Seminar on the Empirical Foundations of Visual Attention
John Palmer and Cathleen Moore
Spring 2007
9 July 2007

This is a summary of a seminar that considered the empirical foundations of visual attention. This document contains the syllabus, weekly question sets, and a summary of the discussion for each week.

Seminar on the Empirical Foundations of Visual Attention
Course Syllabus
John Palmer and Cathleen Moore
9 July 2007

Synopsis

In this seminar, we consider seven research questions in visual attention. These questions are proposed as the empirical departure point for a further understanding of attention. Behavioral measures are emphasized in the readings but we also address the relation between behavioral and neuroscience measures of attention. The readings are drawn from both visual psychophysics and cognitive psychology. Please contact John Palmer by email if you are interested in attending.

General Information

Psychology Graduate Seminar 555B, 2 Credits, CR/NC
Weekly meetings: 3:30 Thursdays in 211 Guthrie Hall.
Entry Codes and other information available from John Palmer and from
http://faculty.washington.edu/jpalmer/files/Seminars/

E-mail: jpalmer@u.washington.edu
Phone: 206-543-0706
Office: 27 Chemistry Library
Address: Psychology, Box 351525, U of W

Course Structure and Responsibilities

John Palmer will jointly lead the discussion with one discussant selected for the week. To help guide discussion, study and discussion questions will be distributed a week ahead. The study questions have answers that can be found in the reading while the discussion questions are open ended. Seminar participants will be expected to read the material in detail before the class meeting. Be sure to understand the study questions and think about the discussion questions. Everyone is expected to be a discussant once or twice. Discussants are expected to meet with John Palmer a few days before class and afterwards to draft a summary of the class discussion. This summary will be distributed to all participants by e-mail before the next meeting.

Reading

We will read selections from Hal Pashler's 1998 book and selected journal articles.

Schedule

Week	Topic
1	Organizational Meeting
2	**Selective attention in perception:** What are the effects of selecting information from one stimulus and not another for simple perceptual tasks?
3	**Divided attention over stimuli:** What are the effects of multiple stimuli for simple perceptual tasks?
4	**Divided attention over tasks:** What are the effects of multiple perceptual tasks?
5	**Time course of selective attention:** What is the time course of switching selective attention from one stimulus to another?
6	**Time course of divided attention:** What are the effects of processing multiple stimuli at the same time versus processing them in sequence (parallel vs. serial processing)?
7	VSS break (9–16 May)
8	**Selective attention in memory:** What are the effects of selecting information from one stimulus and not another for explicit memory tasks?
9	**Divided attention in memory:** What are the effects of attending to multiple stimuli for explicit memory tasks?
10	**Finale:** Reconsider the major issues of the seminar.

Seminar on the Empirical Foundations of Visual Attention
Questions for Week 1: Selective attention in perception
John Palmer and Cathleen Moore
29 March 2007

Goals

What are the effects of selecting information from one stimulus and not another for simple perceptual tasks? To examine this question, we focus on an idealized laboratory paradigm: brief displays to eliminate eye movements; sparse displays to minimize spatial interactions, single-attribute discrimination or detection; accuracy measures of performance; and, cueing paradigms that hold the stimulus constant. The initial question is to demonstrate an effect of attention that is clearly not due to the stimulus. Then we consider whether this effect is due to signal enhancement or noise reduction.

Primary Readings

Pashler pages: 37-39, 53-55, 167-197
Shiu & Pashler, 1994: cueing effects w/ single and multiple stimuli
Smith, 2000: evidence for single-stimulus cueing

Further Readings

Davis, Kramer & Graham, 1983: intro to signal detection models of cueing
Carrasco, Penpeci-Talgar & Eckstein, 2000: evidence against noise reduction
Smith, Wolfgang & Sinclair, 2004: further analysis of single-stimulus cueing

Study Questions

What is cueing? Why is it useful for controlling sensory effects?

What is spatial uncertainty? How does it relate to other manipulations such as set-size effects in search?

In signal detection theory, how does noise limit performance? How is it applied to the tasks in these papers?

According to the signal enhancement hypothesis, what is the effect of attention on perception?

According to the noise reduction hypothesis, what is the effect of attention on perception?

Under what conditions does Smith find single-stimulus cueing effects? Why are these inconsistent with the noise reduction hypothesis?

Discussion Questions

How is an attentional effect defined? How can one establish that an effect is attentional?

What is capacity? How is it operationalized?

What is selection? How is it operationalized?

What are the idealized conditions considered in these experiments and why were they chosen?

What are the pros and cons of using masking in these studies?

What might account for the mask-specific cueing effects found by Smith?

How does one determine if an attentional effect is due to perception or decision?

How might one implement signal enhancement? For example, if one simply amplifies a noisy signal, both the signal and noise are increased.

How might one elaborate the signal enhancement hypothesis to make a more specific prediction?

For the noise reduction hypothesis, what assumptions affect the predicted size of the cueing effects?

How might signal enhancement and noise reduction work together?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for week 1: Selective attention in perception
Heather Knapp, John Palmer and Cathleen Moore
11 April 2007

Papers

Shiu & Pashler, 1994: cueing effects w/ single and multiple stimuli
Smith, 2000: evidence for single-stimulus cueing with masks
Background Pashler book pages: 37-39, 53-55, 167-197

Discussion Summary

Introduction to the organization. John Palmer began with a few words about how the seminar is organized. The underlying agenda is to write a book on attention. Our overarching organization is to consider topics in turn from selective and divided attention phenomena. The beginning topics address simple accuracy paradigms in perception; next we consider the time course of selection and divided attention phenomena; finally, we consider explicit memory tasks as a complement to the previously considered perceptual tasks. The big topic missing from this organization is the distinction between early and late selection. More precisely, for selective attention, which processes occur before selection and which occur after? Similarly, for capacity, which processes have unlimited capacity and which ones don't. These topics are the focus of later chapters in our book. But to begin, we avoid topics that bring in the complexities of all of cognition.

Cueing experiments. We began by discussing selective attention. The approach we emphasized is to hold the physical display and task constant while manipulating attention by instruction. Specifically, a cue provides advanced information about a soon-to-follow stimulus such as its position in space. For example, validly cued trials result in better performance than invalidly cued trials. One can also consider neutral trials which are trials on which the cue provides neither positive nor negative information about the stimulus. They ought to fall somewhere between valid and invalid trials.

A important variation on the cue paradigm is to compare 100% valid cues to a neutral condition. The advantage of this variation is that the degree of validity is completely clear. Otherwise, one must model the subject's interpretation of the degree of validity. Hence, it is much easier to make quantitative predictions with 100% valid cueing.

A third variation of these cueing experiments is to combine cueing with the presentation of a foil. Of particular interest is a foil that is identical to the stimulus except in its location. In this case, the same stimulus that is a target at a relevant location is a foil at an irrelevant location. These experiments can measure the "fate of the unattended stimulus" at an irrelevant location by errors made in responding to a foil. Examples of such foil experiments can be found in visual search (Shiffrin & Schneider, 1977) and detection (Palmer & Moore, 2007).

Noise reduction and signal enhancement. Two hypotheses have been emphasized to account for cueing effects: noise reduction and signal enhancement. By the noise reduction hypothesis, a cue allows the subject to disregard irrelevant sources of information when making a decision about a stimulus. For example, one can ignore locations where the stimulus does not appear. Because irrelevant sources of information can be ignored, noise from those sources is less likely to be confused with the relevant signal. In essence, the noise from irrelevant sources is 'turned down' relative to the relevant signal.

By the signal enhancement hypothesis, the subject can become more sensitive to information from the relevant source. In essence, the volume on the relevant channel is 'turned up' in such a way that it improves the discriminability of a target from a distractor. One possible mechanism for signal enhancement is to allocate some processing resource away from irrelevant sources of information and toward relevant sources. Of course, this presupposes that processing has limited capacity and that the processing of sources of information compete with one another. To be more concrete, consider a sample size model suggested by Shaw (1980) among others. In this model, one can take some number of samples of the world. If one knows the relevant stimulus, then one can just sample the relevant stimulus and get a good estimate of its properties. In contrast, if one is uncertain over many stimuli, then one must spread the fixed number of samples over all of the stimuli. For this case, the number of samples per stimulus is reduced and thus so is the accuracy of the properties of any one stimulus. In the simplest version of this sample size model, accuracy varies with the square root of the number of samples just as with the statistical estimation of a mean from a sample distribution.

Shiu and Pashler (1994). This paper compares cueing effects in multiple and single element displays to adjudicate between the noise reduction and signal enhancement hypotheses. Although cueing effects are quite robust for multiple element displays in which distractors appear with the target, they are less reliably found for single element displays in which only the target is present. This may be especially true for studies in which discrimination accuracy—rather than response time—are the dependent measure. Multiple element displays don't allow one to easily distinguish between noise reduction and signal enhancement, because when both a target and distractors are present simultaneously, cueing can attenuate the distractors, enhance the target, or both. In contrast, single element displays may distinguish these hypotheses because the there are no distractors present. Unless the task is to discriminate a target from empty space, it's hard to argue that cueing turned down the noise from an empty space! Rather, the explanation is naturally constructed around the processing of the target. If there is only a target, then any effect of the cue must be due to enhancing the processing of the target. On the other hand, if cueing effects are absent in single element displays, it is unclear why signal enhancement on the target channel would be absent simply because of the lack of distractors in other channels. Lack of cueing effects in single element displays makes perfect sense for noise reduction models.

An important potential confound is the use of multiple post-stimulus masks. When multiple masks are used with single element displays (i.e., Henderson 1991), one does observe cueing effects. Thus, the introduction of multiple masks may add noise at other locations and thus yield a cueing advantage by noise reduction. The current paper uses a cueing paradigm for single-target displays and compares the effect of the pre-cue on accuracy across conditions with a single or multiple masks.

A brief discussion of the nature of cues. The nature of the cue is also important. John Palmer noted that peripheral cues themselves present difficulties of interpretation. Transient peripheral cues may better localize the target, which makes them more effective as cues. But, because cues also raise the possibility of masking or contrast adaptation which may improve or worsen performance independent of any attentional effect. Lynne Werner pointed out that symbolic and veridical cues often yield same results in audition research (for example, presenting a musical note symbol versus an auditory tone). She wondered about cue duration, which prompts a brief discussion of intervals and sampling. The durations of the transient cues must be quite short. In audition, loudness discrimination is influenced by the range of tested sound intensity. John Palmer described a discussion with Duncan Luce about attention bands in loudness discrimination (Luce, Green & Weber, 1976; Nosofsky, 1983). He argued that this effect probably wasn't attentional. In that case, one had to manipulate the stimuli being presented and cueing did not work. This seemed like a strong hint that something like adaptation was causing the effect rather than selective attention.

A brief discussion about noise reduction. Alec Scharff asked whether noise reduction takes place only after perception on decision processing. John Palmer suggested that noise reduction may be perceptual as well as post-perceptual. In his early work, he followed Marilyn Shaw in interpreting noise reduction as a decision phenomenon. But now he thinks it can affect processing early or late. For example, perceptual processes may combine noise from several sources when uncertain and combine noise from fewer sources when properly cued. Thus, one can consider either a perceptual or decision version of noise reduction.

