given the system properties, what would we expect to find in the neurophysiology? And what linking propositions are involved in the arguments?

22.0.1 L Props: Basic motion perception

The first characteristic of motion perception is that we perceive it, and can distinguish its speed and direction. On the Universal linking proposition, we know that sufficient information to support these perceptual accomplishments is carried in some neural population at every stage of a pathway that extends through all levels of the visual system. Moreover, on the Neuron Doctrine, these properties suggest that somewhere within the visual system there will be individual neurons tuned for the direction and speed of motion. (Indeed, Hubel and Weisel have already shown us that neurons tuned for the direction of motion exist in V1.)

22.0.2 L Props re independent channels

22.0.3 L Props re integration of motion across spatial patterns

22.0.4 L Props re the metamerism of smooth and sampled motion

The metamerism of smooth and sampled motion places an interesting constraint on models of motion perception. As before, arguments arising from the occurrence of metamers usually involve Identity linking propositions.

Smooth and sampled motion lead to different spatiotemporal patterns in the retinal image, and presumably in the photoreceptor outputs and at some of the higher levels of processing. But the metamerism of continuous and sampled motion strongly suggests (some would say implies) that somewhere within the visual system, the neural signals arising from continuous and sampled motion are rendered identical. The fact that continuous and sampled motion are different in the physical world, and different in the retinal image, is lost to perception, just as differences in wavelength composition are lost in the cases of rod vision and trichromatic color vision.

A desirable characteristic of a computational motion model, then, would be one that renders identical the signals arising from continuous and sampled motion. And a desirable characteristic of a candidate neural mechanism would be one in which identical neural signals arise from the two kinds of motion.

22.0.5 L Props re motion after-effects (MAE)

The motion after-effect is an adaptation phenomenon. Models of adaptation effects always assume that the visual system is in a different state after the adaptation than it is before it, so that the same set of inputs yields a different constellation of neural activity than it did before.

What L Prop is involved? Identity or Similarity, maybe? The population code for a stationary object, post adaptation, is similar to the population code for a moving object, pre-adaptation. Therefore, the stationary river bank appears to move.

END CUT SECTION ON L PROPS

22.1 Neurophysiology: A hierarchy of motion codes [Needs integration with rest of chapter.]

We turn now to the neurophysiology of motion processing. Let's begin with some guesses about what we might expect to find. We start with a bumblebees-can-fly argument. Since we can perceive the direction and speed of motion, and see the three dimensional shapes and motions of objects, there must be populations of neurons that carry this information at every level of the visual system.

Moreover, the Neuron Doctrine xx leads us to hypothesize that at some level(s) we will encounter individual neurons that are tuned for the speed and direction of motion. If we are true believers in the Neuron Doctrine, we might also expect that at higher levels of the visual system, individual neurons will be tuned to more complex motion patterns, such as expansion patterns, deformations typical of rigid rotation, or deformations typical of biological motion. Are there neurons that are tuned to these variables? Where do they occur? And how do motion codes change from one processing level to another?

12/16 – ALSO, NEED TO TIE INTO THE ILLUSIONS OF MOTION. WHERE ARE THE EARLIEST NEURONS THAT RESPOND EQUALLY TO MOTION METAMERS? [JUST GOT NEW NEWSOME REFERENCES.] ARE THERE NEURONS THAT ADAPT, TO ACCOUNT FOR MOTION ADAPTATION?? [JUST GOT NEW MOVSHON REFERENCES].

22.1.1 M cells: Responses to temporal change

We ended Chapter 14xx with a summary of the properties of primate retinal M, P, and K cells. Of these three major cell types, M (parasol) cells respond most transiently to stimuli, and have the highest high temporal frequency cut-off. Also, M cells provide the major input to the dorsal stream, which has a well-deserved reputation for motion processing. Thus, in all probability the population of M cells providess the signals that will later be recoded to support motion processing. And in fact, the motion of a spatial pattern across the receptive field of a retinal M cell does yield a burst of firing.

However, just as individual cones do not code the wavelength of light, individual M cells do not carry information about motion nor about the direction of motion. That is, a prototypical M cell would respond equally well to a stimulus that moved across its receptive field in *any* direction; and even more importantly, it would also respond to a light that flickered on and off but did not move. Information about the speed and direction of motion is doubtless carried by the population of M cells, but not by individual primate M cells[1]. LGN cells are similarly not tuned for the direction of motion.

22.1.2 V1 cells: The first appearance of direction selectivity

Direction selectivity in individual neurons first arises within area V1. We have already recounted (Chapter xx, Figure xx) how the vigorous response of a V1 neurons to the moving edge of a slide provided the clue that allowed Hubel and Wiesel to begin to crack the V1 visual code. More recent studies have shown that about 1/4 of V1 neurons are selective for the direction of motion in 2D images. There is a concentration of directionally selective neurons in layer 4B – the layer that receives input from M cells and projects to area MT. Presumably this direction selectivity is built by the elegant neurons shown in Figure xx of Chapter xx, but the circuits that do the building remain a matter of controversy.

The motion tuning properties of V1 cells are relatively coarse and limited. Individual V1 neurons respond only to relatively slow speeds (e.g. 10^o/sec), and their direction selectivity is relatively

[1] The situation is different in other mammalian retinas. In particular, in a very early single unit study, Barlow and Hill (19xx) found directionally-selective ganglion cells in the retinas of rabbits. [Check with Dennis; maybe there are primate retinal cells that respond to motion by now. Xx]

Figure 22.1: Speed tuning in MT neurons

broad. Their receptive fields are small, and there is no evidence that they respond selectively to larger or more complex motion patterns such as expansion or distortion xx. Instead, it is as though V1 neurons signal the presence of small local motions in particular directions in the retinal image, and provide the raw materials out of which sensitivity to more complex motion patterns could be built.

[V2?? V3?? DT – See Rudd lecture for starters.]

22.1.3 Area MT: The classical motion area and its many skills

As discussed previously, the middle temporal area, area MT or V5, is considered a central component of the dorsal processing stream, which has a well-earned reputation for processing motion. Virtually all of the neurons in MT are tuned for the direction and speed of motion, and respond with great enthusiasm to stimuli moving in their preferred direction. They also have a clear "null" direction opposite to the preferred direction of motion, and motion in the null direction often suppresses the spontaneous activity of the neuron.

A direction-of-motion tuning curve for a typical MT neuron is shown in Figure 22.1. Interestingly, although MT neurons seem to be specialized to respond to motion, the direction-of-motion tuning of MT neurons is not particularly sharp, and a typical tuning curve might have a full width at half height of 100 to 120 degrees. Thus, MT neurons do not exhibit a sparse direction-of-motion code; information about the direction of motion must be carried in a distributed population code at this stage.

Several additional and more novel features of MT cells are interesting from the perspective of coding schemes. First, many MT neurons show the presence of *silent surrounds.* That is, MT neurons show *classical receptive fields*: any retinal location to which the neuron responds is part of the classical receptive field, and locations to which it does not respond are outside it. However, moving stimuli in regions outside the classical receptive field – within the silent surround – can

Figure 22.2: Examples of velocity gradients to which MT neurons are tuned

strongly modulate the response of the MT neuron to stimuli within the classical receptive field. Typically, stimuli moving in the direction opposite to the cells' preferred direction enhance the response of the neuron (other forms of interaction also occur). Such neurons seem well suited to responding to patterns of relative motion between an object and its background. [check against Britten xx]

[SWB 144] And second, there is evidence that MT neurons can be tuned for other features of motion patterns. For example, Treue and Andersen (1996) tested MT neurons with random dot stimuli containing local *velocity gradients.* A velocity gradient is a motion pattern within which the velocities of pattern elements vary with spatial location. Within the set of random dots, Treue and Andersen introduced systematic variations of the speeds of dots in different locations; for example, the speed could increase as the dots approached the top (or the bottom) of the field; or could be faster or slower on one side of the field than the other. Although the effects were small, these variations did affect the responses of many MT neurons. Since velocity gradients provide cues to complex aspects of the visual environment and the objects within it, the Neuron Doctrine suggests that we might well find visual neurons that are tuned to velocity gradients, and MT neurons seem to provide the earliest instance of this kind of tuning.

Third – form-cue invariance (Croner and Albright, 1999)

Fourth – global – stochastic motion stimuli [SWB 154] – Newsome of course. Who else earlier?

[Do MT cells respond to expansion patterns? optic flow? Distortion patterns typical of rigid rotation? Distortion patterns typical of biological motion? I haven't found this lit yet. Have a paper from blake that deals with fMRI – haven't looked at it yet. Xx]

22.1.4 Area MSTd and the analysis of motion patterns

Beyond MT, motion signals are further processed at a number of locations, with important transformations of the motion code. For example, area MSTd (the dorsal part of the medial superior temporal region) contains neurons with very large, motion-tuned receptive fields. Many MSTd neurons respond to complex motion patterns such as rotation and/or expansion/contraction, and are tuned to, for example, the direction of rotation of a rotating pattern. Others are tuned for the orientations of planes rotating in depth [Sugahara et al 2002] . Similarly, MSTd neurons respond to *optic flow fields*: large regions of dots or other stimulus elements, moving in patterns that mimic the patterns generated in the retinal image when the subject walks around in the world. Such neurons may reveal the analysis of the shape and locomotion cues contained in complex patterns of retinal image motion.

[Even higher areas? An area that compensates for eye movements?]

22.1.5 Summary: Motion processing as a hierarchical system

Finally, as one moves even farther along the motion processing pathway, and increasingly close to the beginnings of coding for motor actions, one begins to see neurons coded in coordinates appropriate to the production of motor actions. A great deal is known, for example, about the neural codes required for the planned production of particular eye movements, and about the production of the motor movements themselves. Unfortunately, this book has to stop somewhere, and these preparations for motor action cannot be included in more detail.

In summary, and simplifying a highly complex topic, we can imagine the motion processing system as composed of several stages. V1 neurons, with their small receptive fields, are tuned to the directions of motion of small local elements in the 2-dimensional retinal image. To date there is little evidence that they take on more complex tasks, (but it is possible they have just not been asked the question in the right way). In any case, V1 neurons could provide the building blocks out of which neurons tuned to more complex aspects of motion patterns can be built. MT neurons, with their somewhat larger receptive fields, may begin the analysis of simple patterns of motion, such as responding to motion flow fields xx and differences in the direction of motion between objects and their backgrounds. And MSTd neurons might be specialized to respond to more complex motion patterns, perhaps including the distortion patterns that signal the shapes of three-dimensional objects and biological motion xx. MORE. Xx.

A continuing goal of vision scientists is to sort out what aspects of motion are coded by the tuning of individual neurons at each stage, and at which stages the information in various cues are made explicit in the activities of individual neurons.

22.2 A conceptual breakthrough: New criteria for judging causal stories

Prior to the 1990s, speculations about the dependence of visual function on visual structure were based on two main sources of information. As discussed earlier, lesion studies have classically been used to argue that particular anatomical structures are *necessary* for the occurrence of particular behaviors and/or perceptions. And since the advent of single unit studies in the early 1960s, correlations between the tuning characteristics of individual neurons and the characteristics of perceptual and/or behavioral dat, have been used to argue that particular neurons or anatomical structures are *sufficient* to support particular perceptions and/or behaviors. As discussed earlier, these arguments are attractive, but subject to important limitations.

During the early 1990s, a new approach to the neuron/perception question arrived on the scene: single unit recording in awake, behaving monkeys. With this technique, the activity of individual neurons and the perceptions of the animal (as indicated by their behavioral reports) are recorded simultaneously, in response to the same stimuli. With this approach, neural activity and behavior can be compared in detail, on a trial-by-trial basis. Such remarkable techniques potentially bring us closer to the goal of understanding the neural basis of perception. A particularly insightful series of experiments using this paradigm has been carried out by William Newsome and his colleagues.

We begin with a general description of the technique, many aspects of which are novel and complex. To begin with, it takes several months to prepare a monkey to become a subject in these experiments (not a good thesis project!). First, a stainless steel device for stabilizing head position must be attached to the monkey's skull, and a search coil used for recording eye movements must be implanted around one eye. If these surgeries are successful, the monkey then undergoes several months of training as a psychophysical subject. After training is complete, more surgery is undertaken: a stainless steel cylinder that allows recording electrodes to be lowered into the brain is surgically implanted on the skull. If all of these initial steps are successful, and the monkey is still alive and well, the actual experiments can begin.

The stimuli used in many of the experiments from the Newsome lab are fields of dots called *random dot kinematograms*, or *stochastic motion stimuli*, as shown in Figure 21.5. In such a stimulus, each dot can be displaced along a different trajectory between frames. The manipulated parameter is the degree of *correlation* of motion between two frames: the fraction of the dots that are displaced in the same direction. The left, middle, and right panels of Figure 21.5 xx depict correlations of 0%, 50%, and 100% respectively. For human subjects, as the correlation increases from zero, random movements yield to perceptual impressions of an overall directionality of motion, and eventually to clear perceptions of motion.

Stochastic motion stimuli are theoretically attractive for several reasons. Since none of the dots moves continuously, these stimuli do not promote tracking eye movements as edges and gratings do, so the subject can hold his eye relatively still during the experiment. Moreover, to perceive the predominant direction of motion in a stochastic motion stimulus requires combining information across broad retinal regions, and separating one direction of motion from all of the others. Thus, these stimuli are custom designed to isolate a higher level of motion processing than are the edges and gratings initially used to test V1 neurons.

The response measure used by the Newsome group is also novel from our perspective. Rather than pushing levers or buttons to indicate the perceived direction of motion (as we would tend to do in human psychophysics), the monkeys are trained to use *eye movements.* As shown in Figure 22.4A xx, the monkey's task is to fixate a fixation point during the stimulus presentation, and then shift fixation to one of two target LEDs, depending on which direction of motion he wishes to report (e.g. the upper dot for upward motion or the lower dot for downward motion).

In a typical experiment, the percent correlation of motion among the dots is varied from trial to trial. The monkey's task is to report the direction of correlated motion on each trial. Since this is a forced-choice task, there is a right answer, and the monkey is rewarded with a few drops of water or juice after each correct response. The behavioral data, then, take a familiar form: two-alternative forced-choice psychometric functions showing the monkey's percent correct as a function of the percent correlation in the stochastic motion stimulus.

To begin an experimental run, the monkey is seated in a primate chair, and the experimenter attaches all of the necessary interfaces to record both the eye movements and the responses of

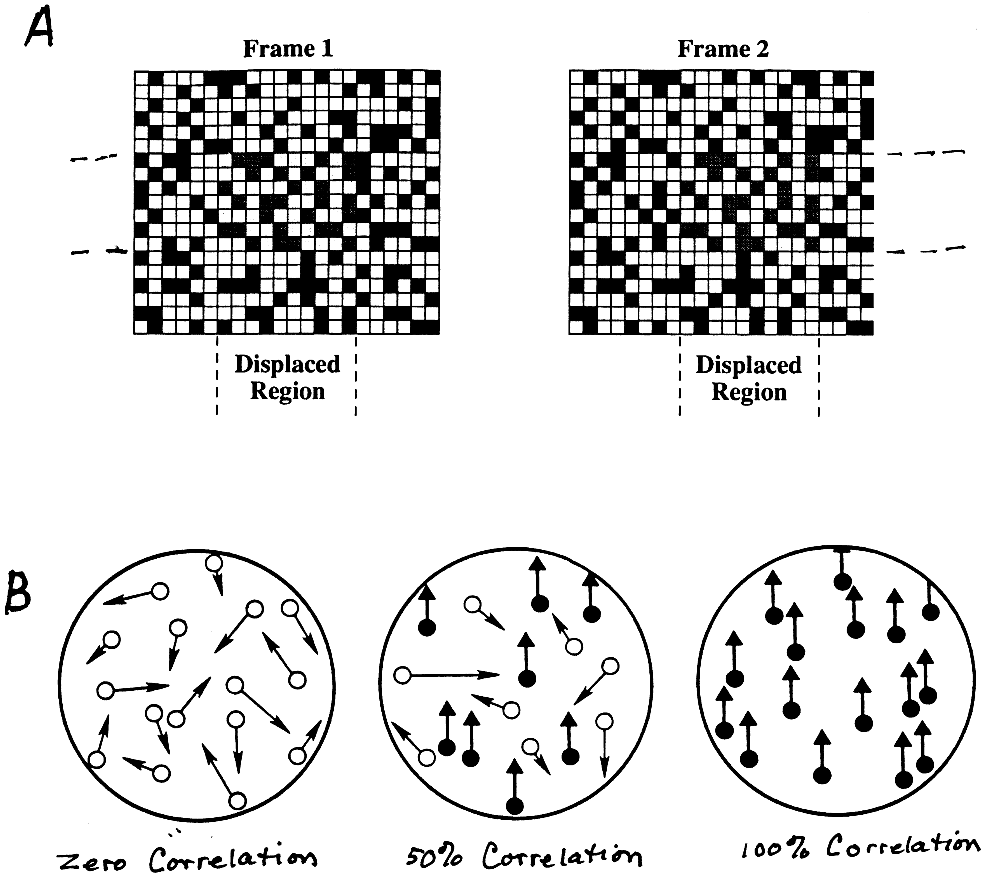


Figure 22.3: Random dot kinematograms. A: A *dense RDK* . The stimulus is composed of randomly arranged, tightly packed small black and white squares. A central region of the small squares, indicated by the region bounded by the dotted lines, is displaced from one location to another (here, by two small squares' distance to the right). The observer's task is to report the direction of displacement of the central figure. (Modified from McKee and Watamaniuk, 1994.) B: A *sparse RDK* (also called a stochastic motion stimulus). In frame 1 the dots have the locations indicated by the circles, and in frame 2 they are replaced by dots at the arrow tips. In the observer's perception, each dot jumps to the position of its nearest neighbor. The experimenter varies the degree of *correlation* between the two frames; that is, the fraction of dots whose nearest neighbors are all displaced in the same direction. The left, middle, and right panels of Figure xxB show correlations of 0%, 50%, and 100% respectively. The observer's task is to report the direction of displacement of the correlated set of dots. (Modified from Salzman et al, 1992).

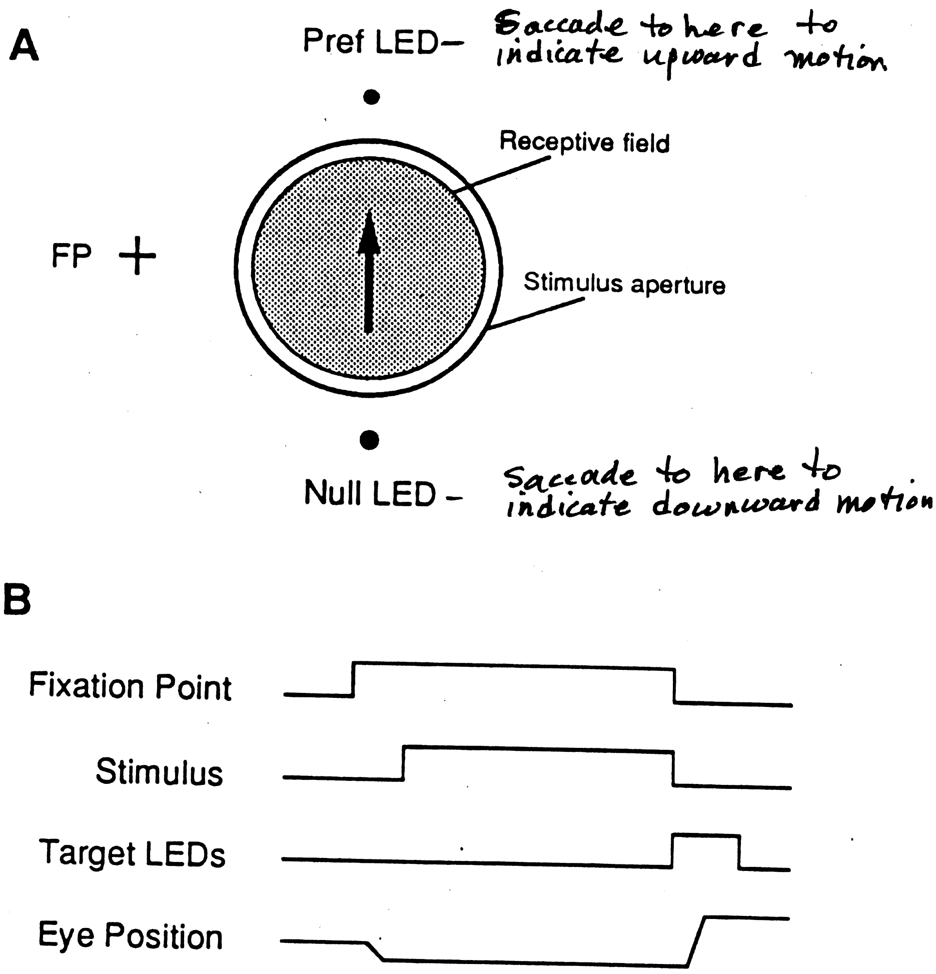


Figure 22.4: Illustration of the spatial and temporal characteristics of a typical experiment from the Newsome laboratory. Panel A shows the spatial configuration used. The monkey fixates the cross at FP. The electrode is advanced into area MT to isolate a neuron that is tuned to the speed and direction of motion. Its receptive field is located, and a stochastic motion stimulus is presented within a stimulus aperture centered on the receptive field. Two target LEDs can be illuminated in line with the preferred and null directions of motion for the neuron. Panel B shows the timing of a trial. The fixation point comes on, and the monkey shifts gaze to fixate it (shown in the eye position trace). The stochastic motion stimulus is then presented for 2 sec. As the stimulus goes off, the target LEDs come on, and the monkey shifts her gaze to one or the other LED to indicate her judgment of the direction of stochastic motion on that trial. The activity of the neuron is also recorded throughout the trial. (Modified from Britten et a, 1992.)

single cells. The experimenter then advances the microelectrode to locate a single neuron that is responsive to motion, in the known vicinity of area MT. Once a neuron has been found, its classical receptive field, direction selectivity, and speed selectivity are determined.

Finally, the experiment proper can begin. The monkey fixates the fixation point. He is presented with a field of stochastic motion stimuli matched to the size and location of the receptive field of the particular neuron under study, moving in either the preferred or the null direction of motion, at the preferred speed (the ultimate custom-designed stimulus!). The experimenter varies the percent correlation within the stimulus and measures the response of the neuron, while the monkey uses eye movements to the target LEDs to make forced-choice judgments concerning the direction of motion of the stimulus. The monkey's behavior can then be compared to the neuron's behavior, either as a statistical average or on a trial by trial basis.

Newsome's and his colleagues' work is also marked by particularly elegant choices of experiments, logic, models, and data analyses, to exploit the full value of the joint physiological and behavioral data. Since the most fundamental concern of this book is the logic of relating system properties to neural activity, these experiments are of special interest. We here review three experiments and critically examine the conclusions that can be drawn from them.

22.2.1 Do individual neurons have sufficient reliability to underlie behavioral data?

First, a study by Britten, Shadlen, Newsome, and Movshon (1992) addressed the question: Do the signals in individual neurons in MT have sufficient reliability to provide the neural basis for the behavioral data?

The behavioral data produced by the awake, behaving monkeys are straightforward – forced-choice psychometric functions. Similarly, the physiological data show the responses of MT neurons to stochastic motion stimuli of various correlations, in the preferred and null directions of the neuron. To allow direct comparisons of the neural to the behavioral data, the data from individual neurons were processed through a signal-detection-based model to yield what Britten et al call a *neurometric function.* The model (which is too complex to recount here, but which involves only the recorded neuron and a postulated identical neuron tuned to the opposite direction) provides an estimate of how well the monkey could perform if he based his responses solely on the activity of two neurons like the particular neuron recorded simultaneously with the behavioral data.

A set of psychometric functions and their corresponding neurometric functions are shown in Figure 22.5. Of course, the results vary from one neuron to the next. Panels A and B show two cases of nearly exact correspondence between simultaneously recorded psychometric and neurometric functions. About half of the neurons conformed to this pattern. Panels C and D show cases in which the monkey was more sensitive than the neuron (panel C) or vice versa (panel D).

Now, what do we make of these data? Britten and his colleagues argue that the close match between psychometric and neurometric functions supports the argument that, under the conditions tested, the information carried by a single MT neuron and its oppositely-tuned twin is *sufficient* to control the monkey's judgments of the direction of motion of the stochastic motion stimuli. The monkey could do as well as he does on the basis of neurons like this one alone. In particular, there is no need to argue that any higher level of the system combines signals across neurons to provide any finer analysis of the direction of motion of the stochastic motion stimulus under the conditions tested. The experiment, then, provides a very sophisticated sufficiency argument for the

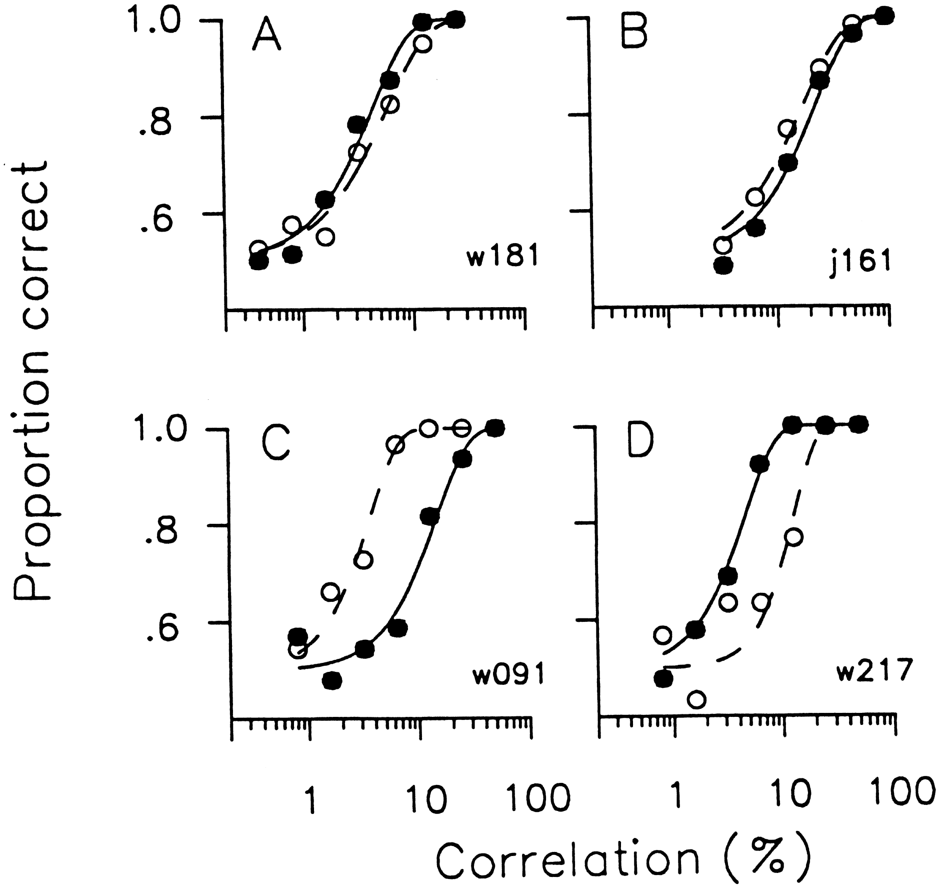


Figure 22.5: Psychometric and neurometric functions. The open symbols are psychometric functions derived from the monkey's behavioral responses. The solid symbols are neurometric functions derived from the responses of the individual neuron. Psychometric and neurometric functions often coincided closely (panels A and B). In other cases, the animal could be more sensitive than the individual neuron (Panel C) or vice versa (Panel D). (Modified from Britten et al, 1992.)

dependence of behavior on neural activity.

22.2.2 Trial-by-trial correlations exist between neural activity and perceptual reports [in the absence of stimulus varialtions xx]

A second question – trial-by-trial correlations between single unit activity and behavior – was addressed by Britten, Newsome, Shadlen, Celebrini and Movshon (1996). The essence of the study stems from the fact that with repeated presentations of the same stimulus, the response of the individual neuron will vary from trial to trial. The question is, keeping the stimulus fixed, will the monkey's behavioral response vary in correlation with the neuron's response? In specific, will the monkey tend to report motion in the neuron's preferred direction on trials on which the neuron happens to fire at a higher rate, and the neuron's null direction on trials on which the neuron happens to fire at a lower rate?

Such an analysis is shown in Figure 22.6. In this particular example, a stimulus with 0% correlation was presented repeatedly within a sequence of trials of other correlations. For 0% correlation, of course, there is no net motion signal created by the stimulus. Yet the neuron produced a variable number of spikes from one trial to the next; and the data show a small but clear tendency for the the monkey's response to vary in correlation with the firing rate of the neuron! Since the correlation is not introduced by the stimulus, it must arise within the monkey's visual system, and it suggests that the activity of these neurons plays a causal role in generating the monkey's perceptual report.

22.2.3 Electrical stimulation biases the reported direction of motion

In a third and most striking experiment, Newsome and his colleagues (Saltzman, Murasugi, Britten, and Newsome, 1992) investigated the effects of cortical microstimulation on the monkey's psychophysical performance. Fortunately, like area V1, area MT has a columnar organization. That is, groups of neurons that respond to the same direction of motion lie near each other in columnar compartments. This anatomical arrangement makes it possible to place a stimulating electrode in one location within MT and stimulate mostly or entirely neurons that respond to a single direction of motion.

To begin their experiment, Saltzman et al first recorded the direction preferences of MT neurons within a local region, and selected regions of MT in which neurons encountered over at least 200 micrometers had similar preferred directions. Electrical stimulation consisted of a series of tiny electrical pulses to this chosen region. Such electrical pulses probably increase the firing rates of neurons in their immediate vicinity. As before, the monkey's task was to view a stochastic motion display and judge the direction of correlated motion. Stimulation and no-stimulation trials – trials on which the monkey's cortex received electrical stimulation, and trials on which it did not – were interleaved in the experiment. The hypothesis was that electrical stimulation should *increase* the probability that the monkey will report motion in the preferred direction of the stimulated MT neurons.

Results from four different stimulation sites are shown in Figure 22.7. In each case, the addition of electrical stimulation shifted the psychometric function leftward. In other words, for each level of correlation in the stimulus, microstimulation influenced the monkey to report the cell's preferred direction of motion on an increased fraction of trials! (Variations in the size of the effect are

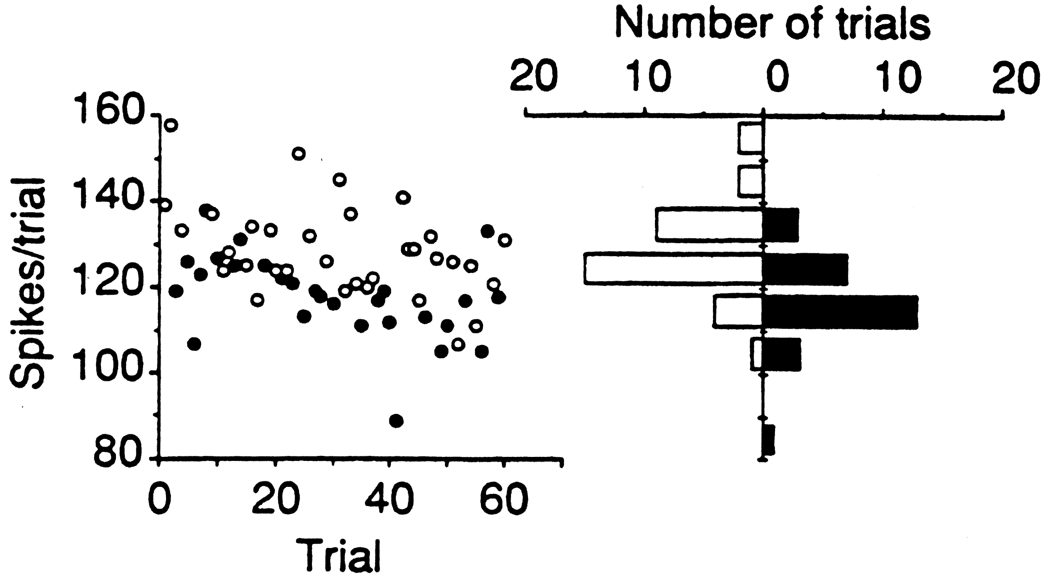


Figure 22.6: Trial-by-trial correlation of neural firing rates and behavioral judgments. Open symbols represent trials on which the monkey reported motion in the neuron's preferred direction, and closed symbols represent trials on which the monkey reported motion in the neuron's null direction. There is a small but clear tendency for the monkey to report motion in the preferred direction on trials on which the neuron fired at the faster rates, and motion in the null direction on trials on which the neuron fired at the slower rates. Other neurons showed similar patterns. (Modified from Newsome et al, 1995.)

Figure 22.7: Effects of cortical microstimulation on behaviorally measured psychometric functions. The horizontal axis shows the percent correlation in the stochastic motion stimulus. The vertical axis, labelled "proportion PD", shows the proportion of trials on which the monkey's report of the direction of motion coincides with the *p*referred *d*irection of motion of the neurons in the column in which the microelectrode is situated. (Modified from Salzman et al, 1992.) The four panels show four different microstimulation sites. The open and closed circles show psychometric functions on no-stimulation and stimulation trials respectively. In each case microstimulation shifts the psychometric function to the left. That is, at each percent correlation of the stimulus, microstimulation increases the percent of trials in which the monkey reports motion in the preferred direction of the stimulated neurons, as though the microstimulation increased the magnitude of the motion signal in the preferred direction.

attributed to variations in the success with which the electrical stimulation was confined to a single compartment, as well as other factors).

What are we to make of these data? Unlike all of the earlier experiments, which are more correlational in nature, microstimulation constitutes an experimental manipulation of the activity of the neurons in area MT. The results show that neurons in MT aren't just tuned to the direction of motion to no effect; externally caused changes in their firing rates cause changes in the behavior of the animal. Thus, this experiment adds particularly compelling evidence supporting the causal role of directionally selective MT neurons in determining the perceptual reports of the animal.

An important remaining question is, what is the mechanism of action of the electrical signals? In particular, do they influence the sensory signal (and thus the animal's perception), or a post-perceptual decision process, or just the motor responses of the animal? To explore the decision process explanation, monkeys were tested with electrical stimulation applied in a different region of the visual field than the region in which the stochastic motion stimuli were presented. Distant microstimulation did not affect the monkeys' psychophysical reports, ruling out a major influence of the stimulation solely on a decision process. To explore the motor explanation, microstimulation was applied out of synchrony – just before the presentation of the stochastic motion stimulus, rather than simultaneously with it. The electrical stimulation was also ineffective in this condition, and no extraneous eye movements correlated with the delivery of the electrical stimulation were seen. Newsome and his colleagues therefore argue that the electrical stimulation had its influence early rather than late: on the animal's perceptions; rather than just on the decision process or the motor response.

22.2.4 Consciousness, or just behavior?

In Chapter 19xx we used the deleterious effects of dorsal stream lesions, and the speed and directional tuning of dorsal stream neurons, to argue that area MT is an important location on the causal chain for neural signals that underlie the perception of motion. We also argued, however, that tuning curves and lesion experiments are rather blunt tools for defining causal relations between perception and behavior.

The Newsome group experiments have gone much further. They have shown that individual MT neurons are sensitive enough – their properties are sufficient – to control the monkey's perceptual reports of the direction of motion in stochastic motion stimuli; that even with a fixed stimulus, the activities of MT neurons can influence the animal's perceptual reports; and that microstimulation of MT neurons influences the monkey's peceptual reports in orderly ways. These experiments clearly probe the question of the neural basis of motion perception more deeply than did earlier lesion studies and tuning studies. Moreover, they greatly strengthen the argument that neurons in MT are important elements along a causal chain leading from moving stimuli to the monkey's behavioral reports of his perception of such stimuli.

But can we make a stronger claim? On the basis of all of these experiments, one might wish to argue that the activities of MT neurons mediate the *conscious perception* of motion. This claim bears careful examination, but for the moment, we will content ourselves with the weaker conclusion: that MT neurons are a major processing stage along the causal chain that leads to the monkey's perceptual reports of motion and its direction. We will return to the question of the neural correlates of consciousness, and how they might be studied, in Chapter 27xx.

22.3 Conclusions: As good as it gets! -NOT WRITTEN YET

In summary, the perception of motion is a classic area of perceptual science. Our description of some of the most historically puzzling aspects of motion perception has barely scratched the surface of this interesting and complex topic.

Unique and illustrative features: Real world constraints ? heuristics Cues in the retinal image Search for neurons that extract those cues Sophisticated analysis of perceptual/neural correlations

A unique feature of the topic of motion processing is the progress computational neuroscientists have made on characterizing the patterns of motion that different kinds of physical motion produce in the retinal image. The translation, rotation and distortion of objects lead to different, complex patterns of retinal image motion, and these patterns can therefore serve as cues to the properties and motions of physical objects.

Some possible neural bases for motion analysis are remarkably well known. There are indeed neurons, early in the dorsal stream, that are selective for the direction of motion; and later in the dorsal stream, for more complex motion patterns.

Moreover, the use of awake, behaving animals has led to a series of studies that provide particularly compelling evidence that these directionally tuned neurons play a major causal role in determining a monkey's reports and perceptions of the direction of motion.

One final cautionary note: Given the experiments reviewed above, you might be temped to believe that MT neurons and other dorsal stream neurons are specialized for the analysis and perception of motion. But remember that it's impossible to test a single neuron on all of the possible stimulus dimensions. In many of the studies cited above, neurons were tested systematically with moving stimuli of various kinds. But there have been fewer studies with stimuli varying in color, or contrast, or distance or depth, and it turns out that MT neurons are also selective within several of these domains. In particular, many MT neurons are selective among stimuli that vary in stereoscopic depth (Chapter xx). If we overgeneralize our characterizations of the roles of particular neurons or structures too early, we may have to eat our words.

Motion: Notes to myself: 10/19/05

Overarching Themes to be illustrated/emphasized by motion chapter:

1. Physiological processing stages: more sophisticated tuning at higher and higher levels;

2. The Neuron Doctrine – tuned neurons as explanations of system properties. [With all its pitfalls.]

2. Illustration of complementarity of cues and heuristics

3. Newsome paradigm – Single unit studies in awake, behaving monkeys; microstimulation. New paradigms open doors to new questions and advances.

4. A moral about role and locus questions (MT neurons respond to motion; do they "do motion") But they also respond to disparity. Don't assign "role" too fast.

***** MOVED 11/02 Finally, it is worth remembering that all of these patterns are subject to illusions if the retinal stimulus is not in fact created by the object/motion most likely to have created it – if a heuristic is misapplied. A nice example is the child who sees the moon following him home at night. This illusion probably arises from a heuristic to the effect that when two parts of the visual field are in motion relative to one another, the larger part will appear to b stationary and the smaller part will appear to move.

21 Motion Figure legends

MOVED 11/04 – Put with new figs that are vector diagrams of motion patterns. [MOVE TO FIG LEGEND?? XXDiagrams such as this are essentially vector diagrams – the directions and lengths of the arrows show the directions and speeds of motion in the retinal image, arising from the various points on the surface of the soccor ball.] END MOVED

Chapter 23

Perception of Brightness and Color

Definitional note, Draft 3: In English, the term color is used in two different ways: Both to include the black/white dimension (What color is your dress? –It's black and white) and to exclude this dimension (is this a color TV? No, it's black and white.)

In this chapter, when we want to refer to the black/white dimension alone, we will use the term "lightness". When we want to refer to just the red/green and blue/yellow dimensions, we will use the term "chromatic". We expect to introduce this distinction earlier in the next draft. Usage may still not be fully consistent in this draft. Up to this point in this book, we have mostly been discussing color vision as described by subjects' responses to isolated disks of light presented against a dark background in the laboratory setting. Use of these simple stimuli has revealed a great deal about the mechanisms that underlie our color vision, particularly the trichromacy of wavelength encoding and the perceptual opponency of colors.

However, when we view more complex displays or real-world scenes the perception of color becomes much more complex. It's time to bite the bullet.

23.1 Lightness contrast and lightness constancy

In the following two sections we provide separate discussions of the lightness (white/black) and chromatic dimensions of color perception. In each case we begin with descriptions of some of the perceptual phenomena, including contrast and constancy, that occur when we move to more complex stimulus fields. We then describe examples of quantitative experiments on these phenomena. At the end of each section, we consider the question of heuristics that might be used by our visual systems to allow lightness constancy and color constancy to occur.

23.1.1 Simultaneous and successive lightness contrast

In Chapter 7 we introduced the concept of metamers by discussing complementary colors: different pairs of wavelengths that, when mixed, are perceived as white. That is to say, the perception of whiteness comes about when we view any of many different mixtures of wavelengths of light from the white metamer set. But so far we have said nothing about how the other achromatic colors – the series of light greys, dark greys, charcoal greys, and black – come about.

We begin by distinguishing between the perceptual terms *brightness* and *lightness*. In perceptual terms, isolated spots of light typically have the appearance of being *self-luminous* – that is, emitting light[1]. The sun and stars, lightbulbs, and a spot of light shone from a projector onto a projection screen in the laboratory, are examples of self-luminous objects, and in general they have the perceptual quality of *brightness.* Varying the intensity of such an isolated spot makes the spot look brighter or dimmer, but it remains apparantly self-luminous. And different self-luminous objects can appear to have different brightnesses. Try this at home by turning on a lamp with a three-way light bulb in an otherwise dark room. As you switch the bulb, its perceived brightness changes, but it always looks bright – never white, grey or black. Where do the achromatic colors – white, grey and black – come from?

It turns out that the achromatic colors come about only when lights of different physical intensities are spatially juxtaposed. For example, let us set up a central disk of one luminance and a surrounding ring (or *annulus*) of another, somewht higher luminance. When we use such displays, we find that adding the annulus introduces a whole new perceptual dimension which vision scientists call the *lightness* dimension. When the spot is surrounded by an annulus of higher luminance than itself, the spot takes on the appearance of a gray surface. Moreover, the luminance of the surrounding region strongly influences the perceived lightness of the central disk: By varying only the luminance of the annulus, we can make the central disk take on any shade of white, grey, or black that we choose. Examples of such *simultaneous lightness contrast* between center and surround regions are shown in Figure 23.1. (More dramatic examples can be produced with projectors, but on the printed page the available range of luminances is limited.)

Successive lightness contrast

Another classic phenomenon of perceived brightness – successive contrast – can also be illustrated with Figure 23.1. If you stare at a black and white pattern such as the right hand panel of Figure 23.1A for 30 seconds or so, and then shift your gaze to the homogeneous white pageabove it, you will see an afterimage of the original picture with the brightnesses reversed. Successive contrast, of course, blends off into the phenomena of light and dark adaptation that we studied in Chapter 10.

23.1.2 Lightness constancy

As we walk around in the world, and go through the day from early dawn to bright sunlight, the amount of light reflected to our eyes from a given object may change by ten million fold. Yet in general we see the lightnesses of objects remain relatively constant across these enormous variations of the incoming signal. Our perception of the shade of white, grey, or black – the achromatic color – of a constant object remains remarkably constant. This is the phenomena of *lightness constancy.* Lightness constancy was also illustrated by the black cat and the white cat in Chapter 10 – the white cat and the black cat retain their lightnesses even when both move from direct sunlight to shadow, or when the illumination changes from dawn to dusk.

So here's the question: to what characteristics of real objects do their perceived lightnesses correspond? What physical characteristics of an object make it be perceived as white, or grey, or black? Physical study reveals that objects perceived as white, grey and black typically reflect light of all wavelengths about equally. But these objects differ in *reflectance* – the *percentage* of incident light reflected from their surfaces. An object that reflects, say, 80% of the incident light is usually

[1] The quality of perceived brightness can also be illusory. For example, most people report that the moon against the dark sky appears as a bright surface – it appears to give off its own light, even though the light it sends to our eyes is actually reflected sunlight.

Figure 23.1: Simultaneous and successive lightness contrast. A: The influence of surrounds. The central squares all have the same reflectance. In perception, the squares surrounded by the higher luminance surrounds look darker than the squares surrounded by the lower luminance surrounds. B. The stair step illusion. Each of the panels has a homogeneous reflectance, with the lowest reflectance panel on the left and the highest reflectance panel on the right. In perception, each panel takes on a bright edge next to its darker neighbor and a dark edge next to its brighter neighbor. ((These figures are not guaranteed to be accurate portrayals. The Xerox process acts to enhance contrast, so the "contrast" effects may be partly present in the stimuli in this instance.))

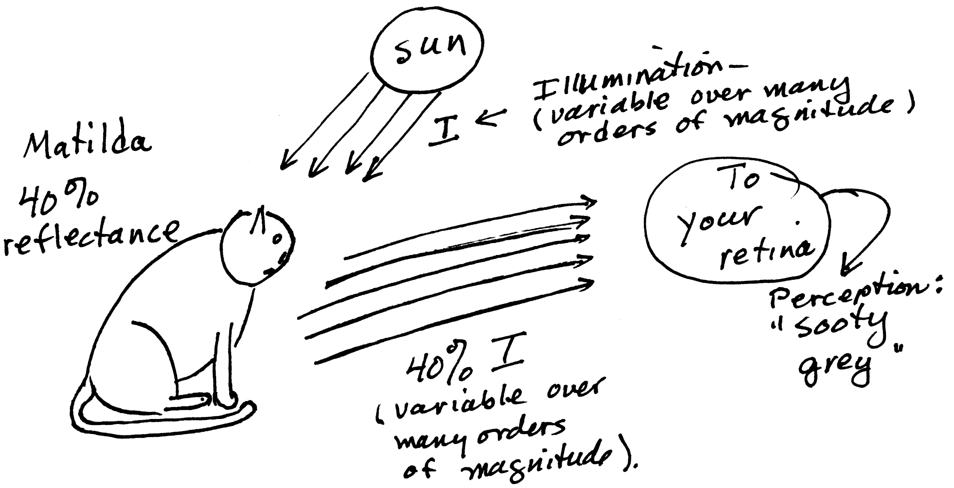


Figure 23.2: The problem of lightness constancy. An object has a surface reflectance value that can range from about 4% to about 80%. The light sent to your eye from the object is the product of the reflectance and the illumination, and completely confounds these two variables. Yet the visual system is able to sort them out, and we perceive an object with a fixed reflectance as having a fixed lightness. Here, Matilda is seen as sooty grey, and her perceived lightness would remain the same even if she goes into the shade.

perceived as white, while an object that reflects, say, 4% is perceived as black; and the different reflectances in between give rise to the different perceived shades of grey.

Lightness constancy, then, refers to our ability to *map variations in the surface reflectances of objects to perceived variations in lightness, despite changes in the overall level of illumination.* Ideally, if human lightness constancy were perfect, the perceived light ness of an object would stay constant as long as its spectral reflectance function stays constant, and change only when it changes. Looked at in this way, light ness constancy is obviously part of the general perceptual talent of *object recognition* – we recognise a particular cat as our cat, Matilda, partly because she is (perceived to be) sooty grey whether she cavorts in the light or sleeps in the shade. But how do we do it, if the light that comes to our eye confounds reflectance with illumination? The problem of lightness constancy is illustrated in Figure 23.2.