Some people in the literature characterize noise reduction as "just" a statistical artifact. John didn't like that view. Lynne Werner wondered if the real question is where is the limit on performance? Is the limiting process one of sensation or decision?

Alec Scharff asked whether noise reduction is like turning down the volume on the unused locations. John Palmer says yes, but it's also about the relative weightings. One could also turn up the volume for the relevant location so that it drowns out the irrelevant. Alec argues, but isn't that the same thing as signal enhancement? John responds that for signal enhancement somehow the signal is improved without increasing the noise at the same location. One way to think about this is to return to the sampling model. More samples can be placed at the relevant location. Alec: Is it like increasing the sampling rate? Is that a metaphor or is it intended as an actual mechanism? John: Usually metaphorical, but once can interpret it literally for cases such as making eye movements or rehearsing items in memory.

Lynne Werner asked about sampling over time versus sampling over space. Sampling over time is clear, but space? John: Consider multiple parts of the scene that have to be parsed and the resulting objects can be sampled for information.

Lynne Werner asked about the reasonableness of proposing that people can narrow their filters. How would this be interpreted? John pointed out that this is one possibility for why peripheral transient cues are especially helpful. Taking another tack, Dosher, Liu, Blair and Lu (2004) considered whether cueing allowed one to more selectively ignore the noise in local regions around a target location. They considered such a change in the weighting of local regions an example of noise reduction. This would also be considered a change in the spatial filter. Lynne: So choosing the right filter (or narrowing the filter) is an example of noise reduction? JP: Yes, but if you didn't have a theory that made the filters explicit, one couldn't tell it apart from signal enhancement.

Lynne Werner filled in some background on the probe–signal paradigm (Scharf, Quigley, Aokin, Peachy & Reeves, 1987; Bargones & Werner, 1994)) that was used to describe the idea of narrowing a filter. Suppose people have narrow filters for auditory frequency. When one presents auditory frequencies or durations that are near threshold outside of their expected range, those stimuli are not detected. Heather interjected: People hear only what the experimenter wants them to hear.

Along those lines, Lynne mentioned that on reading the Shiu and Pashler paper the only hypothesis that she could see as reasonable was noise reduction. John: "Me too". But, in fact, their paper was very controversial at the time. The field remains split about the relative roles of signal enhancement and noise reduction.

Experiment 1. John Palmer moved on to consider some of the results in the paper. To begin, the results of the first experiment shown in Figure 2 show an odd pattern not predicted by either hypothesis. The "no cue" condition with multiple masks is not worse than the valid cue condition. John guessed that it is probably just a random fluke.

Experiment 2: Next consider the results of Experiment 2 with central cues that is shown in Figure 3. The valid cue multiple and single mask conditions don't come together. A common interpretation of this deviation from the predictions is that the central cue isn't fully effective.

Also, the results for the neutral condition was not between the results for the valid and invalid conditions. Lynne asked: A neutral cue is difficult to interpret because you don't know what the subject is doing. For example, are you telling them to look everywhere? Nowhere? Heather: Yeah, you're basically telling them that they need to divide attention over the whole field. Perhaps a neutral cue is really an invalid cue—it's just a different kind of invalid cue. You have 'attend x, right', 'attend x, wrong' and 'attend everywhere'. John: One of the advantages of presenting a neutral cue relative to no cue is that they reduce temporal uncertainty. This discussion is just scratching the surface of the issues around the choice of a neutral cue (see Jonides and Mack, 1984).

A brief discussion of masks. Lynne asked: What is it about the mask that makes you attend to it? Is it because of the onsets? Would it be better to have masks that are somehow present all the time so that the onset flash is not so disruptive? John: This would make the mask like an ongoing camouflage. Perhaps this would be like crowding over time. Briefly discussed crowding over space and its interpretation in terms of mandatory pooling (Parkes, Lund, Angelucci, Solomon & Morgan, 2001). Also mentioned the 4-dot masking paradigm (Enns & Di Lollo, 1997).

Smith et al. (2004). There is an idea in the attention literature that attention enhances weak-signal detection only under conditions of backward masking (mask-dependent cueing). As explained in the text, when signals are masked, detection sensitivity increases for signals at attended locations; when signals are unmasked, detection is unaffected for both attended and unattended locations. This was corroborated in an earlier Smith paper (2000a) in which a single Gabor patch had to be detected from a luminance pedestal rather than a uniform field. Smith found that detection was enhanced only for attended stimuli that were backwardly masked. (A brief search ensues of the Smith paper as to whether the pedestal and signal come on at the same time. Not answered.) John: Moral of story is never use masks! (Laughter....)

The next question is how to interpret a mask-specific effect. What is it about the mask that makes one uncertain? Alternatively, a mask may delay the time to orient. John comments that most people take it on faith that masking simply stops processing without any other changes. Seems unlikely to him.

Lynne commented: Backward masking is very popular in audition because kids with certain kinds of language impairments are particularly susceptible to it. Some think that this masking susceptibility may indicate some underlying sensory processing phenomenon in populations of kids.

Lynne: Why would one build a system that would enhance a single stimulus? Not adaptive. Why not get rid of noise? John launched into an account of one case where amplification at one location does help. Suppose the limiting noise is added late in the sequence of processing. Then amplifying the information before the addition of the noise does improve the final signal-to-noise ratio. We then discussed early versus late noise models. This comparison of early versus late noise may help put the signal enhancement hypothesis in a more specific context for analysis.

John commented: Maunsell, who is a physiologist studying attention in single neurons, suggests that if one amplifies the firing rate of neurons (for example, 100 spikes/sec instead of 50 spikes/sec), then there is an improved signal-to-noise ratio. This is because the noise variance, not the standard deviation, increases with the mean spike rate. As a result, if you increase the mean rate by a factor of 4, the ratio of the mean to the standard deviation increases by a factor of 2. On the other hand, if you increase the spike rate too far, then it would hurt you if attention further increased the spike rate. For neural coding, there appears to be an optimal range for the spike rate. For weak stimuli, increasing the spike rate may result in a gain of sensitivity. In short, it's a dynamic range issue.

Lynne: Cochlear implants do better if actually induce some degree of noise so that the neuron is driven to be more sensitive. It seems the implant has less noise than the normal hearing cells.

John: I wonder if attention can ever reduce activity when neurons are near saturation? He went on to briefly discuss a paper by Huang and Dobkins (2005) that addresses the interaction between attention and adaptation. It compared models of contrast gain and response gain using a contrast pedestal paradigm. Their bottom line is that both kinds of gain are likely to be involved. (Warning: There is an error in the explanation of the predictions in this paper. The contrast gain model predicts a shift both left and down in Figure 1c.)

Seminar on the Empirical Foundations of Visual Attention
Questions for Week 2: Divided attention over stimuli
John Palmer and Cathleen Moore
5 April 2007

NOTE: Late start of 4-5:30 Thursday 12 April 2007

Goals

What are the effects of multiple stimuli for simple perceptual tasks? As with the first question, we focus on an idealized laboratory paradigm: brief displays to eliminate eye movements; sparse displays to minimize spatial interactions, single-attribute discrimination or detection; accuracy measures of performance. Furthermore, we focus on the simultaneous-successive paradigm. The initial question is to demonstrate an effect of divided attention that is clearly not due to the stimulus or non-perceptual factors such as decision making. Then we consider how to generalize this analysis beyond the simultaneous-successive paradigm to visual search accuracy for brief displays.

Primary Readings

Pashler pages: 109-115, 135-144
Huang & Pashler, 1998; introduction to the simultaneous-successive paradigm
Palmer, 1994; generalization to accuracy visual search

Study Questions

What is divided attention?

What is the unlimited-capacity, parallel processing hypothesis?

What is the serial processing hypothesis? What are other relevant hypotheses?

What is the simultaneous-successive paradigm? Why is it a measure of divided attention? What is the prediction of the unlimited-capacity, parallel processing hypothesis?

For what tasks and stimuli are the predictions of unlimited-capacity, parallel processing satisfied for the simultaneous-successive paradigm?

Palmer investigates set-size effects in visual search as another example of divided attention phenomena. How can one distinguish between sensory and attentional effects of set size?

What are the predictions for an unlimited-capacity, parallel model for the accuracy search experiments of Palmer? What assumptions do these predictions depend upon?

For what tasks and stimuli are the predictions of unlimited-capacity, parallel processing satisfied for accuracy search?

What generalizations are suggested for the tasks and stimuli that can be processed by an unlimited-capacity, parallel process? What are alternative suggestions in the literature?

Discussion Questions

What is the significance of masking in the simultaneous-successive paradigm?

Under what conditions might sensory, memory or decision phenomena combine with unlimited-capacity, parallel processing to cause a deviation in the predicted equality between simultaneous and successive conditions?

For the simultaneous-successive paradigm, what is the prediction of a simple version of the serial processing hypothesis?

For the accuracy search, what are the possible residual sensory effects and how might they be measured and/or eliminated?

For accuracy search, how does one distinguish the contributions to set-size effects of perception, attention, memory and decision?

For the accuracy search paradigm, what is the prediction of a simple version of the serial processing hypothesis?

How might one distinguish serial processing from limited-capacity parallel processing?

Is the lack of an effect of divided attention in the simultaneous-successive paradigm compatible with large set-size effects in visual search?

What tasks and stimuli are likely candidates for serial processing?

What other phenomena are considered examples of divided attention across stimuli (e.g. report)?

How can one generalize the measurement of divided attention effects to less idealized conditions (e.g. eye movements)?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for week 2: Divided attention over stimuli
Alec Scharff, John Palmer and Cathleen Moore
18 April 2007

Papers

Pashler pages: 109-115, 135-144
Huang & Pashler, 1998; introduction to the simultaneous-successive paradigm
Palmer, 1994; generalization to accuracy visual search

Discussion Summary

Divided attention phenomena. John Palmer asked Heather Knapp how she typical defines divided attention phenomena in the beginning of her talks. Heather said she likes to use a real world example, like a street scene, with lots going on and an incentive to properly divide attention among stimuli. Driving down a street and navigating properly without hitting things is an example of task that demands divided attention.

Are real-world search tasks a good example of visual attention to use in introducing a talk? Heather felt the search task is confusing because search usually involves the selection of a single item rather than monitoring multiple items. John contended that in search one is required to process many items and reject distracters. Heather countered the best example of divided attention involves monitoring multiple stimuli and combining information from multiple sources as in signing language interpretation. John responded that it's still necessary to reject most other stimuli in order to focus on the signing. Alec Scharff noted that divided attention is essentially about handling multiple information sources, whether specific tasks involve ignoring or interpreting those sources. Heather thought the most typical divided attention tasks performed in labs involved monitoring and combining information from stimuli, as in a dichotic listening task. John thought the distinction was that dichotic listening tasks are a case in which we are very ineffective at dividing attention, while we can be very good at dividing attention in some types of visual search. John also brought up dual-task experiments in which independent responses to separate stimuli are necessary, which we will discuss next week.

Heather expressed frustration about a recent story in the media expounding the significance of work by Scottish researchers who found dual-task interference effects. She opined that vision scientists are generally ineffective in disseminating important information to the public and could make more of an effort in that regard. She mentioned the Dana Foundation's Brain Awareness Week as a potential venue for presenting attention research.

Alternative Hypotheses. We began discussing the standard hypotheses of limited or unlimited capacity and serial or parallel processing. Our initial focus was the operationalization of the term “serial”. Heather understood the term in the sense of serial processing being an extreme example of limited-capacity processing. This seems to be the sense that Broadbent used the term. John contrasted this with an interpretation that focuses on the capacity of the individual processing of a single item. This is how Townsend uses the term in his analyses of alternative search models. Using Townsend's definition, a serial process can have either limited or unlimited capacity. It depends on whether the processing time for a single item is independent of the number of items. The multiple uses of the term capacity is confusing. Perhaps one should always use the term capacity in conjunction with a specification of the relevant process ("unlimited-capacity comparison process", Townsend) or the whole system ("limited-capacity perceptual system", Broadbent)?

Cathleen Moore commented on a draft of the summary that the term "unlimited-capacity serial process" is contradictory for anyone considering the capacity of the whole system (ala Broadbent). Perhaps it would be better to refer to "independent serial processes" because "independent" is a more general term and doesn't have the connotations of necessarily referring to the entire system.

Heather asked which usage was more common. John thought Pashler leaned to the Broadbent usage with serial processing as one of many possible limited-capacity models. Townsend’s language is perhaps more idiosyncratic and has made less impact outside the field of search. Today, we’ll talk about the serial model with unlimited-capacity in the sense of Townsend.

Alec Scharff asked about the ramifications of the fact that some stimuli can be processed in parallel and others only in serial. Suppose one has an information processing sequence in which all stimuli are processed with certain characteristics of stimuli demanding more time consuming processing and being bottlenecked at certain points in the sequence. John acknowledged that stimuli have many attributes and some attributes may make different demands in analysis than others. One can try to treat all attributes the same way but whether this is possible is a fundamental and interesting question.

Heather asked whether specific attributes of stimuli can change their perceptual processing. John gave the example of hearing a tone versus seeing a flash. An information processing approach assumes that all stimuli are processed equivalently as information. Alternatively, a modular processing approach would assume that processing of different attributes can be fundamentally different. For example, a flash has very good spatial representation and relatively poor temporal frequency representation. In contrast, a tone has a poor spatial representation and an excellent temporal frequency representation. Are these differences a sign that the differences are more important than the similarities?

Alec raised the Stroop effect as an example of his conception of interference or limited-capacity processing during visual search. For this effect, perhaps perception is slowed when a task-irrelevant factor causes the relevant process to compete with irrelevant processes. Or, another way to think of it is that the processing of different attributes can interfere with one another. John pointed out that the Stroop effect was dependent on response modality (e.g. Greenwald, 1970; 1972) and that perhaps these effects are not so much about just perception.

After the seminar, John Palmer came to realize that the Stroop and similar effects such as the Simon effect or stimulus-response compatibility may all be examples of the failure of "task selection". This would be analogous to the discussion during the first week of failures of "stimulus selection". The effect is dependent on the task, because it depends on the particular combination of stimulus-response mappings rather than just the stimulus combinations. In the future, John hopes to expand on this view of these paradigms as addressing task selection. This may be a topic as basic and important as stimulus selection.

Alec posed the more general question of why do some attributes require only unlimited-capacity parallel processing while others require either limited capacity or serial processing? John presented his own conception, wherein a “front end” operates across the visual field in parallel and does some types of processing. It can also support the selection of some stimuli for other processing that can only be done in serial. For example, it may be that one can detect and segment word-like objects across the visual field but can only read one word at a time. This is a view that is typical of "two-stage" search theories that go back at least to Hoffman (1979). Heather asked if that is the same as saying that there’s a threshold below which parallel processing is possible and above which it’s not? John reiterated that the front end handles tasks that can be done across the visual field; if you can form a signal that’s relevant to your task at that level, it’s being done in parallel. If that machinery isn’t enough (as in reading), you’ll get a more restricted process, not in parallel.

Special cases such as reading your own name in parallel are not predicted by this model. If those special cases really happen, it would be contrary to this simple "two-stage" model. John argued that the foil experiments we have discussed previously (Palmer & Moore, 2007) are a better example of a failure of selective attention to consider initially. The Stroop paradigm seems to have some non-perceptual elements.

Simultaneous-successive paradigm. Heather began the discussion by reporting her difficulties in describing her work with the simultaneous-successive paradigm. One issue was the possibility that the experiment measured an effect of perception that has nothing to do with attention. Perhaps we are better at perceiving few stimuli than many. John thought the necessary counterargument was a yet to be performed version of the simultaneous-successive paradigm that used cueing so that displayed set sizes and durations were equivalent in both conditions while varying the relevant set size. By keeping the stimuli identical, one can show that the effect of cueing either simultaneous or successive subsets must be attentional.

Heather commented that in the traditional experiment, the simultaneous display introduces additional stimuli and might encourage the use of different strategies in the simultaneous and successive conditions. John agreed, that anything that might be correlated with simultaneous and successive is the complication with interpreting this paradigm. Perhaps the cueing version of the simultaneous-successive experiment is the highest priority for just this reason. With the traditional paradigm, if the simultaneous-successive prediction fails, it could be because of some perceptual effect. This possibility is eliminated with the cueing version of the paradigm.

What is the best way to communicate the predictions of the simultaneous-successive paradigm to others? Heather thought there was something unsatisfactory about the paradigm because of the large amount of explanation necessary compared to the relatively sparse predictions. She called them "lonely little data sets." John agreed that more predictions can and should be made.

John sketched out a schematic demonstration of the simultaneous-successive paradigm for use in presentations. A crude representation follows which shows the processing timelines going down the page for different models with two stimuli (A,B) and two different conditions (SIM, SUC). In each case the processing is assumed to take 3 time steps (three V's).

Parallel Model

SIM		SUC	
A	B	A	B
V	V	V	
V	V	V	
V	V	V	
			V
			V

V

For the parallel model, processing of stimulus A and B occur together for the SIM condition and in sequence for the SUC condition. Thus, performance is the same for both conditions.

Serial Model

SIM	SUC
V	V
V	V
V	V
	V
	V
	V

For the serial model, only processing of stimulus A is completed for the SIM condition while both A and B can be completed in sequence for the SUC condition. Thus, for this model, performance is better for the SUC condition.

Parallel – limited-capacity Model

SIM	SUC
v v	V
v v	V
v v	V
	V
	V
	V

For the limited-capacity parallel model, processing of both stimulus A and B occur for SIM condition. But, this processing is less efficient as represented by the lower-case "v" it the diagram. In contrast, both A and B can be completed in sequence for the SUC condition without the competition found in the SIM condition. Thus, for this model, performance is better for the SUC condition.

We all liked this schematic. Alec thought the graphical demonstration could be usefully elaborated by showing processing occurring in a few discrete steps over time.

Masking in the simultaneous-successive paradigm. John described the division between those that have used the simultaneous-successive paradigm without masking (e.g. Eriksen & Spenser, 1069) versus those that have used the paradigm with masking (e.g. Shiffrin & Gardner, 1972). The use of masks yields different results under some conditions (e.g.

set-size effects, Morgan, Ward & Castet, 1998), so it's clearly an important part of the story. Heather asks, how do you stop processing without masks? Without the mask, processing continues for some period after the display is terminated that cannot be controlled by the experimenter. John agrees that such processing is likely. However, he doesn't think the critical predictions of the simultaneous-successive paradigm are dependent on knowing when processing is terminated. As long as the persistence is the same for simultaneous and successive displays, it won't change the predictions. It does become important when we consider serial predictions in the following section. John issued a challenge to identify predictions that depend on masking?

In Pashler's book (1998, p 118), he argues that "suppose that even in simultaneous displays subjects actually identified each character well after the onset of the mask, perhaps one at a time." In other words, without the time pressure due to the masking, one might resort to a serial process that would eliminate the functional difference between the simultaneous and successive conditions. This is a good start to answering John's challenge. One immediate response is that this point is irrelevant to using a difference between simultaneous and successive conditions to rule out unlimited-capacity processing. Instead, this point is addressing whether equal performance in the simultaneous and successive conditions rules out serial processing.

Here is John's reply to Hal added after the seminar. A key assumption is whether processing is sensitive to the effective duration of the stimulus. Assume that the effective duration is the sum of the physical duration and some fixed persistence time. Furthermore, assume performance improves with increasing effective duration. Finally, let all of the stimuli be presented for a fixed duration that with persistence one can call the "unit duration". For the simultaneous condition, all of the stimuli are available for one unit duration. If there are 4 stimuli, each might be processed serially for ¼ the unit duration. For the successive condition, half the stimuli are available for one unit duration and the other half for the other unit duration. Thus, for the 4 stimulus case, each might be processed serially for ½ the unit duration. Thus, if performance is sensitive to the effective duration, there should be a difference in performance between the simultaneous and successive conditions. The key is not necessarily to limit the duration, but (a) to have the same effective duration for both simultaneous and successive conditions, and (b) to have performance that is sensitive to the effective duration.

Assumptions necessary for the standard prediction. The standard prediction of an unlimited-capacity parallel model for the simultaneous-successive paradigm depends on several assumptions. The first of these is the unlimited-capacity assumption: processing of any one stimulus is independent of the number of other stimuli. Secondly, one must assume processing is independent of being in the simultaneous or successive condition. In other words, processing of a single item is independent of the type of display. One can imagine several ways in which this assumption might not be met (for discussion see Pashler, 1998, p. 118-120). For example, one might be less efficient in beginning the processing of the second display in the successive condition compared to the first display. More specifically, Cathleen pointed out that one must be as flexible in reassigning processing during the second display as in originally assigning processing in the first display. Another possible exception is that the memory for the target must be just as good for simultaneous and successive conditions. This is despite having to remember the target a bit longer from the first interval of the successive condition compared to the second.

One test of this important assumption is to compare performance in the first and second intervals of the successive display. Heather Knapp did this analysis of her own simultaneous-successive experiment comparing presentation in the first and second intervals of her successive display and remembers them being the same. Similar results are reported in the literature for cases that satisfy the predicted equivalence between simultaneous successive conditions (e.g. Shiffrin & Gardner, 1972; Duncan, 1980). John has a memory of reports of a difference for cases where the two conditions differed but cannot find the reference. It was agreed that this sort of analysis is important for interpreting simultaneous-successive experiments.

Predictions of serial processing. Heather asked how the simultaneous-successive paradigm can distinguish between a serial process and a limited-capacity parallel process. John explained the following proposal from Alec Scharff's second experiment in his first year project. The idea is to introduce new condition into the paradigm that allows for a prediction of the simplest serial model. The usual conditions might be a simultaneous condition with 4 stimuli (SIM4) that is compared to a successive condition with 2 stimuli followed by 2 more stimuli (SUC2). The usual prediction of the unlimited-capacity parallel model is for identical performance between SIM4 and SUC2.