23.1.3 Quantitative experiments on lightness perception

In an early experiment on brightness constancy, Hans Wallach (19xx) set up a visual display consisting of a configuration of four fields: two widely separated disks, each with its own surrounding annulus, projected onto a white screen in an otherwise dark room. In each case the surrounding annulus was of higher luminance than the surrounded disk. Wallach set the two annuli to two different luminances – say, 100 units on the left and 200 units on the right. He also set the left disk to a luminance of, say, 25. The question was, to what luminance must the right-hand disk be set, so that the lightnesses of the two disks will appear the same shade of grey? The answer was 50.

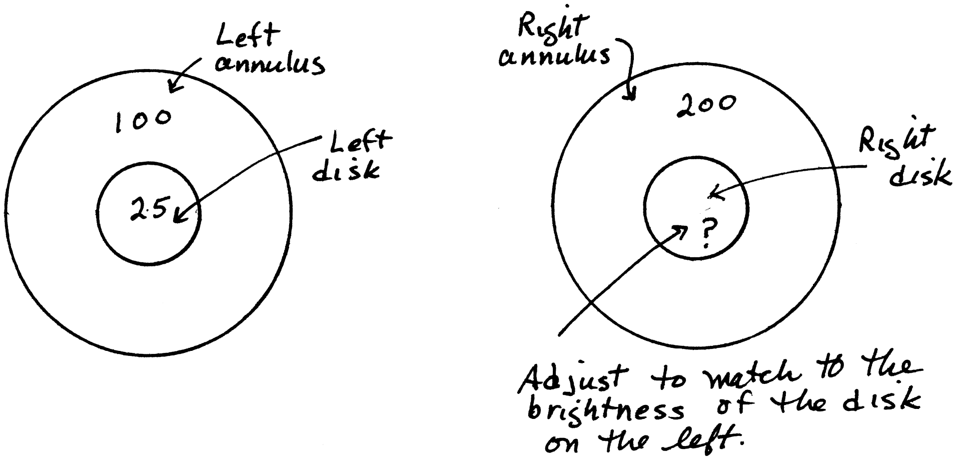


Figure 23.3: Wallach's experiment. This experiment suggested one of the interpretational rules for lightness constancy. The surrounding annuli are always of higher luminance than the central disks. Wallach showed that the two central disks are perceived as the same lightness if they have the same *ratio* to their respective surrounds. In the situation in the Figure, subjects will set the luminance of the right hand disk to very close to 50 units. This experiment can actually be seen as an experiment in either lightness contrast or lightness constancy, and suggests that contrast may have evolved in the service of constancy.

That is, the two disks appeared the same shade of grey when they stood in the same ratio to their respective surrounds (a ratio of 1:4 in this case). Wallach's experiment is shown schematically in Figure 23.3.

Wallach's experiment can actually be seen as an experiment in either lightness contrast or lightness constancy. It demonstrates contrast in the sense that the perceived lightness of the disk is influenced by the luminance of its surround. At the same time, it suggests a heuristic for constancy because it shows that equal ratios yield equal perceived lightnesses. When Matilda moves across the lawn from light to shade, the ratio of reflectances between the cat and the grass remain constant. If equal ratios yield equal perceived lightnesses, Wallach's experiment predicts that lightness constancy should occur under these real-world conditions.

A second experiment by Gelb (19xx) introduces a second consideration. Gelb hung a piece of black cardboard in a doorway, so that he could shine light from a hidden projector onto the cardboard (the excess light around the cardboard shone out of the room at an angle, so it was out of the subjects' visual field). Under these conditions of spotlight illumination, the black cardboard appeared white! Next, Gelb introduced a small spot of white paper in front of the black cardboard. Remarkably, the appearance of the cardboard instantly flips from white to black, and flips back again when the white paper is removed. (Try this yourself.)

Figure 23.4: Lightness constancy of complex objects. For most people, the church appears to be painted a uniform white. The parts that send the lower luminances to the eye are perceived to be in shadow, not painted grey. (From Rock (1984), via Palmer (1999), Fig. 3.3.1, p. 123).

23.1.4 Computational schemes for lightness constancy

It turns out that Wallach's and Gelb's experiments reveal two interpretational rules that our perceptual systems use to sort out reflectance from illumination and allow brightness constancy. The first rule is called the *white anchor point*, and it is suggested by Gelb's experiment: the highest intensity region in the field is assigned the perceptual characteristic of white. The second rule can be called the *ratio rule*, and it is suggested by Wallach's experiment: the lightnesses of other surfaces are assigned on the basis of the ratio of their intensities to the intensity of the highest intensity region. For example, a ratio of 1:2 yields the perception of a white annulus and a bright grey disk, whereas a ratio of 1:20 yields the perception of a white annulus and a deep black disk. Ratios of more than about 1:20 break down these rules, and the higher intensity region comes to be perceived as self-luminous (like the moon against the night sky).

Of course, lightness perception is more complicated than these two simple interpretational rules would suggest. For example, the illumination in the world is not always homogeneous – shadows occur – and we need perceptual rules to classify relatively abrupt changes in light level the retinal image. Is it a change of reflectance, or a shadow falling on an object of a fixed reflectance? Experiments suggest that the sharpest edges tend to be seen either as reflectance changes or the edges of three-dimensional objects, whereas more gradual changes tend to be seen as shadows. And very high level rules also come into play – a region of the retinal image that is seen as a single object tends to be seen as having a single lightness, with variations in light level in the retinal image interpreted as shadows or changes of surface orientation, as shown in Figure 23.4.

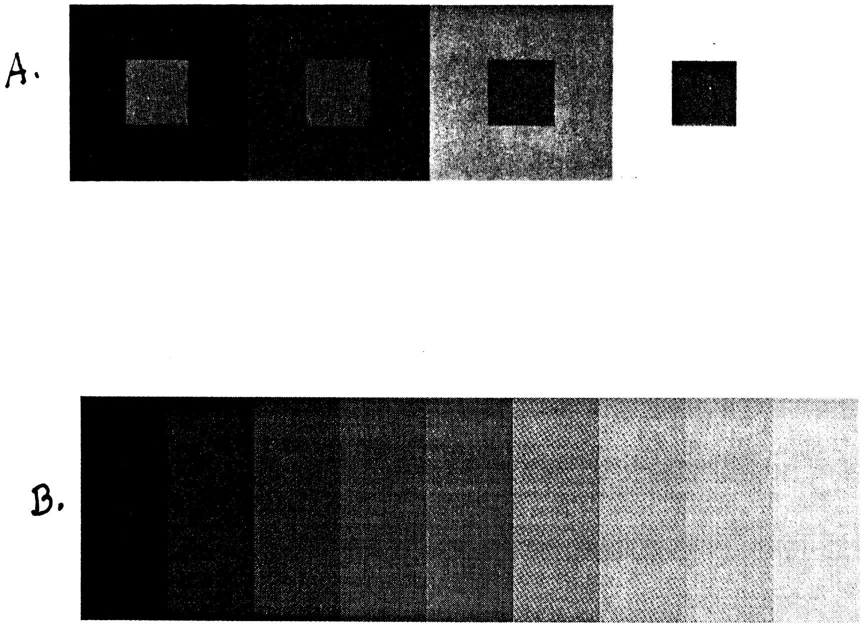


Figure 23.5: COLOR PLATE. Simultaneous and successive color contrast.

23.2 Color contrast and color constancy

23.2.1 Simultaneous and successive color contrast

As in the case of lightness, the perceived chromatic colors of fields of light are also strongly influenced by the wavelength compositions of their surrounds. Some examples of *simultaneous color contrast* are shown in Figure 23.5. The particular perceived colors created by simultaneous contrast are quite predictable, at least in simple situations. Colors induce their complementaries from the opposite side of the hue circle: red regions induce greens, yellow regions induce blues, purples induce yellow-greens, oranges induce blue-greens, and vice versa. More complex patterns, and combinations of luminance and wavelength differences, can also give rise to more dramatic contrast effects, as shown in Figure 23.5.

Successive contrast

Another classic characteristic of perceived color – successive contrast – is illustrated in Figure 23.5. If you stare at a colored pattern for 30 seconds or so, and then shift your gaze to a white field, you will see an afterimage of the original picture with the colors reversed.

What characteristics of the physical stimulus determine perceived color? In earlier chapters we spoke as though (consistent with the lay person's concepts) the perceived color of a light or object is determined by its wavelength composition. But the existence of metamers reminds us that this concept is too simple. We now have a series of other exceptions: simultaneous and successive color contrast bring home the point that *the perceived hue of a light or object does not depend solely on the wavelength composition of the incoming signal from the corresponding region of the retina.* It also depends on the incoming signals from surrounding regions, and on the recent history of retinal stimulation. This point will be critical in developing our understanding of color constancy in the next section.

23.2.2 Color constancy

We now ask, as we asked in regard to lightness: what physical characteristics of an object determine its typical perceived color? The answer is, another characteristic of the surface of the object: its *spectral reflectance function*: The spectral reflectance function is *the percentage of light reflected at each wavelength* from the surface of the object. An object with a maximal or predominant spectral reflectance in the short wavelengths will typically be perceived as blue; a predominance near 500 nm, green; a predominance near 550 nm, yellow-green, and so on.

The term *color constancy* refers to our ability to *perceive an object as having a fixed color, despite changes in the spectrum of the incident light.* Ideally, the perceived color of an object stays constant when the spectral reflectance function of the object stays constant, and changes when it changes. The problem is that (in parallel to the case of lightness and Matilda the cat), the spectrum of the light reaching your eye confounds two things: the spectral reflectance function of the object and the illumination spectrum – the spectrum of the light that illuminates the object. We return to this problem below.

Human color constancy is good, but of limited range. For example, the low pressure sodium lamps that are often used for street lighting emit light of a very narrow spectral band at 589.6 nm. (They typically appear yellow). If you stand under one of these lights at night, with no other source illumination, you will notice that you lose your color vision. Since there is only 589.6 nm light available, all that different objects can do is send 589.6 nm light of different intensities to your eyes. Most people report that objects look distinctly different shades of grey or yellowish grey, but the other colors vanish. Not surprisingly, color constancy is probably quite good over the range of illumination spectra that we encounter in the natural world, and fails outside this range.

23.2.3 Quantitative experiments on color constancy

[To be added: probably Brainard's recent work.]

[Shows that color constancy is good, but not perfect; and better in real-world situations than on a video screen.]

23.2.4 Computational schemes for color constancy

The problem of estimating the spectral reflectance function of an object, however, is the toughest computational challenge we have yet encountered in this book. It is beset by two major complications. First, as stated above, the spectrum of the light that reaches our eyes confounds two factors: the spectral reflectance function of the object and the spectrum of the illuminating light. And second, to specify the spectral reflectance function of a signal completely, one needs a measurement at each wavelength; but in fact our spectral input is limited by our having only three types of cones, each summing broadly over a large region of the spectrum. These problems are beautifully illustrated in Figures 23.6 and 23.7, which were developed by Brian Wandell (1995).

Figure 23.6 shows that the light coming to the eye from an object (which Wandell calls the color signal, but we might prefer to call the incoming spectrum) confounds the spectral reflectance function of the object with the illumination spectrum of the light.

Moreover, the incoming spectrum results only in a particular pattern of quantal absorptions by the S, M, and L cones. Figure 23.7 shows how the pattern of absorption in the three cone types varies, for typical yellow, red, and blue papers illuminated two ways: by a tungsten light bulb and by light from the blue sky. The pattern of quantal absorptions in the L, M, and S cones caused by each paper changes quite dramatically from tungsten to blue sky illumination. Yet if we are to have color constancy, we continue to see the three papers as constant shades of yellow, red, and blue.

The task seems impossible, yet bumblebees can fly: we have reasonably good color constancy. In line with principles developed earlier, the visual system must be using the information available in the incoming stimulus, in combination with interpretational heuristics, to make educated guesses about the spectral reflectance functions of objects. Our task, should we choose to accept it, is to figure out how.

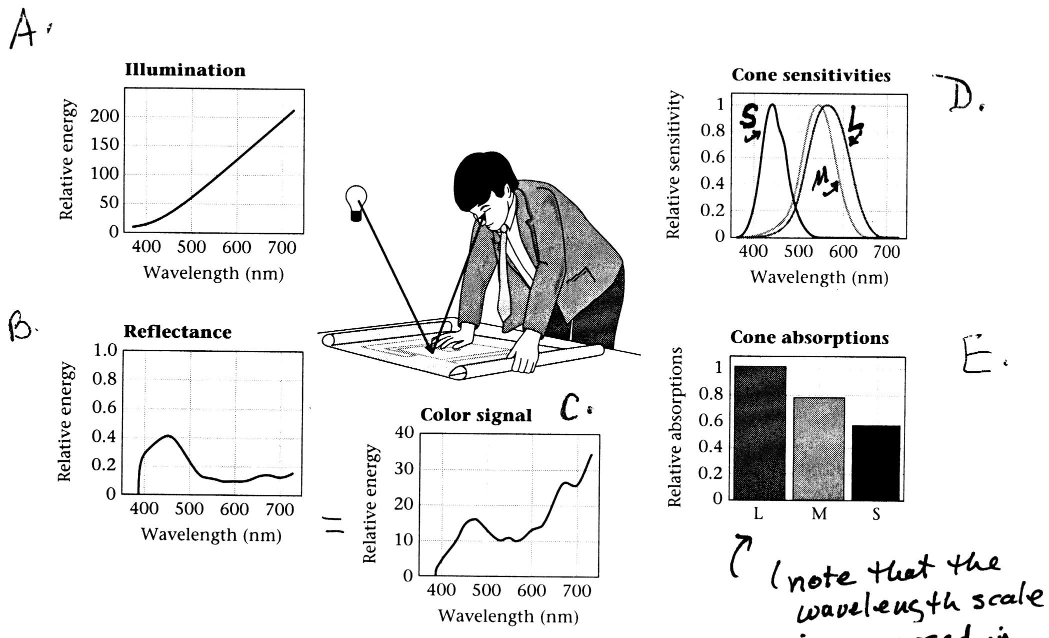
Here's one way. First, suppose we make the (usually reasonable) assumption that the spectrum of the illumination is constant across the whole scene. Then if we could figure out the illumination spectrum just once, we would know it for the whole scene.

If we were building a robot, there would be a simple solution. We could give the robot an extra arm. On the end of the arm we could build in a standard white surface with a flat reflectance spectrum, and we could build knowledge of that reflectance spectrum into the robot's brain. When the robot needed to *estimate the illumination spectrum*, it would extend the arm, take a reading of the incoming quantal catches in its three cone types, and (since it knows the reflectance spectrum of the white surface, and knows that the illumination spectrum is constant across the scene), use yet another set of heuristics to estimate the spectrum of the illumination. It could then apply a constant correction for the illumination to each local set of quantum catches, and derive estimates of the reflection spectra of the objects in the scene.

In a similar vein, it has sometimes been suggested that a person's nose could be used in place of the extra arm and the screen in the robot example. After all, ignoring suntans, each person's nose has a fixed spectral reflectance function, which could be built into that person's computational machinery. If your visual system had a built-in calibration of the spectral reflectance function of your nose, and your nose receives the same illumination spectrum as the visual scene, your nose could act as the standard surface in the robot example, and allow you to have some degree of color constancy. But of course with this scheme your color constancy would fail if your nose receives a different illumination spectrum than the rest of the scene.[2]

Another approach to how our perceptual systems do color constancy was proposed by Gershon Buchsbaum (1980). Buchsbaum suggested what has come to be called the *grey world assumption*;

[2] This leads to a funny story. DT was sitting in her living room one sunny morning, pondering the question of whether the nose contributes to color constancy. There was a prism hanging in the window, twisting gently on its string. As she pondered, a rainbow from the prism came toward her and drifted slowly back and forth across her nose, varying its illumination spectrum. So she seized the empirical moment, and paid very close attention to the perceived colors of objects in the room. The perceived colors remained rock solid, and as others had before her, she rejected the nose-as-a-standard-surface theory of color constancy.



(note that the wavelength scale is reversed in this figure. We might have preferred the S cones on the left.)

Figure 23.6: The dual problem of color constancy, as depicted by Wandell (1995). The lefthand side of the figure shows the first problem: the illumination spectrum of the light falling on an object (A), and the spectral reflectance function (SRF) of the object (B), are confounded in the spectrum of the light coming to the eye (C). The righthand side shows the second problem: The three cone types, the S, M and L cones (D), catch quanta from the incoming light, and make three quantum-catch signals (E). Thus the entire spectrum of the incoming light is reduced to only three signals. The puzzle of color constancy is, how do we back calculate to the reflectance function from the cone quantum catches? It seems impossible. (Modified from Wandell, 1995.)

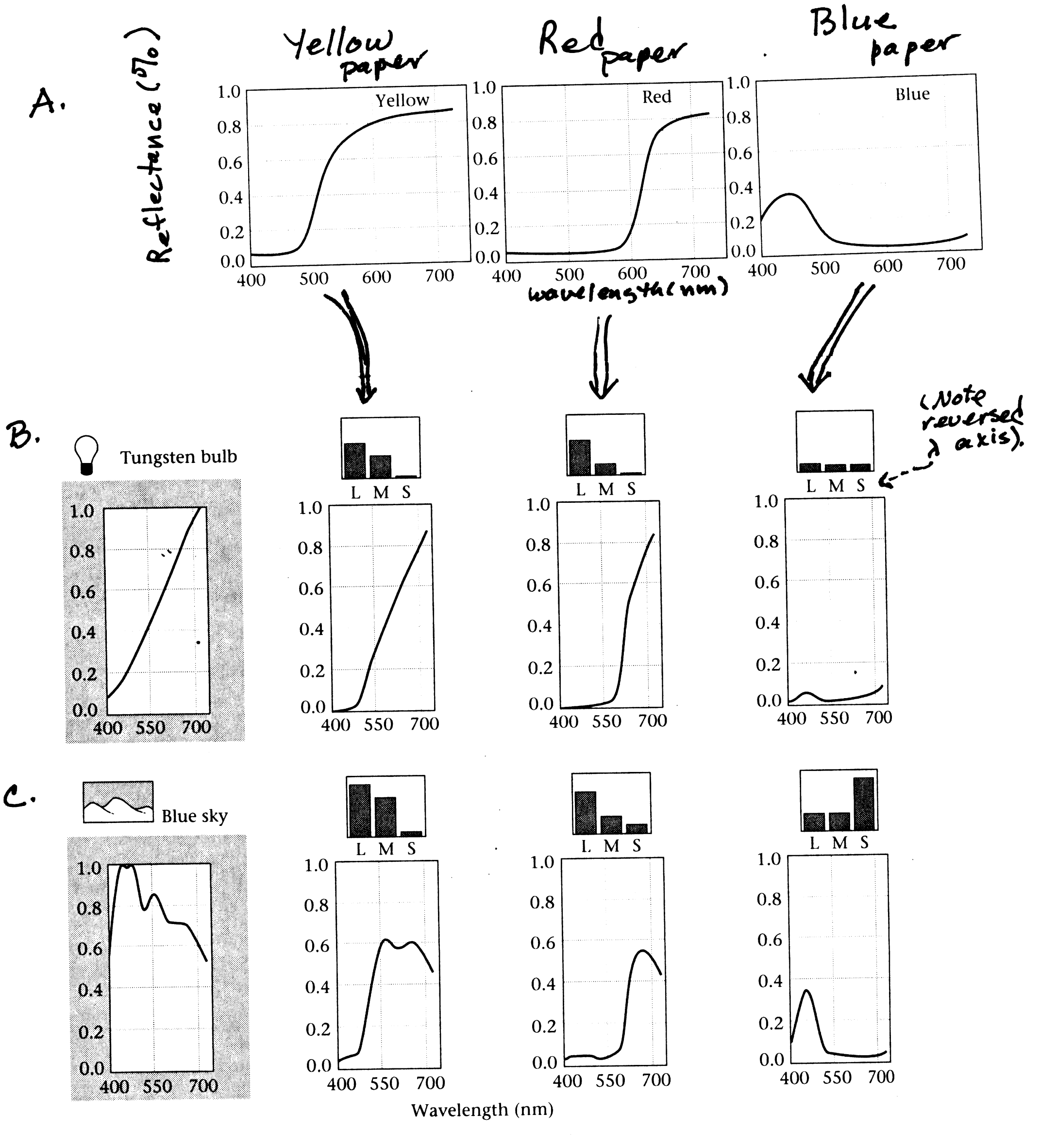


Figure 23.7: Another illustration of the problem of color constancy. Row A shows typical spectral reflectance functions from pieces of paper that are peceived as yellow, red, and blue. Row B shows the illumination spectrum from a tungsten light bulb, and the quantum catch pattern in the L, M, and S cones that results from illuminating each of the three papers with the tungsten bulb. Row C shows the illumination spectrum from the blue sky, and the quantum catch pattern that results from the three papers illuminated with blue skylight. The difference is particularly dramatic with the blue paper. Color constancy allows us to see the yellow paper as yellow, the red paper as red, and the blue paper as blue under both illuminations. (Modified from Wandell, 1995).

that is, the interpretational heuristic that the *average* reflectance spectrum of all the surfaces in the image is flat (as is the reflectance spectrum of a white or grey piece of paper). The illumination spectrum could then be estimated from the spectrum reflected by the average object. It is easy to see that this interpretational rule would probably work in cases in which the average spectral reflectance function actually is flat – for example, in a room filled with objects of many different colors – but would fail in cases in which most of the objects have similar spectral reflectance functions. Color constancy should be ruined in an all-blue bedroom.

A third approach attacks color constancy through the concept of *basis functions*, as illustrated in Figure 23.8. Although a quantitative treatment of this approach is beyond the scope of this book, let's go with some intuitions. The general idea is that, in principle, we can imagine spectral reflectance functions of arbitrarily complex form, as shown in Figure 23.8A. To specify such a reflectance function would take many numbers (say, one for each wavelength, or 300 numbers, just to make up a concrete number). However, in the physical world in which we evolved, the spectral reflectance functions of *natural objects* do not take all possible forms. Instead, the spectral reflectance functions of most natural objects vary smoothly and slowly with wavelength. In consequence, the reflectance functions of most natural objects can be synthesized reasonably well from only a small number – two to six – of underlying spectral reflectance functions.

As an example: if you choose to use three synthesizing functions, it turns out that the three best ones consist of a function characterizing the overall (average) reflectance, and two other functions showing respectively the "blue/yellow" and "red/green" dimensions. These descriptive functions are called *basis functions*. To specify the approximate spectral reflectance function of an object, then, three numbers are enough: the weights of each of the three basis functions that together come closest to synthesizing the spectral reflectance function of the object.

The same kind of analysis also works with illuminants. Instead of (say) 300 numbers to specify the illuminant spectrum, two or three basis functions are sufficient to specify most natural illuminants with reasonable accuracy. So roughly five to six variables, rather than (say) 600, are actually sufficient to specify both the spectral reflectance functions and the illumination spectra encountered in natural scenes. With the use of appropriate heuristics that make use of such basis functions, the problem of identifying surface reflectance functions – color constancy – begins to to seem almost tractable.

Now, let's add one additional assumption – that the illumination spectrum is constant across the scene, so that if it is estimated once, it is known for the whole scene. Without going into more details, you can probably intuit that the problem is now becoming simple enough that the brain might actually be able to solve it when viewing a scene that includes objects of many different colors.

In short, both the simple and the complex computational schemes described above allow estimates of the approximate illumination spectrum and the reflectance spectra of objects in a scene. All of them succeed reasonably well over certain ranges of illumination spectra and visual scenes. But all of them fail in circumstances outside the range over which their particular heuristics are designed to work. For example, they all fail when the illumination consists of only a single wavelength of light. What remains is to work out the quantitative predictions of each heuristic, and to test these heuristics systematically against each other by studying how good human color constancy really is under various conditions.

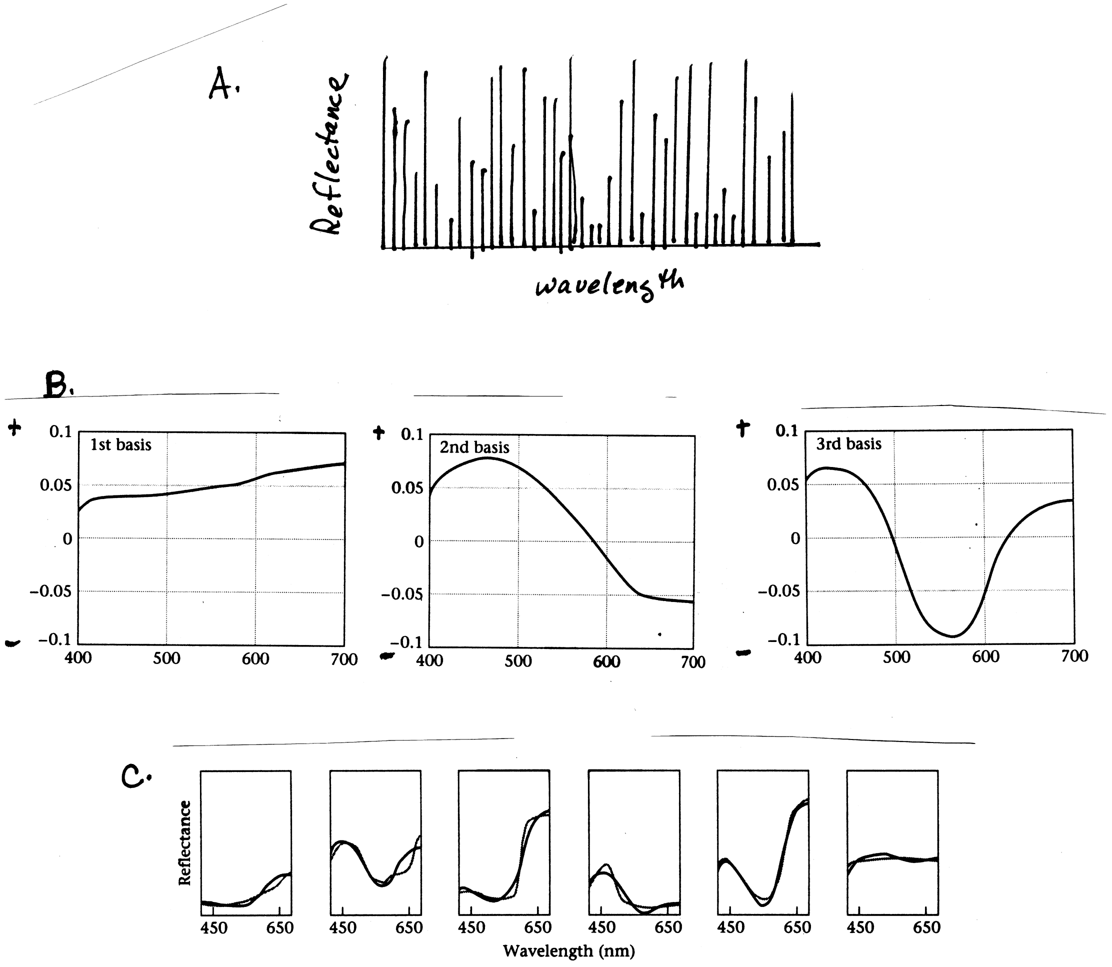


Figure 23.8: Basis functions. A. One can imagine surface spectral reflectance functions that are spiky as in A, but in fact the SRF's of natural objects are smooth like the colored papers in Figure 23.7. B. Basis functions are a set of a small number of functions which, when combined in different proportions, can simulate all of the functions in a much larger set. Here, Wandell presents three optimal basis functions that can be combined to simulate the surface reflectance functions of many different colored surfaces. C. Synthesis of six of the colored surfaces used by Wandell, from optimal combinations of the three basis functions shown in B. The light lines show the spectral reflectance functions of the surfaces. The dark lines show the synthesis of each spectrum. Not too shabby. (Two basis functions would do less well, and four or five would do even better – but three could be good enough to provide us with reasonable color constancy.) (Modified from Wandell (1995), Fig. 9.8, p. 303.

23.3 Physiology: Retinal adaptation as a color constancy mechanism

One final note on color constancy. We showed in Chapter 10 that brightness constancy could be nicely served by the changes of dynamic range inherent in light and dark adaptation. Similarly, it turns out that some of the calculations needed for color constancy can be served by differential chromatic adaptation: by allowing signals from each cone class to light and dark adapt independently, reducing the signal from each cone type in proportion to its average quantum catch. The more long wavelength light in the illumination spectrum, the more the L cone signal would be reduced; and similarly for the other two cone types. And in fact, some of the most recent work from the Dacey lab shows that at least some independent adaptation does occur separately for the three cone types xx.

Color thus provides another example of how an apparently complex aspect of perception – color constancy – might be computed at least partially by a very peripheral part of the visual system. Even though a visual function involves complex heuristics, the relevant calculations can actually occur early or late in visual processing (or partially at each of several levels).

23.4 Central physiology of color

To the extent that lightness constancy and color constancy are not accounted for solely in the retina, we would like to ask, how does central color processing achieve these constancies? The answer is, we don't know. The best we can do at this stage of the science is juxtapose what we do know about central color processing with our account of perceived color, and wait for the insights of the next generation.

But here we come to a problem. In most cases in visual science, experimental results are replicable, and within levels of analysis one scientist can pretty much count on finding results that are compatible with those of his predecessors. In the field of central physiology of color we find a rare exception. This field is an example of a case in which results have often not been replicated, and no consensus view on how color is processed centrally has been achieved as of the turn of the milennium. Given the unsettled nature of the field, it's difficult to decide what to present to students, and it's difficult to guess what any new consensus view will be, or when it will be achieved.

Earlier in this text, we took the position that null plane analysis provides an elegant paradigm for describing color codes. Our choice for this section, therefore, is to center our treatment of central color coding around a study in which null plane analysis has been used. In the section following this one we will summarize briefly some of the historical disagreements that have arisen in this field.

23.4.1 Null plane analysis applied to central neurons

At the end of our chapter on the primate retina, we summarized the physiological responses of the seven (or eight) major kinds of retinal output neurons, and compared their properties to the early color channels postulated by psychophysicists. We found that the Boynton code was roughly but not exactly instantiated at the retinal output. In particular, the null planes of the P (midget) cells are not as well behaved as we would like. The null planes of P cells include or nearly include the tritan axis, but not necessarily the achromatic axis – the null planes of different P cells tilt away from their expected null plane by varying amounts. In other words, P cells respond to both red/green and achromatic modulations, suggesting that they carry (either multiplex or confound) both an achromatic signal and an early r/g signal. The K (small bistratified) cells have similar limitations.

Moreover, whereas M (parasol) cells have null planes that fall near the isoluminant plane, there are relatively few M cells, and they are particularly scarce in the fovea. These factors suggest that if M cells do play a role in the psychophysical color code, they carry at best a low spatial frequency achromatic signal. High spatial frequency achromatic signals are thought to be multiplexed with the early r/g signal in the P cells. On this view, P cells carry both the early r/g and the achromatic channels of the early opponent code.

As described in Chapter 15, the retinal color code is little changed at the LGN. So now the question becomes: what about cortex? When we apply null plane analysis to cortical neurons, will we find that the retinal color code is maintained? Or will we find one or more additional recodings, and if so, at what cortical levels and in which of the cortical streams?

In the study that forms the core of our account, Lennie, Krauskopf and Sclar (1990) extended the use of null plane analysis to neurons in V1 cortex. Each neuron they encountered was first classified as to its receptive field type – non-oriented, simple, or complex – and its best spatial frequency and orientation were determined. Using the best spatial frequency and orientation, the null plane of the cell was then determined, and each neuron was characterized by its *best axis* – the axis normal to its null plane. In addition, Lennie et al were able to document in which layer the majority of their neurons lay, so that they could address the issue of whether the color code is different in the different layers and cell types within V1. These results were then compared to the data from LGN P cells. The format for Lennie et al's graphs (which is very complex) is explained in Figure 23.9, and the data are shown in Figure 23.10.

To make a long story short, Lennie et al found major changes in the color code between LGN P cells and the V1 input neurons, the non-oriented cells in layer 4 of V1. Most of the non-oriented cells lay in layers 4A and 4Cβ, the layers known to receive LGN P cell inputs.

Figure 23.10A shows the best axes of LGN P cells. The best axes of the non-oriented cells, shown in Figure 23.10B, are much more widely distributed than are those of LGN cells, and cover the whole possible range of isoluminant modulations in color space. Thus, the non-oriented V1 cells have widely scattered color preferences, and the two chromatic axes favored in the retina and LGN – early r/g and early b/y – give way to a code in which different cells respond best to many different chromatic modulations. Moreover, unlike the findings of Derrington et al (1984) in the LGN, there are quite a few non-oriented cells with their null planes near the isoluminant plane – responding well to achromatic modulation. An explicit achromatic channel thus begins to appear in the non-oriented cells in V1.

Lennie et al also looked at the chromatic properties of simple and complex cells, mostly in the upper layers (2 and 3) of V1. These data are shown in Figure 23.10C and D. Surprisingly, responsiveness to isoluminant color variations seems to be even less at these levels, and the two early opponent axes are completely lost in the variability of the chromatic axes to which these neurons respond. Most simple cells respond well to achromatic modulation, and poorly to chromatic modulation. In complex cells this trend is even more pronounced, with most complex cells having their null planes quite close to the isoluminant plane and their best axes close to the achromatic axis.[3]

[3] Unlike retinal cells, cortical cells have little or no maintained discharge. This makes it difficult to use an opponent

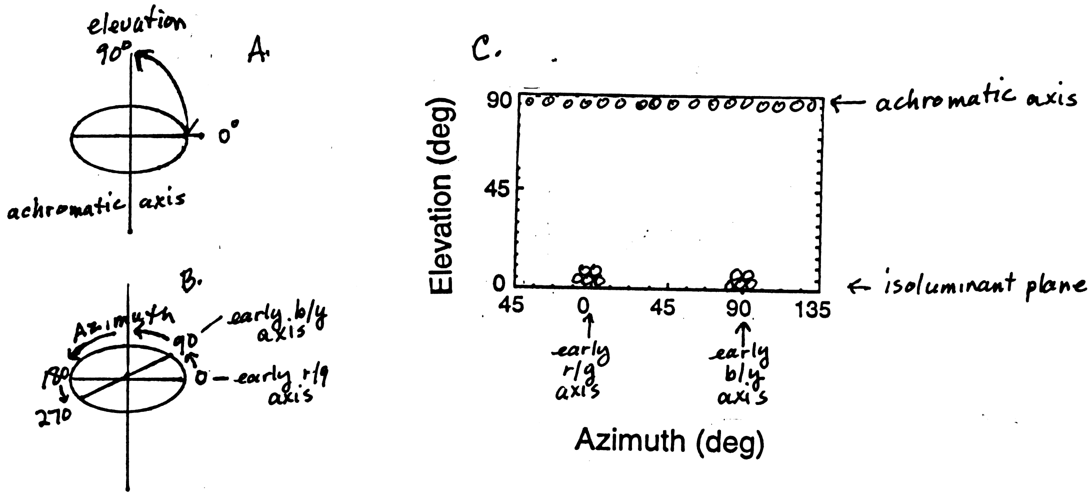


Figure 23.9: An explanation of Lennie et al.'s (1990) data plotting format. A and B: A three dimensional color space in polar coordinates. Lennie et al (following Derrington et al, 1984) specified axes through the white point in three dimensional color space in terms of their *azimuth* and *elevation.* A similar coordinate system is in specifying coordinates on a globe, with the axis of rotation of the earth corresponding to the achromatic axis (elevation), and the plane of the equator corresponding to the isoluminant plane. In Derrington et al's space, the *azimuth* is the angular position around the isoluminant plane (shown in B). The 0 - 180^o axis is assigned to represent the early r/g axis, and the 90-270^o axis represents the early b/y axis. The *elevation* is the angular position above or below the isoluminant plane (shown in A). 0^o elevations lie within the isoluminant plan, and 90^o elevations are assigned to represent the achromatic axis. B: Format of the data plots. In the data plots, *azimuth* is unwrapped along the x axis. The 0^o azimuth is marked *early r/g axis* and the 90^o azimuth is marked early b/y axis. The 45^o and 135^o points on the x axis are redundant. *Elevation* is represented on the y axis. The line at 90^o elevation is marked *achromatic axis.* Lennie et al specify a cell by its *best axis*, which is defined as the axis orthogonal (perpendicular) to its null plane. Each dot in these data plots represents a single neuron. The x-axis (azimuth) value for the neuron shows the orientation of the best axis in three-dimensional color space, projected down to the isoluminant plane. The y-axis (elevation) shows the tilt of the best axis of the neuron above the isoluminant plane – zero elevation indicates that the best axis is within the isoluminant plane, and 90^o elevation indicates that the best axis is the achromatic axis. We are now ready to ask where neurons that conform to the Boynton code would fall in these plots. Neurons that embody an early r/g channel would have best axes with 0^o azimuth (x axis) and 0^o elevation (y axis); neurons that embody an early b/y channel would have best axes at 90^o azimuth and 0^o elevation; and neurons that embody an achromatic channel would have best axes at 90^o elevation and any azimuth (actually, the azimuth is not meaningful if the axis is exactly at 90^o elevation). The three predicted clusters of cells are shown.

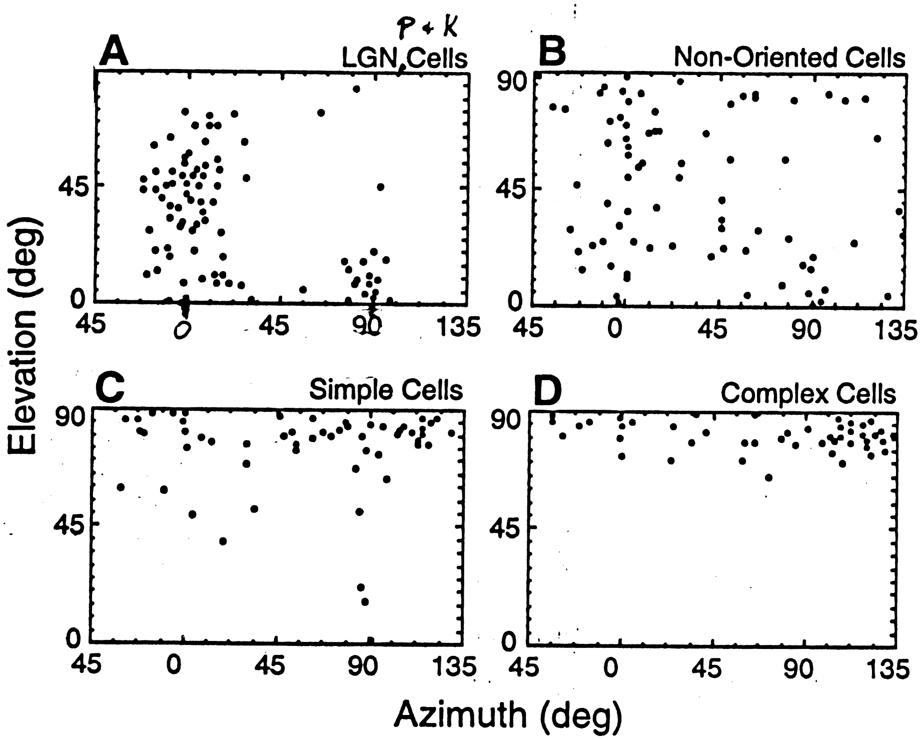


Figure 23.10: Color codes of four different kinds of visual neurons. Data from Lennie, Krauskopf and Sclar (1990), for LGN cells, non-oriented cells in V1, simple cells in V1, and complex cells in V1. [The format of these plots is forbidding, and is explained on the facing page. If too hard, skip to the bottom line.] The bottom line is that the patterns of data points are very different on the four graphs – that is, these four different types of neurons reveal very different color codes. A: *Parvocellular LGN cells (P cells)* have best axes that coincide well with the early r/g (azimuth 0^o) and early b/y (azimuth 90^o) axes, but the cells tuned to the early r/g axis have variable elevations – varying tilts of their null planes (just as we saw in retinal ganglion cells). Very few cells carry a pure achromatic signal (elevation 90). This work was published in 1990. We would now divide these two groups of cells, calling the group at azimuth 0^o P cells, and the group at azimuth 90^o *K cells.* B. *Non-oriented V1 cells* are much less regular in their color tuning – they have many different azimuths. There are no longer just two chromatic axes! Also, there are a fair number of cells tuned approximately to achromatic modulation (elevations near 90^o). C. *Simple V1 cells.* Few simple cells carry purely chromatic signals (i.e. few have their best axes anywhere near the isoluminant plane at elevation 0^o), and many are tuned approximately to achromatic modulation (elevations near 90^o). D. *Complex cells* are even more extreme in responding to achromatic and not chromatic modulation. (Modified from Lennie, Krauskopf, and Sclar (1990).

So in sum, and remarkably, in the simple and complex cells in the upper layers of V1 we find what seems to be a dramatic diminution of chromatic coding. Most simple and complex cells respond well to achromatic variations and poorly to chromatic variations; and those that do respond to chromatic variations exhibit preferences for a wide range of different axes of chromatic modulation. Surprisingly, it looks as though the elegant three-channel color code we saw in the retinal output has been replaced by a multiple channel code at V1. Moreover, color seems to be deemphasized in favor of responses to achromatic modulation. It seems almost as though any simple color coding is lost once the upper layers of V1 are reached. We will ask why this might be so later in this chapter.

After presenting these data, Lennie et al raise the question: How does information about color coding traverse V1 to become available at higher levels of visual processing? Lennie et al found a population of non-oriented, chromatically coded neurons in layer 6 of V1. They note that there are reports of a projection from layer 6 of V1 directly to V2. (Notice that this projection is another violation of the rule that feed-forward projections arise from the upper layers of cortex.) Lennie et al suggest that these neurons may be a major pathway for chromatic signals – perhaps the chromatic signals pass from V1 to V2 via this route, rather than through the upper layers of V1.

What about color coding in cortical areas beyond V1? There are only a few studies of higher cortical areas in which null plane analysis has been used. To date, most descriptions of color coding in V2, V3, and V4 suggest that the color code in these areas is similar to that in V1, with many neurons more responsive to achromatic than to chromatic modulation; and the few neurons that seem color tuned being responsive to a wide variety of directions in three-dimensional color space. [Ref: Kiper, Levitt, and Gegenfurtner – DT must read xx]. Although the data are scarce, there is little evidence concerning any further changes of color code at higher cortical levels.

What about dorsal and ventral streams? Beyond V1, V2, and V3, the general consensus is that color coding – such as it is – is largely or entirely confined to the ventral stream. In fact, it is true that most of the neurons that have been analysed for their responsiveness to chromatic differences lie in the ventral stream: V4 and IT cortex. However, it is important to keep in mind that rumors can be easy to start and hard to eradicate – who wants to do a sophisticated study of the chromatic properties of neurons suspected of not being responsiveness to chromatic differences? The truer answer is that dorsal stream neurons have not often been tested with chromatic stimuli. [But Dobkins and Albright....]

23.4.2 Discordant results in central color processing

The above description of the physiology of central color coding has been confined to studies in which null plane analysis has been used. Earlier studies with other paradigms, however, have revealed a variety of different results. Unfortunately the most interesting of these have not confirmed by more recent studies using the null plane analysis paradigm.

Early reports of central color coding reported a special kind of neuron called a *double opponent* neuron. A double opponent neuron is defined as a neuron that is both spatially and spectrally opponent, and has inputs of opposite sign from two different cone types in both the receptive field center and the surround (e.g. +L- M in the center, and +M - L in the surround). Since these

color code, with a neuron increasing its firing rate in response to to some stimuli and decreasing its firing rate to other stimuli. Following earlier authors, Lennie et al found that these neurons could be described as carrying a *rectified* signal; that is, a signal in which signals that might have been negative are flipped across the abscissa and represented by absolute (positive) values.

neurons would respond best to a disk of light of one wavelength surrounded by an annulus of another, they were hailed as providing a possible neural correlate for simultaneous color contrast. (Of course, the disk would have to be exactly the right size.)

Other early reports suggested that color coded neurons are found predominantly in the blobs and not in the interblobs, and particularly that double opponent neurons were found frequently in the blobs. Similar reports suggested that in V2, the thin stripes (to which the blobs project) are also rich in color-coded, double-opponent neurons.

Other early investigators looked at area V4. Early reports suggested a high concentration of narrowly tuned, chromatically driven cells in area V4 of visual cortex. It was also claimed that these neurons changed their color specificity in concert with the perceived colors seen by human observers viewing the same patterns! These reports generated much excitement, as such cells, if they existed, could provide the guts of a theory of human color constancy. And in fact, these early findings of specialized color cells in V1, V2, and V4 contributed greatly to the emergence of theories of the dorsal and ventral streams having different visual functions, with the ventral stream being further divided into a stream for color and a stream for form, as discussed in Chapter 19xx.

Unfortunately, most of these results have not been replicated in more recent studies, particularly in studies in which null plane analysis has been used. In particular, Lennie et al (1990) saw few double opponent neurons, and little if any concentration of color coding neurons in the blobs. Others have failed to find a concentration of chromatic cells in the blobs or in the thin stripes of V2. The narrow chromatic tuning of V4 cells has also not been seen by others. The only one of these early claims that still evokes consensus is that color coded cells, to the extent that they occur at all, are largely or entirely confined to the ventral stream.

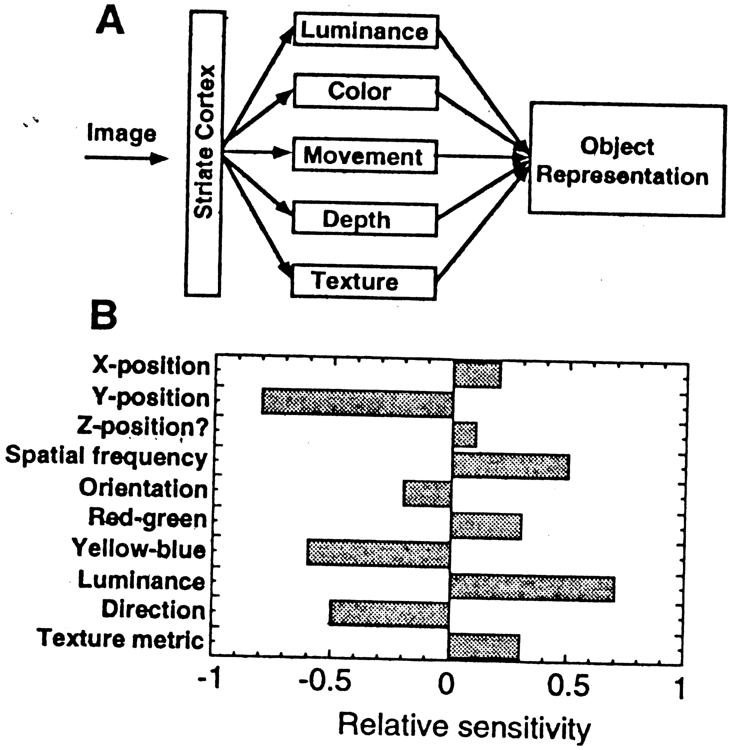
What are scientists to believe when the reports of different laboratories disagree? The are two major possibilities. The first is to believe that some of the empirical reports are simply wrong. The second, more charitable option is to believe that there are known or unknown differences among the studies that lead to the differences in results, and that when we truly understand color coding in cortex our theoretical view will be able to encompass all of the different kinds of data.