The new idea is to introduce a successive condition with 4 stimuli followed by a second display repeating the same 4 stimuli (SUC4). This new condition would contain all of the information of the SUC2 displays with the redundant additional displays. An unlimited-capacity parallel model predicts that performance would be better in the SUC4 than SUC2 because you have more time overall to process each stimuli. In contrast, as long as serial processing can only complete 2 stimuli, SUC4 and SUC2 should yield equal performance. Here the critical auxiliary assumption is that processing of SUC4 and SUC2 are similar in different stimuli are processed serially in the two displays. In other words, the assignment of stimuli to the serial process must be equally optimal for the SUC2 and SUC4 condition. This point might have to be reinforced by instructions and perceptual structure. In summary, this comparison allows for a test of this simple serial model. The more general limited-capacity parallel model makes no such specific prediction.

A discussion followed on what sort of stimuli would be best for finding serial processing. John and Alec had been considering a category search task using word stimuli and asked Heather's opinion. She worried that there is some evidence of parallel processing of words (e.g. Stroop, see MacLeod, 1991), but John didn't think the evidence was convincing compared to the evidence for serial processing in reading (Rayner, Balota & Pollatsek, 1986; Starr & Rayner, 2001) and counterarguments for Stroop (Risko, Stolz & Besner, 2005). Asked her opinion about the semantic categorization task and how word lists should be determined, Heather thought it was a fine task and encouraged the control of word length and frequency.

Seminar on the Empirical Foundations of Visual Attention
Questions for Week 3: Divided attention over tasks
John Palmer and Cathleen Moore
12 April 2007

Goals

What are the effects of multiple perceptual tasks? We focus on dual–task paradigms that minimize the role of memory, decision and response. The initial question is to demonstrate an effect of divided attention that is clearly not due to the stimulus or non–perceptual factors such as decision making. Then we consider what conditions yields such dual–tasks effects of divided attention and what conditions do not.

Readings

Bonnel & Hafter, 1998, example study of dual–task effects

Background Pashler pages: 268–271, (skim 271–310), 311–317
Hafter et al., 1998, pursuing the role of memory in dual–task effects
Pashler, 1992, divided attention for intermittently demanding tasks

Study Questions

What are the defining characteristics of the dual–task paradigm used by Bonnel and Hafter?

What is the "attention operating characteristic" (AOC, Figures 1, 3, 5, 6)? What is the independence point? How is it interpreted? How do you plot the single–task controls? How is the comparison between single–task and dual–task conditions interpreted?

What is the sample size model (unnamed in Bonnel and Hafter but see where they cited Luce, 1977)? What are the predictions of the sample size model (p 179)? What is special about d' squared?

What is a "performance resource function" (Figure 7)? What additional information do they provide?

What hypotheses do Bonnel and Hafter consider for the difference found between detection and identification (e.g. perceptual grouping, response criteria, performance level, response interference, trace memory interference)?

Discussion Questions

Bonnel and Hafter used almost identical stimuli for detection and identification. Is there a way to modify the tasks so the stimuli are identical?

What are the prediction of a all-or-none switching model (p. 183)? Also, given this model, consider the predicted shape of the AOC if plotted in terms of proportion correct (e.g. Sperling and Melchner, 1978).

What are the general ways one can distinguish between switching and sharing models using dual-task paradigms?

Might the divided attention effects for identification found in Bonnel and Hafter (1998) be due to memory requirements? For example, perhaps with identification one must maintain a separate memory standard for the two tasks (see Hafter et al., 1998)?

What general tests for memory or decision interference might one propose?

Might a post-cue indicating the relevant task allow one to only report only one task and perhaps eliminate the interference effects found for identification? What would one have to assume about perception and memory?

Might these "continuous" dual-task paradigms miss intermittent effects of divided attention? Maybe there are hidden capacity limits on detection? Can this be tested by extensions of the "psychological refractory period" (PRP) paradigm (Pashler, 1992; 1989)?

How do the capacity demands of attention for some perceptual tasks relate the PRP effect usually attributed to response selection?

What are the pros and cons of the divided attention paradigms that manipulate whole tasks versus those that manipulate the number or presentation of stimuli?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for week 3: Divided attention over tasks
Heather Knapp, John Palmer and Cathleen Moore
26 April 2007

Papers

Bonnel & Hafter, 1998, example study of dual-task effects
Hafter et al., 1998, pursuing the role of memory in dual-task effects
Pashler, 1992, divided attention for intermittently demanding tasks
Background Pashler pages: 268–271, (skim 271–310), 311–317

Discussion Summary

Dual-task methods and analysis. John Palmer began by drawing attention to how the data are displayed in the Bonnel and Hafter paper. They use an attention operating characteristic (AOC). In such graphs, the performance in one task is plotted against performance in a second task. Thus, both axes are dependent variables. Typically, the different points in the figure are generated by varying the relative effort devoted to one task or the other. The graph also typically includes single-task controls shown on the axes. John Palmer explains that it would actually be nice to use 3 graphs to illustrate an AOC experiment. One graph to show performance (e.g. d') as a function of instructed effort in Task A. Another graph to show performance as a function of the instructed effort in Task B; and the AOC graph as a third graph combining information from the first two. For the first two graphs, the attention instruction is the independent variable (e.g. 20%, 50%, 80%, and 100% the single task control) and performance is the dependent variable. Then in the AOC the two dependent variables are combined eliminating an explicit representation of the instruction manipulation.

Heather Knapp expressed amazement that people can allocate their attention so readily according to experimenter instructions. John Palmer agrees that this isn't always possible. He vaguely recalled dual-task papers using highly automatic tasks (e.g. spatial assignment versus color assignment, see Kantowitz, Elvers and Palmer, 1991 Psychonomics) were subjects could not follow instructions to divide their attention. Lynne Werner asks, "but doesn't it *mean* something that sometime subjects can and sometimes they can't?" The rest of us agree, but are unaware of any analysis of when this happens and when it does not.

Returning to the AOC figure, John points out that these figures loose information about scaling as a function of instruction. This is both attraction and a difficulty in understanding the figures. Alec Scharff asks about why this particular degree of 'resolution' of instruction was chosen? Specifically, why such big steps (20, 50 and 80%)? Is it because the ability to allocate attention in this way is qualitative and would not allow finer resolutions? John replied, perhaps yes. But it is also the obvious place to begin in terms of getting the biggest effects possible.

John prompted for some of the other features of the AOC graph. If the two tasks are independent, where should the data fall. Alec answers that the prediction is the "independence point" which is the point that has the same performance as the single-task controls in the divided attention condition. This point forms the upper right corner of a "box" defined by performance in the two single task controls.

Another issue is the representation of the dependent variable. For example, what are the pros and cons of using d' versus proportion correct? Alec replies that some prefer d'. But d' can be misleading. For example, if Figure 5 was plotted as percent correct performance would be very near the ceiling. The d' measure allows for the appearance of a difference that may not be salient otherwise. Lynne Werner agrees and that in these experiments the performance was very high! Perhaps too high to best distinguish the conditions. Keeping performance around a d' of 1 and 2 may be better.

More generally, using proportion correct, allows the easiest a-theoretic understanding of the results. The more derived d' measure can show off specific theoretical predictions such as the sample size model prediction from Figure 1. One can go even further and plot these results as d'^2. That would show off the prediction of a linear function. But, of course this would be at the cost of a unfamiliar measure. John's general preference is to use the easy to understand proportion correct measure combined with curves for each theory.

Lynne points out that a more fundamental reason to use d' is because of the possibility of bias in a yes-no experiment. The d' measure provides a estimate of sensitivity without bias. Alternatively, one could avoid bias by choice of procedure. For example, using a two-interval-forced-choice procedure minimizes bias without further analysis. On the other hand, that procedure has the disadvantage of requiring two stimulus presentations. This might complicate interpretations for some situations. Obviously, there are many pros and cons to consider for any specific application of these ideas.

Heather Knapp asks whether one can combine detection and identification? John described the 2-by-2 paradigm. It requires both a 2IFC detection response and a 2-choice identification response. This paradigm has a number of interesting features and Norma Graham (1989) treats it at some length in her book. For example, if the two stimuli are detected by independent detectors, the detection and identification performance should be equal. For stimuli that affect overlapping detectors, identification performance should be worse than detection. This method can be used to estimate the properties of the underlying detectors.

Sample size model. Bonnel and Hafter discuss this model on the first page of their paper, but they don't call it by name. They attribute it to Luce (1977) but John suggests a better reference is Shaw (1980; Palmer, 1990). John explains the model treats the perceptual problem as if it was a problem in statistical inference. If you have one thing to measure, obviously all of your samples are directed to that one thing. Given independent samples, the standard error of the mean of the samples is the standard deviation of the sample divided by the square root of the number of samples. In contrast, if one has to measure two things, then the samples must be distributed over the two things. Thus, the sample size per item is cut in half when you must perform two tasks. The number of samples per item is half as large, so the standard error is larger by a factor of the square root of 2.

A related point is why the d'^2 measure appears in these theories. It is described in Lindsay, Taylor and Forbes (1968). This paper analyzes signal detection theory from the point of view on information theory.

Alec wonders why one should consider the theory that detection acts like statistical sampling. John argues that if we think of the problem of perception as estimating noisy things, then statistical sampling a natural place to begin. Lynne adds that there are lots of things in perception that may explicitly act like a discrete sampling process. For example, eye movements. But of course, this theory makes the assumption that processing in dual tasks works according to these principles. This case is not obviously a sampling processes. John says he likes the sample size model because it is very specific compared to the proposal of some kind of abstract mental resources. The sample size model makes specific predictions that one can test.

Alec asks whether this analysis is more accepted for duration effects on single stimuli rather than as a mechanism of dividing attention? John and Lynne agree. In particular, experiments with multiple discrete presentations have been analyzed in detail under the assumptions of independent noise and discrete sampling (Chapter 9, Green & Swets, 1966). Similar analysis of duration effects is also common in the literature (Smith, 1998).

All-or-none mixture model. The all-or-none mixture model assumes that only one task can be performed at a time, but one can change to proportion of trials for which you perform one or another task. This idea is associated with any serial processing model. The all-or-none mixture model was not well treated in Bonnel and Hafter (1998). The model predicts that AOC functions using proportion correct have linear tradeoff functions (see Sperling & Melchner, 1978). It also predicts trial-by-trial negative correlations which are mentioned by Bonnel and Hafter. They argue that such correlations are not seen. It would have also been nice to see the AOC plotted using proportions and both the all-or-none mixture predictions (line) and sample size predictions (curvy).

Different results for detection and identification tasks. The departure point for the Bonnel and Hafter experiment was the unexpected results of their prior experiment with two visual tasks: independence for two detection tasks but not for two identification tasks. In the current paper they find a similar result for the pairing of a visual and auditory task! This shows that the effect of detection versus identification is not specific to two visual tasks. What else could it be?

Alec thought the role of transients was pretty convincing. Their stimulus consisted of a pedestal with an added transient that was to be either detected or identified. Thus, perhaps some of these results are specific to transients. Perhaps one can detect that a transient occurred but not identify the polarity.

These issues were pursued in Hafter et al. (1998). In Experiment 2, they turn the pedestal on and off with the transient and manipulate the gap between the different elements of the display. The ultimate idea is to make the transients ineffective for the task. For the experiment described by Figure 7, they think the transients are ineffective and the result is that both detection and identification show a dual-task effect. Thus, one part of the story is probably the role of transients.

Another part of the story is memory. In a final experiment, they rove the standard for each comparison task. In other words, they change the pedestal trial-by-trial so subjects have to base their comparison on the stimuli within a trial and not some memory for a standard consistent across trials. This manipulation makes the task more difficult so they increase the size of the increments and decrements to stay in the same performance range. This effect of roving the standard suggests that memory for across-trial standards is being relied upon. The amazing result is that with these roved standards they find independence between the two tasks for both detection and identification! Their interpretation is that something about the use of the memory standard (or "context coding mode" to use the term from Durlach and Braida, 1969) is causing the dependence. When the roving standard is used, then the stimuli are perceived and compared to a immediate sensory memory independently for the two tasks. Thus, perception itself is independent after all. This should have been a great paper and not a hidden little chapter.

John argued that the literature doesn't acknowledge the importance of memory standards in these sorts of tasks. For example, if you were performing an orientation judgment would you get the same results if you were judging against vertical as with a random angle intermediate between vertical and horizontal. Lynne added that the variability in the memory representation could be greater (or less than) variability in perceptual representation.

Alec asked that we walk through why having a memory standard would lead to dependence? John answered that in all of these tasks one always must compare the current percept to something. If there is only one standard of comparison to remember, it's pretty easy. Or the task is easy if it is the presence or absence of something as in the detection of transients. But if we have to remember two standards, there may be a limit on the precision of such memories. Even going from one to two stimuli. For example, in my own memory studies (Palmer, 1990) requiring memory for memory for two line lengths resulted in worse performance than requiring a single line. In contrast, perhaps a single trace memory from vision and a single trace from audition can be remembered independently. A further complication is that most task probably involve both a sensory trace comparison and a long-term memory standard. You can rely on the one you think is less noisy. In the experiments here, the long-term memory was less noisy than the immediate sensory memory.

Lynne told of an experiment where giving subjects a tone and testing them the next day. Despite the delay, the subjects were still good at making judgments about that tone. John countered that that may have been based on a memory of a categorical judgment rather than a sensory trace. Lynne said the term in audition was "perceptual anchors". John pursued the point: The danger is that with a study-test procedure, s/he can answer just on the basis of the test (i.e., a loud 'test' is just loud) and doesn't really need to remember anything. One way to defeat this strategy is to hold the test constant and vary only study stimulus (cf. Palmer, 1990).

Returning to Alec's question of how the memory limit works: One difference between the single and dual tasks is how many perceptual anchors the subject must remember. The locus of the effect could be at several different stages: encoding, retrieving, or comparing to the anchors. For example, Hal Pashler proposes that memory retrieval is a critical bottleneck based on his experiments with response selection (1989).

John raised a related point. Dosher has a long history of testing dual task performance in noise, but she always assumes a perceptual interpretation and doesn't consider a memory interpretation in any detail. He would like to revisit her results and critically test the perceptual vs. memory interpretation. For example, does something like roving the standard affect the degree of dual-task interference found in her studies?

Intermittent demands between tasks. A final topic is a different challenge to the usual interpretation of these dual-task results. In our discussion above, we questioned the interpretation of dependence as necessarily perceptual rather than due to something else like memory. Thus, dependent dual tasks don't necessarily imply dependent perceptions. Now, consider the case of independent dual tasks. Are there alternative interpretations?

Perhaps the strongest challenge comes from Hal Pashler. His concern is that these typical dual-task experiments have minimal time pressures. As a result, if the tasks make only intermittent demands on a bottleneck, one can schedule events serially and observe independence. Only if you force people to do these tasks in a short period of time, would the intermittent demand show up as a dependence on behavior.

Lynne asked how can one test for such intermittent demands. One way might be with masks. If the stimulus is only available for a short time, one cannot freely schedule around the processing demands. Another approach used by Pashler is to require a memory retrieval and a speeded response at same time, and show interference when they're directly coordinated (e.g. the psychological refractory paradigm, PRP). Pashler (1989) attempted this with a simple perceptual task and found no sign of interference. But a further difficulty was that he had to argue for a null result.

It would be good to find other lines of evidence that support the arguments made by Hafter and colleagues. Perhaps one can combine the auditory-visual task used by Hafter et al. with the simultaneous-successive paradigm. Think of a tone followed by a flash compared to a tone and a flash occurring simultaneously. Is there a successive advantage when using memory standards? Does it disappear when using roving standards?

The details of this proposal were discussed. Lynne remembers a Sorkin paper that might be related. John worried about masking: how would one lay things out over space (two ears) or perhaps use different frequencies instead of different locations? Lynne thought one could have fixed frequencies from trial-to-trial with one of them modulated.

John asked what is the current thinking about auditory frequency bands? Is the processing independent at the front end? Lynne replied that people think that there are different excitation patterns by channel, and measure how much a tone at a given frequency affects a filter at another frequency. Overall, she believes that these authors found independence.

Alec Scharff asked why not eliminate the memory issue by presenting two auditory stimuli at the same time and ask "same or different," then do two visual at same time and say ask "same or different"? In short, does a same-different task help in interpretation? John replies that many of the same issues still apply. Are there transients? Are there long-term memory standards that can help performance? Does one categorize the stimuli? Ultimately, one needs to get the different methods to agree. For example, the dual-task paper of Bonnel and Hafter and the simultaneous-successive paper of Huang and Pashler both conclude that simple feature detection has unlimited capacity.

Closing comments. John summarized his thinking by arguing that the dual-task paradigm is potentially a good method. But it has weaknesses such as a possible sensitive to memory effects. Whether intermittent demands are important is unclear. In a final question, ask yourself which method of studying divided attention would you invest your money. Hal Pashler has invested in the PRP and simultaneous-successive paradigm. Barbara Dosher has invested in the dual-task paradigm.

Seminar on the Empirical Foundations of Visual Attention

Questions for week 4: Time course of selective attention
John Palmer and Cathleen Moore
19 April 2007

Goals

What is the time course of switching selective attention from one stimulus to another? To address this question, we pursue another variation on the brief display experiments previously discussed. Specifically, the interference between two sequential displays is used to estimate the "attentional dwell time" which may be an upper limit estimate on attention switching more generally. We focus our discussion on the interpretation of this effect as due to perceptual or more central limitations.

Readings

Ward, Duncan & Shapiro, 1996, example study of attentional dwell time
Moore, Egeth, Berglan, & Luck, 1996, counterpoint
Background Pashler pages: 93–94, 235–239

Study Questions

What is attentional dwell time? How does it differ from the time to switch attention?

How do Ward, Duncan and Shapiro estimate attentional dwell time?

What do Ward, Duncan and Shapiro propose is the source of the interference effect?

How does the Moore et al. experiment differ?

What is the Moore et al. estimate of dwell time?

Exactly what stimuli and tasks are used in each of the experiments?

Discussion Questions

How sensitive are these estimates to the use of masking?

How general are these estimates to changes in stimulus or task?

What other ways can one consider the generality of this paradigm?

Are the interpretations in these papers consistent with the common assumptions about the simultaneous–successive paradigm?

For each task, does the time course measured reflect perception or perhaps additional processes evoked by the "two–target effect" (Duncan, 1980)?

Do you expect attentional dwell time to depend on whether the first and/or second task required limited or unlimited capacity in its perceptual processing? What do the common theories predict about such dependence on capacity? What about the effect of other aspects of the task such as the memory encoding of a target?

How do these estimates of attentional dwell time compare to estimates of attention switching using cued partial report (e.g. Shih & Sperling, 2002; Sperling & Reeves, 1980)?

How do these estimates of attentional dwell time compare to estimates of attention switching from the attentional blink (Visser, Bischof & Di Lollo, 1999)?

How can these results from studies of the attentional dwell time be reconciled with estimates of serial processes in visual search? For purposes of this comparison, what are the most convincing cases of serial visual search? How long are the estimated processing times?

What parallels might one make between the paradigm used to estimate attentional dwell time and the psychological refractory period (PRP) paradigm?

How has serial processing times been estimated in search from ERP data (Woodman & Luck, 2003) or single unit data (Bichot, Rossi & Desimone, 2005)? How convincing are these estimates?

Is it possible that eye movements and attention shifts have a similar time course for their preparation and execution in sequence?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for week 4: Time course of selective attention
Serap Yigit, John Palmer and Cathleen Moore
2 May 2007

Papers

Ward, Duncan, and Shapiro, 1996, initial study of attentional dwell time
Moore, Egeth, Berglan, and Luck, 1996, counterpoint
Background Pashler pages: 93–94, 235–239

Discussion Summary

Role of masking. Serap Yigit asked how does a mask play a role in the impairment of performance of the second target in Ward, Duncan, and Shapiro? John Palmer started by suggesting that masking is one way to limit performance. By the most common hypothesis, backward masks such as used here are assumed to interrupt the processing of stimulus. Thus, the mask effectively shortens the duration of the stimulus. Another way to limit performance is the use of brief displays. But what does this have to do with the impairment the second target?

To get at this question, John Palmer first reviewed the main result as shown in Figure 3 from Ward et al. According to the figure, when observers are supposed to identify both stimuli, the first stimulus has an advantage in performance. There is also a decline in performance when both stimuli are presented simultaneously. Poorest performance is with a stimulus onset asynchrony (SOA) of around 200 ms. The full recovery of performance takes around 500 ms.

In the paper by Moore, Egeth, Berglan, and Luck, the first experiment is similar to Ward et al’s first experiment except for the details of masking. In their "difficult condition", Moore et al displayed the mask right after the first target (T1) or in their "easy condition", they delayed the mask until the onset of the second target (T2). Some decline in performance occurs for both easy and difficult conditions. But for the easy condition where there is no mask after T1, performance on T2 is less impaired.

We raised another detail difference between experiments. In the Moore experiment, the mask stays on the screen until the end of the trial, while in the Ward experiment the mask disappears 250 ms after onset. Since the results seem similar between the two experiments (for Moore's difficult condition), keeping the mask on for a long time probably does not further impair performance.

The standard view is that a mask interrupts processing. If ending the processing of T1 early by masking was to limit interference with T2, then Moore's easy condition should have resulted in greater interference. But it actually resulted in less interference. So that cannot be the explanation.

Another possibility is a version of the integration masking hypothesis. By this hypothesis, masking adds noise to the decision process. By reducing the signal-to-noise ratio, one decrease performance. Masking should act like any other manipulation of task difficulty. If making T1 harder increases the interference, then that would predict the effect seen in Moore's experiment. One could also apply this thinking to the interruption hypothesis. Assume the effects are mediated by reducing information from the stimulus rather than terminating processing early.

Yet another possibility is that the mask is processed as if it was a target stimulus. If this mediated the effect, removing the mask from T1 would improve performance as observed in Moore's experiment. But, in the extreme, it would predict a much larger effect when the mask are kept on till the end of the trial as in the difficult condition of the Moore experiment. Another problem with this hypothesis is that the interference effect goes away in the condition where one has to respond to only one stimulus. Thus, it isn't the mask per se that causes the interference, but its action as a decoy for a relevant target.