As confessed earlier, DT's own bias is to put most of our faith in the paradigm of null plane analysis, and we therefore tend to discount the early reports. However, others would disagree, and there is nothing to do but wait and see how these discrepancies are finally resolved.

23.4.3 Close-coupled coding of color with other dimensions?

Why does the color code seem to peter out in V1 cortex? One possible answer, proposed by Peter Lennie (1998), is that color is almost never seen in isolation, but rather as the property of the surface of an object. In functional terms, the goal of encoding the spectral reflectance function of an object is not to encode color per se, but rather to assist in image segmentation and object identification.

A schematic illustration of Lennie's view is shown in Figure 23.11. In Lennie's view, the analysis of color is probably ”close-coupled” to the analysis of form. Lennie points out that V1 neurons are selective on many dimensions simultaneously: spatial frequency, orientation, and motion as well as color. In fact, in most studies of responses of cortical neurons to color, the neuron is first characterized by its spatial frequency and orienation selectivity; and then its chromatic properties are studied with stimuli of the optimal spatial frequency and orientation. One cannot say a V1 cell is a color cell any more than a spatial frequency, or orientation, or motion cell. Instead, each



"Independent"

" close-coupled"

Figure 23.11: Independent vs. "close-coupled" analysis. Panel A shows the idea of independent analysis of different stimulus dimensions, and reunification of the dimensions at the level of object recognition. Panel B shows the idea of "close-coupled" analysis. It shows the hypothetical response profile of a neuron in V1 or post-V1 cortex. The cell is tuned to some degree on each of many stimulus dimensions. Activity in different cells represents different combinations of values on the various dimensions. (From Lennie (1998).

V1 neuron should be seen as representing one relatively confined location in a multi-dimensional stimulus space – e.g. activity in this cell signals a 1 cy/deg, vertical, leftward moving orange grating component. Similarly, one cannot expect to find purely color-coding areas anywhere in cortex, and finding neurons that respond to color does not constitute evidence for the area being a color coding area per se.

23.5 For the philosophers: Are colors objective or subjective?

Students who have taken philosophy will know that philosophers have classically had a fascination with color, and that examples from color are often used to illustrate some of the philosopher's deepest and most puzzling questions. In particular, color provides the basic illustration for a question which has come to be called the Objectivist/Subjectivist question. The question is: What is the fundamental nature of color? Is color a property of objects – their surface spectral reflectance functions – or is it a property of human consciousness – perceived color ? Or more recently, might color be a property of the neurons that underlie the perceptual processes?

From a vision scientist's (DT's) perspective, as of the turn of the milennium, this question has collapsed to a false dichotomy (or trichotomy if one counts the neural option). As you have seen throughout this book, vision scientists have developed a complex, highly articulated conceptual framework in which surface spectral reflectances, neural states, and perceived colors are all necessary and well developed parts. It seems to DT that choosing one option and calling it the "true" nature of color does not do justice to the complexity of the scientific picture. From DT's perspective the question, is color fundamentally subjective or objective, does not survive the vision scientist's elaboration of the concepts necessary to understand what color is. If she had to give a definition, it would be a functional one, and it would include all three levels: color is the brain's and the mind's attempt to represent the surface spectral reflectance functions of objects.

Take a similar but less confusing example – the case of distance. One could ask, is distance a property of objects? Or is it a property of our perceptions? Or is it a property of the neural states that give rise to our perceptions? The answer is, these are three different concepts, and they need different names – let's say physical distance, perceived distance, and the neural representation of perceived distance. Of course, the EDC has provided that physical distance and its neural and perceptual representations will coincide well most of the time! Similarly, in the case of color we need to distinguish spectral reflectance functions from perceived colors from the neural representations of perceived colors. Once we understand that perceived color is just the brain's and the mind's attempt to represent the surface reflectances of objects, and that color constancy is very hard but we are pretty good at it, the philosophical question seems to DT to disappear.

(Of course, DT is also sure that for philosophers the Subjective/Objective question will survive the elaborations of color science, and she's waiting for the thunder of rejoinder from philosophers.)

23.6 Summary: Why do we see colors?

In summary, in this chapter we have described some of the complexities of perceived color. We have introduced lightness constancy as the brain's and the mind's attempt to represent the reflectances of objects in the environment, and color constancy as the attempt to represent their spectral reflectance properties. In terms of neural coding, we have seen that the description of the central coding of color is unusually controversial at the present time, and that (in DT's view) the best experiments seem to show the representation of chromaticity diminishing as we go to higher stages of processing in area V1.

Why do we have color vision? Color vision probably did not evolve so that we could enjoy rainbows, or even because we needed to represent the spectral reflectance functions of objects per se. Color vision probably evolved because spectral reflectance functions sometimes signal important properties of objects or other elements of our environments: whether fruit is ripe or not; whether leaves are tender or not; whether grass is wet or not; whether it's going to rain or not; and so on.

The flip size of this argument, as represented in Lennie's views, is that color is useful to us not as an isolated sensation, but rather as a property of the surface of an object. In functional terms, the goal of encoding the spectral reflectance function of an object is not to encode color per se, but rather to assist in image segmentation and object identification. On this view it is not surprising if the representation of color blends together with the representation of forms and objects, rather than being maintained in a separate neural code at higher levels of visual processing. But it remains to be seen whether or not this view will stand the test of time.

Chapter 24

The Third Dimesion: Distance and Size

In this chapter and the next we return to a problem we first encountered in Chapter 5. The physical world is three-dimensional, but the retinal image is two-dimensional. The First Transformation preserves the two-dimensional layout of the visual field – left and right, up and down – but collapses the third dimension. The facts of the First Transformation would seem to imply that information about the *distance* to an object, as well as its *depth* or *three-dimensional shape*, should be lost to perception.

Moreover, the size of the retinal image of an object depends both upon the size of the object and its distance from the eye; that is,

retinal image size = object size/distance

If information about distance is lost at the level of the optics, then information about size should also be lost, and we should be unable to perceive the sizes of objects accurately.

Yet, bumblebees can fly. Clearly, we do perceive the distances, sizes, and shapes of objects, and we usually perceive them remarkably veridically. So somehow, we are getting the information to allow us to do these things. The question is, how? Part of the answer is that the retinal images contain many features that come about because of variations of distance. In fact, we use these features – *distance cues* – to generate veridical perceptual representations of the distances and sizes of objects. The main purpose of this chapter is to identify the many different distance cues, and to ask how they combine to determine our perceptions of distance.

In addition, the topic of distance perception provides an excellent illustrations of one of our most important themes. As it turns out, there are many different cues to depth and distance. But not one of the cues contains all of the information needed to use that cue as an information source about distance. That is, each of these different cues depends upon the use of its own special heuristics –interpretational rules. A major goal of this chapter is to exercise your understanding of the concept of heuristics, and more generally top-down processing in perception.

24.1 Definitions: Distance vs. depth

We begin with a definitional distinction between the terms *distance* and *depth*. Although these terms are often used interchangeably, we will use them in slightly different ways: *distance* will be used to refer to the physical distance from the observer to some object. *Depth* will be used to refer to the *three-dimensional shape* of an object – for example, whether a given object is a three-dimensional sphere vs. a flat two-dimensional disk, or a three-dimensional human figure vs. a flat human-shaped cut-out. Of course, in a sense the depth of an object is just the difference in distance between its front and back surfaces, and the cues for distance and depth will clearly be closely related and overlapping. However, we believe that the definitional distinction is an important one.

In Chapter 20 we discussed the concept of figure-ground segregation – the perceptual task of determining which part of the retinal image is the figure and which is the ground. Similarly, in three dimensions, our perceptual systems need to work out the task of *scene segmentation*: Which blobs of space are occupied by objects, and which by empty spaces? It seems likely to DT that the eventual representations of objects will be very different from the representations of empty spaces. The analysis of the depths and three dimensional shapes of objects is a major task of object recognition, and it needs to be distinguished from the task of distance perception per se.

A second, related reason to emphasize the distinction comes from our simplest formulation of the functions of dorsal and ventral streams: Where vs. What. To know the distance of an object is to know *where* it is in space. But to know the three-dimensional shape is to begin the task of determining *what* it is; that is, the task of object recognition. If where and what are analysed separately, then distance and depth should be.

In this chapter, we will treat the logic and psychophysics of distance perception. We begin with a review of the cues that seem most likely to provide us with information about distance. We then present a brief discussion of the problem of how distance cues combine. Finally we relate the perception of distance to the perception of size. In the next chapter we will turn to the topic of depth, and to recent research on the quantitative combination of depth cues. We then turn to physiological studies of distance and depth cues together, and ask what is known about the physiological representation of distance and depth in both dorsal and ventral streams.

24.2 Cues to distance

A *distance cue* is a feature of the retinal image that comes about because of variations in distance. Its presence in the retinal image is therefore a potential source of information about distance.

What are the various distance cues? As a matter of logic and speculation, this question fascinated both the artist Leonardo da Vinci and the philosopher George Berkeley. The many cues they discovered, along with one they missed, still form the heart of our understanding of distance perception. In the next few sections we will explore several different kinds of cues to distance: accommodation and convergence; pictorial cues; cues based on motion; and binocular disparity.

24.2.1 Accommodation and convergence

The first two distance cues are created by the motor responses of our visual systems. These cues are sometimes called *ocular* or *physiological cues.*

The first of these two cues is *accommodation* (Figure 4.9). As discussed in Chapter 4, the lens of the eye changes its thickness in order to bring objects at different distances to focus on the retina. In fact, the thickness of the lens required for best focus of a given object varies inversely with the distance of the object. Thus, if your perceptual systems received information about the

Figure 24.1: Convergence. A. When you look at a distant object, and rotate your eyes to place the two retinal images of the object on the two foveas, your two lines of sight are nearly or exactly parallel (left panel). When you look at a close object, you *converge* your two lines of sight to place the two images on the two foveas (right panel). If your visual system kept track of the convergence angle needed to align the two images on the two foveas, and had a look-up table converting convergence angle to distance, you could potentially use convergence angle as a distance cue. B. The convergence angle needed to bring the two images into register on the two foveas, as a function of the distance of the object from the observer. When a near object changes in distance, there is a relatively large change in the convergence angle. But when a far-away object changes in distance by the same amount, the change in convergence is too small to be useful. Hence, variations of convergence are much more useful as distance cues for near than for distant objects. (A by DT; B after Palmer, 1999, p. 206, Fig. 5.2.3).

state of accommodation of your lens, you would have some potentially useful information about the distance of the object – that is, accommodation is a potential distance cue.

The second of these cues, shown in Figure 24.1, is called *convergence*. When you fixate on a distant object, your two eyes point straight ahead, so that your two lines of sight are parallel or nearly so. When you shift your fixation to a nearby object, you rotate your eyeballs inward to *converge* your two lines of sight, so as to bring the two retinal images of the object onto the two foveas. The closer the object, the more the needed convergence. And, as in the case of accommodation, your state of convergence is a potential distance cue. (You can demonstrate convergence by asking a friend to first look at a distant object, and then look at his nose. You can also feel the changes your extraocular muscles make when you do the same thing.)

There are several points to be made about accommodation and convergence as distance cues. First, both accommodation and convergence are cues to *absolute* distance – the state of accommodation or convergence varies in correspondence with the actual distance from your eyes to the viewed object. (As we will see, this is important because most other cues are cues to *relative* rather than absolute distance.)

Second, neither the accommodative state nor the state of convergence *per se* is useful as a distance cue – both need augmentation by other information. To use accommodation as a cue to object distance, the visual system must determine the state of the lens *required for the sharpest retinal image*. Similarly, to use the convergence angle, the visual system must analyse the two incoming retinal images, *determine which sets of contours in the two eyes arise from the same object, and put these two sets of contours in register*. That is to say, these physiological cues are not simple, and they require extensive image analysis before they can be used for distance perception.

But third, once the state of accommodation needed for good focus, or the convergence needed for binocular alignment, have been determined, only minimal heuristics are required. All that is needed is a lookup table – a list – identifying each state of accommodation or convergence with the corresponding distance value. (And again as you will see, this simplicity sets off the ocular cues from most other distance and depth cues, for which much more complex heuristics will be needed.)

And finally, over what range of distance are these two cues useful? Suppose you are looking at two objects that differ in distance by (say) five cm. As shown in Figure 24.1B, at near distances such a small change of distance brings about a substantial change of convergence. But at larger distances, the same change of distance brings about a smaller or even negligible change of covergence. The same argument holds for accommodation. Thus, both of these cues will be most sensitive for signalling differences in distance for nearby objects, and less and less sensitive at greater and greater distances. As a rule of thumb, it is often said that accommodation and convergence are useful distance cues only over the distance range from one's nose to about 2 meters, and not useful beyond that distance. To gain information about greater distances, other cues will be needed.

24.2.2 Pictorial (painters') cues

The second category of cues will be called pictorial cues (often called *painters' cues* or *monocular cues*). Pictorial cues are the static cues available within each retinal image. In contrast to accommodation and convergence, it turns out that the pictorial cues depend heavily on the use of heuristics.

Two very strong pictorial cues, *interposition* (or *occlusion*) and *height in the visual field*, are

Figure 24.2: Two pictorial cues: Interposition, and height in the visual field. A. Interposition. Which object appears closest, and which farthest away? Notice that you get an impression of the *order* in distance, but no sense of the *absolute distances* of the objects. B. Fooling the interposition cue (Rock, 1975, p. 84). The pattern on the left could be created from the physical arrangement on the right. In this case the interposition cue would lead to a false perception of distance. C. Height in the visual field. On the left, which square is farthest away? Which bird? Notice also, on the right, that the distance impressions become more certain when interposition and relative size cues are added. (A and C by DT; B from Rock, 1975, p. 84, Fig. 3-3b).

shown schematically in Figure 24.2. *Interposition* (Figure 24.2A) arises from the physical fact that if there are two or more objects in front of you on overlapping lines of sight, the nearest one will be interposed between you and the farther ones. In the retinal image, the shape of the nearer object will be complete (and typically regular), but the shapes of the farther objects will be incomplete (and typically irregular) because they are occluded by the nearer objects.

The problem, of course, comes with the converse – the heuristic. What kind of heuristic would one need to use, in order to assign distances correctly on the basis of interposition cues? The heuristic must include some kind of assumption about the likely shapes of objects – for example, that objects have simple shapes. The interpretational rule would be something like: if two parts of the image share a common contour, and one of the parts is simpler in shape than the other, the object that creates the simpler retinal shape is nearer than the one that creates the irregular retinal shape, and the nearer object is occluding part of the farther object. Notice that theis heuristic forces your visual system to have a concept of ”simple shape”, which would need further definition. But even more importantly, if the nearer object has a peculiar shape, and there is a coincidental alignment between the contours of the two objects, this heuristic will give rise to a false perception of relative distances, as shown in Figure 24.2B.

Another strong pictorial cue is that of *height in the visual field.* This cue is shown in Figure 24.2C. This cue arises from the geometrical fact that if a set of objects is resting on the ground, the images of nearer objects intersect the ground *lower* in the retinal image than do the images of farther objects. Similarly, for a set of objects floating at equal altitudes above the ground, the image of nearer objects will be *higher* in the retinal image than the images of farther objects. Combining these two rules, we can say that far away objects will be closer to the horizon in the retinal image, whereas nearby objects will be either farther below or farther above the horizon in the retinal image.

The heuristic involved in the use of height in the visual field as a distance cue would be: Order objects in distance according to their closeness to the horizon in the retinal image. This heuristic could, however, give false perceptions for objects that are near the ground but at unequal heights above it, or for objects floating at different altitudes. (For example, the nearer of two birds can be imaged lower in the retinal image than the farther one if it is flying at a lower altitude. And think of a large group of balloons floating at different heights and different distances – the height-in-the-visual-field cue would be totally useless in this case.)

The next set of pictorial cues can be called cues of *perspective*, or *relative size.* Perspective cues are shown in Figure 24.3A. Suppose you are viewing a row of trees extending into distance. If the trees are all about the same physical size, their retinal image sizes will be smaller and smaller the farther away they are. This variation in retinal image size of a set of similar objects is a pictorial distance cue. To make use this cue perceptually, our perceptual systems could use a heuristic something like: if the retinal image contains a set of objects of similar shapes but diminishing sizes, represent them perceptually as a set of objects of a constant size but increasing in distance. (You also saw relative size cues at work in Figure 24.2C.)

A second, similar example of perspective cues is the use of *texture density.* Consider a field of stones, or a lakefull of waves, or a lawnfull of grass blades. At any given distance the element sizes vary – the stones come in a range of sizes. But in many cases the *average* size of the elements is about the same over variations in distance. The distance cue arises because if the world contains repeated elements of a constant average size, the average size of the texture elements in the retinal image gets systematically smaller with increasing distance. The required heuristic is: if the retinal

Figure 24.3: Perspective (relative size) and atmospheric perspective. A. Perspective (relative size) cues. The trees of decreasing size, the stones of decreasing average size, and the convergence of the two sides of the road, are all perspective cues. (Interposition and height in the visual field are also at work in this picture.) B. Atmospheric perspective. (A by DT, to be replaced with famous art. B. To come.)

image contains a set of similar elements of diminishing average retinal image size, represent them perceptually as a set of objects of a constant average size but varying in distance.

A third example of perspective comes about in the case of *parallel lines*, or more generally sets of contours, separated by a constant distance but varying in overall distance from the observer. Examples might be two sides of a street, a set of railroad tracks, or two sides of a winding country road. In cases like these, the retinal image will contain two similar contours, whose separation diminishes regularly over space. (The perceptual heuristic is left for you to figure out.)

A final pictorial cue is *atmospheric* (or *aereal*) *perspective.* This cue applies to the perception of large distances; for example, the relative distances of two or more mountain ranges. Since the atmosphere scatters light, the farther away a mountain range is, the bluer and lower in contrast it will be in the retinal image, and the fuzzier its contours – that is, the more the high spatial frequencies in its contours will be lost. Thus, color, contrast, and contour sharpness can function as distance cues. An example of atmospheric prspective is shown in Figure 24.2B. (Again, we leave the corresponding perceptual heuristic for you to figure out). This completes our list of the most common pictorial cues.

Absolute or relative distance?

Unlike accommodation and covergence, the various pictorial cues do not give information about absolute distances, but rather about distance *relationships.* For example, interposition is a *qualitative cue to relative* distance. It tells you which object is nearer, but gives you no information about *how much* nearer (think of the moon eclipsing the sun.). Perspective cues, on the other hand, are

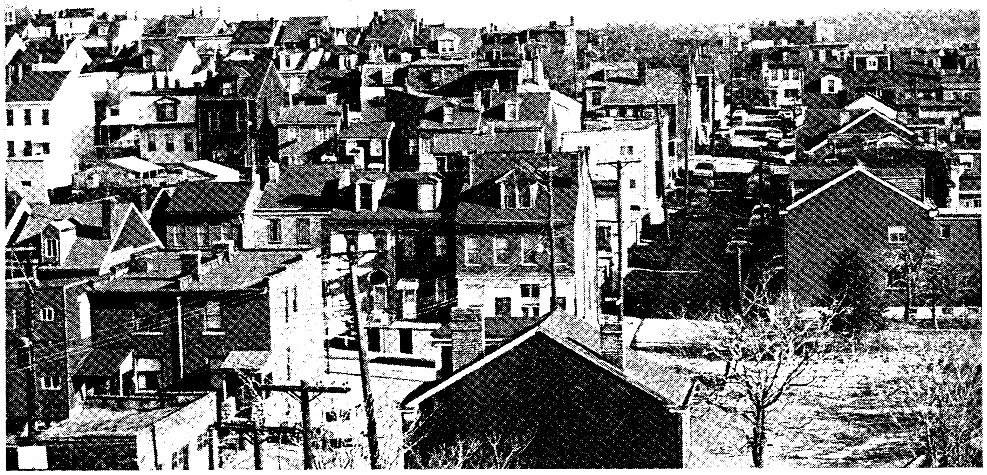


Figure 24.4: The painter's art – fool the eye. [Here will be a famous painting or picture that incorporates many pictorial cues and give a strong impression of distance.] Current version from Goldstein, 1999, p. 216.)

quantitative cues to relative distance. If the image of one tree is twice the size of another in the retinal image, the heuristics dictate that the second is twice as far away; if the distance between two assumedly parallel lines diminishes to 1/2, the distance has doubled, and so on.

The painter's art: fool the eye

The pictorial cues are also sometimes called *painters' cues*, because these are the cues that are available to a painter who wishes to capture the three-dimensional nature of the world on a two-dimensional canvas. Of course no matter what the painter does with his paint, the canvas will remain physically two-dimensional; that is, if he succeeds in creating a picture that we perceive as three-dimensional, he has fooled our eyes (as we have tried to do in Figures 24.2 and 24.3). A picture that uses pictorial cues to yield a strong impression of distance variations is shown in Figure 24.4.

The painter's trick is to tap into the complex set of pictorial cues and interpretational heuristics that we have just reviewed. If he paints two converging lines on his canvas, we may see two parallel lines converging into depth; a row of trees of decreasing size can become a row of trees receeding into depth; a regularly shaped object that has a common contour with an irregularly shaped one may be perceived as occluding it; and an object that intersects the earth higher in the visual field may be seen as farther away.

Thus, the painter's art – the art of creating illusory distance – provides a particularly striking example of the use of heuristics. The pictorial cues to distance are of the logical form: X when occurs in the physical world, Y occurs in the retinal image. The relevant heuristics – Y is in the retinal image, so X must be in the physical world – are converse statements, and not always true.

The eye can be fooled by creating a conventional distance cue on the flat canvas instead of by actually varying distance. It is the same cue, but created by subterfuge. Insofar as it fools the perceptual system into applying a heuristic that is usually worthwhile but false in this instance, it emphasizes the statistical and therefore risky nature of heuristics in perception. But notice also that it is the observer's false application of heuristics that allows painters and photographers to represent three-dimensional space so realistically on two-dimensional surfaces, and delight us with the products.

24.2.3 Motion (cinematographers') cues

A third class of distance cues arises from the *motion* of either the physical object or the observer with respect to the other. A particular form of physical motion will yield a particular pattern of temporal change in the retinal image. These temporal patterns of change are often called *optic flow patterns or motion gradients.* If the perceiver has and uses the appropriate heuristic, she can use the optic flow pattern to set up a (usually) accurate representation of the three-dimensional structure of the world.

One of the simplest examples of an optic flow pattern, a *looming pattern*, arises as an object of a fixed size approaches the observer. In this case, the retinal image of that object will expand in a particular pattern over time, against the fixed images of the surrounding stationary objects. If the object's trajectory will miss the observer's head the expansion pattern in the retinal image will be asymmetrical. But if the object is aimed right at the observer's nose the expansion pattern will be symmetrical, and the observer had better duck! It seems likely that the EDC endowed animals (especially prey) with rapidly initiated heuristics that initiate an immediate escape response to looming patterns.[1]

Optic flow patterns can also arise from motion of the observer. Two of the simplest optic flow patterns are diagrammed in Figure 24.5. The first, *optical expansion*, is similar to looming, but involves an expansion of the whole field of view (Figure 24.5A). As you move forward, the images of the objects in front of you will expand and move toward your visual periphery in the retinal image. The image of the object you are fixating will remain centered on the fovea, and the images of the rest of the objects will expand in a complex pattern that depends on the distances and sizes of each of the objects, and the speed and direction of your motion. (Optical expansion patterns are easiest to demonstrate when you are moving forward rapidly, and looking forward, for example when driving a car.)

Similarly, as you move sideways through the world (e.g. as you look out the side window of a car or train), the images of objects at different distances will move across the retina at different speeds. This motion pattern is often called *motion parallax* (Figure 24.5B). In general, images of nearby objects will move rapidly, whereas far-away objects will move more slowly. The image motion for each object will also depend on the point of fixation – the fixated object, of course, will remain fixed on the fovea, and the images of all of the non-fixated objects will move along particular retinal trajectories. Objects nearer than the fixation point will move in the direction opposite to

[1] When DT was in her first year of graduate school, she and her advisor, Tom Cornsweet, set up an optical system. They needed to vary the sizes of disks of light, so they built an iris diaphragm into the optics. One day Tom was being the subject – he had his teeth on a bite bar and was staring into the light beam that emanated from the optics. Without thinking, DT moved on to the next condition of the experiment, and pushed the lever that expanded the iris, creating a looming pattern on Tom's retina. Tom's startle response was so extreme he almost tore his teeth out getting off the bite bar....but he did survive to serve as a subject again the next day.

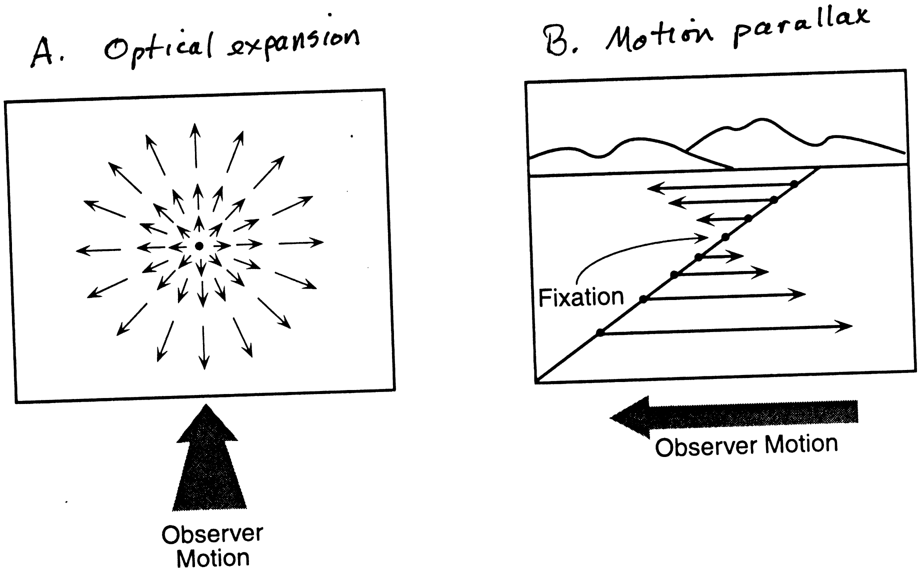


Figure 24.5: Optic flow patterns arising from motion of the observer. A. An optical expansion arising from motion forward with the eyes fixated straight ahead. B. A motion parallax pattern, arising from motion to the side with the eyes fixated at the fixation point shown. These patterns vary in complex ways as the observer's fixation point changes. (Adapted from Palmer, 1999, p. 227, Fig. 5.4.2)

the direction of the observer, whereas objects farther away than the fixation point will move in the same direction as the observer. [You can easily demonstrate motion parallax by closing one eye and moving your head back and forth from left to right. Vary your fixation distance, and notice the direction of motion of objects closer and farther than your point of fixation.]

Again, as with the painters' cues, motion cues must come paired with heuristics. Our visual systems presume that particular patterns of motion in the retinal image are caused by whatever physical motion most typically causes them. [Again, these probable heuristics for motion cues to distance are left for you to work out.]

And of course, just as the pictorial cues can be called painters' cues, the motion cues could be called movie-makers' cues: if the cinematographer creates the usual motion cues to distance on a flat screen, our perceptual system can be fooled into perceiving three-dimensional motion. A particularly interesting example comes from the cinematographer's trick of moving the camera around the scene, in order to simulate a moving observer and thus create motion-based distance cues.

24.2.4 Binocular disparity: The most famous distance cue

Finally, we come to the most well-known distance cue: *binocular disparity.* The binocular disparity cue arises from the fact that you have two eyes, located in two different positions in your head. Your two eyes view the world from slightly different vantage points, with the result that there is a slight difference – a *disparity* – in the configuration of the retinal images of objects in the left vs. right eye. The perception of distance (or depth) resulting from using the binocular disparity cue is called *stereopsis*, or less formally *stereo vision.*

The geometrical principles involved in binocular disparity are shown in Figure 24.6. In Figure 24.6A, suppose there are two objects (perhaps the masts of two sailboats) in front of you. The first one, F, is straight ahead of you at a given distance, and a second one, E, is farther away and slightly to the right. Suppose you fixate F. As shown in the figure, the two retinal images will be slightly different – geometrically, the retinal images of F and E will be closer together in the left eye and farther apart in the right eye.

[Binocular disparity can be demonstrated readily using your two index fingers. Hold the two fingers out in front of you, with the left index finger at (say) 20 cm straight in front of you and the right index finger at (say) 30 cm and displaced slightly to the right. Now fixate your left index finger, and alternately close each eye. With your left eye open, the two fingers should look closer together, while with the right eye open, they should look farther apart.]

Two more examples, in which the two masts are at different distances along the same line of sight, are shown in Figure 24.6B and C. These cases can also be demonstrated with your fingers. Hold the two fingers out in front of you, one directly behind the other, and fixate the *near* finger. With your *left* eye open, the far finger should appear to the *left* of the near finger; with your *right* eye open, the far finger should appear to the *right* of the near finger. This pattern has been named *uncrossed disparity*, and it provides a cue that the fixated finger is the closer finger. Now fixate the *far* finger. The near finger seen with the *left* eye will lie to the *right* of the far finger, whereas the near finger seen with the *right* eye will be seen to the *left* of the far finger. This pattern has been named *crossed disparity*, and it provides a cue that the the fixated finger is the farther finger. These names arise because of the *perceptual* locations of the near and far objects as seen in

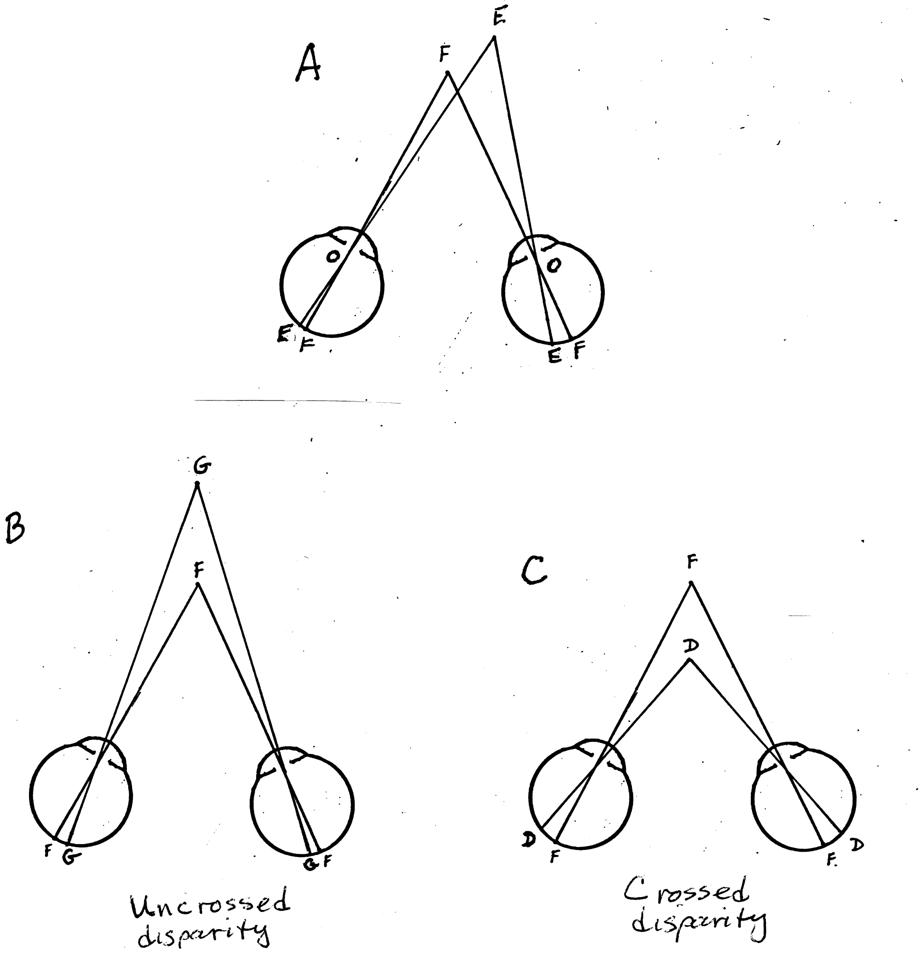


Figure 24.6: The geometry of binocular disparity. The two points in front of the observer represent two vertical lines or objects, such as two sailboat masts. Since the two eyes see the scene from slightly different vantage points, the two retinal images will differ slightly. In each case, the observer fixates at F. A. Two masts in different directions at different distances. The distance EF will be smaller in the left eye than in the right eye. In technical terms, the difference in the two angles subtended by the images in the two eyes, in degrees of visual angle, is the amount of retinal disparity. For example, if the angle EOF in the left eye were 2^o, and in the right eye 1^o, then the disparity would be 1^o. B. Fixating the *nearer* of two masts in the same direction. The two images of G will fall inside the two images of F (toward the nose in each eye). By convention this configuration is called *uncrossed disparity.* C. Fixating the *further* of two masts in the same direction. The two images of D will fall outside the two images of F (away from the nose in each eye). By convention this configuration is called *crossed disparity.* Thus, crossed disparity is a cue that the non-fixated object is nearer than the fixation point; uncrossed disparity is a cue that the non-fixated object is further away than the fixation point.

demonstrations such as this one.[2]

In sum, binocular disparity is a distance cue. The magnitude and direction of the disparity are determined by the relative locations of objects. A complex scene with many objects at different distances yields a complex pair of images with many different disparities, determined by the locations and distances of the objects.

Now, what perceptual heuristics will allow us to make *use* of disparity as a distance cue? The basic heuristic must be that, relative to fixation, a disparity between the locations of figural elements in the two retinal images should be interpreted as a difference in their *distances*. Crossed disparities are interpreted as revealing objects nearer than the fixation point, uncrossed disparities as objects farther than the fixation point, and other specific quantitative variations in disparity as specific quantitative variations in distance.

As an historical note – As you have just seen, the doubled and disparate images from the left and right eyes are easy to become aware of, and the geometry of binocular disparity is not difficult to work out. Yet surprisingly, early authors including both Da Vinci and Berkeley missed this distance cue. The first scientific report of an analysis of binocular disparity is that of Wheatstone in 1838. Wheatstone built a stereoscope – a device for presenting two different pictures separately to the left and right eyes – and showed that artificially produced disparity gives rise to the perception of differences in distance. Thus stereoscopes, like paintings and movies, provide another example of how our visual systems can be fooled when the artist or scientist mimics a distance cue artificially, and the visual system applies the usual heuristic and interprets the cue as having arisen from differences in distance. Examples of several different kinds of stereoscopes are shown in Figure 24.7.

Some of Wheatstone's original stereograms are shown in Figure 24.8. The figure legend provides instructions for free viewing of stereo images to yield the perception of variations in distance and depth.

24.2.5 A tangent: Random dot stereograms

In early thinking about stereopsis, a particular order of sequential processing was often implicitly assumed: That analysis of the patterns in the two eyes must be performed before the two patterns could be "matched up" by the visual system to yield a stereo cue. But the scientist Bela Julesz (1971) created a new custom-designed stimulus called a *random dot stereogram*. Examples of random dot stereograms are shown in Figures 24.9 and 24.10 .

To create the stereogram, Julesz first divided a region of visual space up into very small checks, and colored each check, or *dot*, randomly either black or white. He then duplicated the image. In a central region of the second image, he "picked up" a set of dots from, say, a T-shaped region near the center, moved this whole set of dots, say two dots worth to the left, creating a disparity between the two T-shaped areas. Finally, he filled in the space created at the left edge of the set with new random dots.

[2] In Figure 24.6B and C, of course, the rays "cross" within the eye in both cases, and the nomenclature is confusing. The names arise because in perception, when you fixate the near finger and attend to the far finger, the far finger is perceived as being to the left of the near finger in the left eye, and to the right of the near finger in the right eye, hence "uncrossed". If you fixate the far finger and attend to the near finger, the near finger is to the right in the left eye and to the left in the right eye, hence "crossed". If you remember that the retinal image is left-right reversed with respect to your perception, you can reconcile the "fingers" demonstration with Figure 24.6B and C.

Figure 24.7: Stereoscopes, old and new. A: a mirror stereoscope like that used by Wheatstone. The mirrors rotate to allow easy binocular alignment. B: a parlor stereoscope, popular in the Victorian era. Alignment was achieved with a prism and distance variations. C and D: Old fashioned and modern Viewmasters [trademark xx]. Alignment is achieved by placing the two differing images directly in front of the two eyes.

Notice that in Figure 24.9A the central T – the figure created by the disparity – cannot be seen in either of the two images alone. Yet in fact, when these pictures are viewed by the two eyes separately, the observer can readily see the T, standing out in depth! Julesz thus made the important point that the analysis of the figure need not always precede the application of the stereo processing algorithm in our perceptual machinery – the stereo algorithm can itself analyse the incoming patterns and find the figure. Similarly, these demonstrations show that binocular disparity by itself, without any other cues, is sufficient to yield the perception of variations in distance. We will return to stereopsis and its limitations in our discussion of depth cues in the next chapter.

24.3 How do distance cues combine?

We started this chapter with the intuition that information about distance must be lost in the projection from the three-dimensional physical world to the two dimensional retinal image. Instead, we have found that the two-dimensional retinal images, and the motor responses of the eye to these images, provide a remarkably large number of partial sources of distance information – distance cues. None of these cues is perfect – each one depends on the use of heuristics, with the consequence that each one will fail under conditions in which its individual heuristic fails. But the EDC is opportunistic. Somehow the different cues, working together, manage to cover for each others' weaknesses, and allow us to reconstruct a sufficient approximation of the third dimension to function in the physical world.

This idea brings us to the question of *cue integration.* How do the different distance cues work

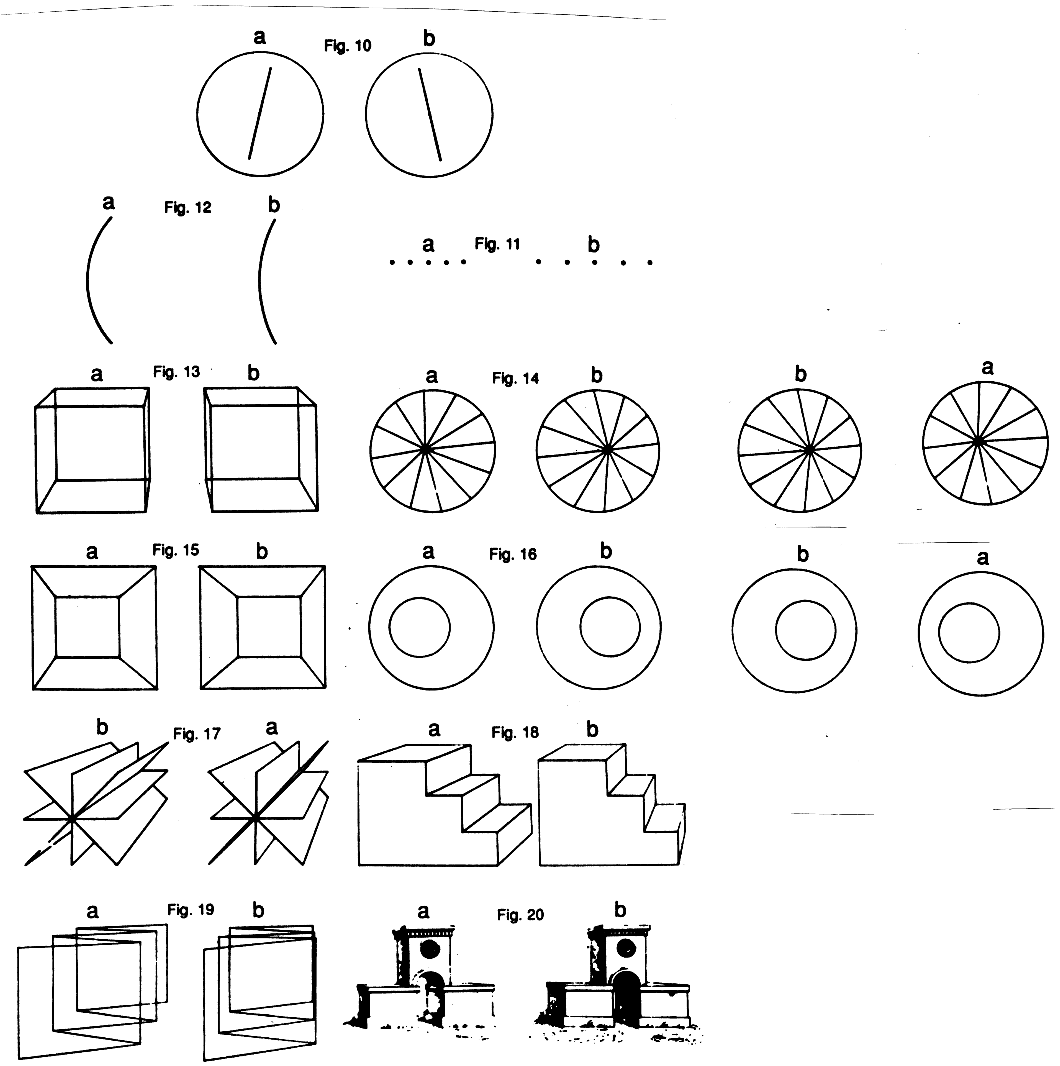


Figure 24.8: Examples of Wheatstone's original stereograms. The four pictures on the right have been made by reversing the left and right eye views of Wheatstone's figures 14 and 16, to change the direction of the disparity. With a little effort, most people can learn to "free fuse" these stereograms. Chose one of the figures (#16 works well). Hold the figure at normal reading distance, but vertically in front of your eyes. Now hold one finger about 2/3 of the way from your face to the page, with its tip aligned with the bottom of the stereogram. Look at the stereogram, and you should see two images of your finger. Move your finger back and forth in depth until one image of the finger is centered under the left half of the stereogram, the other under the right half. Now converge on your finger, but shift your attention back to the stereogram. After a while the two images should merge perceptually, and you should see the smaller circle standing out in front of the larger one. Practice on Wheatstone's other figures, as there are more complicated stereograms to come. (Modified from Gillam, 1995, p. 43, Fig. 7.)

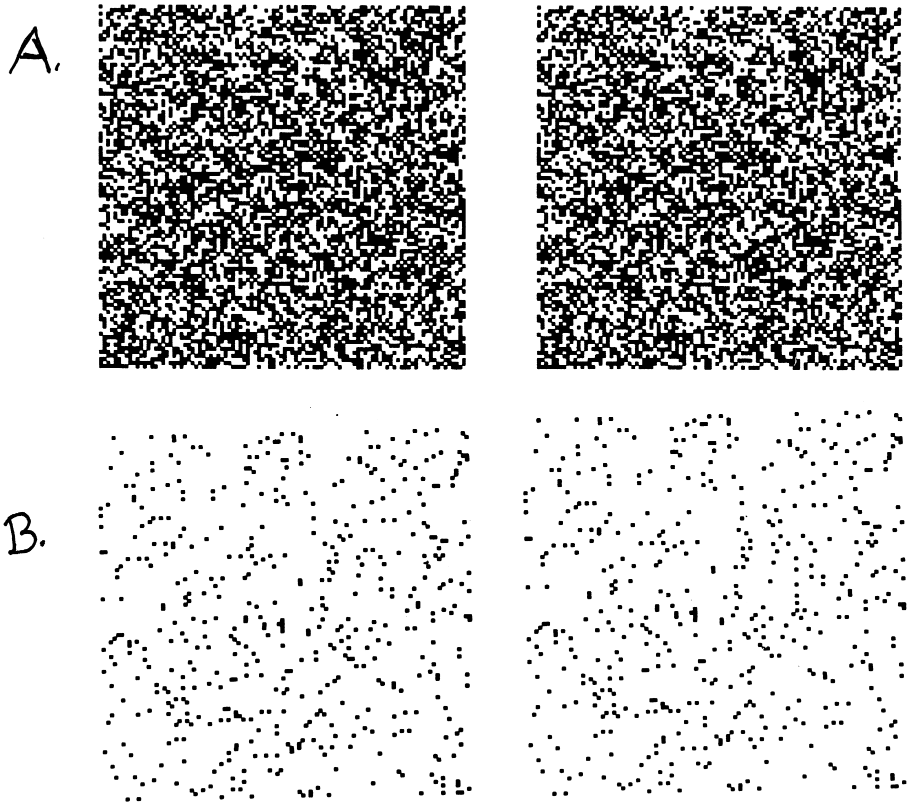


Figure 24.9: Random dot stereograms (RDS). A: a dense RDS, showing a T shaped figure in front of its surround. B: A sparse RDS, showing a square in front of its surround. (From Julesz (1971). A: page 22, Fig 2.4-4. B: page 122, Fig. 4.5-5.)

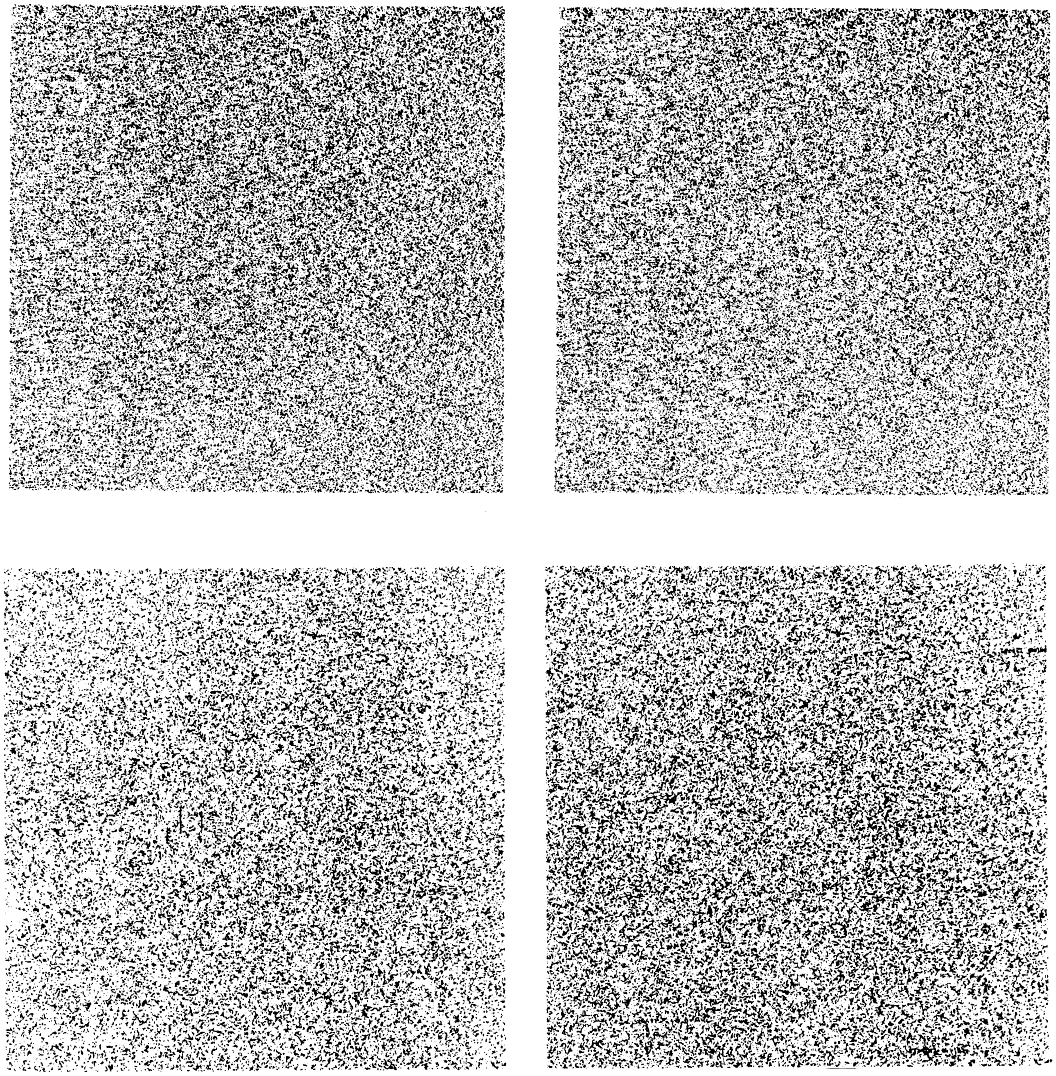


Figure 24.10: Random dot stereograms (RDS). Two more challenging random dot stereograms protraying smooth three-dimensional surfaces. Try to fuse them. It may take several minutes for your stereo system to analyse the two images and reveal the embedded surfaces. (From Julesz (1971); Top, page 122, Fig. 4.5-4; bottom, p. 121, Fig. 4.5.3.)

together? How do they combine to give us a single, unambiguous perception of the distance to an object in the visual field? How is each cue weighted in combining its distance estimate with that from other cues? Are some cues more important than others?

These questions are vital to pose, but they have no simple answers. Part of the problem concerns paradigms. For example, many early perception researchers approached the question by an *isolation of cues paradigm*: try to render all but one of the distance cues ineffective, and test distance perception with that single cue. The problem with cue isolation experiments is that cues set to zero are not neutralized – instead, they are probably signalling that the stimulus consists of a flat plane, so one tends to find that the isolated cue (unavoidably opposed by all the ruled out cues) is not very effective.

Alernatively, we could try an *opposition of cues paradigm*: use stimuli containing one cue that indicates one distance order and another cue that indicates the opposite distance order, and see which one "wins out". A good example would be to take a complex pictorial stereogram, and reverse the pictures presented to the two eyes. This maneuver will reverse the stereo cue while maintaining the pictorial cues in their original state. Which will dominate? Most people tend to assume that the stereo cue will dominate. Remarkably, most studies report that the pictorial cues dominate in this situation.

Many theorists working in the field of distance perception, however, would argue that neither of these paradigms has much value. The argument is that in perception in real-world situations, most of the distance cues vary together, and most of the time all of the distance cues signal much the same distance. That is, tests of the *cooperation* of cues may be a more intelligent paradigm than opposition or isolation of cues.

In short, many studies have been carried out on the question of distance cues and their combinations. In general, the more the distance cues the more accurate the perception of distance. However, there is no simple summary of the outcomes of these studies, and no simple combination rule has emerged. We will return to the question of cue integration, and take on a more searching analysis of it, when we discuss the topic of depth (three-dimensional form) in the next chapter.

24.3.1 Cutting and Vishton's analysis

Instead, we turn to an interesting summary and synthesis of information provided recently by James Cutting and Peter Vishton (1995). Cutting and Vishton combined logical (geometrical) analyses of the various distance cues with the results of empirical studies, and came up with some interesting educated guesses about the values and roles of different distance cues.

In our review of distance cues, we emphasized that different distance cues have different properties. Some cues – particularly accommodation and convergence – provide information about absolute distances. Others, like height in the visual field and the various perspective cues, provide information about relative distances. And yet others, particularly interposition, provide us only with ordinal information – they tell us about the order of objects in depth, but not the distances between them.

In addition to the absolute vs. relative vs. ordinal distinction, Cutting and Vishton emphasize that different cues have different *levels of sensitivity*, and different *distance ranges* over which they yield sufficient information to be useful in the calculation of distance. Their estimates of the sensitivities and useful distance ranges of the various cues are shown in Figure 24.11. (Unfortunately we don't have space to justify these estimates in detail, but you can derive at least some of them

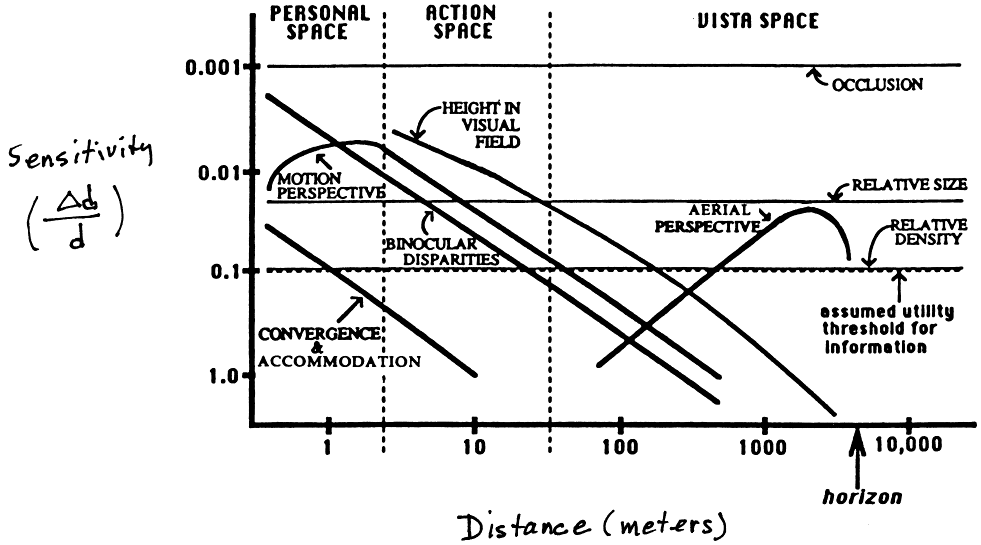


Figure 24.11: Cutting and Vishton's (1995) analysis. Summary of the sensitivities of the different distance cues, and how these sensitivities vary with the distance of an object. The horizontal axis in this graph is absolute distance – how far the object is from the observer. The vertical axis is sensitivity to changes in distance – Δd, the smallest detectable difference in distance, normalized to the absolute distance (d). By this measure, a difference threshold of 0.1 meter (10 cm) at 10 meters, and 1 meter at 100 meters, would both be plotted as sensitivities of 0.01. Cutting and Vishton also introduce the assumption, shown by the dashed horizontal line, that below a sensitivity of 0.1 a cue no longer provides useful information about distance. (Modified from Cutting and Vishton (1995), p. 80, Fig. 1.)

from the principles of geometry).

On the basis of their summary, Cutting and Vishton point out that there are basically three patterns of variation of sensitivity with distance. In the first category are cues that are most sensitive at short distances, and diminish in sensitivity with increasing distance. As shown earlier in Figure 24.1B, convergence provides an example of this pattern; so does accommodation. In fact the ocular cues are never highly sensitive, and they become ineffective at a distance of perhaps two meters. Motion perspective and binocular disparity also show this pattern. They are both highly sensitive distance cues at near distances. But like accommodation and convergence, the sensitivities of motion perspective and binocular disparity also fall with distance, to become ineffective beyond about 10 to 30 meters. Height in the visual field is also in this category.

In the second category are distance cues that are equally effective at all distances. These cues are represented by horizontal lines in Cutting and Vishton's graph. The cue of interposition is the best example. As we argued earlier, interposition is a unique distance cue in that it occurs any time one object lies in front of another. Hence it is a highly sensitive cue at all distance, as it can signal a miniscule difference in distance at any distance. (The downside is that it gives only ordinal information). Similarly, the cues of relative size and relative density (which we have called perspective cues) are also shown as horizontal lines. Cutting and Vishton have placed them at lower sensitivity levels than occlusion to indicate that perspective cues are not as sensitive to small differences in distance as occlusion is.

And in the third category, there is one cue whose efficacy actually increases with distance – the cue of aereal perspective.

Cutting and Vishton's graph has the advantage that it summarize and represents information about many distance cues in a form that makes them commensurate, and should be helpful in leading to more sophisticated theorizing in the area of distance perception and cue integration. In their terms, the assembled information provides a powerful network of constraints on models of distance perception.

At the end of their analysis, Cutting and Vishton introduce the interesting argument that different ranges of distance are of different functional importance to the perceiving organism. As shown at the top of Figure 24.11, they divide distance into three ranges. The first is *personal space* (out to perhaps 2 meters) – the space that immediately surrounds each person. The second is *action space*, the space within which we move and within which we direct our actions toward objects. And the third is *vista space*, the space at larger distance, beyond the range at which evolving humans could expect their actions to have any immediate impact. Cutting and Vishton suggest that environmental demands are different in these different distance ranges, and that different constellations of cues have been selected to combine optimally to serve perception and action in these different distance ranges. They suggest that studies of cue integration might profitably take the different distance ranges into account. Perhaps if studies of distance cue integration were sorted out by distance range we would find simpler rules of combination.

DT is also curious as to whether one could make a similar analysis in the time domain. That is, there are some situations in which rapid response to a motion cue would be critical to survival – the expansion pattern caused by the shadow of a descending hawk again comes to mind. It seems possible that different distance cues might be weighted differently depending on the immediacy of the task the observer is asked to perform. Taking off from Cutting and Vishton and from Milner and Goodale, perhaps different distance cues are used in different weights for action than for perception.

Finally, we note that, despite all of the available distance cues, human distance perception is not always as good as modern man might wish, and we provide ourselves with many crutches to augment our perception of distance. For example, cameras are equipped with range finders precisely because our perception of distance is not accurate enough to guide optimal focusing of our cameras. And golfers have distance scopes that use a graded scale to judge the visual angle of the flag at the next hole, and in that way judge its distance.

24.4 The perception of size

24.4.1 The size/distance invariance hypothesis

We turn now to the perception of size. Under ordinary viewing conditions, our perception of the sizes of objects is usually quite veridical: as stated earlier, we have a good degree of *size constancy.* The question is, how can this be? Since the size of the retinal image of an object varies inversely with its distance, the retinal image cannot supply us with information about the object's physical size. The question of how we know about object size has been a classic problem in visual perception.

The usual answer has been to suppose that the relatively veridical perception of size comes about because of the presence of distance cues and the ensuing relatively veridical perception of distance. That is, in physical terms, the size of an object can be calculated if its retinal image size and its distance are known:

object size = retinal image size x distance

A modified version of this equation, using perceived rather than physical variables, has commonly also been assumed to hold:

perceived size = retinal image size x perceived distance

This equation is often called the *size-distance invariance hypothesis.* An implicit assumption in the size-distance invariance hypothesis is that the visual system calculates perceived size by first calculating a unified estimate of perceived distance, and then using the estimate of perceived distance in the calculation of perceived size.

24.4.2 Holway and Boring's experiment

Within the context of the size/distance invariance hypothesis, many classical studies of size constancy take the approach of asking what distance cues are necessary for the veridical perception of size. A particularly nice experiment on distance cues and size constancy was carried out by A. H. Holway and Edwin Boring (1941). The setup they used is shown in Figure 24.12.

Holway and Boring's subjects were seated at the intersection of two corridors, and asked to compare the sizes of disks of light that appeared at various distances down the two corridors. Down the right corridoor at a series of different distances, from 10 to 120 feet, were a series of different test stimuli, S. The sizes of the test stimuli, Ss, varied with their distances, Ds, so that each one subtended the same visual angle of 1^o, and hence each one produced the same retinal image size. Ten feet down the left corridoor was a comparison disk C, whose size, Sc, could be adjusted by the subject. On each trial, the experimenter presented one of the test disks. The subject's task was to vary the size of the comparison disk until he judged that test and comparison disks were matched in physical size.

In the first condition of the experiment, Holway and Boring allowed the subjects to view the disks with natural binocular viewing with the lights on. In subsequent conditions different distance cues were sequentially eliminated. In the second condition, subjects viewed the disks monocularly, eliminating both retinal disparity and convergence cues. In the third condition, the accommodation cue was removed by having subjects view the disks through a small artificial pupil. And in the fourth condition, a "reduction tunnel" – baffles – were used to eliminate reflections from the walls (perspective cues) .

The results of the experiment are shown in Figure 24.12B. With all distance cues available, the subjects matched the physical sizes of the disks very closely at all test distances – they showed good size constancy. Interestingly, constancy was preserved under monocular viewing, suggesting that neither binocular disparity nor convergence is necessary for size constancy. However, when accommodation was also eliminated, size constancy broke down rather dramatically, and the subjects' matches began to approach matches based on retinal image size. And in the final condition, in which reflections from the walls were eliminated, the subjects' matches conformed to matches of retinal image size. This experiment thus shows very clearly the dependence of size constancy on

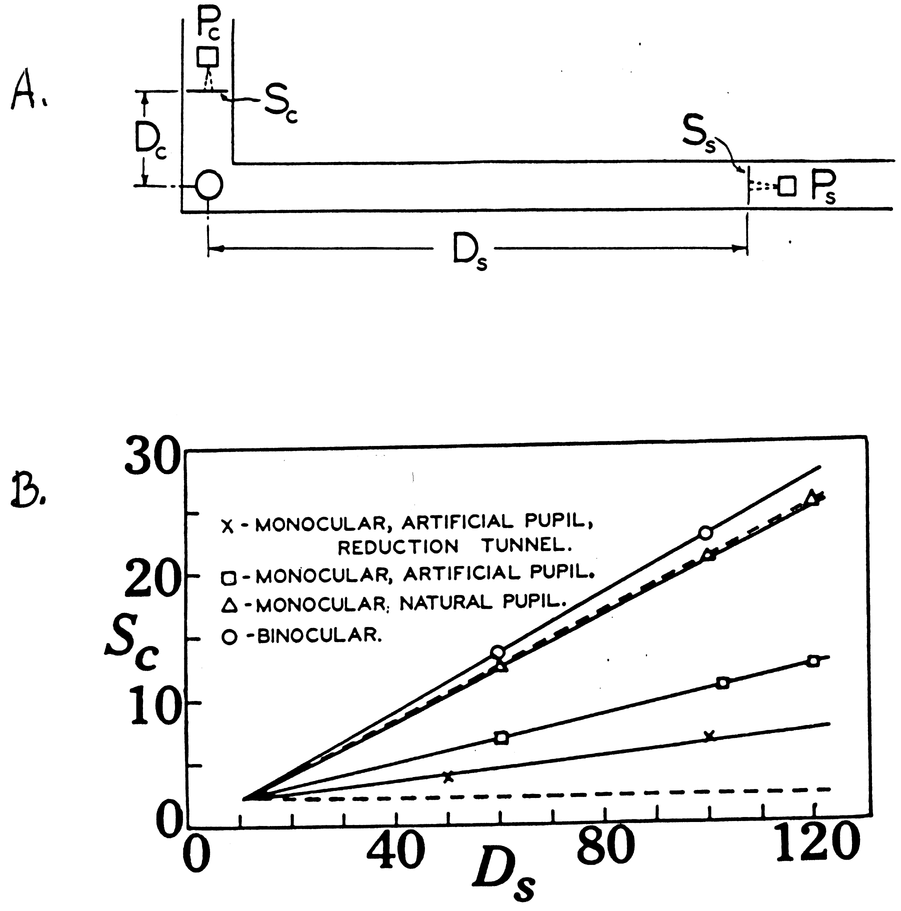


Figure 24.12: Holway and Boring's experiment. A: The experimental set-up. B: The data. The diagonal dashed line shows the predicted size matches if the subject has perfect size constancy. The horizontal dashed line shows the prediction if the subject matches the visual angles subtended by the two stimuli. (From Holway and Boring (1941). A from page 24, Fig. 2. B from page.)

one particular combination of classical distance cues.

There are three more points to be made about size constancy. First, we need to be careful with our conclusions. It might be tempting to conclude that accommodation and perspective cues are the distance cues most important to size constancy. Instead, the more valid conclusion is that these cues are *sufficient* to allow size constancy. In other words, had Holway and Boring done the experiment in the opposite direction, eliminating accommodation and perspective cues first, they probably would have found that the combination of disparity and convergence was also sufficient to yield good size constancy. In all probability size perception is opportunistic, using whatever distance cues are available in any given situation.

The second point is a subtle one about causality. It is tempting to conclude that since size constancy depends upon the presence of distance cues, it must be dependent upon a neural calculation of distance per se. This theoretical perspective is implicit in the equations at the beginning of this section. But in fact, Holway and Boring's experiment has only shown that the *same* cues that contribute to the perception of distance also contribute to the perception of size. Whether distance is computed first and then enters into the calculation of size, or whether size and distance are calculated independently from the same base of distance cue information, is a difficult problem to sort out. DT suggests an open mind on this issue.

The third point concerns the cues we use to perceive the sizes of very distant objects, for which the only available distance cues are interposition, perspective and atmospheric perspective. Most people report that size constancy is non-existent at large distances: that everything looks tiny. Many authors suggest that we can see distant objects in either of two ways: as very small (that is, size constancy breaks down), or as having their normal sizes (that is, size constancy still holds). In the latter case, we probably use the more cognitively based cue of *familiar size.* That is, to have size constancy, we only need to perceive a distant house to be about the size that houses usually are, a distant person to be the size that people usually are, and so on. Such a cue necesarily comes with a heuristic: Things are their normal sizes. [Think about it. When would size constancy governed by this heuristic break down?]

24.5 Summary: Cues and heuristics

In summary: we started with the puzzle of how we can know about the distances of objects. We have found that the visual system has available perhaps a dozen different distance cues, each with its own peculiar set of limitations, and many with complex and fallible heuristics required for their use. Yet we can unite the various distance cues to judge the distances and sizes of objects with reasonable accuracy.

At the meta-level, this topic has given us a chance to work with the concept of cues and heuristics, and more generally the necessity for believing in top-down processing. The long and short of distance perception is: there are cues contained within the incoming signal, but there is insufficient information to allow veridical distance estimates to be made. To yield up its secrets, each distance cue must be paired with the appropriate heuristic. Logically, these heuristics must be stored within our visual brains, and called upon when needed.

What a remarkable system! The entire scheme seems to DT so complex that it's hard to imagine it being instantiated in an ensemble of neurons. The number and variety of cues, and the necessary heuristics, are mind-boggling. Yet bumblebees can fly, and DT at least can only marvel at this remarkably jury-rigged but remarkably effective analysis of the visual scene. Another round of congratulations to the EDC for doing a job that at first glance semed to be impossible.

In the next chapter we turn to the question of depth perception. We then review the available evidence for the analysis of depth and distance cues within the visual cortex.

Chapter 25

Depth and the Physiology of Depth and Distance

In this chapter we continue our exploration of perception of the third dimension. We concentrate on two main topics. The first is the perception of depth –the three-dimensional shapes of objects. As was the case with distance, there are a variety of different cues to depth. We first explore four of these cues, along with their accompanying heuristics and their limitations. We then explore the way in which the visual system combines depth cues to provide a unified perception of the three-dimensional shapes of objects.

The second major topic concerns physiological studies of the tuning of single neurons in various parts of the visual cortex for variations in distance and depth. We review the neurophysiology of distance and depth together because there are few (if any) neurophysiological studies in which the distinction between distance and depth has been made.

Our review of distance and depth coding will bring us back to a question we have seen before: What form do we expect particular kinds of visual information to take within the brain? If we perceive the distances, sizes, and depths of objects in a single unified perception of a scene, what does this imply (or what do we implicitly assume) about neural representations? Do we expect to find a separate representation of each distance and depth cue, followed by a single representation in which all of the depth or distance cues are combined? If not, what are the options? We return to these themes later in the chapter.

25.1 Cues to depth

In the previous chapter, we made a distinction between distance and depth, reserving the term *depth* to refer to the *three-dimensional shape of an object*. Of course, it's likely that distance and depth perception will depend on overlapping sets of cues. However, there are some cues that provide information about distance but not about three-dimensional shape, and vice versa. Let's go back to the beginning, and discuss four cues to *depth*: shading, texture, motion, and disparity. In much of the current literature, three-dimensional shape is referred to simply as *shape*, and the use of the various cues is called *shape-from-X: shape-from-shading, shape-from-texture, shape-from-motion*, and *shape-from-disparity*. We will adopt this terminology. (There are other shape cues, including shape-from-contour and shape-from-highlights, but we will omit these for simplicity.)

25.1.1 Shape from shading

The first depth cue is *shading* – lighting and shadows. In the physical world, if there is only one source of illumination (say, the sun), then the surface of an object that faces toward the sun will have the highest illumination level, whereas the surface away from the sun will be in shadow, and there will be a pattern of shading from light to dark in between. The retinal illuminance in the image of the object will change regularly over the differently angled surfaces of the object, and the pattern of change will be different for objects of different three-dimensional shapes – think of a cylindrical silo vs. a square barn. These shading patterns are a potential cue to three dimensional shape. Examples of shading patterns that provide cues to depth are shown in Figure 25.1A.

There are three things to notice about the shading cue. First, you may notice that shading is a new cue in this section – it was not part of our list of distance cues. In fact, shading is a cue to depth, but not a cue to distance. That is, three cylinders of different sizes will yield three shading patterns that are scaled in size but otherwise identical. Nothing in the shading pattern provides information about the distance or the size of the cylinder. Shape from shading is a pure shape cue – the shading pattern provides information about shape and shape alone.

Second, shading is a pictorial cue; and as with all pictorial cues, we should suspect that heuristics will play a major role in the use of shape from shading cues. The basic assumption that we introduced above – that the surface reflectance *is* constant over the object, is of course a heuristic. Usually the surface reflectance is constant; but if the paint on the silo were shaded cleverly from white to grey, the shape-from-shading mechanism would be defeated. The heuristics involved in using shading cues are doubtless complex, but in general they would be something like: a specific gradual shading pattern should be perceived as a cylinder; a different gradual pattern as a sphere; a sharp change in illuminance as a corner; and so on.

25.1.2 Shape from texture compression

A second pictorial cue to three-dimensional shape is the cue of *texture compression*. We mentioned texture cues earlier – they are a major pictorial cue to distance. However, it turns out that there are several aspect to texture. The aspect that matters to distance perception is variations in the sizes (or average sizes) of the texture elements (look back at Figure 25.1xx). But a different aspect of texture – the relative compression of texture elements – turns out to provide an important depth cue. Examples of texture compression patterns that yield perceptions of depth are shown in Figure 25.1B.

Like shading, texture compression patterns are a cue to depth but not to size or distance. Notice that the portrayed surface is covered with a texture of circles and ovals. In the left-hand figure, the entire figure is covered with circles. In the center picture, the texture at the center is composed of circles, but as you approach the left and right edges, the circles give way to more and more compressed ovals. The compression is exaggerated in the rightmost picture. Most observers report that the lefthand diagram looks like a flat surface, the middle as a bowed surface, and the righthand surface as having a more exaggerated bow. [The general heuristic involved is easy to figure out.]

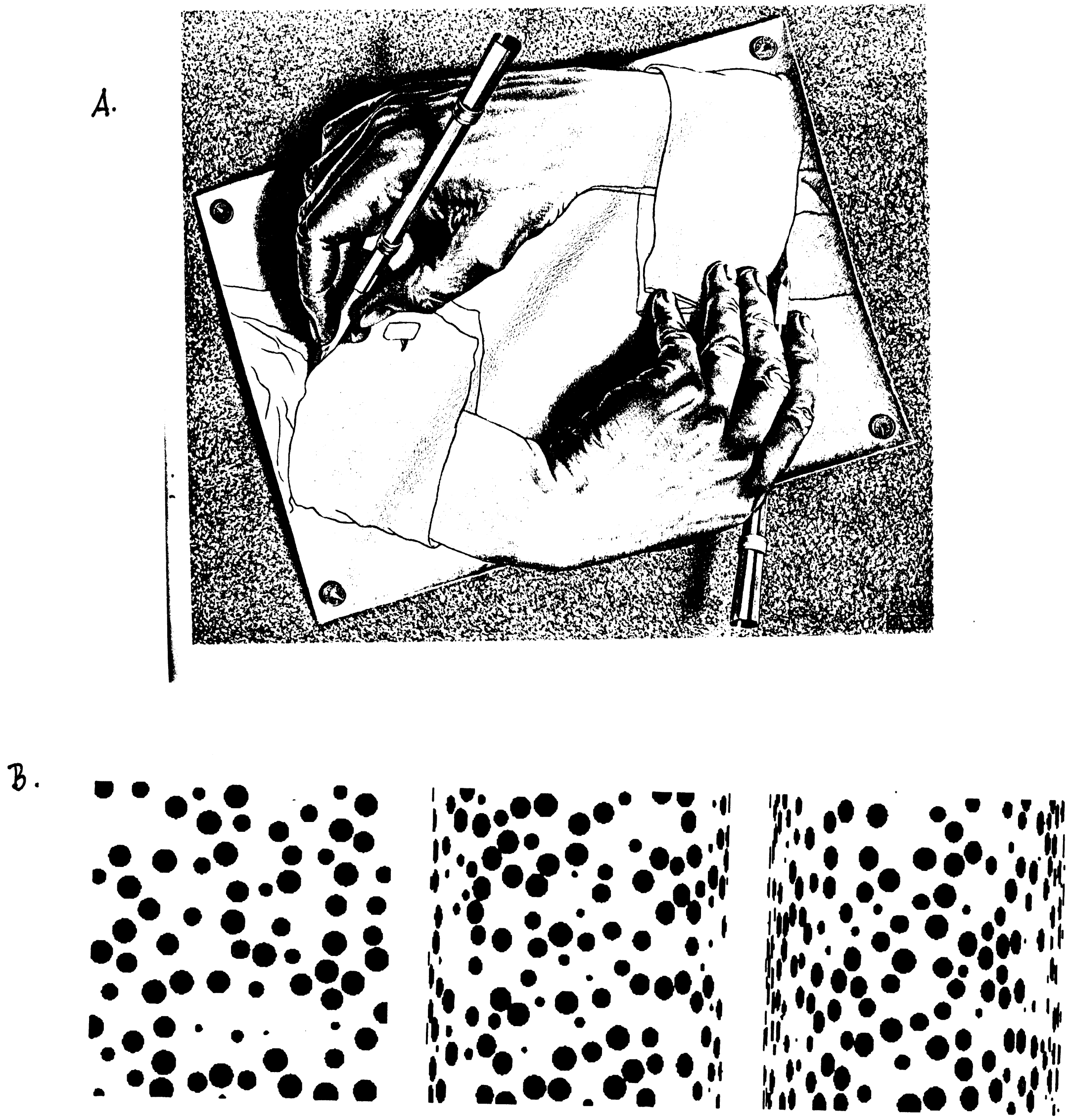


Figure 25.1: Pictorial cues to three-dimensional shape. A: Shape-from-shading. B: Shape-from-texture-compression. The pattern on the left has no texture compression. The middle pattern has texture compression appropriate to represent a cylinder. The pattern on the right is appropriate to a bowed cylinder, longer in depth than in width. Notice that the middle and right patterns give quite strong perceptual impressions of depth, even though other cues such as disparity must be signalling a flat plane in each case. (A. from Escher; B from Young, Landy and Maloney, 1993, P. 2687, Fig. 2).

25.1.3 Shape from motion: The kinetic depth effect (KDE)

As we discussed in Chapter 21xx, particular kinds of temporal sequences in the retinal image are cues to the perception of physical motion. A third depth cue arises from motion sequences generated by rotating objects. This cue is called *shape-from-motion*, and the resulting perception is called the *kinetic depth effect (KDE).* That is, a physical object of a particular three-dimensional shape, rotating in front of you, creates a particular temporal pattern of two-dimensional images on your retinas; and objects of different shapes will create different temporal sequences. These sequences are potential depth cues, and in fact they provide vivid impressions of objects of particular three-dimensional shapes, rotating in depth. The images in Figure 25.2A and B, presented sequentially, yield strong perceptions of objects rotating in depth.

You can demonstrate the kinetic depth effect informally by rotating a three-dimensional object in such a way as to cast a temporal sequence of two-dimensional shadows. Almost any object will do, but an object that is not immediately recognizable from a single shadow will make a more convincing demonstration. For example, try bending a large paper clip into a novel (but fairly flat) three-dimensional shape, with a handle sticking out that you can hold to rotate it with. Place the wire just above the display plane of an overhead projector, with its handle pointed toward you. Its three-dimensional shape will not be easy to discern. Now rotate the bent wire by its handle, just above the display plane of the projector. Even though it is a novel object, the shadows of the bent wire will create a vivid perception of three-dimensional shape, and wires with different bending patterns can be recognized by the temporal sequences of their shadows. [These shadow patterns, of course, show that a heuristic is involved in perceiving shape from motion, because they are actually just sequences of two-dimensional patterns arranged to mimic the temporal sequences of retinal images that would come from real objects. If you take an analytical perspective, you can also just rotate any irregular object, and notice the sequence of shapes it presents.]

You can probably also imagine that if you varied the distance from the projector to the screen, the sizes of all of the images would be scaled up, but the temporal sequence would remain the same. That is, like shading patterns and texture compression patterns, temporal sequences arising from object rotation yield cues to shape but not to distance or size.

25.1.4 Shape from binocular disparity

Finally, as a fourth cue to depth, we return to binocular disparity. Like two objects at different distances (Figure 25.6xx), any three-dimensional object produces slightly different images in your left and right eyes, and this binocular disparity provides a potential depth cue. A striking example of the use of disparity to produce the perception of three-dimensional shape is shown in Figure 25.3.

25.1.5 Each cue has its limitations

We now come to one final but important point about depth cues. We have said that the first three of these cues shape-from-shading, shape-from-texture-compression, and shape-from-motion–are purely depth (and not distance or size) cues, because the patterns they depend on scale up proportionally with distance. In other words, if we used these cues alone, we could perceive the 3-D *shapes* of objects, but not their sizes or distances. These cues would need to be augmented

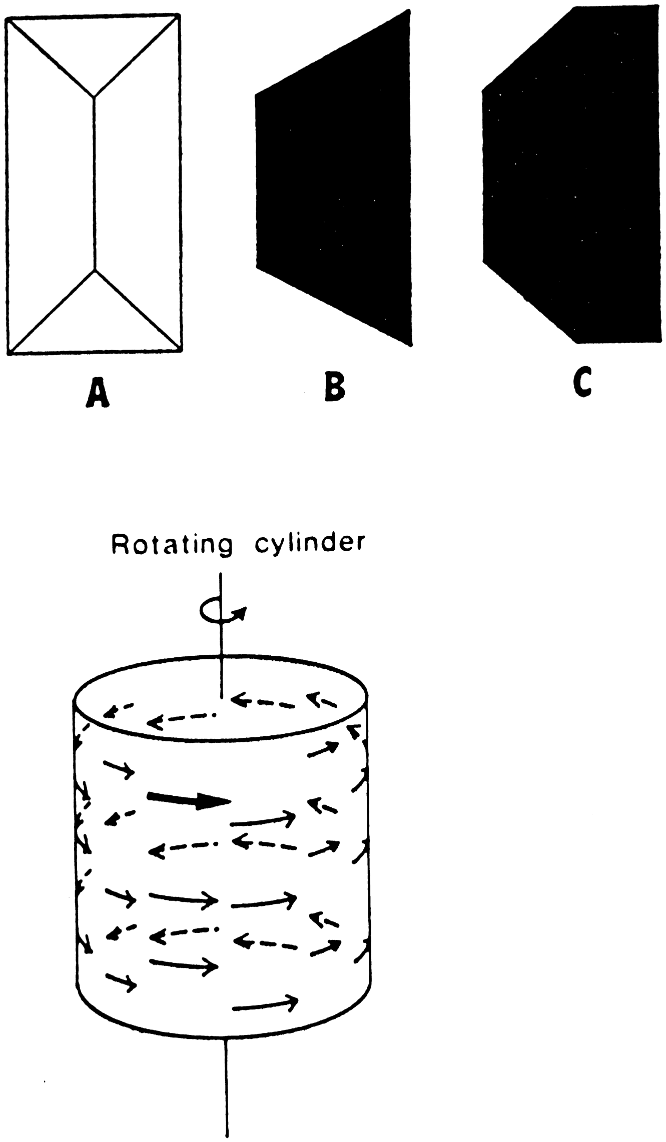


Figure 25.2: The kinetic depth effect: shape-from-motion. A: The line drawing at the left is intended to show a solid three-dimensional object. When this object is rotated, it generates a predictable sequence of two-dimensional patterns; two of these patterns are shown at the right. (Try to fill in a few more.) From Wallach and O'Connell (1953). B. This figure represents a transparent, rotating cylinder with dots on its surface. The lengths of the arrows indicate the directions and speeds of the dots in the sequence of images generated by the cylinder as it rotates. Remarkably, subects are able to group the dots on the two surfaces separately, and perceive the temporal sequence of images as a transparent, rotating cylinder. (A from Wallach and O'Connell, 1953, Fig. 1, p 207; B from Siegel and Anderson, 1988, p. 260)

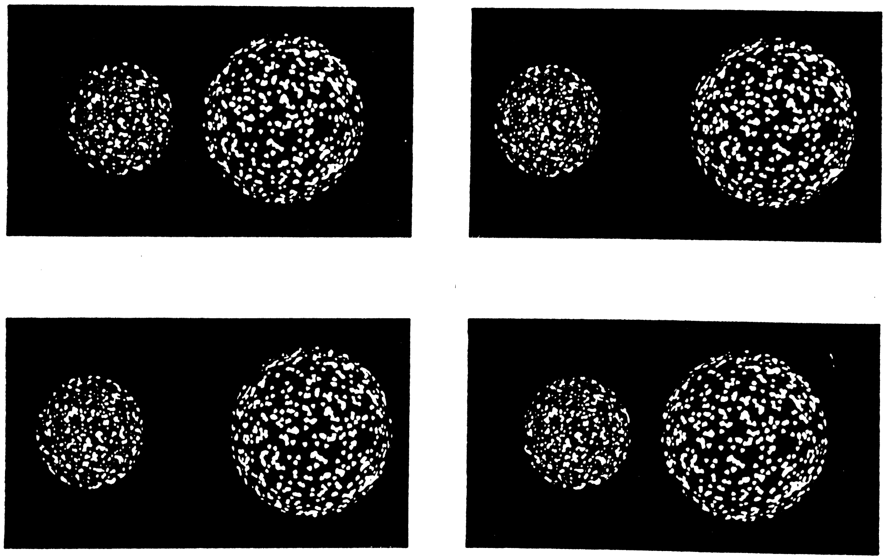


Figure 25.3: Shape-from-disparity. The bottom stereogram is left-right reversed from the top stereogram. The figures also contain texture compression cues. (From Brenner and Landy, 1999, p. 3836, Fig. 3)

by absolute cues to distance (such as accommodation and convergence) before we could know the sizes and distances of objects.

It turns out that binocular disparity also has a weakness, but its weakness differs from the weakness displayed in common by the first three cues. We learned in the section on distance that the size of the retinal image produced by an object of a given size depends not only on the size of the object but also on the viewing distance. In consequence, the retinal image size is not a reliable cue to the physical size unless it is augmented by a separate estimate of distance. Similarly, it turns out that the binocular disparity produced by an object of a given depth depends not only on the depth of the object, but also on the distance at which the object is viewed[1]. In consequence, the binocular disparity produced by an object is not by itself a reliable cue to the 3-D shape of the object. This is a critical flaw – even if the visual system calculates the disparity, the disparity by itself does not provide sufficient information to specify the depth of the object. This problem is sometimes referred to as the *stereo scaling problem.*

The ambiguity of disparity as a depth cue is nicely illustrated in a study by Elizabeth Johnston (1991). In her study, subjects were shown a random dot display with disparity as the only available cue to depth. The display depicted a half-cylinder standing out in front of the plane of fixation, like that shown in the stereogram of Figure 25.4A, and as shown schematically in Figure 25.4B. The subject could control the amount of disparity in the display. Johnston used a task she called the *apparently circular cylinder* (ACC) task, shown in Figure 25.4C. In this task, the subject was asked to vary the perceived ratio of the cylinder's half diameter, a, to its depth, b, until the cylinder appeared to have a circular cross-section. Most people report that the cylinder in Figure 25.4A appears flattened in depth, and would increase the disparity in the display to produce an apparently circular cylinder.

Using the ACC task, Johnston asked subjects to make ACC judgments at three distances: 53.5, 107, and 214 cm. The experiment showed that the settings depended on viewing distance, and were often strongly non-veridical. Johnston's data are shown in Figure 25.5. At the short viewing distance, a circular cylinder appeared elongated, and a decrease of disparity was required to make it perceptually cylindrical; whereas at the long viewing distance the opposite was true. Only at the intermediate viewing distance was depth perception approximately veridical. The problem revealed is that without information about viewing distance the disparity cue cannot reveal a veridical depth value.

Moreover, the pattern of results suggests a novel and highly interesting kind of heuristic. When it receives no information about absolute distance, the visual system apparently adopts a default distance value somewhere near 100 cm, and interprets the amount of disparity by this default distance to come up with a perception of depth. If the distance is less than the default distance, the physical depth is overestimated, whereas if the distance is greater than the default distance, the physical depth is underestimated. The suggested heuristic is: in the absence of a certain bit of information, make a particular pre-specified guess, and get on with the perceptual task. Here, the perceptual guess is that the distance of an object that provides no distance cues is about

[1] In fact, the binocular disparity produced by an object of a given depth decreases with the *square* of the distance from the observer to the object. That is, an object A of a fixed depth produces four times the disparity at 10 cm than it produces at 20 cm; and a second object B of four times the depth produces the same disparity at 20 cm that A produces at 10 cm. Try to work out this geometry. Hint – a) with retinal images of a fixed size, disparity would decrease with distance (there is no binocular disparity created by the moon); and b) the sizes of the retinal image also decrease with distance, as always. Together these two effects cause binocular disparity to decrease with the square of the viewing distance.

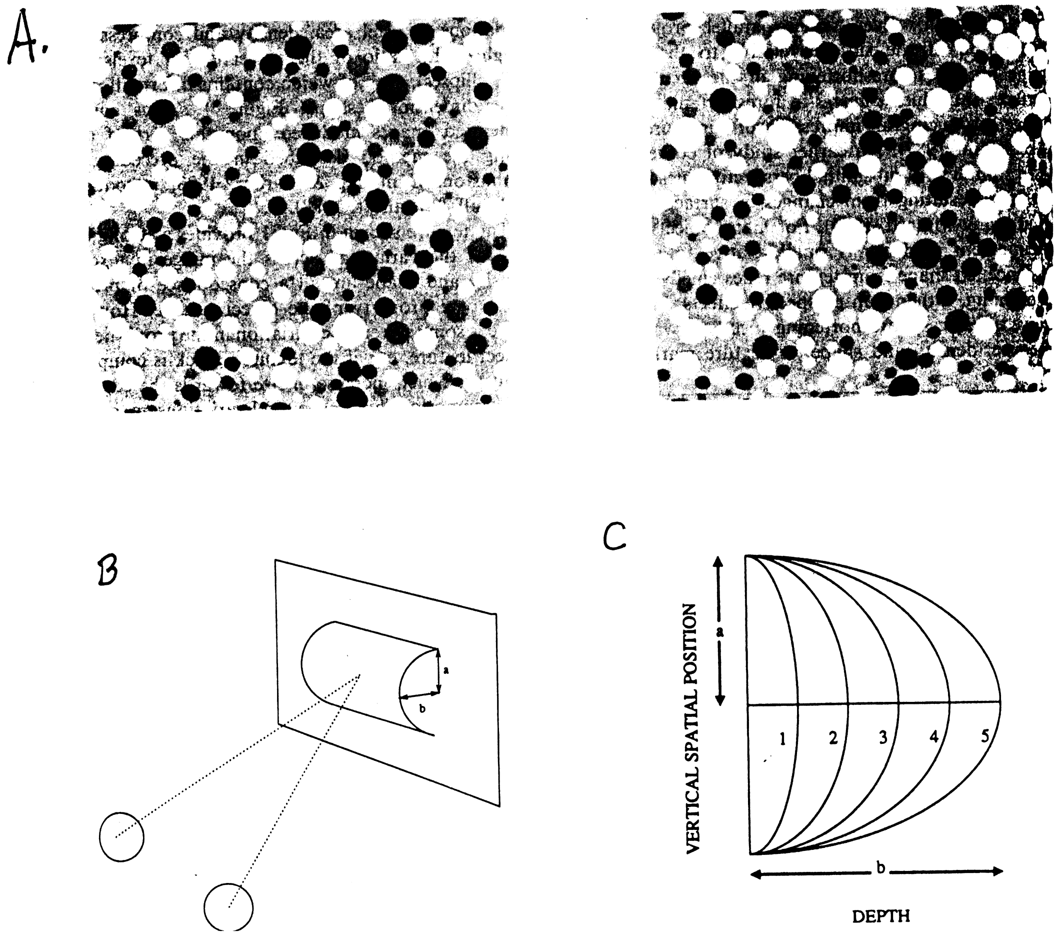


Figure 25.4: The apparently circular cylinder (ACC) task. A: A stereogram of a cylinder similar to that used by Johnston (1991). B: A schematic view of the stimulus, showing a circular cylinder, with its depth, b, equal to its halfwidth, a. C: Schematics of various flattened and elongated cylinders. Cylinder #3 is the circular cylinder. Most subjects judge the binocularly fused cylinder in A to be flattened, like cylinder #1 or #2, and would increase the amount of disparity to produce an apparently circular cylinder. (A from Cumming, Johnston, and Parker, 1993, p. 828, Fig. 1B. B from Johnston, 1991, p. 1353, Fig. 1; C from Johnston, 1991, p. 1355, Fig. 3.)

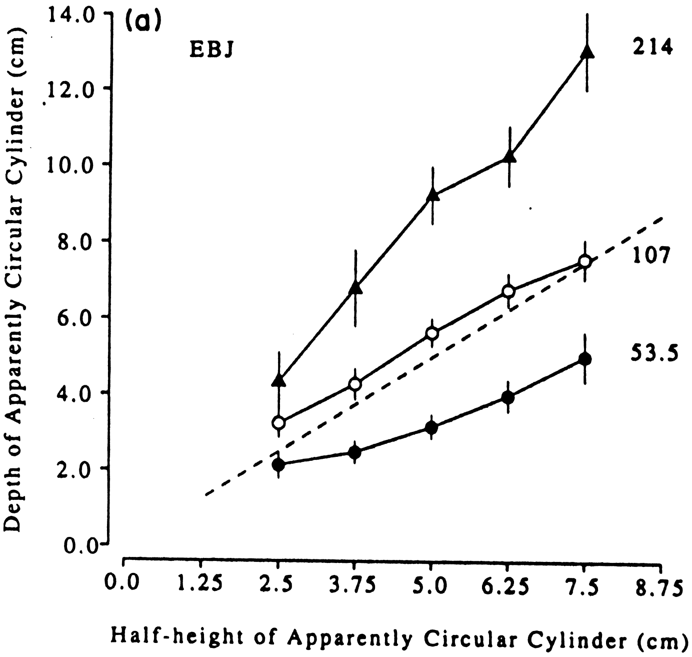


Figure 25.5: Data from Johnston's experiment (1991). The abscissa and ordinate show the half-heights and the depths of the cylinders respectively. For physically circular cylinders, the depth equals the half height, so veridically perceived circular cylinders would fall along the dashed line. Data points show combinations of half-height and depth that yielded apparently circular cylinders. The results of the experiment depended on viewing distance. Close cylinders were seen as elongated in depth, and disparity had to be decreased to make them perceptually circular. Distant cylinders were seen as flattened, and disparity had to be increased. At the intermediate distance of 107 cm, the cylinders were nearly veridically perceived. (From Johnston, 1991, p. 1355, Fig. 4A.)

100 cm. In DT's mind, it is only a short jump to a very general and very interesting depth and distance heuristic: the physical world, and the objects in it, are three-dimensional (!), and should be interpreted as such unless there is evidence to the contrary.

25.2 How do depth cues combine?

As discussed above for the case of distance, a long-standing question in psychophysics has concerned *cue integration*: how the visual system integrates several potential cues to yield a unified perception. In the case of distance perception, there were so many cues that we did not explore their variuos combinations, but instead relied on Cutting and Vishton's analysis. The case of depth perception provides a more tractible example, especially because we can concentrate on only four cues, and also because it has been an area of recent theoretical advances and perceptual experiments on cue integration.

As discussed above, different cues to depth have different weaknesses. Binocular disparity produces a depth cue, but one that cannot be used without the addition of a veridical distance cue; and shading, texture compression, and the kinetic depth effect each allow estimation of the three dimensional shape of an object, but not of its absolute distance nor its absolute size. Yet somehow we piece together veridical perceptions of the three-dimensional shapes of objects most of the time. How does the system work?

When an observer is viewing an ordinary object under ordinary viewing conditions, all four of these depth cues – shading, KDE, disparity, and texture compression – if they were present, all indicate a consistent depth for a single object. And in general, as shown in Figure 25.6, the greater the number of consistent depth cues the better the perception of depth.

But in artificial laboratory situations, we can create stimuli that incorporate only, say, two of these depth cues. We can then perturb the degree of depth portrayed by one of the cues, and find the degree to which the other must be perturbed in the opposite direction to re-establish the original perceived depth. The question is, how do two cues trade off? That is, by what quantitative rule does the visual system combine the information from the two different cues?

There are actually two parts to this question. The first question is, are the different cues initially analysed separately, or are they analysed together with some joint optimization algorithm? Some theorists have argued that the visual system analyses all available depth information jointly, and comes out with a single best estimate of the depth of the object (a theoretical position called *strong fusion*)[2]. Other theorists have argued that the visual system has a set of separate modules for initial processing of the different depth cues separately, and combines them only later (a theoretical position called *weak fusion*). The second question is, if the different depth cues are analysed separately – if weak fusion prevails – by what rules are the different depth estimates then combined to yield a single perceived depth? Do we use a linear weighted sum, or a more complicated combination rule? And if a linear weighted sum, what determines the weights?