Based on these points, John Palmer summarized three possible hypotheses for the role of masking in these papers:

a) masking by interruption,

b) masking by integration, or

c) mask as a distractor stimulus.

Alec Scharff wondered how much of the effect described by Moore is due to the details of their procedure for fitting the data. John Palmer argued that it wasn't critical. But one must recognize that Moore didn't have the control condition of attending to only one task that made the effect in Ward easy to see. To make a comparable analysis, must assume a constant baseline against which to compare performance. This seems reasonable given the constant performance found in Experiment 1's control (but not Experiment 4). We were a little confused by comparing Moore's easy and difficult conditions. But that isn't the right comparison.

Temporal uncertainty. Lynne Werner pointed out that in all of these experiments, the observers do not know when T2 will appear. In other words, different SOAs are intermixed randomly, so observers cannot predict when it will appear. In short, there is temporal uncertainty about T2. John Palmer agreed and contrasted this with the simultaneous-successive paradigm in which observers typically know the stimulus sequence exactly. Might temporal uncertainty contribute to these effects?

Parallel and serial hypotheses. Serap asked if Ward et al. are assuming parallel processing of their stimuli. John agreed that they mention parallel processing and at the very least imply that is the preferred account.

But does this mean that if the processing was serial, the same effect would not be seen? John argued that there is also a perfectly fine serial interpretation of the effect. Suppose one presents two stimuli with some SOA. If the SOA is long, then the first stimulus appears and it is processed to completion. Next, the second stimulus appears and it is also processed to completion. For a short SOA, such as 100 ms, the processing of the second stimulus is affected. Assuming serial processing, observers must finish processing the first stimulus, which will delay the beginning of processing of the second stimulus. This is similar to the bottleneck of the PRP paradigm. Now consider the role of masking. The first stimulus is processed until interrupted by the mask. Then one must switch to processing T2. The results suggest that this switching time is quite long.

Generality to conditions with likely parallel or serial processing. John wondered whether the results would be the same if one measured conditions that clearly required serial processing or clearly can be done by unlimited-capacity, parallel processing. The current experiments all involve letters which are a debated case. Some experiments find results supporting serial processing while the others find results supporting unlimited-capacity, parallel processing. For example, Heather Knapp's dissertation and Pashler and Badgio (1987) are consistent with letters being processed by an unlimited-capacity, parallel process. But there are other simultaneous-successive experiments that do not come to the same conclusions (Kliess and Lane, 1986). Would we get the same results as Ward et al if we used words? Would we get the same result if we used luminance increments? Probably not.

Relation to the attentional blink. Alec asked what is different between this paradigm and the attentional blink. In the attentional blink paradigm, rejecting distractors does not cause a blink. The effect on the second target (T2) is specific to processing a target in the first task (T1). There are many distractors on all of the other RSVP displays that do not seem to limit performance. In contrast, Ward et al found that distractors do cause a performance impairment.

Experiment 4. Experiment 4 of Ward et al. is a yes-no search task. The target is the letter 'L' and the distractors are L's that are rotated 90 degrees to the left or the right. Four possible display locations are grouped into two sets: for the vertical set the stimuli appeared either above or below fixation; for the horizontal set, they appeared either to the left or right of fixation. For each trial, one stimulus appeared in one of the vertical locations and one stimulus appeared in one of the horizontal locations. For the 'attend both' condition, both stimuli were relevant to the search task. For the 'attend one' condition, the target can appear in only one of the two sets of locations.

After some confusion, John figured out that this is not a foil experiment because they never put a target in an unattended location.

The key results of this experiment are shown in Figure 9. This figure plots the percent hits as a function of SOA. Negative SOAs are for cases where the target was first. There is essentially no effect on either the attend-one or the attend-both conditions for these negative SOAs. The interesting case is for the positive SOAs where a distractor was presented before the target. For the attend-both condition there is a decline in performance for simultaneous presentation and even more for short positive SOAs (100-300). Some effect persists even out to 600 ms or beyond. If the effect was specific to targets, there should have been no effect of a distractor preceding the target in this experiment.

One complication to this story is that there is also an effect of SOA for the attend-one condition. One must argue that the instructions to ignore some locations were not followed perfectly.

What makes this interference effect different than the attentional blink? Obviously, distractors have no effect on the blink task. There are differences in the paradigm. For example, in the Ward et al task, observers never attend to the same location while in the RSVP procedure of the blink paradigm they attend to the same locations over and over again. The Ward et al paradigm is closer to the sequential-simultaneous paradigm.

Alec asked what is interfering for the case where distractors are presented before the target? Is it that observers do not have time to shift their attention from where the distractor appeared? Maybe this makes sense for a task with limited-capacity processing. However, in many search studies, observers seem to process distractor stimuli very lightly. They certainly don't remember them. Alec suggested that if distractors can be rejected pre-attentively, then one should not find performance impairment on the target in Experiment 4. Again we don't know what would happen if simpler stimuli were used.

John wondered how the effect of distractors was eliminated for the attend-one condition. Such selection would be easy to account for by an early selection theory. However, Duncan usually favors a late selection theory. From this point of view, the irrelevant location would still be processed and attend-one should also yield interference effects. Another puzzle.

How can one improve this design to compare target processing with distractor processing? If there were two targets instead of one, then one could compare the effect of a first target vs. a distractor on a following target. Perhaps the interference would be even larger with a target?

John wished that Moore and colleagues had conducted experiments similar to Ward's Experiment 4 but without masking. It seems just possible that the interference effect might disappear for this combination of conditions.

More general discussion. Serap asked what was the "two process account" raised at the beginning of the discussion section of Ward et al. John answered that he considered it as a kind of catch-all hypothesis. Perhaps there is both a high-speed switching phenomena underlying set-size effects in visual search and some slower interference effect that was measured here. Clearly experiments like Experiment 4 that begin to minimize the differences between this paradigm and traditional search paradigms are needed to link the two phenomena.

Serap asked if there are ERP studies about attentional dwell time. John Palmer cited the Woodman and Luck (2003) studies trying to demonstrate serial processing from ERP data. They estimated that 100 ms. required to switch from one target to another. Woodman is currently pursuing similar studies, but we do not know much about them. Another line of relevant physiology is direct measures of attention switching by Herrington and Assad (2007, SFN talk).

John indicated that eye movements are also relevant to attention switching. If one moves his/her eyes form one stimulus to another in a search task, it is a very good candidate for serial processing.

About the generality of this interference effect, John said that Moore et al paper makes it clear that there is the interference effect is not of the same magnitude for all tasks and stimuli. John also pointed out a particular pair of data points in Figure 3. It seems that conditions with SOAs of 0 and 500 ms have the same amount of interference. This made John and Alec to think twice about their simultaneous-sequential experiment that used just those values. They had assumed that 500 ms was long enough to avoid any lingering interference effect.

Serap briefly described her ERP study on attentional blink. She said that the blink was reduced when subjects listened to music while they carried out the task. One explanation was that processing of the music diluted the commitment of resources to T1. Perhaps less commitment of resources to T1 results in less interference?

John said that perhaps covert attention switching phenomena is similar to executing a sequence of eye movements. Attention shifts and eye movements might require the same time to prepare and execute. He thought this a good default hypothesis.

Seminar on the Empirical Foundations of Visual Attention
Questions for week 5: Time course of divided attention
John Palmer and Cathleen Moore
26 April 2007

Goals

What are the effects of processing multiple stimuli at the same time versus processing them in sequence (parallel vs. serial processing)? Here we consider comparisons of parallel and serial models using the *speed-accuracy-tradeoff* (SAT) methods championed by Barbara Dosher. She uses a response-signal paradigm to measure accuracy at different times following the presentation of a stimulus. Thus, this can be applied to a brief display to measure how much performance improves with time. We will read two papers, one supporting parallel processing for two kinds of visual search tasks; the second supporting serial processing of retrieving order information from short-term memory. (We know of no SAT studies of visual stimuli that support a serial model.)

Readings

Dosher, Han & Lu, 2004; SAT experiments supporting parallel model of search
McElree & Dosher, 1993; SAT experiments supporting serial model of memory
Background Pashler pages: 116-124, 135-136, 141-147, 275-284

Study Questions

What is the response-signal paradigm for measuring the speed-accuracy tradeoff?

What are the characteristics of SAT functions shown by Figure 1?

What is the probabilistic serial search model (Figure 3)?

What is the probabilistic parallel search model (Figure 4)?

What are search asymmetries and how are they measured in the Dosher paper (Figure 5)?

How do the fits of the serial and parallel models compare (Figures 9 & 10)?

Turning to McElree and Dosher (1993), what order memory task are they evaluating (Figure 2)?

How do the SAT functions of Figure 7 in McElree and Dosher (1993) qualitatively differ from those observed in Dosher et al. (2004)?

Discussion Questions

How do Dosher's conclusions about simple visual search being parallel compare to the conclusions drawn from other approaches to simple visual search? With whom do they disagree?

How does the response-signal paradigm used by Dosher and others (Reed, 1973) compare with other SAT methods (e.g. Wickelgren, 1977)?

What are advantages and disadvantages of analyzing the SAT using an x-axis of lag time plus response time (also called "total processing time")? How do we interpret effects on response time in this situation?

For SAT data, can one derive and test more general properties predicted by parallel and serial models?

How is performance with a response-signal paradigm related to varying the duration of the stimulus and measuring accuracy?

The visual search SAT experiments in Dosher et al. were done with brief displays (e.g. 100 ms). How can one generalize these experiments to longer duration displays and cases with eye movements?

How can the SAT paradigm and analysis be generalized to response time (e.g. Thornton & Gilden, 2007)?

Does the SAT method show more promise for distinguishing parallel and serial processes than other methods devised for conventional response time experiments (e.g. Townsend & Wenger, 2004)?

More specifically, how are the shapes of the SAT functions related to the shape of a response time distribution and the shape of the conditional accuracy function (accuracy conditional on the response time)?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for Week 5: Time course of divided attention
Alec Scharff, John Palmer and Cathleen Moore
17 May 2007

Papers

Dosher, Han & Lu, 2004; SAT experiments supporting parallel model of search
McElree & Dosher, 1993; SAT experiments supporting serial model of memory
Background Pashler pages: 116-124, 135-136, 141-147, 275-284

Discussion Summary

Motivation. The motivation of the first Dosher paper is the desire for a method of distinguishing serial and parallel processing. Serial and parallel processing models make predictions about the time course of stimulus processing. These papers claim that serial processes have an slower initial time course than a parallel model with a similar mean completion time. This paper goes beyond previous accuracy studies because it takes into account both processing time and accuracy. While fairly difficult, this paper is probably easier to understand than the corresponding response time models (e.g. Thornton and Gilden, 2007).

Response signal paradigm. Brief displays were followed by a variable cue delay before a response signal. In these experiments, the cue was a tone that indicated the subject must respond immediately. The subjects received feedback on response time to the tone, and attempted to minimize this response time. In the first paper, this paradigm was applied to a typical yes-no visual search task. In previous work, Dosher and other colleagues applied it to memory search tasks.