A number of cue combinations have been investigated recently. For example, Johnston, Cumming and Landy (1994) followed up Johnston's (1991) study of disparity that we decribed above, but they added a motion cue to the disparity cue. Using the ACC task, they studied the veridicality of disparity and motion (KDE) cues separately and in combination. Their results are shown in

[2] In the literature on binocular disparity, the term "fusion" is traditionally used to refer to the fact that the left and right eye images images "fuse" perceptually to yield a single perceived depth or distance. Here we are using the term differently, to refer to the *combination of different depth cues* to yield a single perceived depth.

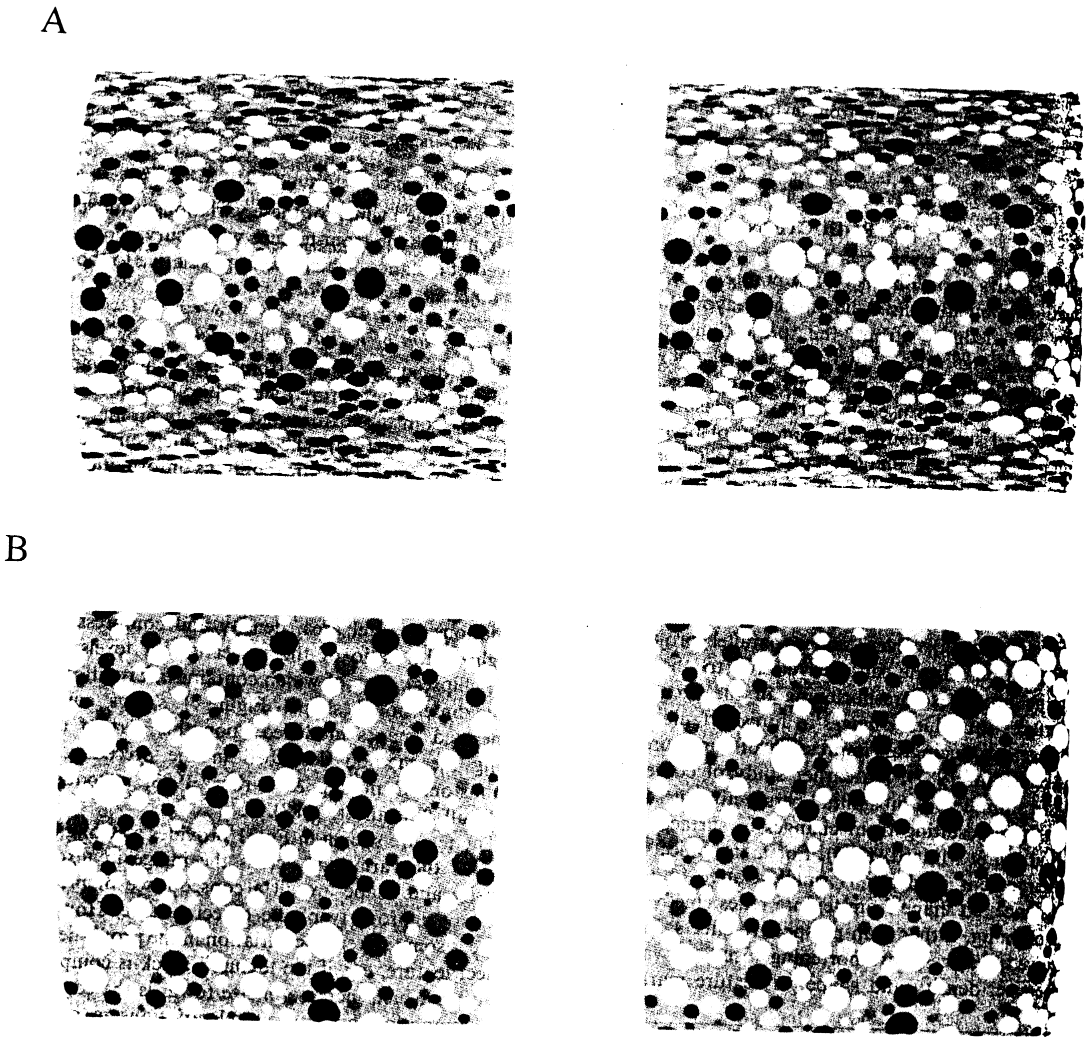


Figure 25.6: The combination of depth cues: texture compression and disparity. A: The upper images, directly viewed, show the effectiveness of texture compression alone as a depth cue. B: The lower images, viewed stereoscopically, show the effectiveness of binocular disparity alone. Stereo viewing of pair A shows the effectiveness of both cues together. Most subjects report that the combination of depth cues gives a stronger impression of depth than either cue alone. This result implies that depth cues are not *averaged*; rather, they are *added* together in some kind of a weighted sum. (From Cumming, Johnston, and Parker, 1993, p. 828, Fig. 1.)

Figure 25.7A. As Johnston had shown before, the subject's settings varied with viewing distance. At 50 cm, the disparity cue yielded a small over-estimation of depth, whereas KDE gave reasonably veridical estimates of depth, as did both cues together. But at 200 cm, the disparity cue yielded large systematic under-estimations of depth, whereas KDE alone and both KDE and stereo together yielded quite veridical depth estimates. This initial result, particularly at 200 cm, would seem to suggest that the KDE cue simply dominates the disparity cue, but that 's too simple.

These authors did a second experiment, in which they made the two cues discrepant – disparity indicated one depth and KDE indicated another. These results are shown in Figure 25.7B. They found that the amount of disparity present influenced the amount of KDE needed for the subject to perceive the apparently circular cylinder! The greater the disparity, the less KDE-based depth was needed. That is, the notion that disparity is not used is incorrect. Moreover, the two cues traded off in a remarkably regular way, consistent with the hypothesis that they were combined as a weighted linear sum. But the weightings were not fixed, but varied with the viewing distance.

Finally, these authors found that if they weakened the motion cue by reducing the length of the motion sequence to two frames, its weight was diminished while the weight of the disparity cue was increased. These and other studies have shown that the weightings given to two cues can vary with circumstances and across the scene. For example, if noise is added to the texture compression cue, it is weighted less heavily; and similar findings have been presented for other cues and cue combinations.

From findings such as these, Landy, Maloney, Johnston, and Young (1995) have proposed an intermediate theory of depth cue combination, which they call *modified weak fusion*, or *MWF*. They build on the fact that each depth cue is imperfect – each is missing the specification of one or more parameters (as we said earlier, the disparity cue is missing a specification of absolute distance, and shading, texture compression, and motion are all missing a specification of absolute size). Landy and his colleagues suggest that the cue combination process acts in three stages. First, each cue is analysed separately, as far as it can be, given its missing parameters. Second, the cues interact in a very limited way – each provides the missing parameters for the others. In their terms, the cues *promote* each other – each cue process gleans what it can from the other cues to fill in its missing parameters. And third, information from the different cues is then combined in a weighted linear sum to yield an overall estimate of the depth of a particular object. Finally, the values of the weights vary depending on the reliability of each particular cue provided by the particular object in the particular situation.

Landy and his colleagues argue that MWF theory has advantages over the strong fusion and weak fusion theories, particularly that it is internally consistent and specific enough to be falsifiable. Attempts to test the notions in the MWF model – the combination rules for different pairs and sets of cues, cue promotion among the various cues, the generality of the weighted linear summing rule, and the factors that influence the weightings of the different cues in different situations – are a major current focus of work in depth perception.

25.3 Physiology of distance and depth

We now come back to one of the fundamental themes of this book – the art and logic of forging links between perceptual and physiological events. Suppose you are a neurophysiologist, and you are interested in studying the encoding of distance, size, and depth. You want to attack the question of what parts of the visual brain carry out the analysis of distance and depth cues, and

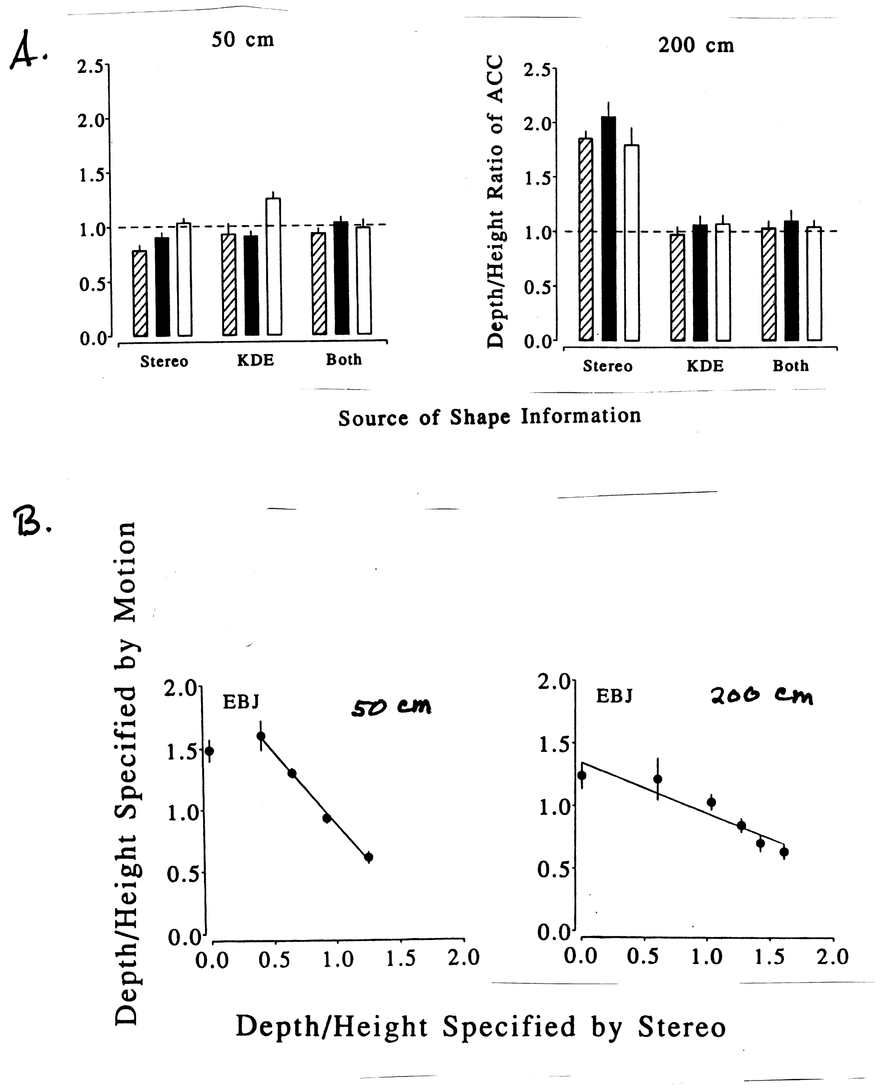


Figure 25.7: The results of Johnston, Cumming, and Landy (1994) concerning the combination of motion ("KDE") and disparity ("stereo") cues. A: In Experiment I, the two cues were tested separately and together. The results varied with viewing distance. At 50 cm, depth perception was nearly veridical under all conditions. But at 200 cm, the stereo cue alone yielded systematic underestimations of depth. KDE yielded quite veridical estimates, as did both cues together. B: In Experiment II, the two cues were traded off , and the apparently circular cylinder task was used to find pairs of cue values that all yielded apparently circular cylinders. The data fit straight lines, providing evidence that the cues are combined by a weighted linear sum. The slope of the line indicates the weightings of the two cues. The weightings are not fixed, but vary with the viewing distance. (All from Johnston, Cumming and Landy, 1994. A, Fig. 5, p. 2266; B, modified from Fig. 6B, p. 2268.)

by what sequence of information transformations. What hints can we glean from combining logic with system properties, to mount an intelligent attack on this question? Again, this is the point at which the neuroscientist places his or her bets, on the experiments he or she thinks will best reveal the fundamental processing features of the visual system.

Perceptual studies tell us that distance and depth information comes in many different forms – cues based on the analysis of pictures, the analysis of motion, and so on. Given the idea of parallel processing, we tend to think of some of these different cues – motion and form, for example – as being analysed separately in early visual processing. On this line of thinking, one line of attack would be to look for neurons in early visual processing that are tuned to each of the different distance cues separately, and then for other neurons later in the visual system that respond to various combinations of cues or even provide a single unified representation of the distances of objects. Similarly, following on Landy et al's modified weak fusion (MWF) theory, perhaps we should look for three stages of processing. We might look for an early, separate processing of the different depth cues – shading, texture, motion patterns, disparity; a second stage at which each of these cues is promoted; and a third and final stage at which a combined representation of the three-dimensional shapes of objects emerges.

The fact that there are so many distance and depth cues suggests that the neural analyses underlying veridical distance and depth perception will probably be complex. Moreover, the depths and distance of objects are important both to our perception of *where* objects are and to our recongition of *what* they are, so taking the Ungerleider and Mishkin characterization of visual streams, we might expect to see representations of depth and distance in both the dorsal and the ventral stream. Taking the Milner and Goodale action vs. perception approach, we might expect that both distance and depth should be represented in the dorsal stream, as we need both to guide our motor responses (think of reaching for an object and picking it up). We might also expect that depth will be represented in the ventral stream, since the three-dimensional shape of an object is critical to object recognition.

Curiously, there has been very little physiological work motivated by these lines of thinking. Instead, until very recently, most physiologists have concentrated almost exclusively on looking for neurons that are tuned for a single depth cue – binocular disparity – and almost exclusively in the dorsal stream[3].

In this section we begin with studies of disparity tuning in individual neurons in visual cortex. We will see that disparity tuning emerges first at early cortical levels – V1 and V2. Moving to the dorsal stream, we will find that MT neurons, which we previously learned were tuned to the direction and speed of motion, are also tuned to disparity. We will then move to a few recent studies of the responses of dorsal stream neurons to depth cues other than disparity. And finally, we will review some very recent work on responses to depth cues in neurons in the ventral stream.

[3] It is an interesting historical question to ask, when the multiplicity of distance and depth cues has been appreciated for so long, why have physiologists concentrated so exclusively on binocular disparity, rather than exploring the many other cues to depth and distance? Early comments in the physiological literature even sometimes suggest that binocular disparity is the *only* depth or distance cue. DT's speculation is that disparity is easy to comprehend from the geometry of having two eyes, especially if the disparity scaling problem is set aside. Perhaps it is harder to be convinced of the value of cues that depend more obviously on (fallible) heuristics.

25.3.1 Disparity tuning in V1 and V2

In the earliest published study of disparity tuning in cortical neurons, Horace Barlow, Colin Blakemore, and Jack Pettigrew (1967) recorded from V1 neurons in the anesthetized cat. The orientation tuning and the binocularity of most cortical cells had been established by Hubel and Wiesel a few years earlier. Building on this work, Barlow et al used moving black or white bars, or black/white edges of various orientations, in each eye separately, to locate the receptive fields of individual cortical cells in each eye. They then placed the optimal stimulus in both the left and right eyes, and varied the horizontal *displacement* between the two stimuli, thus varying the binocular disparity of the combined stimulus. They found that many cat V1 neurons were tuned for binocular disparity, and even more importantly, that different V1 neurons responded optimally to different disparities.

Rather than plotting the disparity tuning of their cells per se, Barlow et al presented their data in terms of distance coding. For purposes of illustration, they "solved" the stereo scaling problem by assuming a fixed viewing distance of 50 cm. They then back calculated from each neuron's optimal disparity to find the *distance* (relative to 50 cm) that the stimulus would have had to be from the cat, in order to elicit the best response from the neuron.

Barlow et al's results are shown in Figure 25.8. These data showed for the first time that lines and edges at different relative distances from the cat will excite different neurons in the cat's visual cortex. In other words, *activity in this ensemble of neurons could represent the presence of lines and edges at different relative distances.* These data thus provide us with our first glimpse of a possible neural representation of distance and/or depth in the mammalian cortex.

A decade later, Gian Poggio and B. xx Fischer (1977) undertook a study of disparity tuning in V1 and V2 neurons in awake, behaving macaque monkeys. The monkeys were trained to fixate a fixation point while the researchers isolated a single neuron, and located its receptive field in each eye. Moving bright and dark bars were used as stimuli. Like Barlow et al (1967), Poggio and Fischer then varied the relative locations of the stimuli within the two receptive fields – the binocular disparity – and plotted the neuron's disparity tuning curve. They found that more than 80% of the neurons in both V1 and V2 were tuned for binocular disparity. Moreover, both simple cells and complex cells were disparity tuned in about equal numbers.

Some of the patterns of disparity tuning seen in V1 and V2 in this and later papers are shown in Figure 25.9.[4] Some V1 and V2 neurons are narrowly tuned for binocular disparity. The most common type of neuron is the *tuned excitatory* neuron, which responds best over a narrow range of disparities, and is inhibited by disparities outside this range. Different tuned excitatory neurons respond best to negative (crossed) disparities, to disparities near zero, or to positive (uncrossed) disparities. Similarly, *tuned inhibitory* neurons are inhibited across a narrow range of near-zero disparities. Other neurons are much more broadly tuned, and are either excited over a broad range of near disparities and inhibited over a broad range of far disparities (*near cells*) or vice versa (*far cells*). More recent authors have suggested that these cell types are not mutually exclusive, and that there is probably a continuous variation of disparity tuning among V1 and V2 cells. Thus, at least some analysis of binocular disparity apparently occurs at the earliest levels of processing in visual cortex.

[4] In this and later figures, the abscissa shows the disparity between the stimuli presented to the left and right eyes. By convention, crossed disparity – which arises from objects closer than the fixation point in natural viewing – is labelled with a (-) sign, and neurons that respond best to crossed (-) disparity are often labelled *near* neurons. Uncrossed disparity – which arises from objects farther away than the fixation point in natural viewing – is labelled with a (+) sign, and neurons that respond best to uncrossed disparity are often labelled *far* neurons.

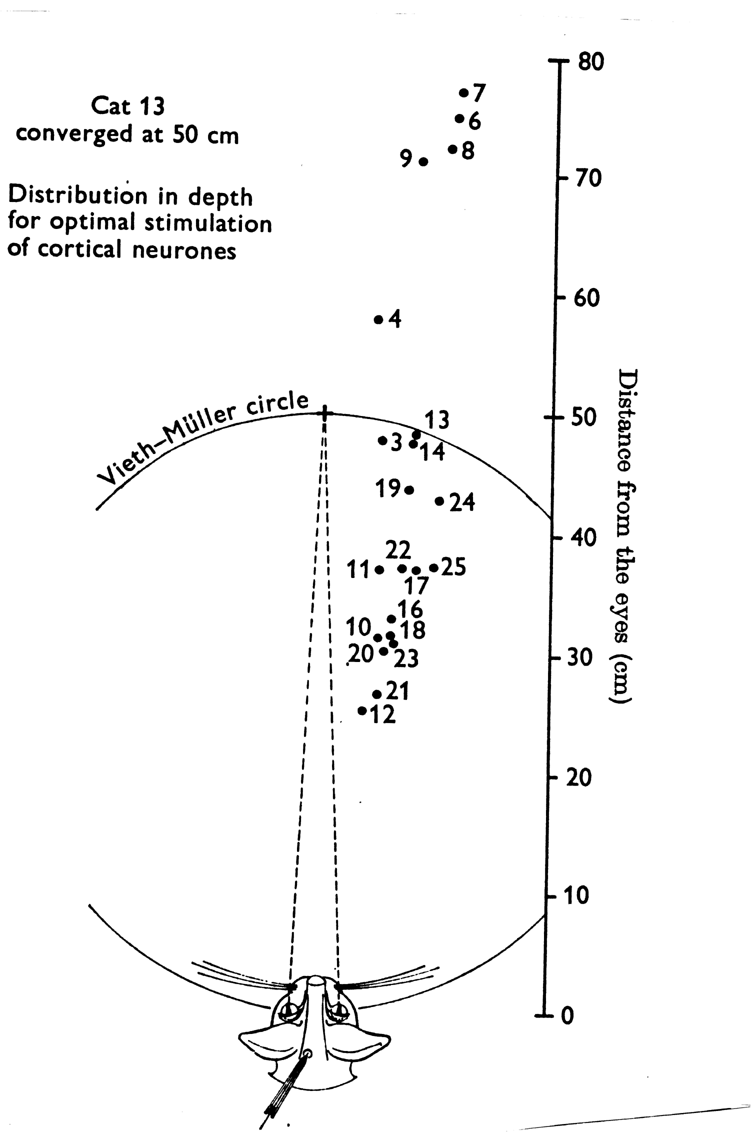


Figure 25.8: A code for relative distance in cat V1 cells. In this figure, the cat is assumed to be fixated at 50 cm. (The Vieth-Muller circle is the locus of all point that yield zero disparity for the 50 cm fixation distance.) Each of the numbers at the right of the figure corresponds to a different cortical neuron. The numbers are plotted at the distances that would have led to optimal firing for that neuron. Different neurons respond optimally to lines or edges at different distances. These data were the first to reveal a possible neural representation of distance in mammalian cortex. (From Barlow, Blakemore and Pettigrew, 1967, p. 339, Fig. 6)

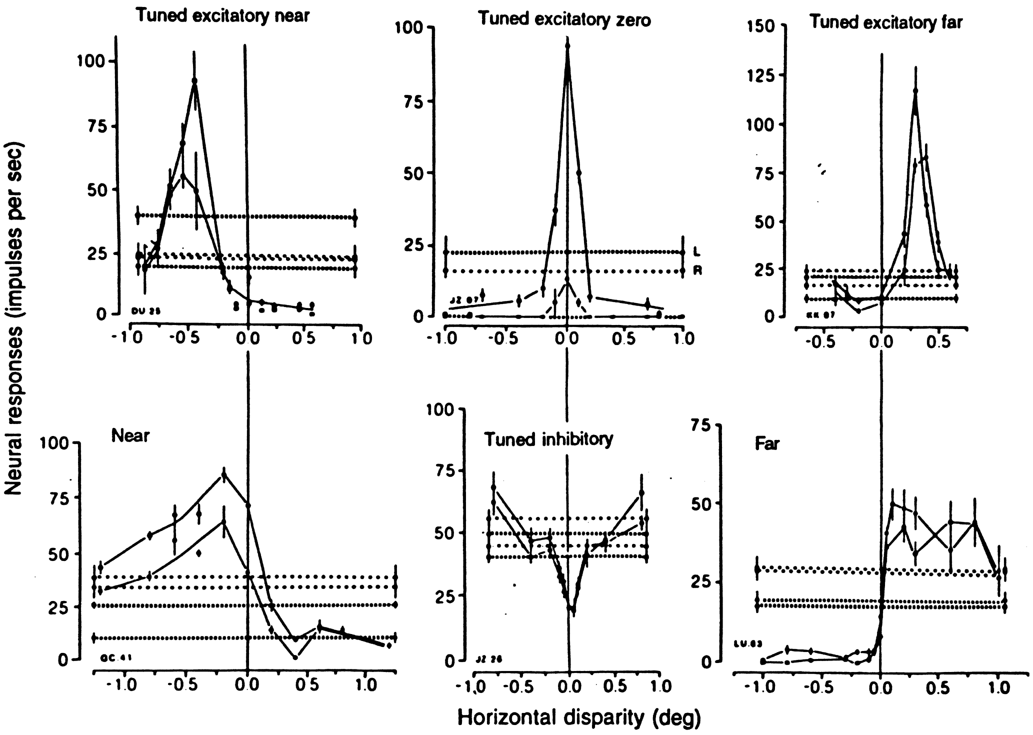


Figure 25.9: Six prototypical neurons with different types of disparity coding. In each panel the abscissa shows the amount of disparity between the stimuli in the two eyes. By convention, negative disparity values (-) represent crossed disparities (created by near objects in normal viewing) and positive disparity values represent uncrossed disparities (created by far objects in normal viewing). The ordinate shows the firing rate of the neuron. The two traces in each panel show responses to two opposite directions of motion. The horizontal dashed and dotted lines show the responses to left and right eye stimuli alone. The top row of panels shows three tuned excitatory neurons, tuned for negative ("tuned excitatory near"), zero ("tuned excitatory zero"), and positive ("tuned excitatory far") disparities respectively. The bottom center panel shows a tuned inhibitory neuron. More broadly tuned neurons responsive to negative disparities ("near" cells) and to positive disparities ("far" cells) are shown at the bottom left and right respectively. (from Poggio, 1991).

In later studies, Poggio and his colleagues also tested V1 and V2 neurons with random dot stereograms (RDS). Interestingly, simple cells were generally unresponsive to RDS. A few complex cells did respond, and some of these were tuned to the amount and direction of disparity in the RDS (reviewed in Poggio, 1991). Thus, as Julesz would have wanted us to believe from the psychophysical data, at least a few neurons in early cortical processing can signal the amount of binocular disparity even in the absence of other cues to the shape of the stimulus. We will need to wait to see, however, whether these relatively infrequently encountered early cortical neurons form a major basis of RDS analysis.

25.3.2 Disparity tuning in MT

In Chapter 21xx we learned that neurons in cortical area MT, located at the beginning of the dorsal stream, are tuned for the direction and speed of motion, and we concluded there that MT plays a major role in the processing of motion signals. However, it has been known since the early 1980s that MT neurons are also tuned for binocular disparity.

Recently, Greg DeAngeles, William Newsome and their colleages have carried out a series of studies of the disparity tuning of MT neurons (De Angeles, Cumming and Newsome, 2000). The logic of the work is similar to that we have already seen for studies of motion from the Newsome laboratory (see Chapter 21xx), and combines single unit recording, behavioral testing, and microstimulation.

To study disparity tuning in area MT, De Angeles et al used sparse, moving random dot patterns similar to those used in their earlier studies of motion, but adding binocular disparity as a major independent variable. In the disparity studies, the authors chose to record both the activity of a single neuron, and the background activity of *multiple units* in the vicinity of that single neuron. The locations and sizes of receptive fields of the individual neurons and multi-unit recording sites were characterized, along with the preferred direction of motion and the preferred speed of the neurons. In most cases, the responses of the single neuron and the multiunit response had similar tuning on all of these dimensions.

The authors then varied the amount of binocular disparity between left and right eye inputs within the RF of the neuron. Figure 25.10 shows six examples of disparity tuned neurons in area MT. In these experiments, the binocular disparity was varied in small steps from *near* (negative) to *far* (positive). Both single unit and multiunit responses are shown. As in the case of the V1 and V2 neurons recorded by Poggio and Fischer (1977), MT neurons show clear disparity tuning, but vary in their sharpness of tuning. In each case, both the single unit activity and the multi-unit activity varied together with the amount of binocular disparity, suggesting a patchy or columnar organization, with cells with similar disparity tuning curves grouped together anatomically.

Microstimulation and behavioral studies

. A patchy or columnar structure, of course, is very advantageous experimentally, as it makes microstimulation studies possible. Since neurons of the same disparity tuning occured in clusters, the Newsome group proceeded to carry out microstimulation studies. In these studies they introduce a new stimulus variable: the *percent binocular correlation* of the dots.

The stimuli and the response paradigms used in these studies are shown in Figure 25.11. The dots outside the receptive field of the neuron under study were all set to zero disparity, and appeared as a plane of twinkling dots to a human observer. Within the RF, a chosen fraction (say, 50%)

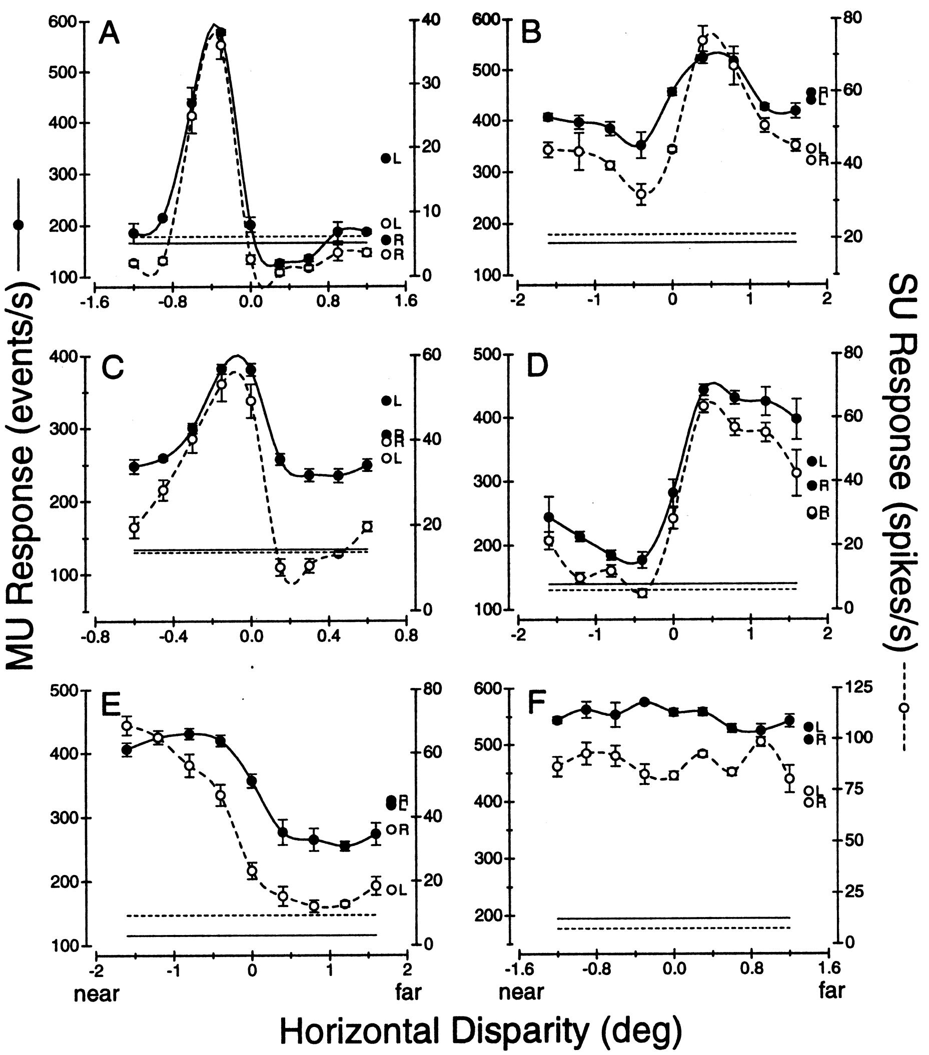


Figure 25.10: Six patterns of disparity tuning in MT neurons. Each panel shows the responses of a single neuron (open circles) and the simultaneous multi-unit activity (closed circles) recorded at a particular site in MT. The single unit and multi-unit activity correlate well, suggesting that neurons in local regions all have similar disparity tuning patterns. As described by Poggio and Fischer (1977) for V1 cells, the sharpness of disparity tuning, and the tuning for "near", "far", or zero disparity, vary from one recording site to the next. Panel F shows a recording site at which neurons showed much enhanced firing rates to binocular stimulation, but were not appreciably tuned for binocular disparity. The dashed and solid horizontal l ines show spontaneous firing rates for single unit and multiunit responses, respectively. The isolated symbols marked L and R show responses to left or right eye stimulation alone. (From De Angeles and Newsome (1999), Fig. 4, p. 1402.)

of the dots had a common disparity, while the other dots had randomly selected disparities. This fraction is the percent binocular correlation of the dot display. The dots with the common disparity were presented with either a positive or a negative disparity from one trial to the next, so that to a human observer the stimulus appeared as a region in which some of the dots were displaced either in front of or behind the surrounding plane of dots. The percentage of binocularly correlated dots was varied from one trial to the next, making the task harder or easier. In the psychophysical experiments, the monkey's task was to report whether the dots with the common disparity appeared nearer or father than the plane of fixation.

The results of the study are shown in Figure 25.12. Both of the data sets shown were obtained from multi-unit activity. In both cases the multiunit activity was tuned for binocular disparity; and in both cases microstimulation increased the proportion of trials on which the monkey reported depth in the direction to which the local neurons were tuned. Thus, in all likelihood MT neurons are on the causal chain that controls our perceptual access to binocular disparity cues.

A final note – we have now learned that most MT neurons are tuned for the direction and speed of motion, and that some are tuned for binocular disparity. However, little is yet known about how these two types of tuning relate to each other.

25.3.3 MT neurons tuned for tilt

In Chapter 24 we introduced the concept of *velocity gradients* or *motion flow fields* – the idea that differential motion between an observer and objects in the physical world gives rise to complex but highly specific patterns of motion in the retinal images; and that analysis of these patterns could provide information about the motions, shapes and locations of objects. The next question is, are there neurons that are specifically tuned to variations in velocity gradients?

When a textured surface is *tilted* in depth and moved laterally in front of the observer, the texture elements at the near edge of the plane move relatively rapidly across the retina, whereas those at the far edge move relatively slowly, and there is a velocity gradient in between. For example, imagine a plane tilted around a horizontal axis, so that the top is farther away from you and the bottom is nearer. As the plane moves, the texture elements at the bottom will move more rapidly in the retinal image, those at the top more slowly, and there will be a regular velocity gradient in between. Moreover, planes with different tilts will yield velocity gradients with different orientations (top slow, bottom fast; bottom slow, top fast; left slow, right fast; etc.)

Guy Orban and his associates (Xiao, Marcar, Raiguel, and Orban (1997)) asked whether or not MT neurons were tuned for such variations in tilt. To find out, they tested MT neurons with two-dimensional random dot patterns. They first determined the cell's preferred direction and speed of motion and its preferred binocular disparity. Patterns with the preferred motion and the preferred disparity were then presented to the cell, but the patterns also contained velocity gradients indicative of tilted planes. Xiao et al found that some individual MT neurons were indeed tuned for the direction of tilt, with different neurons tuned to different tilts. An example of one such neuron is shown in Figure 25.13. This finding is interesting because it suggests that these neurons have properties useful for representing the orientations of tilted surfaces, and thus helping to represent the spatial layout of the visual world.

These experiments are important ones, because they open the door to the neurophysiological analysis of distance and depth cues other than binocular disparity, and they immediately raise many questions. It would fascinating to know whether or not neurons like these respond to the *rotations*

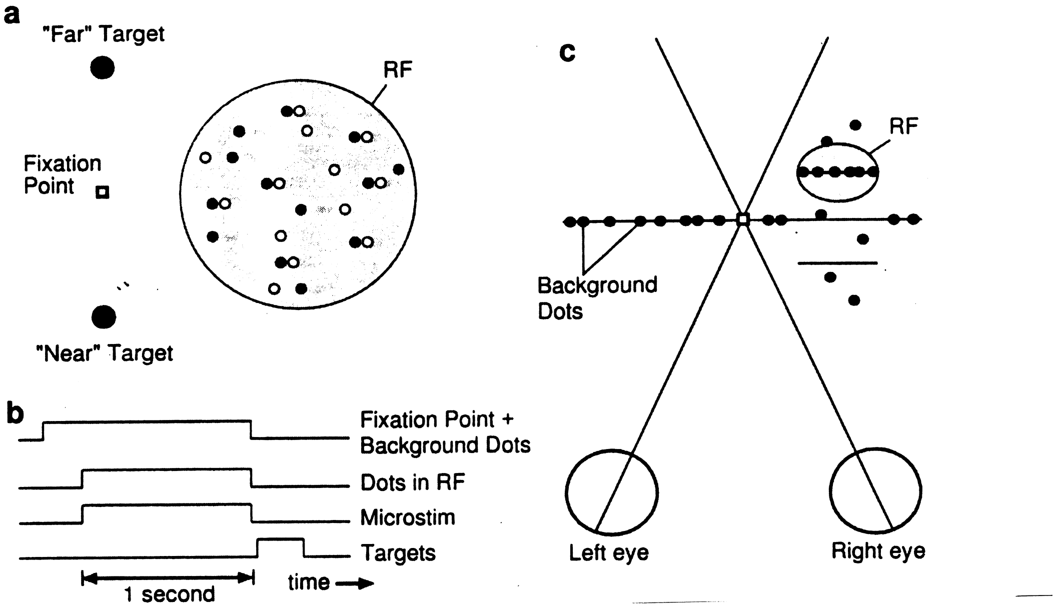


Figure 25.11: The stimuli used for microstimulation studies of disparity tuning. A: The stimulus field. The field of dots shown was placed over the receptive field (RF) of the cell. Filled and open dots show the patterns presented to the left and right eyes respectively. A variable percentage of the dots (here, 50%) had a common binocular disparity (*percentbinocular correlation*), whereas the other dots within the receptive field had random disparities, and the dots outside the RF had zero disparity. The dots labelled "near target" and "far target" were the target lights to which the monkey was trained to shift fixation, in order to indicate his perception of whether the correlated dots were perceived to be in front of ("near") vs. behind the plane of fixation ("far"). B: Sequence of events for each trial. The monkey first fixated the fixation point. The field of dots then appeared for 1 sec. In the microstimulation experiments, microstimulation was also applied during the same 1 sec. interval. After the field of dots went off, the monkey refixated to either the "far target" or the "near target" light to indicate his psychophysical response. C: A bird's-eye schematic of the visual stimulus, with 50% binocular correlation. In the region of the RF of the cell, half of the dots – the six dots in the oval – share a common disparity, in this case indicating a "far" target. The other six dots within the RF have random disparities, shown as dots at various distances. The background dots all have zero disparity. The horizontal line shows the plane in which the correlated dots would appear on "near" trials. (From De Angeles, Cumming, and Newsome, 2000, p. 310, Fig. 21.4)

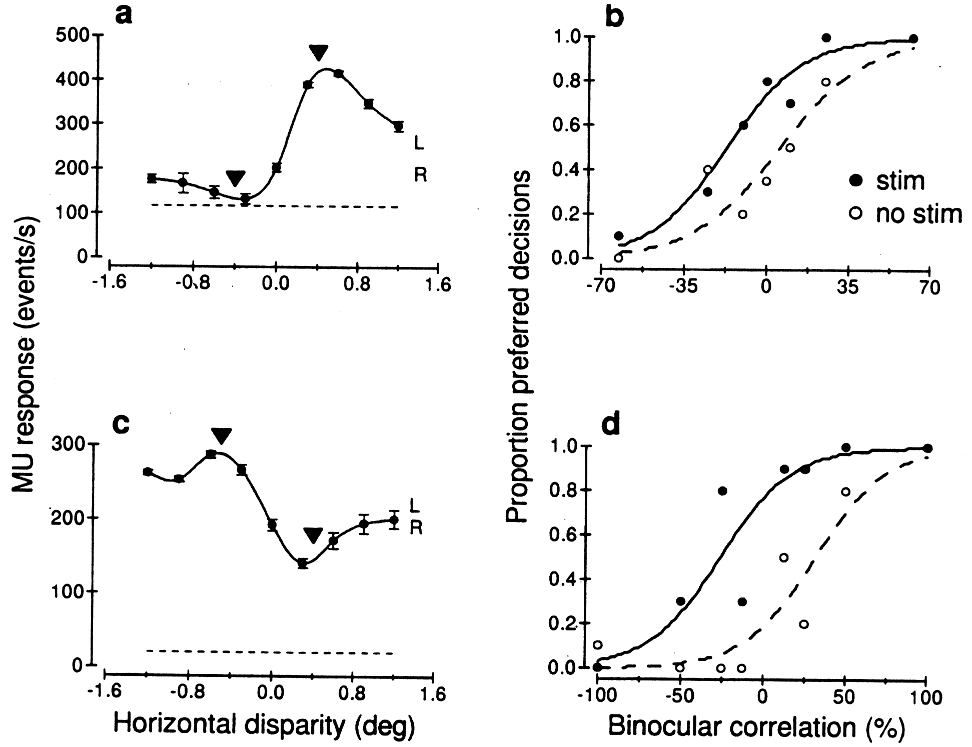


Figure 25.12: The effect of microstimulation on the monkey's reports of perceived depth. A: Disparity tuning of a multiunit site in area MT. The site is maximally reponsive to far (+) disparities. The arrows show the two disparities chosen for behavioral testing. B: Results of the behavioral tests. The abscissa shows the percent binocular correlation in the test field. The ordinate shows the "proportion preferred decisions"; that is, the proportion of trials on which the monkey chose the preferred disparity (far, or +) of the neurons at the site under test. Without microstimulation (open circles), the monkey's psychometric function passes through chance at zero percent binocular correlation; at higher binocular correlations (either + or -) the monkey is increasingly able to report the direction (near vs. far) of the binocular correlation in the stimulus. With microstimulation (closed circles), the monkey's psychometric function is shifted leftward, indicating that he saw the stimulus as "far" on an increased number of trials. That is, microstimulation influences the monkey's perceptual reports, increasing the proportion of trials on which the monkey reports depth in the direction to which the local neurons are tuned. C and D: A similar experiment at a site tuned to "near". (From De Angeles, Cumming, and Newsome, 2000, p. 311, Fig. 21.5).

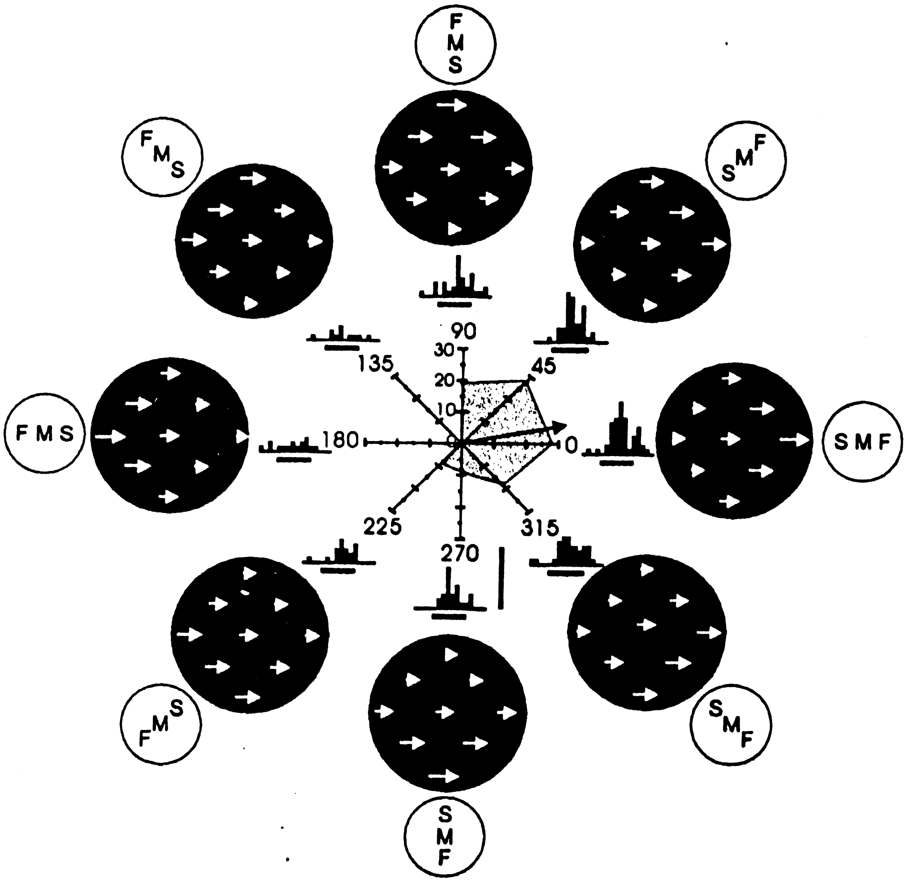


Figure 25.13: Tilt tuning in MT neurons. The black patches with arrows show the patterns of motion of the stimuli, intended to simulate the motion flow fields that arise from tilted planes moving laterally in front of the observer. The lengths of the arrows represent the different speeds of motion. The circles marked S, M, F denote the locations of the different speeds (slow, medium, fast), across the stimulus. The histograms near the center show the responses of a single neuron to the different motion flow patterns. The graph in the center shows the response in a polar plot – the direction out from the center represents the tilt of the motion flow field, and the distance out from the center shows the magnitude of the response. This neuron shows a maximum response to the flow field shown at the right (0^o), and a minimal reponse to the flow field shown at the left (180o). These neurons could participate in representing the tilts of planes in three-dimensional space. (From Xiao, Marcar, Raiguel, and Orban (1997), Fig. 2, p. 958.

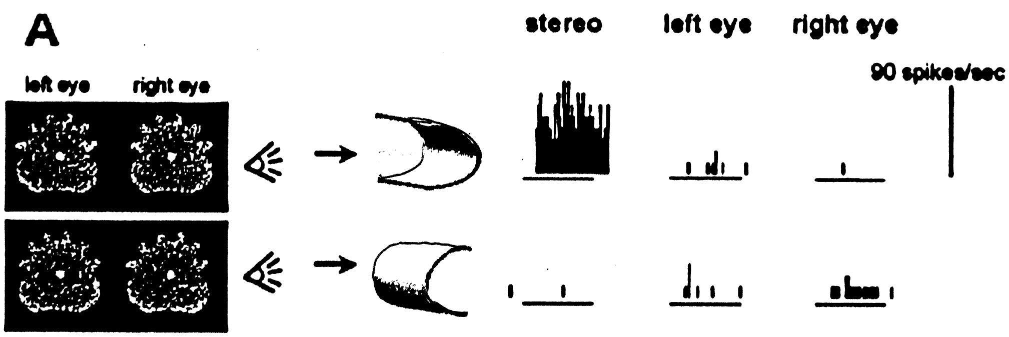


Figure 25.14: An IT neuron tuned for three-dimensional shape. The left panel shows an example of the stimuli used by Janssen et al (1999, 2000). The middle panel shows the perceived three-dimensional sstructure (concave vs. convex) for the two left eye-right eye configurations. The right panel shows the response of an IT neuron to two binocular (”stereo”) and four monocular stimulus configurations. The neuron responds only to the stimuli representing the convex shape, and not to the other five stimuli. (Adapted from Janssen, Vogels, and Orban (2000), p. 2055, Fig. 2).

of three-dimensional objects in depth, such that they could underlie our use of structure-from-motion cues. It would also be fascinating to know whether some individual MT neurons are tuned to disparity and others to motion, or whether the same neuron might even be tuned jointly to the same depth value coded by both motion and depth. If so, individual MT neurons might integrate two different shape-from-X cues – shape-from-disparity and shape-from-motion – to represent the depth of an object. Quantitative analysis of neurons like these could allow us to test Landy et al's modified weak fusion theory, or lead to new ideas about the ways in which depth cues combine.

25.3.4 Ventral stream neurons tuned for depth cues

The above review of the tuning of individual neurons to distance and depth cues has been confined to the early levels of visual processing and to the dorsal stream. But what of the ventral stream? We argued early in this chapter that, insofar is the ventral stream ”does” pattern recognition, it too should have neurons tuned to the three-dimensional shapes of objects. This question has been addressed only very recently. (We are a little ahead of our story here – we will return more systematically to the properties of ventral stream neurons in later chapters.)

Orban and his colleagues (Janssen, Vogels, and Orban, 1999, 2000) used random dot patterns and complex shape contours to create pairs of stimuli, A and B, that differed in binocular disparity. These stereo pairs could be presented in two configurations – A to the left eye, B to the right, or vice versa. Human observers viewing these stimuli perceive surfaces that are wavy in three dimensions, with the pattern of waves reversing from concave to convex depending on which stimulus goes to which eye. An example of such a stimulus pair is shown in Figure 25.14.

Janssen et al presented these stimulus pairs in both configurations to individual neurons in inferotemporal (IT) cortex – the classical terminus of the ventral stream. They found that more than 1/3 of neurons in IT cortex were tuned for such stimuli, responding preferentially to the stereo pair in one left eye-right eye configuration as opposed to the other. Responses of a neuron tuned for these differences in binocular disparity is shown in Figure 25.14. These data provide the first evidence that some IT neurons are indeed capable of representing the three-dimensional structure of objects. Janssen et al (2000) also showed that disparity tuned neurons are confined to one subdivision of IT cortex – area STS, a subpart of area TE that lies in the lower bank of the superior temporal sulcus.

Of course, these results raise many questions. Do these neurons respond to other depth cues? Do they respond to the same aspects of three-dimensional shape to which neurons in MT respond, or do MT neurons respond only to the tilts of planes while STS neurons respond only to the three-dimensional shapes of objects? Perhaps the next few years will provide a full-scale attack on the coding of distance vs. depth in dorsal and ventral streams.

Finally, there is one report that some IT neurons are tuned for monocular depth cues such as shading (Ito, Fujita, Tamura, And Tanaka, 1994, Cerebral Cortex 4: 499). [DT must get.]

25.4 Summary: The "role" of an anatomical structure

In Chapter 21xx we learned that cortical area MT contains neurons tuned for the direction and speed of motion, and that microstimulation of area MT changes a monkey's psychophysical reports of the perceived direction of stimulus motion. Based on these data, we argued that MT has the right characteristics to play a major role in coding the direction and speed of stimulus motion. But now we also know that many MT neurons are tuned for binocular disparity, and that microstimulation of MT changes a monkey's reports of perceived distance (and/or depth). Given the new knowledge, and the same logic, we must also argue that MT is poised to play a major role in coding the distance (and/or depth) of visual stimuli. Thus, we must recharacterize our view of the "role" of area MT. Perhaps rather than coding motion per se, MT neurons respond to both shape from motion and shape from disparity, and are setting up a *cue-invariant* representation of the three-dimensional shapes of objects.

[Define cue invariance here if not earlier.]

The moral of the story is that we are still at the beginning of our characterization of the physiological properties of cortical neurons, particularly in the cortical processing areas beyond V1. Moreover, the data we do have on any area is largely dependent on the whims (or theories) of the individual investigators who have been drawn to study it, rather than on any systematic survey of the properties. We know, for example, that MT neurons are tuned for the tilts of planes, and that some TE neurons are tuned for binocular disparity; but we don't know whether or not the reverse is true. Thus, we should not be too confident in our current characterizations of the "roles" of anatomically and physiologically defined structures in the visual brain. We can expect that our descriptions of neurons and structures at all levels of the visual system will be forced to change as visual scientists confront them with new questions, new paradigms and new stimuli.

[Add here if not earlier]

Wandell (1995) has suggested a different way of putting the question. Rather than speaking of the "role" of an anatomical structure, he suggests that we ask, what computations are accomplished within this structure? This is probably a better (clearer) way of thinking.

[The stereograms in this version are too degraded to work.]

Chapter 26

The Perception of Objects

We now come to the most high-level visual function we will consider in this book: the perception and recognition of objects. This topic brings us back to the phenomenon we have called *object constancy*: the capacity to perceive an object as being the same object over a broad range of perspectives and viewing conditions.

From a nave realist's perspective, the perception of objects seems effortless. We readily and correctly recognize objects, and we do so across changes of many extrinsic variables, including retinal location, retinal size, and two-dimensional and (especially) three-dimensional rotation. I know that the teapot is a teapot regardless of its viewing distance or the viewpoint from which I view it. At the same time, we know that we can also perceive the orientations and distances of objects, and respond to them appropriately – I know that the teapot is within reaching distance, and that the handle is toward my hand. In short, our observations as perceiving organisms tell us that in the overwhelming majority of cases, visual perception of an object is *view-invariant* – constant over changes in three-dimensional viewpoint and other extrinsic variables.

From a sensory perspective, however, veridical object perception poses a formidable computational challenge. The challenge becomes apparent when we remember that the optics of the eye render a three-dimensional object as a two-dimensional retinal image. Moreover, as the object varies in its extrinsic properties location, size, and two- and three-dimensional rotation – it creates a wide range of different retinal images. We process the information in the retinal image through the retina and within the primary visual cortex, V1, ending up with a code in which local features (or local Fourier components) are coded in the firing rates of individual V1 neurons. But as extrinsic variables change, the pattern of activity in V1 changes drastically. For example, as the object changes its location, the image of the object falls on different retinal locations, and as it rotates in three-dimensional space, the shape of the image morphs through a series of major changes. Each different view of the object yields a different pattern of activity in V1 neurons. In the face of such a hodgepodge of incoming signals, how can our object perception possibly be view-invariant? And how can we also preserve information about its size, location, and orientation? Such system properties boggle the mind.

The outline of this chapter is as follows. First, we will briefly accept the nave realist's view – human observers have excellent object recognition – and explore a common working model of object recognition. Second, we will reconsider the nave realist's view, and ask whether the properties the nave realist takes for granted actually hold up in the laboratory. We will center our discussion around the question of whether object perception is really as view-invariant as it seems, and explore the conditions under which it does and does not occur. Third, we will look at the properties of cortical neurons in a particular part of the ventral visual stream, the inferotemporal (IT) cortex, and ask whether these neurons have the properties suggested by the system properties of object recognition. Fourth, we will briefly review some modern fMRI studies of object recognition, and explore the popular view that different classes of objects are represented in different cortical subregions. Finally, we make brief mention of the topic of object categorization, and relate it to object perception.

26.1 The common working model: An object sets up an invariant neural representation

Let us begin by adopting and exploring a *common working model* of visual object recognition. Lets assume that there exists within the visual system a stage of neural processing that is critical to object recognition. Lets also assume that signals from all of the different retinal images that arise from a fixed physical object converge at this stage of neural processing, in such a way that *all of the different views of a fixed physical object set up the same unique pattern of neural activity.* Each different object would set up its own unique pattern of neural activity. The recognition of an object would depend on the activation of the particular pattern of neural activity unique to this object, at this critical stage of neural processing. This unique pattern of neural activity would represent the object in abstract terms, devoid of information about size, location, or two- or three-dimensional orientation. The naive realist's belief about object recognition is shown in Figure 26.1A, and the common working model is shown in Figure 26.1B. The computational problem occasioned by the common working model is shown in Figure 26.1C.

26.1.1 Implicit assumptions and linking propositions

Notice that although the working model is a statement relating physical objects to neural states, it is occasioned by reasoning from perceptual states to neural states by means of an implicit linking proposition. The linking proposition is shown by the dashed line in Figure 26.1B. It is that the *perception of a fixed object suggests or implies the presence of a unique pattern of neural activity.* The linking proposition is a Converse one, xx, because the inference runs from the perception to the physiology. However, it is probably weaker than a Converse Identity proposition, such as we saw in trichromacy, in which we argued that metamers – perceptually identical stimuli – implied identical neural signals somewhere within the visual system. Why? Because the claim is not really that different views of a fixed object are indiscriminable – we know they are not – but just that they evoke the perception of a fixed object.

Digging a little deeper, many of us are probably attracted to the common working model because it allows the mapping from neural states to perceptual states to be 1:1. The alternative is a set of many:1 mappings between neural states and perceptual states. For example, if we assumed that the perception of objects arose directly from the pattern of activity in V1 neurons, we would be stuck with the fact that a large and disagreeably heterogeneous set of different neural patterns maps to the perception of one object, a second large set to the perception of a second object, and so on. In short, there is a trade-off between the complexity of the linking proposition and the required complexity of the neural processing. A many:1 mapping allows a simpler model of neural processing (V1 neurons will do); a 1:1 mapping necessitates a model that includes complex neural

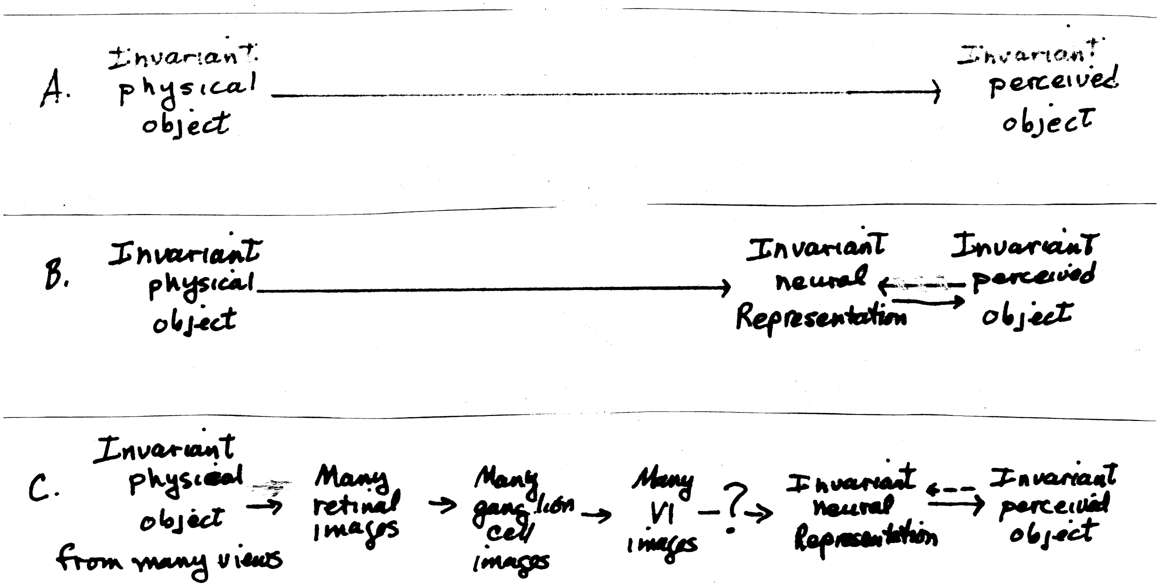

A.
Invariant physical object
Invariant perceived object
B.
Invariant physical object
Invariant neural Representation
Invariant perceived object
C.
Invariant physical object from many views
Many retinal images
Many ganglion cell images
Many V1 images
?
Invariant neural Representation
Invariant perceived object

Figure 26.1: The common working model. A: The nave realist's belief about object perception: a fixed (invariant) physical object gives rise to a fixed (invariant) perceived object. B. The common linking proposition (denoted by the dashed arrow): An invariant perceived object implies (or suggests) the existence of an invariant neural representation at some stage of processing. C. The problem of object perception. An invariant physical object can be seen from many views – many distances, many retinal locations, and many three-dimensional rotations (views). These different views give rise to very different retinal images, photoreceptor images, ganglion cell images, and V1 images. How, then, is the putative invariant neural representation of the object created?

processing (we need a unique and invariant neural code for each perceived object). The common working model presumes the latter, endorsing a simple 1:1 linking proposition, and thereby setting our appetites to find some interestingly complex neural processing at higher levels of the visual system.

A second important element in the common working model is that there will exist a critical processing level for object recognition, within which the neural code for each object will be relatively simple. This desired simplicity probably has two parts: the code should be both *sparse* and *direct*. As defined in Chapter xx, a code is *sparse* if only one cell, or a few cells, or a few hundred cells, are active in the representation of a single object, rather than most or all of the neurons in the relevant population. And a *direct* code, as defined by Horace Barlow, is a code in which "...the active high level neurons directly and simply cause the elements of our perceptions"; that is, a code that is based on simple and recognizable isomorphisms between the activity of individual neurons and the natural elements of perception. Although the choice of natural elements would have to be defended, one might assume for example that there exists a set of neurons that respond to a set of three-dimensional geometric forms, and that the representation of a complex object is made up of, or arises from, activity in the subset of these neurons that correspond to its geometric parts.

Notice, by the way, that regardless of the degree of sparseness or directness assumed, the original assumption of the common working model – that a fixed physical object will have a fixed neural code – implies that each neuron involved in the representation of a given object should have a constant response to that object across a wide range of extrinsic variables. If we are looking for the neural signals that represent an object, we will be looking for neurons with equivalence classes that encompass all views of that object, whether they are Grandmother cells or parts of small or large ensembles.

In sum, the common working model of object constancy stems from nave realism. It leads us to look for high level visual neurons with two novel and remarkable kinds of equivalence classes. First, we would like to find individual neurons that respond to the natural features or parts of objects (whatever they are). And second, we would like these neurons, or perhaps neurons at the next higher level, to give invariant responses to fixed objects across variations in extrinsic factors such as size, location, and viewpoint. The next two questions are, first, is nave realism right? And second, do such neurons exist?

26.2 A controversy: Is object perception really view-invariant?

We now return to the effects of three-dimensional rotations on object recognition. From the nave realists perspective, object perception is view-invariant – I recognize my rocking chair as readily from the back or side as from the front. But in fact, in the laboratory the effect of 3-D rotation on object recognition has been embroiled in controversy. Some authors claim that object recognition is *view-invariant* does not vary with object rotation – whereas others claim it is *view-dependent*, falling off rapidly with object rotation. Although the effects of three-dimensional rotation were initially cast as a critical experiment with deep and dichotomous theoretical implications (to be discussed later), it now seems likely that either outcome can be obtained, depending on the choice of stimuli.

26.2.1 The visual priming paradigm: Evidence for view-invariance?

One of the paradigms used in studies of object perception is that of *visual priming*: the speeding of perception of one visual stimulus after prior exposure to another. Priming is interesting because it suggests that somewhere within the visual system, the first exposure to a picture speeds the physiological processing of the second. Moreover, it turns out that this facilitation generalizes across some changes in the picture but not others. The pattern of generalization is in some respects counterintuitive, and some scientists argue that it provides a tool with which we can isolate and explore one of the processing stages involved in object perception.

A typical priming experiment has two phases, the *priming phase* and the *test phase*, with the two phases classically separated by several minutes. In the priming phase a subject is shown a series of (say) pictures of objects. The subjects task is to name the pictures as quickly and accurately as possible for example, to say shoe when presented with a picture of a shoe. The subjects reaction times and error rates are recorded. Then, in the test phase, the subject is again tested with a variety of pictures, some identical to the priming stimuli and others differing from them in various ways. Again the subject is asked to name the pictures, and again reaction times and error rates are recorded.

Two basic phenomena emerge from visual priming studies. First, for pictures that are identical in priming and test phases, priming occurs. Reaction times are lower in the test phase than in the priming phase, sometimes by as much as about 100 msec. But second, priming is more effective between two identical pictures than between pictures of two *exemplars* (examples) from the same category of object, such as a high-heeled shoe and a walking shoe. Even though subjects spontaneously call each of them by the same verbal label shoe– there is a larger reduction in reaction time if the visual picture stays the same.

Now, the next interesting question is, what will be the pattern of generalization of priming effects? Over what range of stimulus variations will priming be effective? If one is used to thinking in terms of retinal images, one might expect that only stimuli that make the identical retinal image would prime each other. But remarkably, this is not the case.

Classic examples of priming and its range of invariance were provided by a series of papers by Irving Biederman and his colleagues in the early 1990s. For example, an experiment by Biederman and Cooper (1992) both illustrates the basic priming effect and examines the generalization of priming across exemplars and across retinal image size. The stimuli from Biederman and Cooper's experiment are shown in Figure 26.2A. For example, in the priming phase the subject might see a grand piano. In the test phase, the subject would be shown a grand piano of the same or different size, or an upright piano, also of the same or a different size.

Reaction time data from Biederman and Coopers experiment are shown in Figure 26.2B. The leftmost bar of the graph shows reaction times from the priming phase of the experiment. The other four bars show the data from the test phase, for four different combinations of test conditions – same exemplar, same size; same exemplar, different size; different exemplar, same size; and different exemplar, different size. For all four test conditions, reaction times are reduced by about 100 msec in the test phase – priming occurs. But reaction times are reduced more in the same-exemplar than in the different-exemplar conditions – as discussed above, changing the exemplar interferes with the priming effect. In contrast, the size of the test stimulus doesn't matter – priming generalizes fully across variations in retinal image size! And other studies by the same investigators show that priming generalizes across variations of retinal location and left/right (mirror image) reversal of

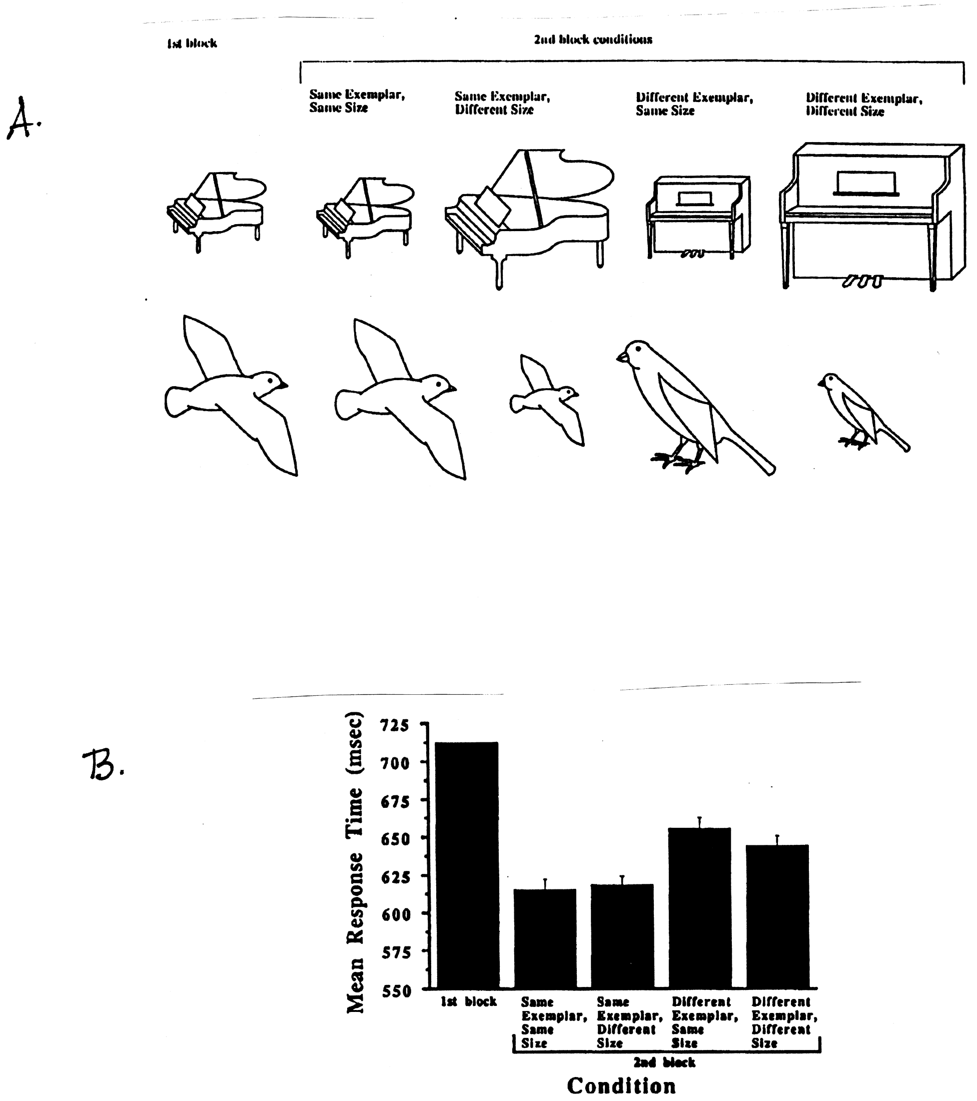


Figure 26.2: The priming experiment of Biederman and Cooper (1992). A: Examples of the stimuli. The left column shows stimuli used in the first (priming) block. The right four columns show stimuli used in the second (test) block. Stimuli that are the same exemplar, same size; same exemplar, different size; different exemplar, same size; and different exemplar, different size, are shown in the four right columns respectively. B: Reaction time (response time) data from the same experiment. The left bar shows response time for the first block of trials. The right four bars show response times for the four kinds of exemplars. Response times are shorter for all stimuli on the second block – this is the general priming effect. In addition, response times are shorter for the two same-exemplar conditions than for the two different-exemplar conditions. Priming is more effective for the same than for a different exemplar, but is invariant across changes of size. [From Biederman and Cooper, 1992, A: p. 125; B: p. 126.]

the stimulus as well as retinal image size.

In addition, Biederman and his colleagues have shown that priming generalizes across specific object contours and across spatial frequency bands. A stimulus composed of half of the contours in the picture will prime a stimulus composed of the other half of the contours, and an object depicted with a band of low spatial frequencies will prime the same object depicted with a non-overlapping band of higher spatial frequencies. In sum, priming is tied neither to the size and location of the retinal image nor to the activity of particular V1 neurons.

Most importantly, what about viewpoint? Biederman and Gerhardstein (1993) studied the effects of priming on the perception of pictures of familiar objects such as desk lamps and flashlights. After priming by such objects, the stimuli in the test phase were pictures of the same or different exemplars of these objects, rotated by 0, 67.5 or 135^o, in such a way that the same parts of the object remained visible in all rotations. They found the typical effect of exemplars – a change of exemplar reduced the magnitude of priming. But as long as the same parts of the object remained visible, priming was nearly view-invariant – rotation in the third dimension had little effect on the magnitude of priming.

Now, what do the system properties revealed by visual priming suggest about the visual system? First, if the signals initiated by the priming and test stimuli are to interact if one stimulus is to prime another some aspect of the two signals must both be present the same neural locus at a single point in time. Since the priming stimulus is over long before the test stimulus is presented, the priming stimulus must leave a *trace* a change in the state of the visual system that persists for at least several minutes after the offset of the priming stimulus, and through the presentation of all intervening stimuli. Second, the test stimulus must access the neural circuit whose processing is affected by the presence of the trace, and have its processing speeded thereby. And third, the processing of many different images of the same object, even those originating in different retinal locations, traversing different V1 neurons, and viewed from different viewpoints, will all be speeded by the trace of the priming stimulus.

At the systems level, it is only a short step to the suggestion that ”.priming depends on a perceptual representation systeminvolved in processing information about form and structure, but not the meaning and associative properties, of words and objects.” (Schachter and Buckner (1998, p. 188). That is, consistent with the common working model, perhaps at some level of the visual system each individual object is represented by a perceptual representation that is specific to that object but invariant across size, location, spatial frequency and viewpoint. And perhaps it is the access to these representations that is speeded by the priming stimulus. Different exemplars are not primed because they activate different visual representations.

At the neural level, the properties of visual priming suggest the existence of a neural representation that is abstracted cut loose – from retinal image variables. Given priming between two stimuli, we would not be surprised to find a neuron or set of neurons at high levels of the visual system that gave the same (or highly similar) responses to the two stimuli. Signals initiated by all of the retinal images that would be produced by the same object, or more precisely, signals initiated by all of the retinal images that will lead to the *perception* of the same object, would produce the same activity pattern in the same set of neurons. Moreover, this high-level representation would be a natural substrate for object perception.

If the priming paradigm does isolate and give access to a high-level perceptual object representation, it would be fun to explore more of the properties of this representation. For example, does priming generalize across color? Does it generalize across three-dimensional shape cues – does an object defined by binocular disparity prime the same figure defined by shading or motion? Does it generalize across figure-ground reversal? How much dissimilarity is required for priming to break down? Doesit break down with illusory changes of shape or size? Does it show plasticity with stimulus exposure? Can it be influenced by instructions?

26.2.2 Same-different judgments: Evidence for view-dependence?

But other laboratories approach the problem from different perspectives and find different answers. For example, in the mid1990s N. K. Logothetis, and his colleagues (Logothetis, Pauls, Bulthoff, and Poggio, 1994; Logothetis and Pauls, 1995) undertook a set of studies of the effects of viewpoint, using macaque monkeys as subjects. These studies are exemplary in that three different kinds of test stimuli were used, so the effects of stimulus variables could be examined. Moreover, single unit recording was carried out on the same monkeys with the same stimuli, sometimes during the actual behavioral testing, so that detailed comparisons between psychophysical and physiological data could be made. (The single unit data will be reviewed below in the section on IT cortex.)

Examples of the three kinds of stimuli used by Logothetis and his colleagues are shown in Figure 26.3A. In each case one test object is shown in the left box of the figure. The first row shows a wire like test object (sometimes called a bent paper clip), in five different rotations about its vertical axis. The monkeys basic task was to discriminate this wire object, in all of its rotational positions, from a set of *distractor* objects. At the right is an example from the set of distractors a view of one of many different but highly similar wire objects. The second row shows a second kind of test object, which the authors call a spheroidal, or amoeba-like object, again in five rotational views, together with an example from the set of spheroidal distractors. And the third row shows a test object selected from the human world a teapot in five views, together with an example a space ship – from the set of other human-world objects used as distractors. One striking thing to notice is the apparent high level of difficulty of the discrimination between test objects and distractors in the first two sets. Objects in the third set bear a much closer resemblance to those used by Biederman and his colleagues, and are readily discriminable from each other (at least for human subjects)[1].

All stimuli were presented on a video monitor. Shading was used to provide a three-dimensional cue; and in addition, during training the stimuli were rotated by means of an animation sequence. In consequence the monkey could use both motion and shading cues to perceive the three-dimensional structure of the (virtual) objects.

In a prolonged training period that lasted many months, the monkeys were trained in a *same/different task*: to pull one lever when they saw a single target (say, a particular wire object at a fixed rotational view) and another lever when they saw any of a large number of distractors. They were then trained to respond positively to the target object presented at increasingly large rotations from the initially trained view, while continuing to respond negatively to the distractors. When the monkeys had mastered this task with one target, they were trained with other similar targets, and eventually with targets from the other two classes (say, spheroidal objects and human-world objects). At this point, the experiment proper could begin.

[1] The wire-like and amoeba-like objects are sometimes called *novel* objects, and the human-world objects are sometimes called *familiar* objects. This nomenclature arose in the literature on human subjects, for which these names might be appropriate. However, as described below, the monkeys practiced for months with the objects, so none are really novel; and of course the objects from the human world were not initially familiar to the monkeys.

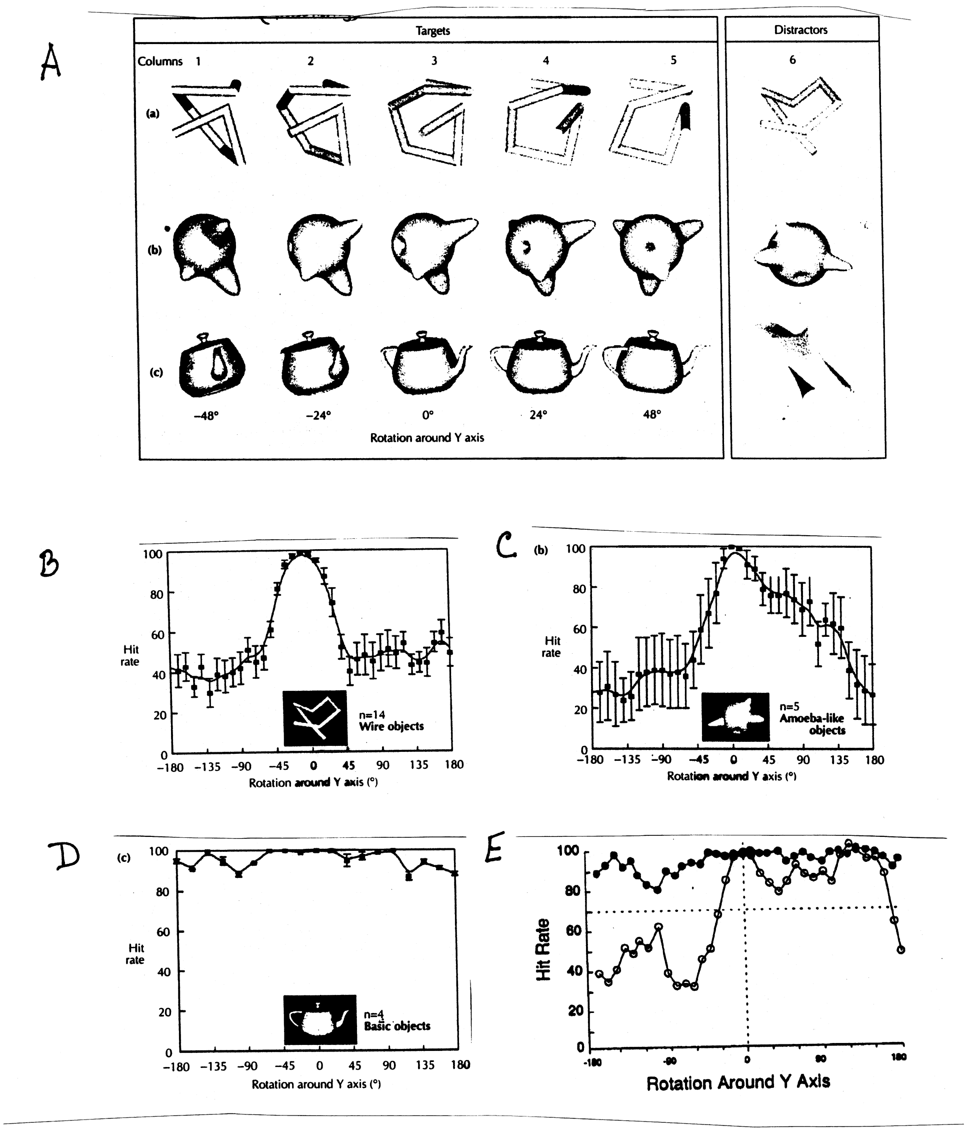


Figure 26.3: The behavioral experiments of Logothetis et al (1994), and Logothetis and Pauls (1995) on monkey subjects. A: The stimuli used. The first row shows the wire objects; the second row shows the amoeboid objects; and the third row shows the common (basic) objects. B-E: The results. For wire objects (B) and amoebas (C), performance fell off dramatically with rotations of ±45^o. For common (basic) objects (D), performance was view-invariant over the full 360^o of rotation. E: Effects of training. Results after training with two views (0 and 120^o, open circles) of a wire object, and then a third view (60^o, closed circles). After training with three views, the monkey's performance was completely view-invariant, even for this difficult target. Both the characteristics of the objects and distractors used, and the amount of prior training or familiarity with them, affect the amount of view-invariance seen. [A-D from Logothetis et al, 1994. A: Fig. 1, p. 402; B-D: Fig. 5, p. 405. E from Logothetis and Pauls, 1995, Fig. 6, p. 278.]

The experiment proper was divided into a repeated sequence of *learning* and *test phases.* During each learning phase, the monkey viewed a target object from a single view for about 2 seconds, and presumably recognized that it was one of many familiar objects. The learning phase was followed by a sequence of ten test trials. On each trial, either the target object was presented at one of several rotations away from that shown in the learning phase, or one of the distractor objects was presented.

The results are shown in Figure 26.3B-D. Panels B, C, and D show the data for the wire objects, amoebas, and human-world objects respectively. Notice that both view-dependence and view-invariance are seen, depending on the choice of test and distractor objects. For wire-like objects with wire-like distractors, and spheroidal objects with spheroidal distractors, the monkeys recognition performance clearly diminished rapidly with the angle of rotation of the test stimuli away from the trained view – view-dependent behavior was seen. But for human-world objects with highly discriminable distractors, performance remained near 100% across the whole range of object rotations the monkeys displayed near-perfect view-invariance. [Do you think it was the abstract vs. human-world nature of the stimuli, or the similarity vs. dissimilarity between the targets and the distractors, that made the difference? What would happen if both the targets and the distractors were teapots, or if the targets were wire objects and the distractors were amoebas?]

In a second paper using the same monkey subjects, Logothetis and Pauls (1995) report additional behavioral data. They showed that even with wire figures or amoebas, if the monkey were trained with several views of the objects, the monkeys behavior became view-invariant. Typically three to five views of an object, rotated around a vertical axis, were sufficient to produce view-invariant recognition of the object for all rotations around the vertical axis. Data of this kind are presented in Figure 26.3E (see also Figure 26.8C).

26.2.3 What experimental conditions create view-invariance?

What shall we make of the apparently discrepant results from the two different traditions? First, we might note that the differences in results are not as great as the authors sometimes make them out to be. Biederman and his colleagues emphasize their findings of view-invariance, but limit the prediction of view-invariance to cases in which the same object parts remain visible. But on average, the larger the rotation the less likely that parts will not be occluded, so on average, one might expect that the larger the rotation the less the view-invariance would be. Simiiarly, Logothetis and his colleagues emphasize their findings of view-dependence, but concede that view-invariance occurs for their set of human-centered objects and also for their wire objects and amoebas after practice with several views.

Let's use logic and common sense, along with the data, to speculate as to the characteristics of objects that lead to view-invariance and view-dependence. Let's begin by remembering that, in combination with the incoming signal, the visual system employs heuristics. And let's take the perspective that maybe the cases in which reasonable heuristics might be meployed will be cases in which view-invariance is seen.

First, perhaps view-invariance will prevail when the shapes of the objects are simple; for example, when they are made up of parts with simple shapes and/or have axes of symmetry. A heuristic that made guesses based on such properties could be stored in the visual system. In contrast, when

The objects are best described by their physical properties, and not as novel or familiar.

stimuli are created by purposely making random twists in a wire, or adding random blobs to a sphere, no heuristic is possible, and view-dependence seems likely.

A second factor might be the use of real three-dimensional objects, and the presence of a rich variety of consistent cues to the three-dimensional shape of the object. In fact, in most of the studies in this field two-dimensional drawings rather than three-dimensional objects are used as stimuli, yet the perception of three dimensions is at stake. That is, contradictory cues – for example, from binocular disparity – were present in the stimuli. In the real world we have many cues, and they are usually mutually consistent; so our initial intuition that we have view-invariant object perception may still be correct.

As agreed by all, a third factor is the familiarity of the objects. It really is possible that a truly novel object has no highly developed perceptual representation, and that time and/or practice are required before three-dimensional view-invariance is achieved. [But seems unlikely from a depth-perception-from-multiple-cues perspective.] THINK.

A fourth probable factor is the perceptual context – the nature of the other objects from which an object must be discriminated. In ordinary perception, when I perceive my rocking chair, I am aided greatly by what might be called a continuity heuristic – if it's in my living room, and it's a rocking chair, it's probably my rocking chair. At the other extreme, when a wire object must be discriminated from another wire object, the task is difficult and perhaps no general guessing rule can be of any help. [The usual case is to perceive a set of quite different objects – view-dependence may be a lab phenomenon, occurring only under highly artificial conditions.] THINK

Finally, some of the differences between studies could be due to differences in tasks – for example, priming vs. same/different judgments. There are a few studies that suggest that different tasks can yield different degrees of view-invariance, but the problem is little studied.

As we will see, the controversy over view-invariance was occasioned by, and engendered, continuing differences in theoretical perspectives and models of object perception. It has been argued, for example, that view-dependence is a signature for the existence of feature-based stages of processing and view-dependent object representations, and view-invariance a signature for three-dimensional parts-based stages of processing and view-invariant object representations. But in fact, both kinds of behavior occur, and we will need a model consistent with both. In the meantime, we turn to neurophysiology.

26.3 Inferotemporal (IT) cortex and its role in object recognition

It is commonly believed that a particular part of the visual cortex, the inferotemporal (IT) cortex, is heavily involved in the processing required for object recognition. This belief is supported by several kinds of evidence. First, IT cortex has inputs from earlier visual processing areas, along the ventral processing stream (Figure xx in earlier chapter). Second, lesions of IT cortex interfere with object recognition but have little effect on low level visual capacities. And third and most strikingly, many IT neurons respond selectively to complex visual patterns. Qualitatively, IT cortex has the right machinery to play a major role pattern recognition.

26.3.1 Anatomy of IT cortex

The inferotemporal cortex lies in the lower part of the temporal lobe, as shown in Figure 26.4. With single unit recording, the most striking pattern specializations are seen in the anterior 2/3 of

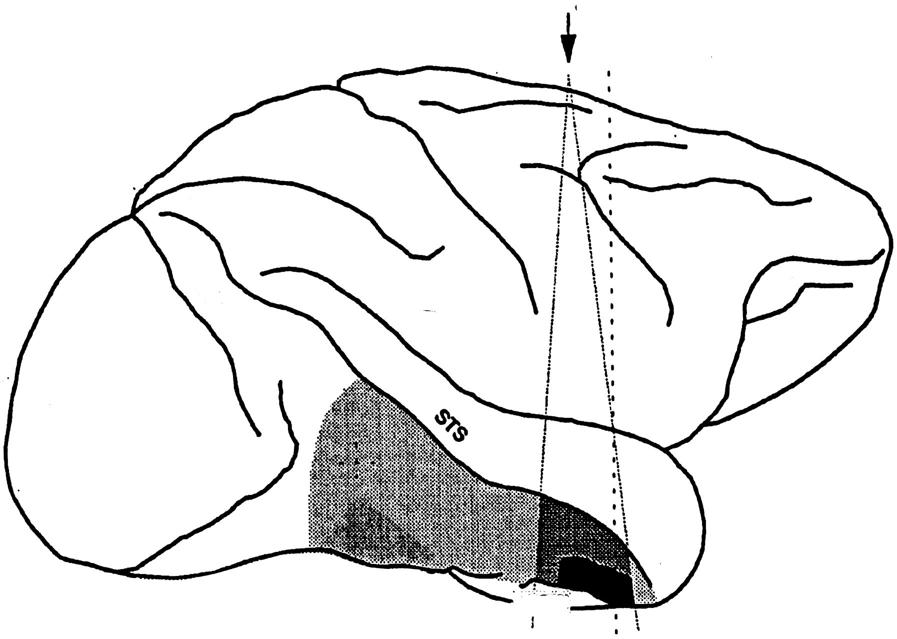


Figure 26.4: Anatomical location of inferotemporal (IT) cortex in the macaque monkey. The light grey shading shows the approximate location of IT cortex. Cells that respond to complex visual patterns and objects are most often found in the anterior 2/3 of this region. Face cells are most common closer to and within the superior temporal sulcus (STS). The black area shows the area recorded from by Logothetis and Pauls (1995). [Adapted from Logothetis and Pauls, 1995, Fig. 1, p. 272.]

IT cortex – the *anterior IT cortex*, or *aIT*. In the following discussion, most of the data we report are taken from aIT.

26.3.2 Many IT neurons are selective for complex image features

The modern story of IT cortex begins with the work of Charles Gross and his colleagues in the early 1970s. By this time Hubel and Wiesels discovery of simple and complex cells in V1 was well known. Hubel and Wiesels hierarchical processing model provided a scheme for the creation of neurons selective for increasingly specific aspects of the stimulus. Simple cells could be created by summing inputs from a row of cells with center-surround receptive fields, and complex cells could be created by combining inputs from simple cells across a range of locations. It was tempting to speculate that the visual system might continue such a hierarchical processing scheme, creating neurons that could respond to increasingly more specific visual stimuli at higher and higher levels of processing.

In the early 1970s, Charles Gross and his colleagues (Gross, Rocha-Miranda, and Bender, 1972) began to explore the properties of the neurons in IT cortex. Their stimulus set-up was a translucent screen illuminated from behind. Incremental (light) stimuli could be presented with a projector, and decremental (dark) stimuli could be created by placing black cut-outs against the back of the screen. One of their first discoveries was that the receptive fields of IT neurons were very large anywhere from about 10 x 10 to about 50 x 50^o – and almost always included the fovea. When a stimulus was discovered that excited the cell, it did so over the whole receptive field of the cell, providing an analogy to the fact that we can recognize objects across a broad range of retinal locations.

Following Hubel and Wiesels lead, Gross and his colleagues initially tested IT cells with a range of relatively simple stimuli dark and light moving and stationary bars of various colors. IT neurons did not respond well to these simple stimuli, and it gradually became apparent that more complex stimulus patterns yielded more vigorous responses. In particular, Gross and his colleagues stumbled upon an IT neuron that responded best to the shadow of a monkey hand. As they reported, The use of the (monkey hand) was begun one day when, having failed to drive a unit with any light stimulus, we waved a hand at the stimulus screen and elicited a very vigorous response from the previously unresponsive neuron. We then spent the next 12 hr testing various paper cutouts in an attempt the find the trigger feature for this unit. When the entire set of stimuli used were ranked according to the strength of the response that they produced, we could not find a simple physical dimension that correlated with this rank order. However, the rank order of adequate stimuli did correlate with similarity (for us) to the shadow of the monkey hand. (Gross et al, pp. 103-104). The responses of this neuron are shown in Figure 26.5A. This remarkable finding was the first indication that IT cells might respond selectively to complex visual patterns.

”Face cells” – neurons that respond vigorously to faces, and much less well to all other stimuli – were soon identified by several laboratories, and generated much interest and excitement. An example of the response of an IT cell to a variety of face and non-face stimuli is shown in Figure 26.5B. The abscissa of this figure represents the series of stimuli used, but ordered by the magnitude of the neurons response. Stimuli labeled P are profiles of faces; stimuli labeled F are frontal views of faces; B stands for body parts; and non-labeled stimuli are other objects. Clearly, this neuron responds more vigorously to most face stimuli than to most non-face stimuli, and in this sense activity in this neuron can be said to signal the presence of a face with high probability. In some subregions of IT cortex, as many as 20% of the neurons are selective for faces.

The properties of IT neurons were further characterized by Keiji Tanaka and his colleagues in the 1990s (Tanaka, 1996). The approach of this group is to search for the stimulus that elicits the highest firing rate in eachindividual neuron. Each neuron is tested with a heterogeneous set of two- and three-dimensional objects, including lines, spots, geometric patterns, laboratory tools, three-dimensional models of plants and animals, and the experimenters hands, body and face. From this large range of stimuli, the stimulus that elicits the highest firing rate from the cell is selected for further study. This stimulus is then presented on a video monitor, and *reduced* –simplified in various ways – in order to search out the simplest stimulus feature or combination of features that continues to excite the cell.

Examples of two IT neurons and the ranges of stimuli that excited them are shown in Figure 26.6. In the first case, the neuron responded to stimuli composed of a large black disk with a smaller rounded protruberance situated at the upper left. Other subparts and configurations of the pattern the disk alone, or the protruberance alone, or changes in the configuration of the two

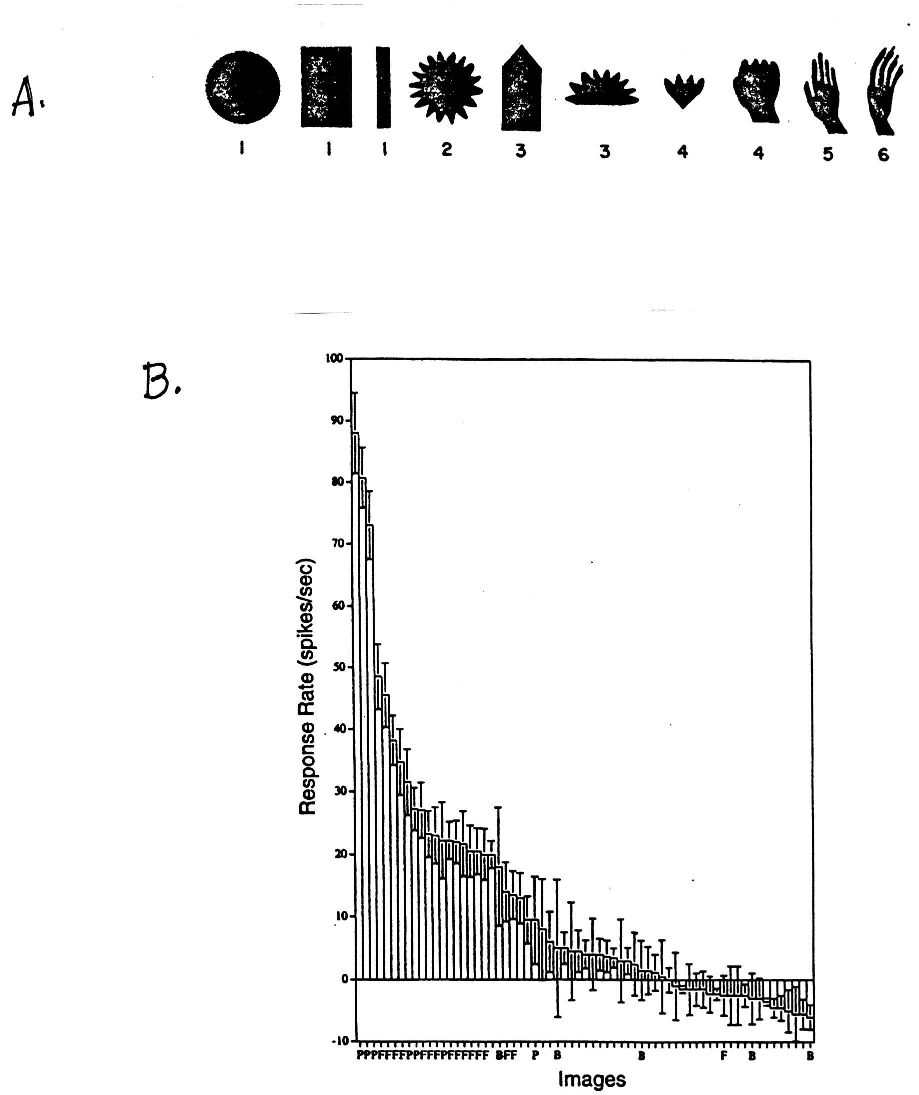


Figure 26.5: Responses of neurons highly selective for hands and for faces. A: The famous monkey-hand cell from Gross et al (1972). The stimuli are ordered by the degree to which the excited the cell, and the numbers are the experimenters' ratings of the responses. B: A face cell. The abscissa shows a set of stimuli, ordered by the degree to which they excite the cell. P = profile face; F = front view of face; B = body part; non-labelled: other pictures. The cell responds to many faces. On average, it responds much more to faces than to non-face stimuli. [A from Gross et al, 1972, p. 104; B from Rolls, 2000, p. 207.]

parts – did not excite the cell. In the second case, the cell responded to the face of a toy monkey. It also responded to simplified versions of the face that included both eye and mouth features, but did not respond unless both the eye-like and mouth-like parts were present.

Tanaka and his colleagues have studied many hundreds of IT neurons, and found a wide variety of forms of stimulus selectivity and invariance. For some of these neurons, shape was the only critical feature, whereas others were tuned to particular combinations of shape, color, and texture.