Heather Knapp wondered whether the response signal can influence the subjects' strategy. Might subjects change what they are doing when they know that they will be interrupted? John Palmer described Meyer, Irwin, Osman and Kouios (1988) in which response signal and response time trials were randomly intermixed. In this mixture of conditions, subjects cannot expect the tone on every trial and are instructed to try to respond accurately as in a typical response time task. But if they hear the tone, they must respond immediately. There are pros and cons to this task. The major advantage of this task is that it puts subjects in a common state of readiness for both tasks. John also pointed out that in many psychophysical tasks performed at threshold, the subject is asked to respond with incomplete information, which only a small step from guessing. In this context, John didn't think a response signal would necessarily disrupt a process that is already akin to guessing at times. Dosher has criticized this variation on the response signal task but John doesn't remember the reasons.

Speed-accuracy tradeoff functions. Figure 1 illustrates the approach of describing the speed-accuracy tradeoff (SAT) functions using parameters of asymptote, rate, and intercept. The ‘dynamics’ of the curve refer to the intercept and rate. The asymptote indicates the accuracy with unlimited time. An experimental manipulation can be assessed in terms of the parameters that it affects.

Figure 2 illustrates the predicted set-size effects for a simple serial search model. A larger set size is more difficult, as indicated by the lower d’ asymptote. In addition, the time course of the rise to the asymptote is also slower for the larger set size. This decline is more than a proportional drop due to the asymptote. This illustrates how the dynamics differ for the serial model. A parallel model yields functions with different asymptotes but more similar dynamics.

Equation 12 provides the three-parameter description of the SAT function. It has an initial delay followed by an exponential approach to the asymptotic accuracy. One parameter for the initial delay, one for the exponential "rate" and one for the asymptote. The exponential form of this equation is a reasonable starting point, as it reflects the unique property of a process that is "memoryless". In other words, it's future change can be determined by its current state without any "memory" for its prior states including the orignal state.

Set-size dynamics. The key claim in the paper is that parallel and serial models make different predictions about the dynamics of SAT functions. Parallel models predict that dynamics change little as set size increases. Serial models predict that the dynamics do change. In particular, the effect of set size is relatively larger at short processing times because there is delay during which the observer has no information about a subset of the stimuli. On initial analysis, SAT curves of target-absent trials show off the largest effects on the dynamics. It is not so clear whether the larger changes in the dynamics are in the intercept or rate parameters.

An interesting question not taken up in the Dosher paper is how does a simple manipulation of discriminability affect the SAT functions. The simplest idea is that only the asymptote of the SAT function changes in response to varying the difficulty of discrimination (e.g., by narrowing the gap in the C in the C-among-Os search task). Alec asked if the intercept wouldn’t change as well. From his response time modeling paper with Huk and Shadlen, John was confident that one estimate of nondecision time (also called ‘residual time’) was sufficient for all levels of difficulty within the same task. So at least for such simple motion discriminations, it seems plausible that SAT dynamics would not change with difficulty. This basic question needs to be addressed experimentally.

John critiqued Dosher's decision to combine the cue delay and the response time into one variable that she called the "total processing time". By Dosher's admission, response time was not completely independent of the cue delay. Typically, response time is a bit longer for the shortest cue delays. Moreover, John warned against replacing an independent variable with a variable that included a part that was dependent on the response. This undermines the independence assumptions necessary for almost all approaches for modeling. For example, a regression model describes the effect of "independent" variables on the dependent variable. In general, it cannot be applied to the relation between two dependent variables. Despite these complaints, this issue will probably not change any significant interpretation.

John mused that perhaps the effect on response time was due to a detail of the design of these experiments. Dosher used equal probabilities of cue delays. Under such a design, the conditional probability of the response signal increases as the delay increases. Thus short delays are in this sense less likely so the subject in not as prepared for them. In experiments that vary the foreperiod of a response time discrimination, this has the consequence of producing results with longer responses at short foreperiods (see Luce, 1986). This effect can be reduced by increasing the probability of the short delays. The 'poor man's version' of this design is to use cue delay times that have intervals that increase proportionally over time (e.g., 20 ms, 40 ms, 80 ms, 160 ms, etc.) so that the conditional probability per unit time is equal. Indeed, Dosher used irregularly spaced cue delays that approximated this pattern (0, 0.05, 0.15, 0.30, 0.50, 1.15, 1.80 s). This is not far from the regular pattern: 0.05, 0.1, 0.2, 0.4, 0.8, 1.6 s. But perhaps this modest increase in the number of short delays would increase the response time for those conditions.

Results. Figure 8 reveals similar dynamics for the two set sizes. This was found for both O–in–C and C–in–O searches. This result is consistent with a parallel search model. John noted that it would nice to see a standard error of the rates for the different set size conditions in order to make it completely clear that they were not reliably different and are sensitive enough to reject the magnitude of effect expected from a serial model.

Results consistent with the serial prediction of changing dynamics between set sizes can be found in the McElree and Dosher paper. In this task, subjects saw consecutively presented letters and then had to respond which of a pair of letters had been presented more recently. Figure 7 in this paper reveals that dynamics do indeed slow with the serial position in the list. In particular, compare Panel A that had the most recent test item at the end of the list with Panel D that had the more recent test item near the start of the list. This paper also included an item recognition memory experiment where set size had little effect on dynamics (Figure 10).

Figure 7 in the second Dosher paper was not easy to read. More convincing information can be extracted from the relevant table of estimated parameters. But perhaps it would also be more effective if the curves were also shown scaled to the asymptote.

Detailed models. Figures 3 and 4 of the first Dosher paper are schematics of the serial and parallel models. In both cases, each stimulus discrimination is associated with a gamma distribution of completion time and a probability being correctly resolved. Both are "race models" in the sense that the first 'yes' state generates a response. A "no" responses occur when all processes fail to find a target. The time course to process an individual stimulus is characterized by a right-skewed gamma function. Gamma functions are the sum of component exponential distributions and are a common starting point for simple processing time models. If a response is demanded before the observer moves into either the 'yes' or the 'no' state, the subject guesses with some fixed probability whether there is a target present or not. The models differ in that processing of all stimuli occurs concurrently for the parallel model, while each stimulus is processed one after another in random order for the serial model.

Seminar on the Empirical Foundations of Visual Attention
Questions for week 6: Selective attention in memory
John Palmer and Cathleen Moore
3 May 2007

Goals

What are the effects of selecting information from one stimulus and not another for explicit memory tasks? Again we focus on an idealized laboratory paradigm: brief displays, a few widely separated stimuli, single-attribute discriminations and accuracy measures. The initial question is to demonstrate an effect of attention that is clearly not due to the stimulus. We then consider papers that argue for perceptual versus memorial accounts of this selection process.

Readings

Palmer, 1990, especially Exp 7; cueing of a memory task with simple stimuli
Rock & Gutman, 1981; especially discussion on perceptual interpretation
Moore, 2001, especially commentary on percept. versus mem. interpretations
Background Pashler pages: 55-60; background on memory 319-356

Study Questions

What is the memory paradigm in Palmer (1990)?

How are set-size effects measured in Palmer (1990)?

How are cueing effects measured in Palmer (1990)?

What is the perceptual interpretation of such effects as proposed by Rock and Gutman (1981)?

What is the memory interpretation of such effects as proposed by Palmer (1990)?

How does Moore (2001) describe this distinction?

Discussion Questions

What hypotheses can be ruled out by showing that set-size effects are consistent over the entire psychometric function (Palmer, 1990, Experiment 2)?

What hypotheses can be ruled out by showing that set-size effects are consistent over different retention intervals (Palmer, 1990, Experiment 4)?

What hypotheses can be ruled out by showing that set-size effects are consistent for experiments that vary the display or uses cues while holding the display constant (Palmer, 1990, Experiment 1, 5 and 6)?

What can one make of the exceptionally poor performance for the invalidly cued stimuli in Experiment 7 of Palmer (1990)?

More generally, how might one distinguish perceptual and memory interpretations of cueing effects in explicit memory tasks such as in Palmer (1990)?

Can one distinguish between memory limits due to component memory processes such as encoding, storage and retrieval?

What is the role of conscious awareness in the debate between perceptual and memory interpretations (see Moore's discussion of Mack and Rock)?

How does selection affect indirect measures of memory (Moore & Egeth, 1997; Moore, Lleras, Grosjean & Marrara, 2004)? Do these measures allow one to better distinguish perceptual and memory interpretations?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for Week 6: Selective attention in memory
Heather Knapp, John Palmer and Cathleen Moore
24 May 2007

Papers

Palmer, 1990, especially Exp 7; cueing of a memory task with simple stimuli
Rock & Gutman, 1981; especially discussion on perceptual interpretation
Moore, 2001, especially commentary on percept. versus mem. interpretations
Background Pashler pages: 55-60; background on memory 319-356

Discussion Summary

Cueing experiments in memory. We first considered the sixth experiment in Palmer's (1990) visual memory paper. In Figure 10 of that paper is a measure of cueing effects on a short-term visual memory task.

Serap Yigit asked about the difference between the terms "compound" vs. "concurrent" tasks cited in the paper. John Palmer explained that this terminology comes from reviews by Sperling and Dosher (1986). In a typical concurrent task, divided attention is manipulated by comparing single and dual task as discussed a few weeks earlier in this seminar. In the dual task condition, more than one task is done at the same time, but for each component task there is an independent one-to-one relation between each the stimulus and response. In contrast, a compound task manipulates the number of stimuli while not varying the number of possible responses. The prime example is a visual search task in which a variable number of stimuli map into a fixed set of responses. The complication cited in the Palmer (1990) paper is the role of the cue. For these experiments, the cue provides a more complex relation between the stimulus (and cue) and the response that is consistent with the compound terminology. Once the cue is accurately perceived, the "compound" nature of this task is removed.

We next had a brief discussion of whether cues are really doing what we think they are doing. Alec Scharff suggested that there are two chances for an influence from cuing—both the presence of the precue itself and the influence of the precue on the relevant target. For example, the cue may reduce temporal uncertainty and/or spatial uncertainty. John replied that the key ingredient is the nature of retrieval in these memory tasks. If memory retrieval requires a search process, it may be a compound task, but if there is direct access to the memory then the other components are not relevant. The usual interpretation of Sperling's partial report paradigm is that one can directly access the relevant information by location. Thus, the cue allows this direct access and eliminates competition from memories of other stimuli.

Unfortunately, there is little direct evidence for this account over a search account.

Alec wondered about using a change detection paradigm. John hadn't thought about this option but is enthusiastic. Alec pointed out that change detection is a kind of a recognition. John said it's pretty unconstrained, though. That may make it more like a recall task. One has to remember everything from the stimulus in order to perform the task. Alec asked what if you just cue one portion of display such as the upper left-hand corner. Then one can compare: precue in display 1, precue in display 2, post cue in display 1, post cue in display 2. John recalled about a talk by Melissa Beck at VSS in 2005 (Beck, Peterson, Angelone, in press). He believed that she tested change detection for familiar things vs. unfamiliar things, and found that subjects detect a change better with familiar than unfamiliar items. She systematically laid out the alternative hypotheses that might account for this finding including distinctions about retrieval.