Tanaka and his colleagues have also studied the correlations between the responses of IT neurons in local neighborhoods in IT cortex. They showed that near neighbors tend to respond to the same or overlapping sets of stimuli. These and more recent optical imaging studies have suggested that IT cortex has a patchy or columnar structure, like that seen in V1 and MT.

What can we conclude from these observations? First, we have clearly encountered another major transformation of the form of the visual code. At the level of IT cortex, the visual system is no longer coding the retinal input in terms of the retinal locations of edges or local Fourier components. The responses of IT neurons are specific to moderately complex stimulus features or combinations of features, over large retinal regions.

Second, how might this code transformation come about? Although less is known about neurons in the intermediate stages of processing (V2, V3, and the posterior IT cortex), we can presume that some of the code changes take place there. At the same time, the columnar structure of IT cortex suggests that further hierarchical processing may be going on within IT cortex itself. Based on the principle that neurons engaged in intense joint computations should be located near each other, the columnar organization suggests that indeed, intense computations about objects and their features is being performed in IT.

Third, what do these observations tell us about the form of the code? Clearly, IT neurons are not so specific that they respond only to a single face or object they are not grandmother cells. Tanaka and his colleagues suggest that .simultaneous activation of a few to a few tens of cells is required to indicate the concept of a particular object.Images of objects are thus coded by combinations of active cells each of which represents the presence of a particular partial feature. (Tanaka et al, 1991, p. 187). In other words, the speculation is that in IT cortex objects are represented by a sparse population code. However, the neurons encountered are so heterogeneous that to date it is impossible to guess the nature of the primitives, or to determine the directness of the code.

26.3.3 Some IT neurons are location-, size-, and/or cue-invariant

The next question is, do IT neurons have the invariances we might expect, given the common working model of object recognition? The answer is that only some IT neurons show the appropriate invariances, but some show them to a remarkable degree. Suffice it to say that, in parallel to priming studies, many IT neurons give relatively invariant responses over variations of retinal image size, retinal image location, rotations in the two-dimensional plane, and mirror image reversals. Studies of two more invariances cue-invariance and spatial frequency invariance – will be discussed in a little more detail.

Cue-invariance

Are the responses of IT neurons specific to particular shapes, and invariant across the cues used to define these shapes? Sary, Vogels, and Orban (1993) tested IT neurons with spatial patterns

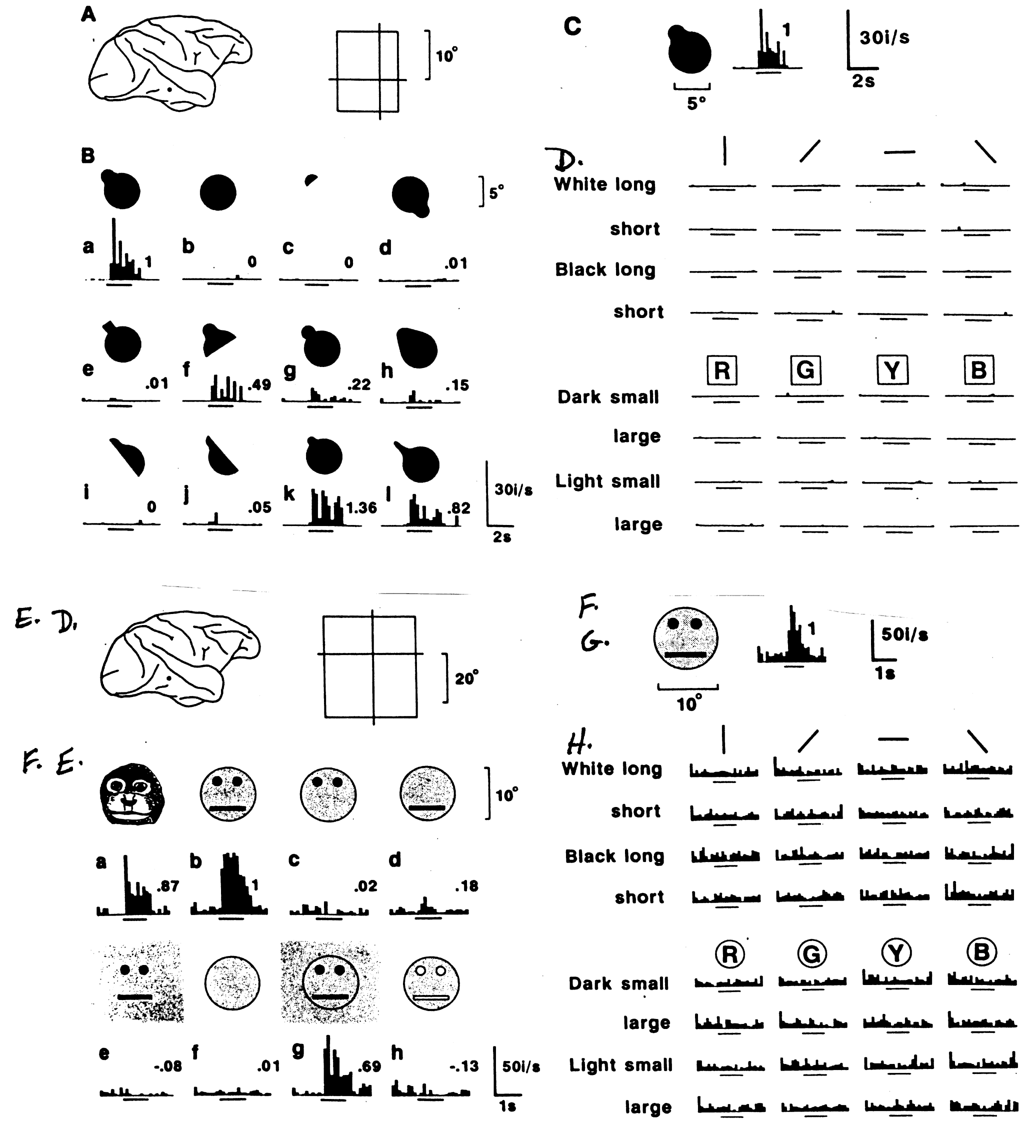


Figure 26.6: Responses of IT neurons recorded by Kobataki and Tanaka, 1994. Two different neurons are shown in the upper and lower parts of the figure. A, D: Recording sites. B, E. Responses of the neuron to optimal stimuli and stimulus parts. C, F. Stimulus size and calibration data. D, H. Responses of these neurons to a variety of lines and simple chromatic stimuli. The neurons respond only to the complex stimuli. [From Kobataki and Tanaka, 1994, Fig. 3, 858; Fig 4, p. 859.]

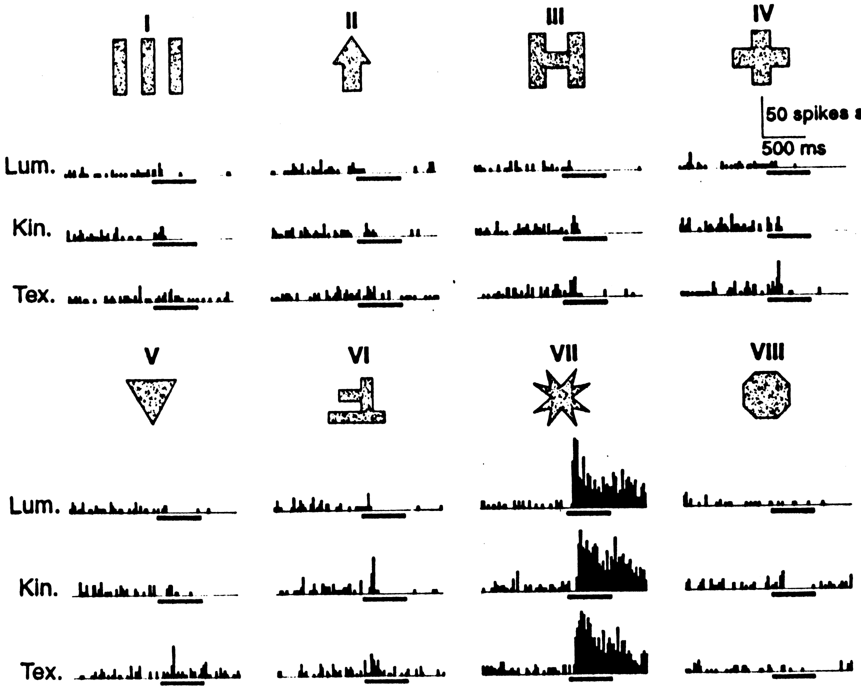


Figure 26.7: Cue invariant shape selectivity in an IT neuron, from Sary et al, 1993. The forms labeled I-VII show the shapes that were presented to the neurons. Each shape was defined by three different form cues – a luminance difference Lum.), a kinetic (motion) difference (Kin.), and a texture difference (Tex.). The neuron responded to the star (shape VII) defined by all three form cues, and did not respond to any of the other stimuli. [From Sari et al, 1993, p.996.]

wose shapes were defined by three different cues: a luminance difference, a difference in texture, or a difference in relative motion. Eight different two-dimensional shapes were defined by each of the three cues. A population of neurons was selected that responded to all three cues. Within this population, the responses of neurons were strongly tied to the particular shape of the stimulus. The shape selectivity was largely invariant across variations in stimulus size and retinal location. Most interestingly, the shape selectivity was also largely invariant across the choice of cue. Sample results from Sary et als study are shown in Figure 26.7. Neurons like these neurons seem to abstract the two-dimensional shape properties of visual patterns across a remarkably wide range of stimulus variations.

Spatial frequency invariance

In a particularly interesting study of cue invariance, Rolls, Baylis and Leonard (1985) studied the responses of face-selective neurons to faces that had been low- and high-pass filtered. They found that many neurons would respond both to low-pass and to high-pass-filtered images of faces, even though the low-pass and high-pass versions of the face had no spatial frequency components in common. At the same time, these neurons did not respond to simple gratings of any spatial frequency. The important stimulus characteristic is apparently a constellation of spatial frequencies at a set of orientations and phases that arises from faces, and not constellations arising from other objects. This invariance is particularly interesting because it suggests that signals arising from the same face, but traversing entirely different sets of V1 neurons, still converge on the same IT neuron.

26.3.4 A few IT neurons are view-invariant

Are there neurons in IT cortex that respond to an object across all viewpoints across all of the retinal images that result from rotations in the third dimension? Interestingly, two different studies seem to yield different answers to this question.

As discussed above (Figure 26.3), Logothetis and his colleagues carried out a behavioral study of view-invariance in two monkey subjects, finding view-invariance under some conditions and view-dependence under others. At the end of behavioral testing, Logothetis and Pauls (1995) also recorded from neurons in IT cortex in the same two monkeys. They found a small population of IT neurons about 9% of their sample – that responded to the objects with which the monkeys had been trained. Moreover, most of these neurons were view-dependent they responded to individual views of the objects the monkeys had been trained to recognize, and not to other views of the same object. In addition to the view-dependent neurons, Logothetis and Pauls also found eight neurons about 1% of the neurons tested that showed view-invariant responses to a single target object.

On three occasions, Logothetis and Pauls were lucky enough to record from several neurons that were all tuned to views of the same object. An example is shown in Figure 26.8. In this case, the four neurons responded to four views of a wire object, separated by an average of about 60 degrees. At the same time, the monkey showed view-invariant behavioral responding, correctly signaling same for all views of the object. Setting aside the view-invariant neurons, the clear implicit message of Figure 26.8 is that a set of view-dependent neurons would be sufficient to provide a coarse code for the view-invariant recognition of the object.

A more recent study by Michael Booth and Edmund Rolls (1998) addresses the same question. Booth and Rolls began by choosing two sets of ten (human) everyday objects, and placing them in the home cages of two different monkeys, so that the monkeys could play with the objects for several weeks before the start of single unit recording. During recording, four different views of each of the objects were used as test stimuli. Figure 26.9A and B show the sets of forty stimuli (ten objects x four views) that were presented to each monkey during recording.

Booth and Rolls found that over half of the IT neurons responded to one or more views of one or more of the objects that had been placed in their home cages. But importantly, a few neurons (14% of the neurons reported) showed relatively view-invariant responses to one or a few of these stimuli. Figure 26.9C-F show the responses of four neurons to each of the 40 stimuli. The neuron in Figure 26.9C consistently responded more strongly to object #9 an apple corer than to any other object, and the neuron in Figure 26.9F responded relatively invariantly to two of the ten objects. Such consistent responses occurred even though there was much similarity across particular specific

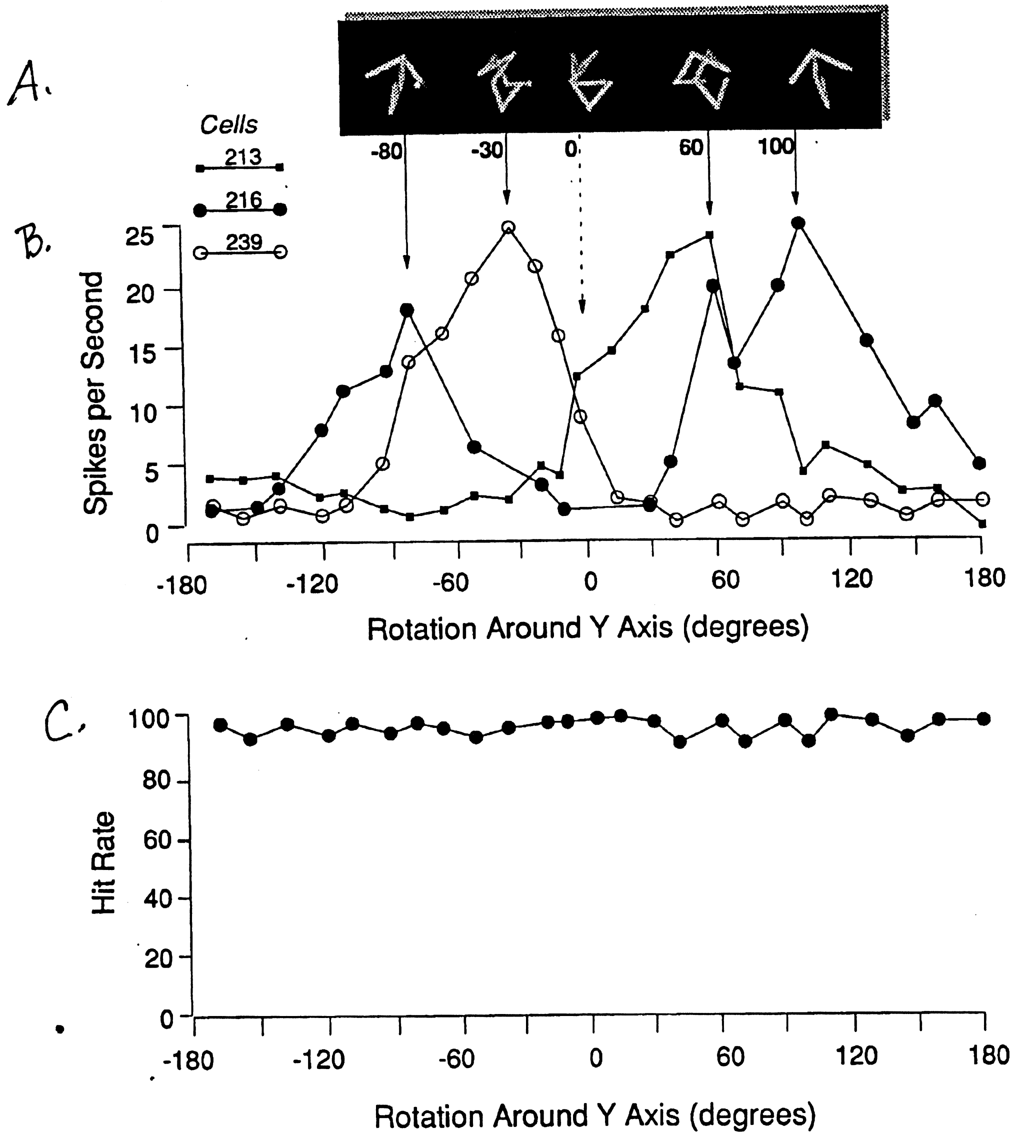


Figure 26.8: An example of four neurons tuned to different views of the same object. A: Five views of a wire object. B: The responses of four IT neurons to these objects. Different neurons give view-dependent responses across different ranges of rotation. C: The behavioral responses of the monkey in discriminating this wire object from other wire objects. The monkey had been trained with several views. The monkey's performance was view-invariant. The argument implicit in the figure is that the view-invariant response could be generated by a combination of activity in the set of view-dependent neurons. [From Logothetis and Pauls, 1995, Fig. 14, p. 286.]

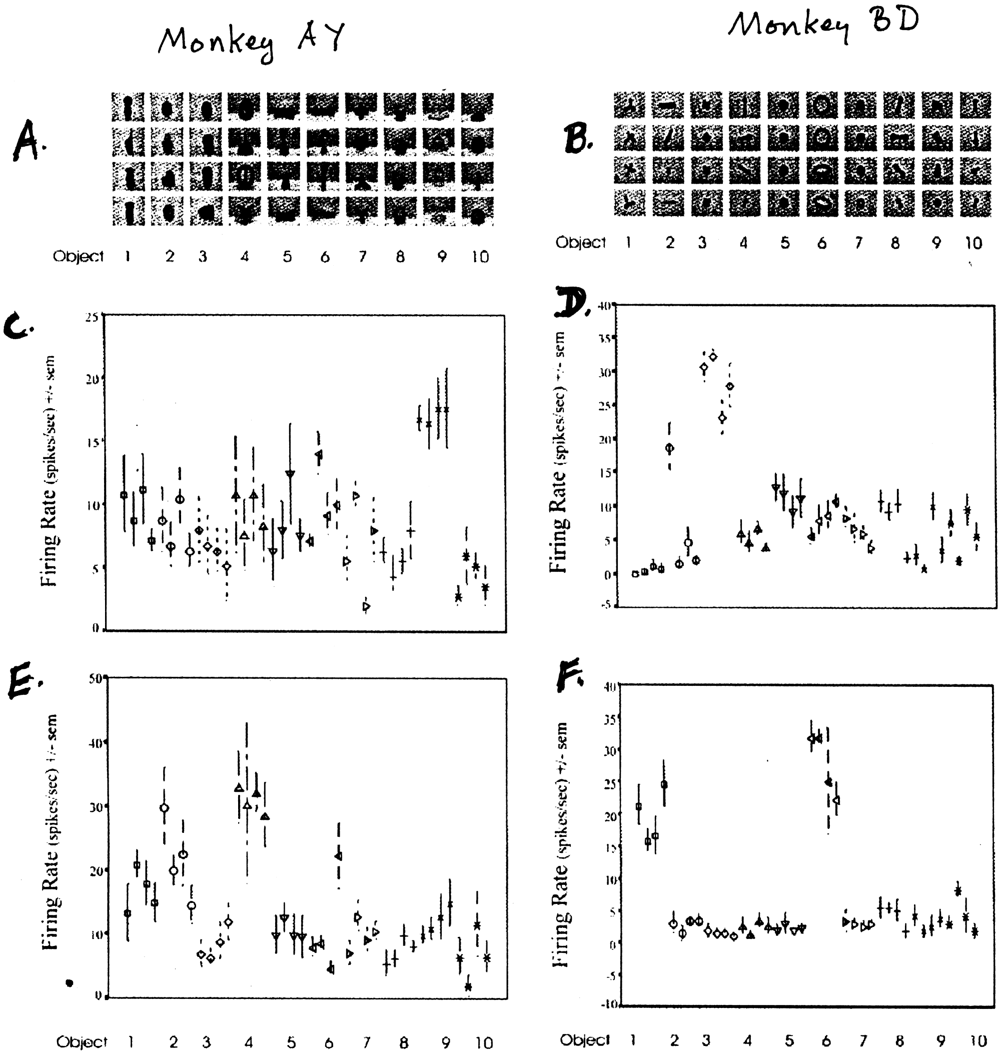


Figure 26.9: Examples of neurons with approximate view-invariance. The left and right columns show experiments on two different monkeys. A, B: Four views of each of the 10 common objects with which each monkey had home-cage experience. The objects are numbered 1-10 for each animal. C-F: Responses of individual neurons tested with the four views of each object. The objects are indicated with numbers along the abscissae; the four views of each object are plotted in sequence. C shows a neuron that responded strongly and consistently only to object #9 (the apple corer). F shows a neuron that responded strongly but less consistently to each of two objects. D and E show intermediate cases. [From Booth and Rolls, 1998, p. 514.]

views of several objects. Booth and Rolls argue that these nearly view-invariant neurons code the presence of particular objects.

Now, notice that in both of these studies, both view-dependent and view-invariant neurons were found. The biggest difference is in numbers – about 1% of the neurons reported by Logothetis and Pauls, and about 14% of those reported by Booth and Rolls, were view-invariant. As we have seen, Logothetis and Pauls emphasize the view-dependent neurons, and argue that the joint activations of view-dependent neurons provide the neural substrate for view-invariant object recognition. In contrast, Booth and Rolls emphasize the view-invariant neurons, and argue that they do the job. We have here an interesting case in which the conclusions favored by the two sets of scientists differ more than the experimental data.

26.3.5 Plasticity on IT neurons

A final and troublingly complex aspect of IT neurons is their plasticity. It has often been observed that the responses of IT neurons diminish rapidly to repeated presentations of the same stimulus, making it difficult to study the responses to individual stimuli with repeated presentations. In addition, as we saw in the Logothetis and Pauls study, when monkeys are trained behaviorally to respond to particular stimuli, large numbers of IT neurons come to respond to those stimuli.

Moreover, specific training isnt even necessary, as shown by the Booth and Rolls study. Monkeys exposed to a set of man-made objects in their home cages for a few weeks developed neurons that responded to these particular objects, and less to other man-made objects. These observations suggest that the response properties of at least some IT neurons change with the visual environment. Monkeys, and presumably human beings, develop IT neurons that are tuned to respond to the objects present in their environments.

This plasticity is doubtless useful to the organism. But if one is used to thinking in terms of fixed mappings between neural activity and perception, this plasticity is troubling indeed. If a single neuron changes from responding to one object to responding to another, does a given response from a given neuron signal one object on one occasion, and another object on another occasion? If so, how do we know the meaning of any particular cortical activity pattern? The plasticity seen in IT neurons would seem to rule out any model of object perception that depends on the presumption that IT neurons have fixed properties. THINK xx

26.3.6 Areas beyond IT

Since IT cortex is the last stage of processing that is mostly or entirely visual, we end our account here. Suffice it to say that there are neurons at higher levels that also respond invariantly to objects and faces, and to more complex aspects of faces such as emotional expressions.

26.4 Models of object recognition

The properties of IT neurons that we have just described leave us with several intriguing questions. The first of these is, how are neurons with these properties formed within the visual system? What stages of computation could take us from the V1 code to a code in which single neurons show the kinds of properties seen in IT neurons?

26.4.1 Hierarchical models

Hierarchical models are models of visual processing in which the properties of neurons at each higher level are created by summing signals from carefully selected sets of neurons at one or more lower levels. As discussed in Chapter xx, hierarchical models were first introduced into the vision literature by Hubel and Wiesel in 19xx in their pioneering studies of V1 cortical neurons. In their model, the outputs from rows of cells with center/surround receptive fields were summed to make simple cells – cells that responded best to a line of a particular orientation. Similarly, the outputs from simple cells with the same orientation preferences at neighboring retinal locations were summed to make complex cells – cells that responded invariantly to a line of a fixed orientation over a range of retinal locations. In hierarchical models, this selective resampling strategy is repeated at a series of increasingly high-level stages.

To be concrete, let's walk through a hierarchical model consisting of simple cells, complex cells, and three levels of processing beyond them. The model, proposed by Riesenhuber and Poggio (1999) is shown in Figure 26.10. The picture at the lowest level shows the stimulus– a wire object. This object will excite a set of V1 simple cells (a different set for each three-dimensional view). Recall that our update of Hubel and Wiesel's model suggested that V1 simple cells with the same orientation tuning could be summed with a MAX (or winner-take-all) summation rule across local cortical regions. This processing could create the V1 complex cells – cells that are translation-invariant in that they respond to lines of a fixed orientation across an expanded retinal region. This processing forms the first two stages of Riesenhuber and Poggio's model.

Beyond the simple and complex cells, Riesenhuber and Poggio posit the weighted linear summation of inputs from two or more complex cells tuned to different orientations, to make *composite feature cells* that respond best to combinations of lines at the two or more orientations. Composite feature cells responsive to the same combinations of features at different locations are then summed with a MAX rule to make *complex composite cells* that are translation-invariant. Finally, complex composite cells are summed with simple summation to yield a set of *view-tuned cells* responsive to particular views of the original wire object. Of course, the whole hierarchical process would have to be repeated a very large number of times in parallel, to create high-level neurons with view-dependent responses to all of the different objects we (or the monkey) can see.

During the 1990s, hierarchical models of object recognition were developed by several different research groups. The models differ in the proposed numbers and sequences of computational stages, and especially in the kinds of features [primitives] – e.g. views of objects, or simple three-dimensional volumes, or complex mathematical functions – for which neurons were presumed to be selective at each stage. These models also differ in many quantitative computational details, such as the summation rules governing the various transformations. Over time these models have become increasingly sophisticated in terms of computational specificity, and some have been implemented in computational form. When tested with a pre-specified set of stimuli, several of these models are now capable of providing view-invariant responses to individual objects, and different responses to different objects. That is, the output of the model contains information sufficient to determine which of a predetermined set of objects has been presented.

What are the limitations of hierarchical models? First, hierarchical models obviously have available many degrees of freedom. Second, one has the sense that once the key features have been discovered, many different hierarchical models will all be equally capable of producing the right behavioral specificities and invariances. Third, some authors argue that feedforward, or bottom-up,

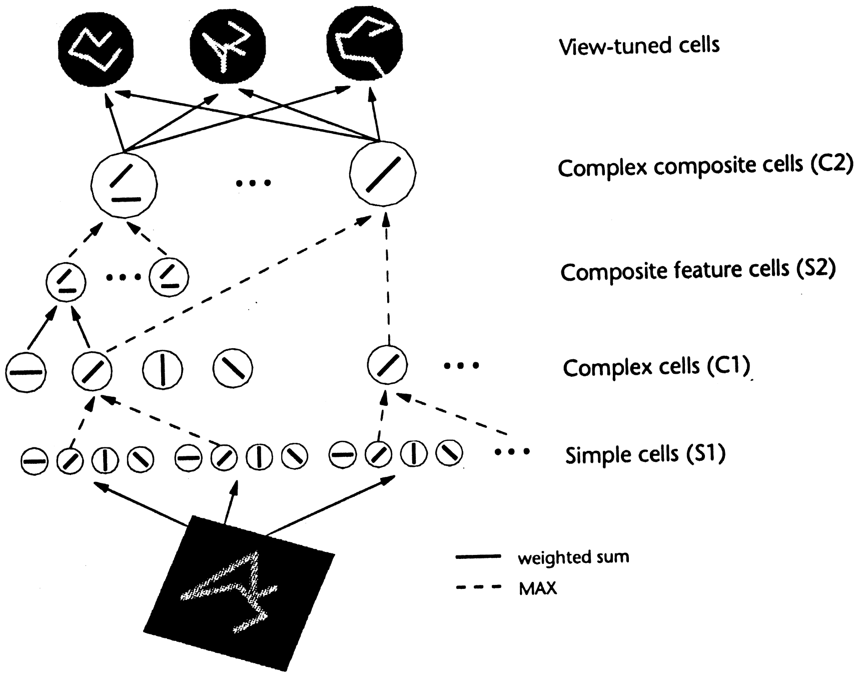


Figure 26.10: An example of a hierarchical model, as proposed by Riesenhuber and Poggio (1999). The picture at the bottom shows the stimulus – a wire object – and the pictures at the top show a set of model cells tuned to different views of the stimulus. The model has five stages: simple cells, complex cells, composite feature cells, complex composite cells, and finally view-tuned cells. Two summation rules are used, simple weighted sums (solid lines), and a MAX or "winner-take-all" rule (dashed lines)(see earlier Chapter on cortex). Only a few of the many cells needed are shown at each stage. View-tuned cells for other objects would be created by similar hierarchical networks, starting from the simple cells. Notice the absence of a stage consisting of view-invariant (or object-tuned) cells, which could readily be added at the top of the hierarchy. This stage is added in a later version of the model (Riesenhuber and Poggio, 2000). [From Riesenhuber and Poggio, 1999, p. 1021.]

processes of the kinds we have discussed are not sufficient to create the desired high level neurons, and that feedback, or top-down, processes will have to be included in the model. And fourth, the output layers of hierarchical models are not always clearly specified. For example, the Riesenhuber and Poggio model in Figure 26.10 has view-dependent neurons as its top level, but a similar model proposed by the same authors (Riesenhuber and Poggio, 2000) adds a layer of *object tuned units* that give view-invariant responses to individual objects.

In addition, as described above, some authors emphasize that IT neurons are known to be plastic. Their models incorporate the idea that new neurons are created as the monkey experiences new objects, and make assumptions as to which stages of the hierarchy are involved in this plasticity. In fact, assumptions about how the high-level neurons ”learn” their response characteristics are built into most hierarchical models from the beginning.

26.4.2 What are the primitives?

Suppose that we endorse a hierarchical model in principle. Clearly, there is still a very large number of possible hierarchical models. One of the ways in which hierarchical models differ is in the choice of the characteristics of the intermediate units (sometimes called *hidden units*, or *primitives*) that intervene between, say, the stage that represents V1 and the stage that represents IT.

One of the most interesting approaches to specifying primitives was popularized by the computational theorist David Marr in 1982. Marr argued that many different objects could be considered to be made up of sets of *generalized cones* (also called generalized cylinders). A generalized cone is a three-dimensional shape created by moving a fixed two-dimensional form along an axis. The form can be of constant or variable size, and the axis can be straight or curved. So for example, a cylinder, a football, and a forearm, shin, or torso, can all be seen as (at least roughly) generalized cones. More recently, Irving Biederman (1987) has proposed a variant of this approach, in which the primitives are a particular set of generalized cones that he called *geons*. An example of a small set of geons is shown in Figure 26.11. The essence of these theories is that the primitives for form perception could be *three-dimensional* shapes rather than combinations of lnes or two-dimensional features, and that many complex objects could be represented as combinations of a small set of geometrically simple three-dimensional forms.

When Marr and Biederman proposed their models, they were working in the domain of cognitive science, and did not need to commit themselves to any particular physiological instantiation of these models. But today we can ask, as we asked about the mathematical model of trichromacy, is there any evidence that such a model is instantiated in the human visual system? If we were physiologists hot on the trail of object perception, we might start looking to see whether there exists a processing stage at which different individual neurons respond optimally to different geons, and give invariant responses to a fixed geon across extrinsic variables such as its location, size, and three-dimensional rotation. Sadly, no such experiments have been published, although Biederman and his colleagues have shown that some IT neurons are *not* invariant across changes in the geon composition of the stimulus.

26.4.3 How sparse is the code?

We now return to the question of sparseness. Early in the chapter, we set out three options for the neural code for objects. These options occupy three locations along a continuum: an object could be represented by activity in a single neuron, with all other neurons silent – a grandmother

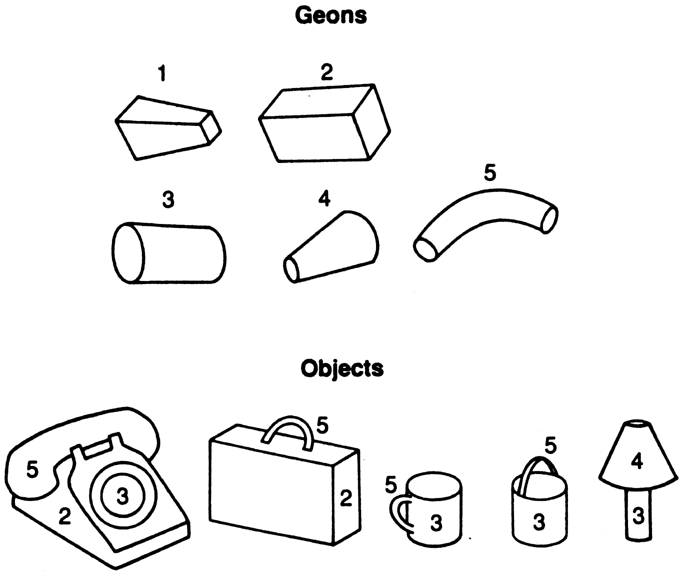


Figure 26.11: Examples of five geons and some objects that can be constructed from them. [From Palmer, 1999, p. 434; after Biederman, 1987.]

cell code. Alternatively, IT could use a sparse ensemble code only a few neurons out of a large population might be active in the representation of any given object; or, at the other extreme, a factorial code, in which the activity level of every neuron in the population is important to the representation of every object. Does the current knowledge of IT physiology allow us to narrow the range of options?

Based on the evidence from IT cortex, most scientists in this field now reject the option of a grandmother cell code. A major reason for this conclusion can be seen in Figure 26.5B. The particular face cell depicted there responds much more to faces than to other stimuli, and much more to some faces than to others. However, it does respond to some degree to many different faces. Turning the argument around, we presume that many different neurons respond to some degree to any particular face. Thus, it is not the case that all but one of the IT neurons is silent – the facts do not fit the definition of a grandmother cell code. Similarly, the view-invariant neurons that could be posited to exist at the highest level of hierarchical models are probably not grandmother cells – that depends not on the cell's view-invariance, but on whether or not it is the *only* cell that responds to some particular object. Studies of neurons at higher levels, beyond IT cortex, also do not reveal cells with the level of specificity required for a grandmother cell code.

[The following paragraph is due largely to Rolls. I would like to base it on Young and Yamane, but I don't understand the paper.]

On the other hand, many scientists in this field endorse the likelihood of a sparse rather than an extensive population code. Support for this perspective comes from studies in which a set of neurons are tested with a specified set of stimuli say, 20 neurons tested with 30 faces. Say that each neuron gives a pattern like that in Figure 26.5, with different neurons responding best to different faces, but all neurons responding to several of the faces to different degrees. The argument is that if you knew the pattern of activity across this set of neurons, you could infer with a fair degree of accuracy which of the faces was presented. Since a small number of neurons is sufficient to carry so much information abut the stimulus set, one can argue that a sparse code a few neurons – could be in force in IT cortex.

Although there is much to be said for this argument, some caveats should be noted. First, all that we have said is that this set of neurons could in principle allow the observer to tell which of the 30 arbitrarily chosen faces was presented. But this is not the usual task of the perceiving organism, who must recognize a face or an object out of a very large and unspecified set. And second, there is no evidence to date that the neurons recorded in these studies [Young and Yamane] are invariant over size, location, or viewpoint. If they are not, the whole applecart is upset, as variations of firing rate due to changes of viewpoint will be confounded with variations due to changes of the stimulus object.

26.4.4 And what of the common working model?

We began this chapter by defining a common working model of object perception: the notion that whenever we perceive a particular object, there will exist an invariant pattern of neural activity at a critical locus within our visual system. The common working model sent us on a search for neurons that demonstrate invariances consistent with it, particularly and most challengingly viewpoint invariance.

The biggest challenge to the common working model comes from hierarchical models that posit that the invariant perception of an object arises from one of a set of view-dependent neurons – for example, like those in Figure 26.8, or in the output layer of Figure 26.11. A hierarchical model might posit that a set of neurons with coarse codes like these is the highest level needed by the object recognition system, and omit any higher level at which one might have expected to find view-invariant neurons. This situation would be a contradiction of the common working model, because the perception of a common object would not arise from a common pattern of neural activity, but rather from activity in any one of a set of different neurons.

It is sometimes difficult, however, to discern whether the putative view-invariant level is really left out of the model. It may, in fact, be there implicitly, embedded in the rules needed to interpret the outputs of the view-dependent cells[2].

26.5 fMRI studies

A maor new thrust in the field of object recognition has come from imaging techniques, particularly fMRI. As we have discussed, the major contribution of imaging studies is that they can reveal cortical locations that are relatively highly activated by one set of stimuli with respect to another (e.g. faces vs. non-face-like objects). Over the past few years, these studies reveal increasing sophistication of experimental design, both for the choice of stimuli and for the choice of tasks.

For example, one of the problems of early imaging work was that the subjects attention was not well controlled. That is, one might ask a subject to look passively at, say, a set of faces and a set of non-face-like objects; and one might find that a particular brain area was more active for faces than for objects. However, this difference might come about, not because faces are faces, but because faces are (let us say) inherently more exciting or attention-grabbing. In more recent imaging studies, the subject is asked to do a task that requires attention to all of the stimuli. For example, in a *consecutive matching* or *one-back task*, the subject is asked to judge whether each stimulus is new, or is a repetition of the one seen on the previous trial. The same task is used with each different class of objects, so that, at least to a good first approximation, attention is controlled across tasks.

A second increase in sophistication comes about via the definition of regions of interest (ROI) and the use of multiple stimulus classes. For example, one might measure fMRI responses to faces and to objects, subtract the two, and determine the brain region perhaps a very small number of pixels in the fMRI image that responds more strongly to faces than to objects. This ROI can be determined within individual subjects. One can then go on to test the responses of the same ROI, in the same subject, to more refined classes of stimuli. In this way, one can begin to define the specificity or generality of the response in a particular highly specified brain region. The higher the selectivity for faces the greater the range of classes of stimuli that do not yield a response the stronger the argument that this region is specialized for the processing of faces.

26.5.1 The modular view: Faces, places, and body parts

In recent years, Nancy Kanwisher and her colleagues have carried out a series of studies designed to discover regions of the inferotemporal (IT) cortex that respond selectively to different classes

[2] Here's a golden oldie, just for you, faithful reader of this draft. DT's mother used to tell a joke that was very risqu for her time. What happens to your lap when you stand up? It runs around back and pops up under an assumed name! DT suspects that view-dependent neurons have a way of popping up under the guise of readout rules. Is she right?

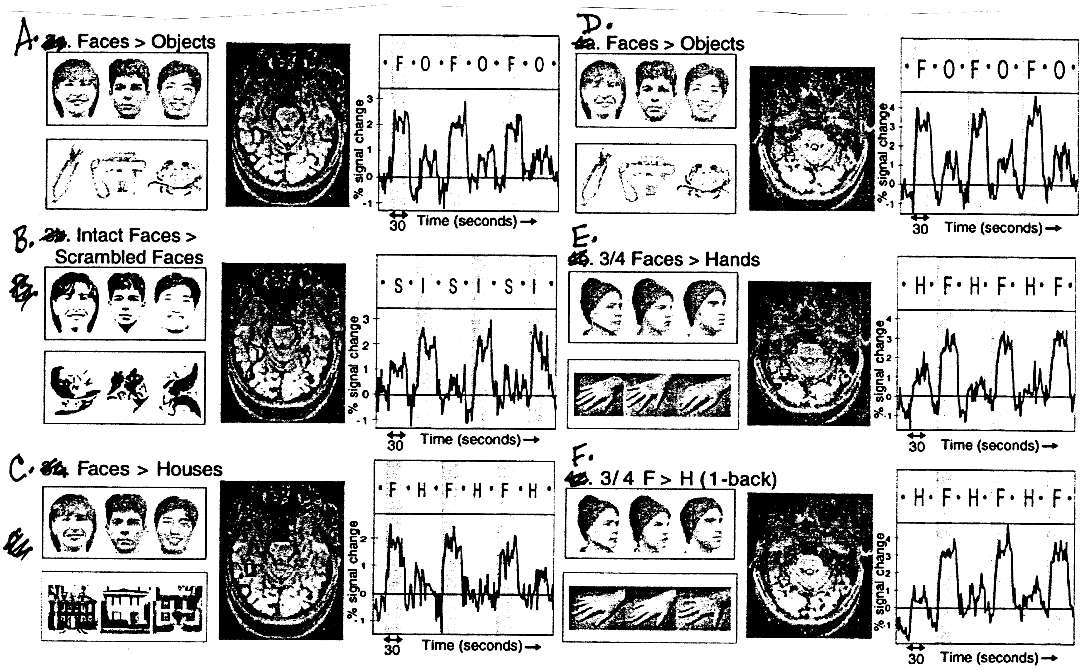


Figure 26.12: fMRI study defining the fusiform face area (FFA). See text for details. [From Kanwisher et al, 1997, Figs. 3 and 4, p. 4307.]

of stimuli. In one of the earlier studies, Kanwisher, McDermott, and Chun (1997) tested subjects with a variety of face stimuli vs. a variety of other stimuli. The stimuli, paradigm, and data from this study are laid out in Figure 26.12.

Kanwisher et al began by using passive viewing of grey-level faces vs. objects to define a ROI for faces in each of two groups of subjects. The data from this phase of the study are shown in the top row of Figure 26.12. The left and right halves of the top row (Figure 26.12A and D) show these initial data from the first and second groups of subjects respectively. Within each half of the figure, the left column shows the stimuli; the middle column shows the brain images; and the right column shows the stimulus timing and the fMRI responses. An examplar of one class of stimulus (e.g. a face) was presented for 30 sec, followed by a blank screen for xx sec, an examplar of the other class of stimuli for 30 seconds, and a final blank screen for xx sec. The cycle was then repeated xx times. Each graph shows the fMRI signal the percent change in xx – for three stimulus cycles, averaged across the individual ROIs of five subjects. In each case, as shown in the middle column, in a small region in the fusiform gyrus, the faces produced a far larger signal than the objects. This region was defined as the ROI for the remainder of the study.

The two groups of subjects were then used to make different kinds of additional comparisons within the defined ROI. The first group was tested with two-tone (black-and-white) faces vs. scrambled faces (B), and with faces vs. houses (C). The second group was tested with 3/4 views of faces vs. hands, both with passive viewing (E) and with a one-back task (F). As can be seen, in the

ROI the fMRI signal was at least a factor of two larger for each type of face stimulus than for the alternative classes of stimuli. As a consequence of this and other similar studies, this brain region was named the *fusiform face area*, or *FFA*.

Kanwisher and her colleagues have since gone on to find two additional regions of IT cortex that are highly selective for other particular classes of stimuli. One of these, named the parahippocampal place area (PPA) is selective for places or scenes; the other, named the extra-striate body area (EBA) is selective for human body parts (but not for faces). In addition, these scientists explored IT cortex xx extensively, with many different comparisons of classes of objects, and have not yet located any other highly specialized areas. A fourth area, named the lateral occipital area, or LO xx, responds rather generally to objects. Taken literally and at face value, these data suggest the view that object recognition could be carried out with a set of more or less dissociable neural modules, located in different sub-areas [of IT cortex? Xx].

26.5.2 Challenges to the modular view

Two major challenges to the perspective of Kanwisher et al have been voiced in the fMRI literature. The first is the argument that these different brain areas do not respond all-or-none to classes of stimuli. Rather, each class of stimulus sets up a pattern of activity across such brain regions; and it is that pattern of activity that provides the neural correlate for a particular class of stimuli. For example, Leslie Ungerleider, James Haxby, and their colleagues (e.g. Ishai, Ungerleider, Martin and Haxby, 2000) have shown clear differential patterns of response for three classes of stimuli faces, houses, and chairs. They argue that classes of objects tend to share stimulus attributes, and would therefore generate responses in similar or overlapping columns of IT neurons. These common responses could be mistaken for highly selective modules, when in fact it is the pattern across areas that codes the class of stimulus.

A second challenge has been aimed at the functional specificity of the face area, FFA. Isabel Gauthier and her colleagues have challenged the conclusion that the FFA is actually specialized to process faces. For example, Gauthier, Skudlarski, Gore and Anderson (2000) studied a group of bird enthusiasts who could discriminate readily among species of bids, and a group of car enthusiasts who could discriminate readily among makes and models of cars. The fusiform face area, FFA, was defined as the ROI. The bird experts showed a relatively greater signal in the FFA when viewing birds than when viewing cars, and the car experts showed the opposite pattern. On the basis of these and other data, Gauthier and her colleagues argue that the so-called face area is, instead, an area that is called into play whenever an expert is making a difficult, learned and highly practiced within-category discrimination.

26.5.3 Closing the gap between paradigms

How do fMRI studies relate to our understanding of the properties of IT neurons? This gap is difficult to bridge at the present time, for several reasons. First, single unit studies are carried out on monkeys, whereas fMRI studies are carried out on human subjects. Since the two kinds of brains are quite different in their overall conformations, it has not yet been possible to equate with certainty particular regions defined with fMRI studies to those defined by anatomy and neurophysiology. Second, the two techniques yield different kinds of information. At best, fMRI studies tell us that intense computations concerning particular aspects of visual function are occurring in particular sub-areas of cortex; and single unit studies tell us the properties of specificity and invariance for at least a few individual neurons in particular cortical sub-areas, along particular electrode tracks.

In short, the two kinds of studies do not readily address the same questions. So, for example, it is difficult to glean much from fMRI studies about our original questions: how do we recognize objects? What is the form of the neural code at the object recognition stage? How does it get to be that way, and what are the primitives at various stages of processing? The answer that different cortical subregions do intensive computations on different kinds of objects, does not scratch the original itch. However, the two fields will doubtless draw closer together as hints and challenges cross back and forth between the two techniques. Of special note are the new studies of fMRI images from monkey subjects.

26.6 A brief comment on categorization

Categorization is the capacity to identify an object as belonging to a particular category or class of objects dogs vs. cats, for example. In categorization, the emphasis is on the placing of objects into functional categories rather than just on the abstract perception of their three-dimensional shapes. The task is difficult because the physical characteristics of the members of the class need not be very similar, and may overlap extensively between different categories of objects: dogs vs. cats, for example, or tables vs. chairs. Moreover, in object categorization, the emphasis is increasingly on top-down processing: the accessing and use of stored information concerning the functional characteristics of the object and other members of its class.

Categorization is usually thought of as a more cognitive task than object perception, and is beyond the scope of this book. However, it can be noted that many of the same ideas that apply to object perception also apply to categorization. In object recognition, under the common working model, the goal of neurophysiological processing is to create neurons that are invariant across viewpoint, size, location, cues and so on. In categorization, one might argue that what is needed is to create a set of neurons that is also invariant across all instances of members of the category. The object recognition stage would sum inputs from lower level neurons to create an invariant representationof (say) a poodle. The categorization stage would go one step further in hierarchical modeling, and sum inputs from many dog-breed-specific neurons to create a neuron invariant across all species of dogs.

26.7 Summary

In summary, we began this chapter by accepting our nave realists' belief that objects are usually perceived veridically, and that a constant object is perceived as such most of the time. Based on this premise, we articulated the common working model: A constant perceived object suggests the presence of a unique visual code at some critical level of object processing.

We then delved further into the psychophysics of object perception, to see whether or not object constancy is as good as the nave realist assumes it to be. We found a major controversy concerning the degree to which our perception of objects is or is not view-invariant. Extrapolating beyond the available data, we suggested that this topic ties back to the logical availability of heuristics. Perhaps the simpler and more symmetrical the object, the greater the likelihood that workable heuristics will be available to generate accurate guesses as to its three dimensional shape, and therefore the greater the probable degree of view-invariance.

We then had a look at the properties of single neurons in inferotemporal cortex. We found that many IT neurons respond well to complex two-dimensional shapes, or to combinations of shapes, colors and textures. Other neurons respond invariantly to different two-dimensional view of objects, and a few neurons even respond invariantly to particular three-dimensional objects, and more to one object than to others. However, the available data are insufficient to allow us to even guess at a set of primitives on which object perception might be based. Moreover, the plasticity of IT neurons raises the specter that the neurons we record will tune themselves to respond to the stimuli we have chosen to use, and we will always get what we are looking for.

We next explored the characteristics of hierarchical models of object recognition. We noted that different hierarchical models could differ in the assumed number of layers of processing, the summation rules that connect each layer to the next, and in the features or primitives assumed to be computed and represented at each layer. Finally, we provided a brief exploration of the newly emerging field of fMRI studies related to object perception.

It will be interesting to watch the different models of object perception compete in modeling the actual system properties of object recognition, as revealed by careful and specific psychophysical studies with custom-designed stimuli. It will also be interesting to watch the continuing exchange of hints among IT neurophysiology, fMRI studies, psychophysics and computational modeling in this exciting field.

Chapter 27

Visual Consciousness

THIS CHAPTER IS STILL PRETTY ROUGH.

IN DRAFT 3, I AM MOVING LINKING PROPOSITIONS FROM THIS CHAPTER TO SEVERAL EARLIER LOCATIONS IN THE BOOK. BUT THE CHANGE IS NOT CONSISTENTLY INTEGRATED IN THIS DRAFT. SO I'M LEAVING THE SYSTEMATIC TREATMENT OF LINKING PROPOSITIONS HERE TO SUMMARIZE THE MATERIAL. AGAIN, FEEDBACK APPRECIATED AS TO WHETHER IT IS BEST TREATED ONLY AT THE END, OR THROUGHOUT THE BOOK, OR IN BOTH PLACES.

Throughout this book it has seemed useful to include references to conscious perceptual states. We have also made much use of perceptual demonstrations as a way to call on the reader's own perceptual experience, in order to convey an understanding of the questions and answers that make up the discipline.

Yet visual science has traditionally been rather schizophrenic in regard to consciousness. From a natural science perspective, consciousness would seem to be too hard to define, too impossible to measure, and just plain too ephemeral to incorporate into the discipline. Moreover, it seems likely that visual anatomy and physiology can be pursued by confining oneself to questions about neural circuitry: information goes in through the eye, around and around through the brain, and comes out through one or another response system – can't we make life easier, and leave it at that? In fact, the whole move toward behaviorism in psychology in the first half of this century can be seen as a manouver to avoid having to think about consciousness.

On the other hand, many visual scientists were probably initially drawn to the discipline because of excitement or questions raised by their own visual perceptions. What is happening in my brain when I have the sensation of red, or when I perceive a tree? Class B experiments especially seem difficult to rephrase in behavioristic terms. It seems to DT to be more honest – more true to our own knowledge base – to acknowledge the existence of conscious states, and to make room for them them in our scientific discipline. If accepting the challence of consciousness makes the discipline conceptually more complex, so be it.

There has been a recent resurgence of interest in the concept of consciousness. This new wave of interest is best embodied by the xx Society, which holds a biennial meeting in Tucson, Arizona. The meeting is called *Toward a Science of Consciousness* xx. It is attended by scientists and philosophers from many disciplines, specifically including behavioral neuroscience and vision. Some neuroscientists and other enthusiasts argue forcefully that scientific paradigms exist which will allow us to discover the neural basis of consciousness. In the present chapter we will delve into these ideas.

27.1 What is visual consciousness?

The concept of consciousness has been notoriously difficult to define. David Chalmers (of whom more later) quotes the following defeatist definition from *The International Dictionary of Psychology*:

Consciousness: The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means....Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it (Sutherland, 1989).

To narrow the topic a little, we can talk about just *visual consciousness.* We define visual consciousness as visual awareness; the having of visual perceptions.

27.2 Mappings between brain states and conscious states

The minute we include consciousness in our scientific discipline, we inherit the classical philosophical problem called the mind/body or mind/brain problem. How are mind and brain related? Shall we think of the physical universe as fundamental, and mind as arising from particular states of the physical world; or mind as fundamental, and the physical world as a creation of each individual mind? We will short-circuit this question the same way that Hering did in 18xx. His argument was that we conld be agnostic on this question, and just posit that the relationship between brain states and conscios states is *lawful.* Then the question becomes, what are the lawful relationships – the mapping rules – that hold between brain states and conscious states? Can we ever say that when brain state B happens, conscious state C also happens? Can any such laws be worked out?

The philosopher David Chalmers (1996) has argued recently that the mapping rules between brain and consciousness are fundamental laws of the universe, and that discovery of these laws is the most important set of questions remaining to be solved by modern science. This argument warms the hearts (or swells the heads?) of visual scientists, because as you have seen from this book, one of the goals of visual science is to discover these mapping rules.

One philosophical approach is to think of the mapping rules as akin to *bridge laws.* In the history of science it can happen that aspects of the same problem are studied by scientists in two different disciplines, each of which develops its own set of concepts and technical terms to describe the same set of phenomena. Later, when the two disciplines merge, some of the concepts in one discipline will be seen to be the same as some of the concepts in the other. What is required then is to work out the bridge laws – the identities of entities – between the two disciplines. For example, the statement that invaginating bipolars are ON-center bipolars and flat bipolars are OFF-center bipolars is an example of a bridge law; and similarly for the statement that parasol cells are M cells and midget cells are P cells.

In regard to the mind/brain problem, the argument would be that perceptual states and brain states are two manifestations of the same entity (which Patricia Churchland (19xx) christens the *mind/brain*). This perspective is called the Identity theory of mind: the mind *is* the brain. Another classical statement of this perspective is that we can see the mind/brain from two different perspectives: from the outside (the physical brain) or from the inside (the conscious perceptions). In these terms, the goal would be to work out bridge laws to establish which brain states (as defined by physiological experiments) map to which perceptual states (as defined by perceptual experiments). A tall order!

27.2.1 The neural correlates of consciousness (NCC)

Another common argument in the philosophy of consciousness is that not all brain states are relevant to perceptual states. That is, one can argue that the states of only some, but not all, of the neurons or neural circuits in the visual system matter to our conscious perceptions.

The subset of visual neurons whose states map to conscious states can be called the *bridge locus* (Teller and Pugh, 1983), or alternatively the *neural correlate of consciousness* or NCC (Crick and Koch, 19xx). The question then becomes, which parts of the visual processing system are within the NCC? To realize the full impact of the question, look back at Figure 19.3xx (the van Essen wiring diagram), and think of trying to lay a bet on the location(s) of the NCC.

It turns out that at present, different visual scientists and phlosophers have very widely differing views on this question. At one extreme, the philosopher Rene Descartes (16xx) argued that the mind interacts with the brain only at one small point – the pineal gland. Similarly, Crick and Koch (19xx) argue that it is an effective strategy to believe that the NCC is confined to a small set of neurons [all in one place? All projecting to one place? All within a single layer of cortex? A few, in many different places?] xx

In contrast, others think that the NCC is very broadly distributed, and includes much of the visual system. The computational scientist David Marr believed that the NCC encompassed a broad range of neurons within the visual system (but probably not the retinal neurons). The philosopher Daniel Dennett (1991) holds a similar but more baroque view in his dynamic ”multiple drafts” model of consciousness. As we have seen, Milner and Goodale (19xx) argue that the NCC resides in the ventral, and not in the dorsal stream. And in a conversation with DT in 1998, the vision scientist David Williams was heard to argue that perhaps all of the visual neurons lie within the NCC. The thought experiment was, if you could silence one of your cones, but keep the states of all the other neurons the same, don't you think your visual perception would change, just even an eensy weensy bit?

27.3 Kinds of evidence

Given this wide range of opinions on the identity and extent of the NCC, actual evidence would seem welcome. But what kinds of evidence can one possibly bring to bear? To date, there there seem to be three lines of evidence that scientists interested in consciousness are calling upon to support hypotheses concerning the set of neurons that comprises the NCC. Let's have a look at them, and make evaluations if we can.

27.3.1 Patients with brain injuries and monkeys with lesions

The first line of evidence comes from the reports of patients who have suffered loalized injury to different parts of the visual system. As we learned in Chapter 19, different brain injury patients report different changes in their perceptions, and have different patterns of visual losses.

One question that immediately arises is, can we believe these patients' reports? How can a person know that he is not conscious? Isn't this like a patient who is colorblind – wouldn't a colorblind person be unable to know that he is colorblind? The affirming argument is that a patient who has previously been visually conscious, but now is no longer so, has the knowledge to report the change in the available range of conscious states. For example, a person who loses his color vision can report this change.

As we reported in Chapter 19, patients who have lost function in Area V1 report the phenomenon of blindsight: they often retain some aspects of visuomotor function, yet report that they are no longer visually conscious. The theory is that the residual visual capacities come from a signal through the superior colliculus and pulvinar to join the dorsal stream. If things are simple, these cases suggest that the NCC lies in cortex, either in V1 or in locations to which V1 provides critical input.

Milner and Goodale (19xx) refine this argument. They argue that patients with damage to the ventral stream (or its major input, V1) often report losses of conscious perception, while patients with dorsal stream losses retain conscious perception. Cases of this kind suggest that the NCC lies within the ventral stream. [Consider these cases and this logic more carefully.]

Must we wait for nature's accidents, and learn only from human patients? Or might we be able to learn more about the locus of the NCC from surgical lesions in monkeys? A potential extension of the lesion paradigm would be to do further studies of patients who report the loss of visual consciousness. One could continue to tease out more and more exact sets of visual tasks that are (and are not) lost in conjunction with human patients' reports of their losses of conscious experience. Perhaps one would find that patients who report loss of visual consciousness would still be able to carry out tasks X, Y and Z, but not tasks A, B, and C. One could then test lesioned monkeys with these carefully worked out sets of tasks. The logic here is: In humans, reported losses of consciousness are tightly correlated with losses on specific tasks X, Y, and Z, but not others A, B, and C. In monkeys, this pattern of losses comes about only with lesion L. Therefore, lesion L has destroyed the NCC. [How tight is this argument? Do you buy it? Why or why not?]

27.3.2 Neurons that "follow the percept"

A second line of evidence comes from using ambiguous figures – stimuli that while remaining physically unchanged, are perceived to alternate between two or more perceptual configurations. Examples of ambiguous figures, like the wife and the mother-in-law, were shown in Figure **??**xx. Nikos Logothetis and his associates have used another kind of perceptual alternation called *binocular rivalry*: Binocular rivalry occurs when a subject is presented with two different patterns (such as gratings of different orientations) in the two eyes. Although the stimulus is unchanging, human subjects report that their perception alternates every few seconds back and forth between the two grating orientations.

The logic involved in these experiments is that the firing rates of neurons in the early stages of the visual system will probably remain unchanged, while neurons later in the system will *follow the percept – change their firing rates in correlation with the changes in perceived configuration.* The argument is that NCC neurons would be expected to follow the percept, and therefore that neurons that follow the percept are either within the NCC or on the anatomical route to it; and moreover, that neurons that do not follow the percept must not be part of the NCC. [What is the linking proposition involved here? xx]

An example of the use of binocular rivalry to search for the NCC is presented in a paper by Leopold and Logothetis (1996). Monkeys were first presented with gratings of identical orientations in both eyes, and trained to use lever presses to report the orientations of the gratings. When they had learned this task, they were presented with gratings of orientations that differed in the two eyes so as to create binocular rivalry. The monkeys were able to learn to alternate their lever presses over time in a manner that (in conjunction with various control conditions) suggested rather convincingly that the monkeys experienced binocular rivalry, and that they were reporting the perceived orientation of the grating on a moment to moment basis. For example, the monkeys' lever presses alternated at rates typically associated with perceptual reversals in humans.

Leopold and Logothetis then recorded from single neurons in areas V1/V2 and V4 of the monkeys' cortices, while the monkeys continued with their behavioral task[1]. They isolated a single cell, and mapped its receptive field and its preferred spatial frequency and orientation. They then presented the monkey with rivalrous stimuli – the preferred orientation in one eye and the orthogonal (so-called *null*) orientation in the other – and recorded the responses of the neurons. Finally, they constructed peristimulus time histograms averaged with respect to the monkey's report of a change of perceived orientation.

Leopold and Logothetis's results are shown in Figure 27.1. The figure shows responses of a V4 neuron during binocular rivalry. The neuron *increases* its firing rate just before the monkey reports a change of perceived orientation from the null to the preferred direction, and *decreases* its firing rate just before the monkey reports a change of perceived orientation from the preferred to the null direction. These authors report that 20% of neurons in V1/V2 and 40% of neurons in V4 show this behavior – follow the percept!

More recently, Sheinberg and Logothetis (1997) have carried out similar experiments on neurons in inferior temporal (IT) cortex. They report that "in contrast (to recording at earlier processing levels), the activity of almost all neurons in IT cortex and the visual areas of the cortex of superior temporal sulcus was found to ge contingent upon the perceptual dominance of an effective visual stimulus. These areas thus appear to represent a stage of processing beyond the resolution of ambiguities [stemming from rivalry] – and thus beyond the processes of perceptual grouping and image segmentation – where neural activity reflects the brain's internal view of objects, rather than the effects of the retinal stimulus on cells encoding simple visual features or shape primitives...." (p. 3408). [DT must read xx.]

Based on data like these, what shall we conclude about the locus of the NCC? Both neurons in V1/V2 and neurons in V4 follow the percept, so the data do not favor V1/V2 as opposed to V4 as the NCC. Perhaps it will turn out that a certain subset of the neurons in both areas of cortex are the ones that follow the percept. Or perhaps the V1/V2 neurons drive the V4 neurons, and the V4 neurons drive later neurons that form the NCC. Or perhaps the neurons in IT will carry the day. The eventual conclusions remains to be seen in continuing use of this interesting paradigm.

[A note to ponder – as we have seen, perceptual constancies constitute cases in which perception is more closely tied to the properties of physical objects than to the properties of the retinal image. Might it be true that each case of constancy provides a paradigm for testing the NCC? That is, e.g. for color constancy, at the NCC neural responses should correspond more closely to perceived color than to retinal wavelength composition; for size constancy, at the NCC the neural response should correspond more closely to physical size than to retinal size; and so on. Think about this.

[1] Since the monkeys were still undergoing more behavioral testing at the time of the report, exact locations of the neurons could not be reported.)

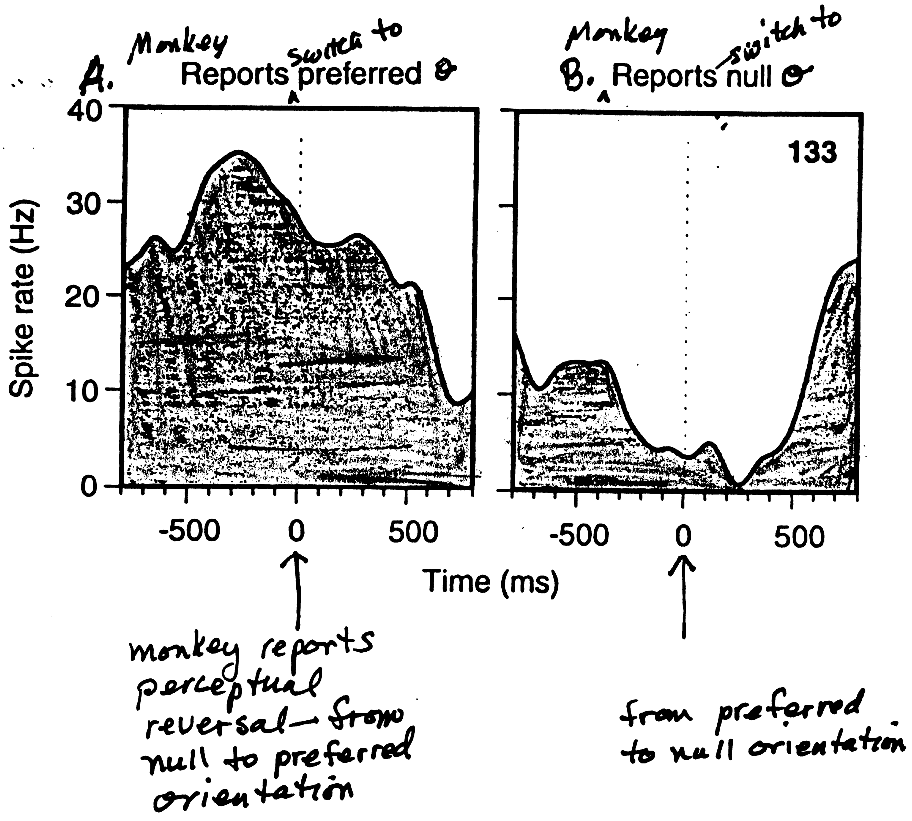


Figure 27.1: A V4 neuron "following the percept": Leopold and Logothetis (1996). The abscissae represent time, with 0 representing the moment at which the monkey reported a perceptual reversal. The left histogram shows spike rates around the time that the monkey reported a perceptual reversal from the null to the preferred orientation for this cell. The right histogram shows trials on which the monkey reported a reversal from the preferred to the null direction. The argument is that this neuron "follows the percept" and therefore is a candidate to be included in the NCC. (After Leopold and Logothetis, 1996, p. 552.)

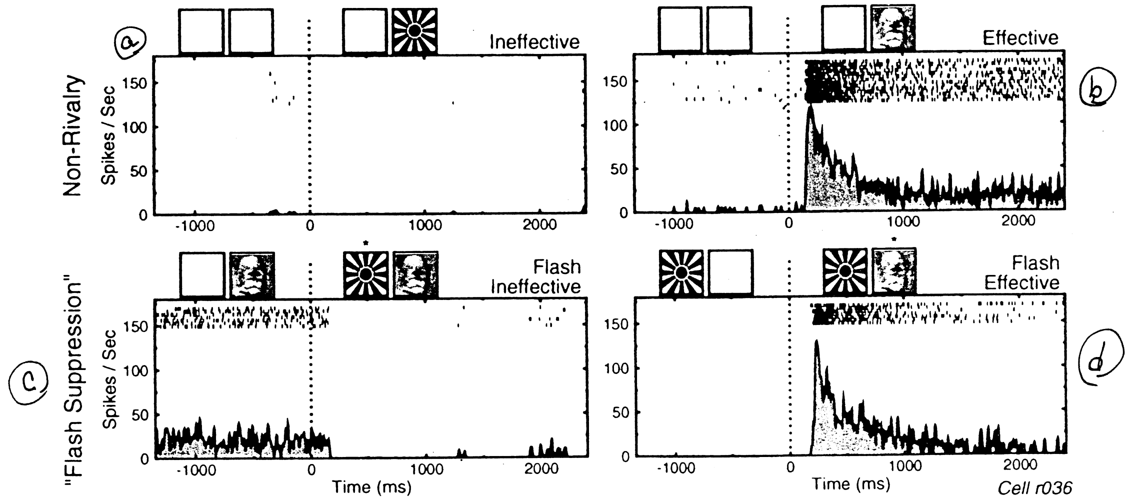


Figure 27.2: Shinerg and Logothetis' (1997) experiments on IT cells.

Design some experiments. If this line of thinking works out, "following the percept" could turn out to be a very general paradigm.]

[Add Tong et al fMRI study. xx]

27.3.3 Activity in location X is "not sufficient for consciousness"

The third paradigm is examplified by several experiments in which the claim is made, directly or indirectly, that activity in V1 neurons sometimes occurs without consciousness of that explicit activity; and therefore that V1 is not the NCC. [Of course the same paradigm would apply to any cortical area, but V1 is the popular site at present.]

Example #1: In a physiological experiment of this type, Moshe Gur and Max Snodderly (1997) showed that some color-opponent V1 neurons follow isoluminant chromatic flicker up to flicker rates as high as 30 Hz. You will recall from our discussion of flicker photometry that human subjects cannot resolve chromatic flicker for flicker rates this high; and neither can macaque monkeys. Gur and Snodderly's data thus provide a dissociation between V1 activity and the perception of chromatic flicker, suggesting that the neural limitation on flicker resolution is provided at a later site. Thus, they would argue that the NCC occurs beyond V1.

Example #2: Certain psychophysical experiments augmented by prior physiological knowledge also enter into this kind of paradigm. For example: suppose it is true that the earliest neurons in the visual system that are tuned to respond to moving lines are in V1, and that adaptation to moving lines are brought about by changes in the sensitivity of those neurons. As in adaptation experiments we have seen earlier, adaptation to one direction of motion yeads to an elevated threshold for test stimuli moving in that direction – a phenomenon called the *motion aftereffect*. Under this premise, He, Cavanagh and Intriligator (1996) carried out an interesting psychophysical experiment on human subjects.

He et al's paradigm is shown in Figure 27.3A. In the first (*single*) condition of the experiment,

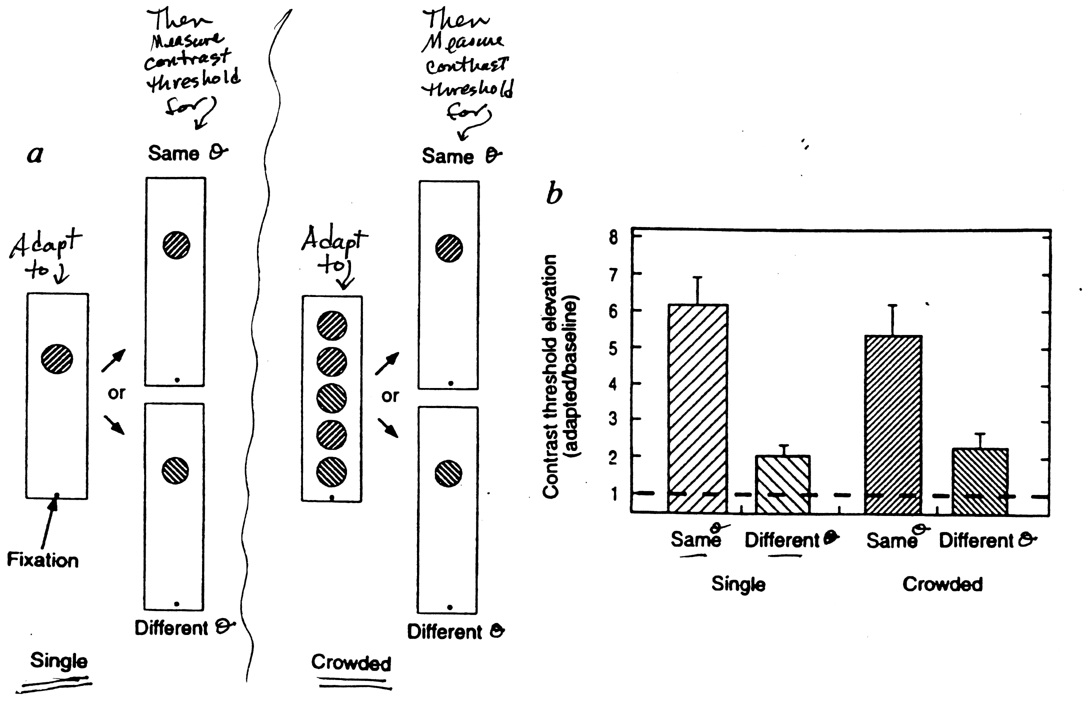


Figure 27.3: Some activity in V1 is not conscious: He, Cavanagh, and Intriligator (1996).

shown in the left two panels, He et al had their subjects adapt to a moving grating at a retinal location 25^o above the fixation point. Subjects reported that they were conscious of the moving grating during the adaptation phase, and they could report its direction of motion. Then the subjects were asked to set a detection threshold for each of two test targets, moving either in the same direction as the adapting grating or in the orthogonal direction.

In the second (*crowded*) condition of the expriment, as shown in the right two panels of Figure 27.3A, He et al. exposed the subject to the original adaptation grating *plus* several other adapting gratings moving in various directions. The subjects reported that they were no longer conscious of the direction of motion of the moving grating, could they no longer report the direction of motion. The subjects were then again asked to set detection thresholds for each of the two test targets

The results are shown in Figure 27.3B. As shown in the left two panels, the contrast threshold was elevated for a test target moving in the same direction as the adapting grating, but not much changed for a test target moving in the opposite direction. That is, when the motion of the adapting grating was visible to the subject, the adapting grating produced an orientation-specific motion aftereffect at its location in the visual field.

The question now is: in the crowded condition, when the motion of the adapting grating is no longer visible to the subject, will the adapting grating still produced a motion aftereffect? As shown in the right-hand panels of Figure 27.3B, the answer is yes. Even though the subject is not conscious of the direction of motion of the adpating grating, it still differentially elevates the threshold for gratings moving in the adapted direction.

Since the crowded adapting field still elevated the threshold for the subsequent test target, the authors argue that the adapting field must still have created activity in V1 neurons. But since the subject was not aware of the direction of motion of the adapting field, the authors argue that *there is activity in V1 of which we are not conscious.*

On the basis of these and other experiments and arguments, Crick and Koch adopt the position that V1 is not within the NCC; that is, the neurons that form the immediate neural substrate of visual consciousness lie at higher levels of the visual system. [Of course we must note the logicall fallacy involved in their speculation: just because *some* activity in V1 fails to support conscious perception, it doesn't follow that *all* activity in V1 fails to support conscious perception. Just because one Martian is green, it doesn't follow that all Martians are green.

[Christoph Koch tells DT there are now several experiments of this kind, and that together they implicate more central as opposed to more peripheral cortical areas as the NCC. Must find out about them. Question: will the conclusions all be internally self-consistent, or will different experiments yield opposite conclusions? Inconsistencies seem to me likely. That is, to DT it's highly unlikely that we will be conscious of all activity in any particular cortical locus. She is is betting that there is *some* activity in every area, of which we are not conscious. But this outcome would be disastroous for the paradigm!]

In sum, there are currently at least three different lines of experimentation that are claimed to provide evidence concerning the locus of the NCC. Different vision scientists doubtless are more or less persuaded by the different kinds of arguments. Aside from the human brain injury evidence, the other two kinds are new, and haven't yet had time to be debated in the vision literature. The logic surely has hidden assumptions, and the paradigms have not been fully evaluated. The playing out of these paradigms, and the task of uniting the conclusions of each use of the paradigm into a single conclusion about the NCC, are tasks for the next generation.

Of course, some people may think that all arguments about the neural basis of consciousness are misguided. This stance is OK, but it's interesting to try to figure out exactly why they think so.

27.4 Linking propositions [Redundant with earlier sections of the book]

But here a second question arises. It's all well and good to ask which neurons lie within the NCC. But the second and in some ways more interesting question is, *which states of these neurons map to which states of perception*? Neurons within the NCC, like any other neurons, presumably have a wide range of states and firing rates, and it is reasonable to assume that different firing rates and constellations of firing rates should map to different perceptual states. Are there any general principles that govern the mappings from phsyiological to peceptual states?

[Since writing Chapter 1, I have thought of some nice examples to add there. There are some candidate linking propositions that we would doubtless all reject. For example, we would deny that the representation of the three-dimensional world must be literally three-dimensional withinthe head. We would deny that a neuron that signals redness/greenness or yellowness/blueness would have to be literally red and green or yellow and blue. And we would deny that when we perceive a dance, our neurons dance. But the question is, are there other linking propositions that we accept readily; or accept as speculations; or draw implicitly and unnoticed into our arguments about the neural basis of perception?]

As discussed in Chapter refchapter:intro, the early German psychophysicists, wishing to draw physiological conclusions from psychophysical data, outlined a set of linking statements that they called *axioms of psychophysical correspondence.* More recently DT (Teller, 1984) has worked out a more extensive analysis of linking statements, which she calls *linking propositions.* A linking proposition is defined as *a claim that a particular mapping occurs, or a particular mapping principle applies, between perceptual and physiological states* (Teller, 1984, p. 1235). The early psychophysicists thought of linking propositions as axioms – that is, as unproveable but necessary assumptions that were needed if they were to infer brain states from perceptual states. But it probably makes more sense to conceive of them as propositions – that is, as statements whose truths and usefulnesses need to be individually evaluated.

27.5 General Linking Propositions

[Big organizational question: Should linking propositions go here, at the end, as a summary of the logic of relating psychophysics to physiology; or in Ch 2A and 2B, so students could be thinking about them all along? Let me know if you have an opinion.]

Let us begin with general linking propositions. Since we want to reason both from psychophysics to physiology and *vice versa*, we will argue that linking propositions come in families, with different members for different directions of inference.

27.5.1 Family structure

Each family of general linking propositions has four members that relate to each other as shown in Table 27.1.

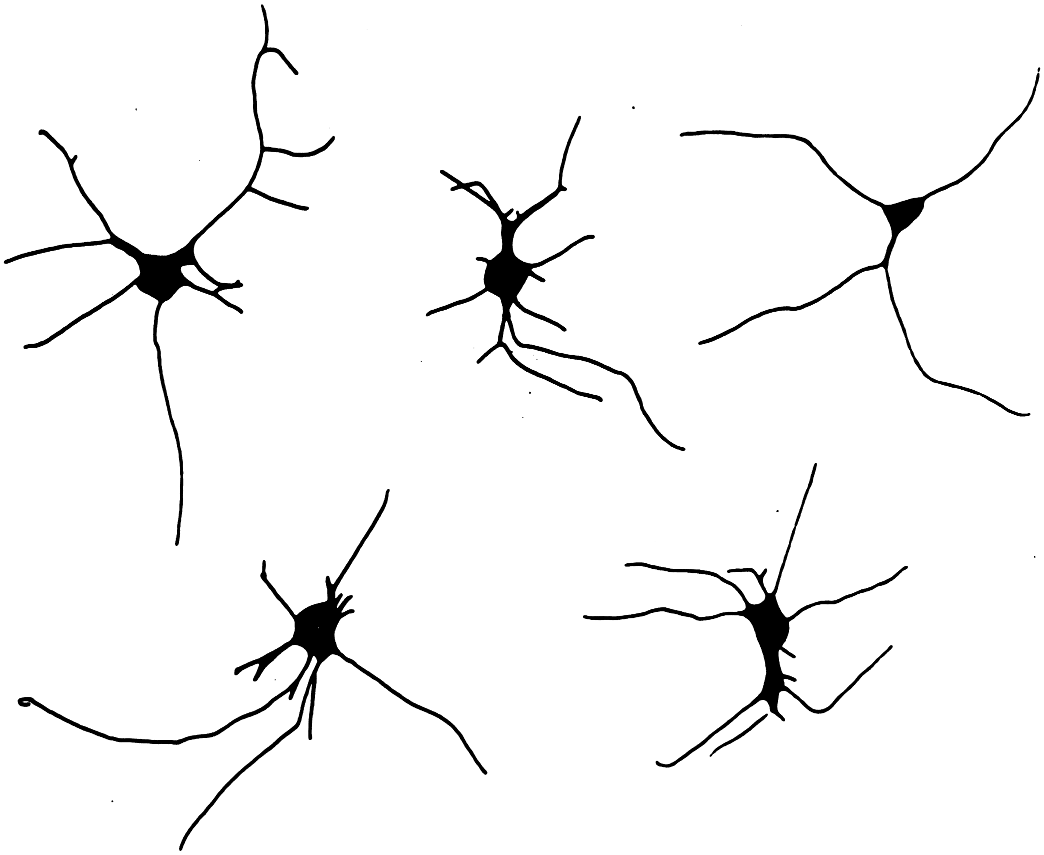


Figure 27.4: Dancing neurons?

1.	Initial Proposition	A → B
2*.	Contraspositive	not-B → not-A
3*.	Converse	B → A
4.	Converse Contrapositive	not-A → not-B

Table 27.1: Family structure of linking propositions.

For students who haven't studied logic, Table 27.1 requires some explanation. We are discussing two events or items, A and B. The question is, how do the four statements about A and B in Table 27.1 relate to each other? The arrow is to be read as *implies*; so the *initial* proposition, A –¿ B, is to be read as, *A implies B.* Essentially this statement says that whenever A occurs, B occurs.

Statement 2 is known as the *contrapositive* of statement 1. Statement 2 (not-B → not-A) states that if B idoes not ocur, then A does not occur. Now, an initial statement and its contrapositive are logically equivalent: if we accept the initial statement, we must also accept its contrapositive. For example, the initial statement might be, if it rains, the ground is wet. If we accept this statement we must also accept the contrapositive: if the ground isn't wet, it isn't raining.

How does statement 3, the *converse*, relate to the initial proposition? Importantly, a statement and its converse are not logically equivalent. Even if A implies B, that doesn't mean that B implies A. If the grass is wet, that doesn't necessarily mean that it is raining – the sprinklers could be on. As another example – a fellow graduate student of DT's once said to her, I don't mind being misunderstood, because to be great is to be misunderstood. DT, who had had introductory logic, was pleased to be able point out that unfortunately, to be misunderstood is not necessarily to be great!

Of course, just like the initial proposition and its contrapositive, the converse and its contrapositive, statement 4, (not-A implies not-B) are logically equivalent.

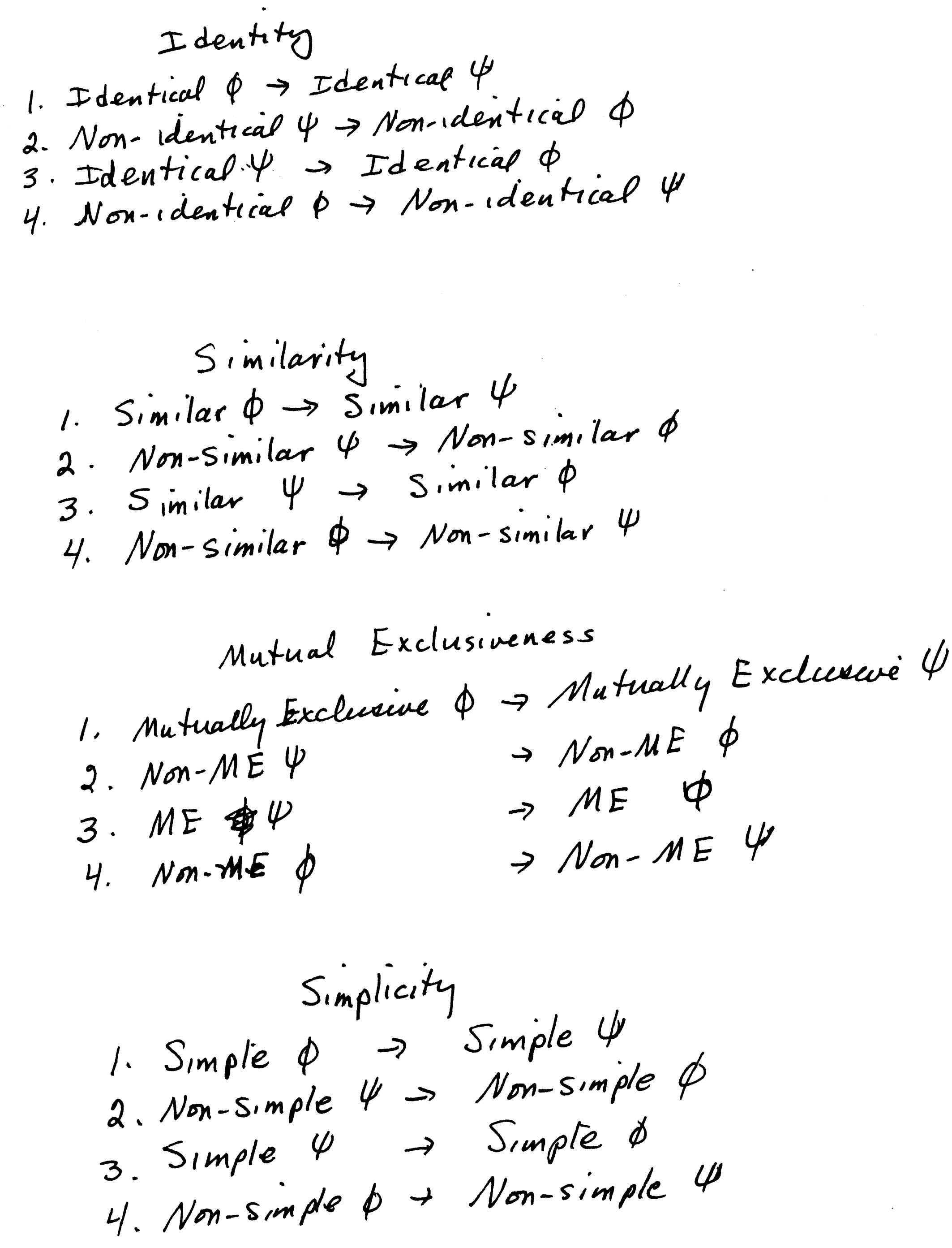
One final comment about the general form of the families of linking propositions: let A be a physiological statement and B be a psychophysical statement . Then two of the propositions in the family, the initial proposition (1) and the converse contrapositive (4), are proposed inferences from physiological states to perceptual states; while the other two, the contrapositive (2) and the converse (3) are proposed inferences from perceptual states to physiological states. The two that are starred in Table 27.1, statements 2 and 3, will be those that visual scientists use in arguing from perceptual to physiological states. Notice that if you are doing a perceptual experiment, from which you will want to draw physiological conclusions, you know ahead of time that either a contrapositive or a converse linking proposition will be involved in your argument; but you don't know which, because you don't know ahead of time how the experiment will come out .

27.5.2 The Identity family

Let's apply this logical structure to one of Mueller's axioms; the axiom that *identities of material states imply identities of perceptual states and vice versa.* The resulting family of linking propositions is shown at the top of Figure 27.5.

Let's walk through the Identity family and evaluate whether or not we think each proposition is true. For similicity, let's just think of the neurons that are within the NCC. The initial proposition in the Identity family is: Identical physiological states imply identical perceptual states. Do you believe it? It has been argued that this statement is in fact true – not an axiom, but a necessary or *a prori* truth – that it wouldn't be possible to have two identical states of neurons at the NCC, and not have identical perceptions. If we accept this argument, it is very important, because *it gives us at least one of Chalmers' mapping laws between physiological and perceptual states.* Our foot is in the door!

The converse identity proposition is: Identical perceptual states imply identical physiological states. Do you believe thiis one? In fact, this proposition is only true if the mapings from physiological to perceptual states within the NCC is 1:1, rather than many:1. Most people are not willing



Identity

1. Identical ϕ → Identical ψ
2. Non-identical ψ → Non-identical ϕ
3. Identical ψ → Identical ϕ
4. Non-identical ϕ → Non-identical ψ

Similarity

1. Similar ϕ → Similar ψ
2. Non-similar ψ → Non-similar ϕ
3. Similar ψ → Similar ϕ
4. Non-similar ϕ → Non-similar ψ

Mutual Exclusiveness

1. Mutually Exclusive ϕ → Mutually Exclusive ψ
2. Non-ME ψ → Non-ME ϕ
3. ME ~~ϕ~~ ψ → ME ϕ
4. Non-ME ϕ → Non-ME ψ

Simplicity

1. Simple ϕ → Simple ψ
2. Non-simple ψ → Non-simple ϕ
3. Simple ψ → Simple ϕ
4. Non-simple ϕ → Non-simple ψ

Figure 27.5: Four families of general linking propositions

to concede the 1:1 mapping, but instead argue that many different states of the NCC neurons could map to a single perceptual state. That is, most people think that the converse identity proposition is not true *a priori*, and will have to be assumed anew as part of every argument into which it enters.

Trichromacy provides a good example of the use of a converse Identity proposition. The argument began with the psychophysical observation of trichromacy. The inference (or speculation) was that identical perceptual states (metamers) implied identical physiological states – equal quantum catches in each of the three cone types. Notice that the speculation was not about neurons at the NCC, but all the way back at the photoreceptors. Not a logical inference, but a very lucky guess, and a hint exchanged across disciplines.

The important point is that in visual science one often does psychophysical detection or discriminaton experiments – experiments that depend on a subject's judgment of whether two things are the same or different. The argument is that in all such cases, if the goal is to draw physiological conclusions, the experimental logic will always involve an Identity proposition, either the contrapositive or the converse.

27.5.3 Three more families: Similarity, mutual exclusiveness, and simplicity

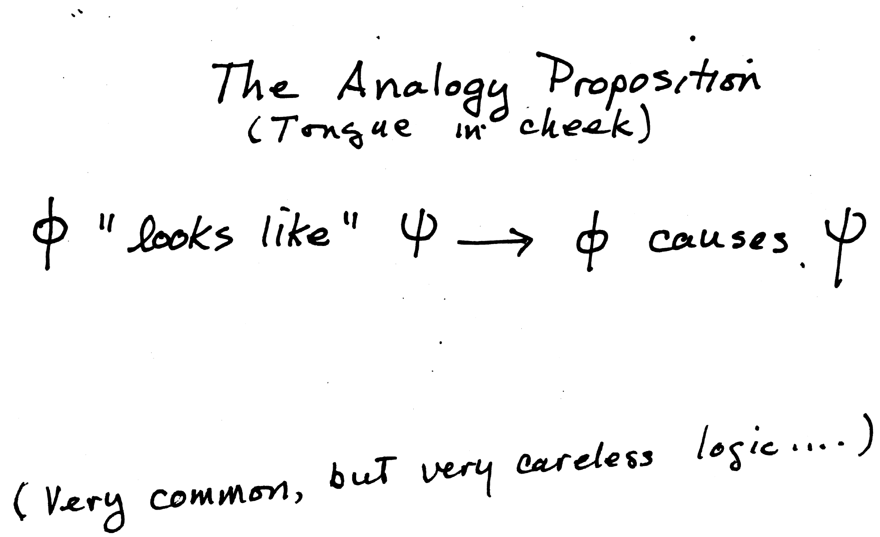
Three other families of general linking propositions are also shown in Figure 27.5: Similarity, Mutual Exclusiveness, and Simplicity. In general, none of the propositions in the other families is thought to be true a priori; all play the role of assumptions in any argument in which they are used.

Rather than walk through all of these propositions, we here collect up some specific examples that should be familiar to you from earlier chapters of the book. A Converse Similarity proposition is in use when we ask, why is the color circle a circle? That is, why do long wavelength (red) and short wavelength (violet) lights share a similar reddishness? The answer visual scientists inevitably give – the linking proposition – is that the physiological processes created by these two lights must somehow be similar to each other; and this assumption has often influenced the choice of post-receptoral codes in color theories.

A Converse Mutual Exclusiveness proposition is in use when ask, why are redness and greenness perceptually mutually exclusive? A neo-Heringian answer is, because the neural states set up by these two lights share a mutually exclusive code – redness is coded by (say) an increase in firing rate of neurons in one of two chromatic channels and greenness is coded by (say) a decrease in firing rate of the same neurons in the same channel. And a Converse Simplicity proposition is also in use in Hering's theory. Why are there both unique (simple) and binary (more complex) hues? A neo-Heringian answer is that unique hues occur when only *one* of the color channels is active and the other is at its resting state; and binary hues occur when both channels deviate together from their resting states.

27.6 The Analogy proposition

Finally, there's another general linking proposition that DT calls the Analogy proposition. It's the one that usulally lies implicit in arguments in which two similar curve shapes are presumed to indicate a causal relationship between psychophysical and perceptual realms. DT has been asking you to think hard about such analogies throughout this book. One example is the Mach band analogy in Chapter 11xx. [List and review others.]



The Analogy Proposition
(Tongue in cheek)

ϕ "looks like" $\psi \longrightarrow \phi$ causes ψ

(Very common, but very careless logic....)

Figure 27.6: The analogy proposition

27.7 Specific Linking Hypotheses

It is interesting to notice that linking propositions have been giving us trouble throughout some of the most difficult questions of this book. In particular, we asked in Chapter 18xx, how is the perception of an object coded? What form does the physiological code take, and where, when we see and recognize an object? We can now see that *a large element of our quest involves trying to find a linking proposition that we can believe in* – we perceive grandmother when a grandmother neuron increases its activity; or when a certain pattern code occurs, or when a percept "emerges" from a pattern code; or, it makes my head hurt to think about it (Don't worry, be happy). As you review the course, you will find many examples of arguments that include implicit linking propositions; and you may find it interesting to try to figure out how to make these various linking propositions explicit.

Finally, here are four questions that you might have asked during the past quarter or semester. All involve the perception of two-dimensional or three-dimensional space. Some of these questions are generally considered to be pseudo questions, while others are still thought to be real. DT argues that all of them involve implicit linking propositions, and that making the linking propositions explicit helps understand the problem or pseudoproblem better. Your task, should you choose to accept it, is to identify the implicit linking propositions in each of these four questions.

1. Why do we see the world right side up when the retinal images are upside down?

2. Why do we see the world non-distorted when the maps in LGN and V1 "magnify" the representation of the fovea?

3. Why do we see a continuous visual scene, when the left and right hemifields are represented

in two separated regions of V1?

4. Where in the brain are the various depth cues integrated to give an integrated perception of depth?

27.8 Summary

In this chapter, we have tried to place visual science within the context of the philosophy of science and the philosophy of mind. We have argued that visual scientists can get by with finessing the mind/body problem, by taking the position that the mappings between perceptual states and brain states are *lawful* and remaining agnostic as to causal relations between them.

We then defined the concept of the neural correlate of consciousness, or NCC, and asked whether it encompasses all or only part of the visual system. We expored three kinds of evidence that might bear on this question, but noted that scientists and philosophers are all over the map in the answers they endorse at the beginning of the 21st century. It will be interesting to watch these paradigms either congeal or evaporate.

Finally, we asked more specifically, what kinds of regularities might we expect in the mappings between brain states and perceptual states. We examined the concept of a linking proposition – a proposed mapping rule between perceptual states and neural states – and identified several general families and specific instances of linking propositions. We leave you with the challenge of using linking propositions carefully, and making them explicit, whenever you make arguments from physiological to perceptual data or vice versa.

[DT – see more notes on computer at the end of this Chapter.]

Needed in Ch on central processing, or in the present chapter.:

What are we to make of the argument that if a neuron carries information about x it does x? Barlow's doctrines may claim this.–neuron signals the presence of the stimulus that makes it fire fastest. van Essen – we must face the fact that neurons may signal many things. The De Valoises address this too, – a cell that responds differentially to different colors may not be ”coding color”. So does Lennie.

Carry a color signal, doesn't mean codes for color or is the bridge locus for color. You see these assumptions often; watch out.

But: ”carries a signal sufficient for X, therefore either the neural correlate of X or on the pathway to it, until shown otherwise” would be a wise working rule.

What working linking propositions shall we use? If a set of neurons Can do X, it does do X, until proven otherwise.

Bibliography