Serap asked about effects of set size and retention interval as measured in Experiment 4 of Palmer (1990). She wondered why there isn't a difference due to the retention interval if memory is the underlying constraint on performance. John pointed out that the observed retention interval effects are relatively small, and that the main points are that the set-size effect is present across a range of intervals, and that the log-log slopes are similar on these log axes. This suggests that the loss of information occurs early, and doesn't accrue over time as a function of load. In other words, it is consistent with the cue affecting the encoding into memory rather than storage.

One possible complication is a categorization strategy. Can one reduce the load on memory by recoding an analog memory to a discrete categorical memory representation. In short, can one use verbal STM to mediate these memories? If so, this may not be a fair test of the characteristics of visual memory.

Serap wondered if John could say more about switching versus sharing as discussed on page 347. John replied that switching versus sharing is about whether attention must switch between different stimulus representations, or whether it can be shared concurrently across multiple stimulus representations. These alternatives have been developed in the context of alternative decision models (e.g. Shaw, 1980). Palmer (1990) illustrates a analysis to distinguish these possibilities in Figure 14. In this task, a subject has 4 items to remember. Suppose the subject has x probability of remembering one item. The question is whether one can distinguish the following two models:

a. A model with three slots and complete information about three of the items, but nothing about the fourth item.

b. A model with some, perhaps incomplete information about all four of the items.

For model (a), the maximum possible proportion correct is .875 (1.0 for 3 items and .5 for 1). Observed performance is better than this bound. So it must be the case that some info is obtained about all 4 items.

At this point, Steve Luck's poster at VSS was briefly discussed. John described a similar analysis that showed partial information was obtained for up to 4 items but not for larger numbers such as 8. Alec asked whether that result is general for all kinds of stimuli. John said that Luck was studied color; Palmer (1990) looked at several simple attributes; and Awh and colleagues have examined letters, Chinese characters and certain confusable geometric forms. Alec pressed on whether we were sure that the magic number 4 held up over all of these case? John wasn't sure. He thought it possible that one can choose to encode one item very well at the expense of any info among other items. Alternatively, we thought that Awh's interpretation is that detailed information is encoded about some items and coarse information about others. Thus, some information about all four always gets through. Experiment 7 in Palmer (1990) cast doubt on Awh's interpretation.

John talked about extensions to Palmer (1990) that pushed the task to 8 items and then to 32 items. The task became impossible to interpret at 32 items because of evidence for sensory limitations (e.g. crowding).

Alternative hypothesis. To switch gears to a discussion of the possible hypotheses, we listed several on the board. Each of these aim to account for aspects of when one can select or fail to select the relevant stimuli to preserve in a memory task.

Perceptual Hypotheses

a. **Perception of features**. Does selection control what features are processed? This would be an early version of early selection.

b. **Object recognition or categorization**. Does selection control which stimuli are categorized.

c. **Sensory interactions**. Does selection fail because before one can select there is already some kind of sensory interaction. Examples may include crowding, configural effects between stimuli, texture judgments that combine stimuli. This possibility can be controlled using identical stimuli and manipulating selection using cues for subsets of the stimuli.

d. **Perceptual awareness (consciousness).** Does selection in perception result in awareness? If so, such awareness measures are relevant to understanding selection.

Memory Hypotheses:

a. **Encoding**. Does selection control what is encoded into a durable memory?

b. **Storage**. Does selection act to improve storage of some items over others?

c. **Retrieval**. Does selection allow one to retrieve some items and not others?

Other Hypotheses

a. **Decision**. Does selection modify the inputs to the decision processes? It probably does, but one can minimize this contribution using post cues that follow perception and memory but precede the decision based on the memory.

John described an experiment by McLean, Palmer, and Loftus (unpublished) that tried to compare capacity limits (and selection) in a search and memory task. For the search condition, a target was presented first (e.g. letter A) followed by a study array of letters (e.g. ABCD). For the memory condition, the order of the two displays was reversed. First one saw the study array, then the target (now termed "probe"). Thus this was comparing a search task to a probe recognition task. Our surprise was that the capacity limits were very similar for the two tasks. We expected much larger set-size effects with the probe recognition task.

In a follow-up experiment, McLean (1999) used simpler visual stimuli (geometric forms varying in size, shape, position and contrast). In this study the set-size effects were larger for probe recognition than for visual search. This is consistent with a bottleneck at memory rather than perception. Alec asks what is it about stimuli that resulted in this effect? John said he suspects that something about categorization is critical. Letters are processed much more categorically than typical visual materials.

Alec suggested that one could exploit the simultaneous-successive paradigm in memory. If the limit is on encoding, it would be an advantage for the successive over the simultaneous condition. John thought this is a great idea and doesn't think it has been done. But he also pointed out that most everyone assumes that memory encoding is serial so an advantage for successive would not be surprising. But it would be really surprising to see otherwise.

Serap pointed out that priming without awareness is an often reported finding. Such priming affords perception, but no memory. John replied that this goes along with the idea that some information gets in but one isn't necessarily aware of it. Alec described the Moore and Egeth paper (1997) in which they ask if potentially unattended stimuli modify size judgments that were attended. The experiment used the Muller-Lyer illusion among "texture" elements in the background of a display with other stimuli to which subjects were attending. He summarizes the Rock catch phrase: “There’s no awareness without attention”. But might it be better to say there’s no awareness w/o short-term memory?

John talked briefly about implicit memory: Are we more likely to find a word in an anagram if we’re primed with that word. Serap pointed out that you can learn much and not be aware of it. Memory for musical phrases are often learned, for example, but not remembered consciously (or able to be retrieved) until they’re presented again, and recognition turns into recall. Alec: Yes, perceptual learning, right? When you learn things and you don’t know you’ve learned them. John: there’s this interesting hierarchy of what it takes to learn something, what it takes to remember something, and what provides savings when you learn it again.

Serap asked, so what about the effects of using a Christmas tree shape and other familiar figures in Experiments 3 and 4 of Rock and Gutman (1981)? John said he thinks both perceptual and memory interpretations remain viable. Perhaps if you see it and don't immediately encode it, the probability of eventually encoding it into memory is more likely because it’s a familiar object. But then he waxed on about oddball stimuli being more memorable. Perhaps the key is not familiarity, but something novel.

DeSchepper and Triesman (1996) pursued the stimuli used in Rock and Gutman to further examine the fate of unattended information. The basic idea is that when a ignored item on a previous trial appears as a target in a subsequent trial, it is more difficult to access that representation. This is an example of negative priming. Loula, Kourtzi and Shriffrar (2000) have pursed this further.

Seminar on the Empirical Foundations of Visual Attention
Questions for week 7: Divided attention in memory
John Palmer and Cathleen Moore
17 May 2007

Goals

What are the effects of attending to multiple stimuli for explicit memory tasks? Again we focus on an idealized laboratory paradigm: brief displays, a few widely separated stimuli, single-attribute discriminations and accuracy measures. We consider on how memory depends on both the number of objects and sensitivity to the relevant discriminations. A key further question is whether the capacity is much more limited for memory compared to perception.

Readings

Luck & Vogel, 1997; memory as a function of the number of objects
Palmer, 1990; memory as a function of sensitivity to the relevant attributes
Background Pashler pages: 102-109; background on memory 319-356

Study Questions

What is the memory paradigm of Luck and Vogel (1997)?

How are set-size effects measured in Luck and Vogel (1997)?

How are contributions of verbal codes estimated and minimized?

How are effects of decision minimized?

How is the contribution of perception estimated and minimized? Are you convinced?

Can they compare performance for different features?

How do they compare memory for features and conjunctions?

Turning to Palmer (1990), how are set-size effects measured?

Luck and Palmer interpret the limits on memory very differently. What are the differences in their experiments that might contribute?

Discussion Questions

In each paper, how do they model the memory limits?

In each paper, are the limits on memory consistent with sharing or switching accounts?

In each paper, are the memory limits due to memory encoding, storage or retrieval?

Can these results be combined in a model that takes into account both the number of objects and the resolution needed for the discrimination of particular attributes?

How do these limits on memory compare to those found in search tasks? In particular, are there attributes for which perception has unlimited capacity while memory shows a fixed capacity limit of one kind or another? Both Palmer and Luck touch on this point. Several unpublished Palmer papers also pursue this point of comparison (Palmer, McLean & Loftus, unpublished; McLean, 1999).

What are the effects of perceptual structure on the observed memory limits? What kinds of theories predict these effects?

How do these limits on memory compare to those found in response selection and memory retrieval tasks?

How can one generalize from these idealized conditions? In particular, consider measurements across eye movements?

Seminar on the Empirical Foundations of Visual Attention
Summary of discussion for Week 7: Divided attention in memory
Serap Yigit, John Palmer and Cathleen Moore
31 May 2007

Papers

Luck & Vogel, 1997; memory as a function of the number of objects
Palmer, 1990; memory as a function of sensitivity to the relevant attributes
Background Pashler pages: 102-109; background on memory 319-356

Discussion Summary

Capacity in terms of objects or features? Luck and Vogel (1997) argued that visual working memory stores integrated objects instead of single features. Specifically, they manipulated both the number of objects and the number of relevant features within an object and found that the errors were determined by the number of objects rather than the number of features. Alec Scharff raised concerns that Luck and Vogel's results have not been replicated by others. He didn't remember the details. Possibly they don't generalize to other attributes and objects. The most specific concern is that Luck and Vogel always used very easy discriminations between features. Their results may well not generalize to stimuli that are more difficult to discriminate.

Luck and Vogel found an effect of set size when they increased the number of objects but not when they increased the number of features and held the number of objects constant. For comparison, Palmer's (1990) experiments do not address this specific point. Palmer did not vary the number of features independent of the number of objects. He did vary the number of features and objects together and found large effects of set size even over the range of 1 to 4 objects/features. This result differs from that in Luck and Vogel. The likely critical difference between the two experiment is the level of discriminability.

John described new results from Luck that appear to resolve this discrepancy. These were described to him by Luck at VSS this year. Luck has measured color match values for a cued stimulus as a function of set size. These matches can be analyzed in terms of the variability near an accurate match versus random responding. He found that varying set size from 1 to 4 increased variability near the correct match. But increasing set size beyond 4 did not increase this variability any further. Instead, the proportion of random responses increased. He interpreted this as evidence for limited feature memory for high resolution information for fewer than 4 objects AND a limit on the number of object that can be remembered at all. This interpretation is consistent with the results in both Luck and Vogel (1997) and Palmer (1990).

John briefly mentioned that he conducted a follow-up study of Palmer (1990) with Eric Ruthruff when Eric was an undergraduate at UW. John remembered that they generalized the threshold versus set-size result to 8 stimuli. After hunting up the old data, he found they did measure entire psychometric functions for set size 8 in the fashion of the Palmer (1990) experiment. Indeed, at least for a subject or two, performance reached 95% correct or more for very large stimulus changes. On the face of it, this means that 7 to 8 stimuli must have contributed to performance. This would be an exception to the limit of 4 objects suggested by Luck and Vogel.

Role of verbal memory. One of the questions that Luck and Vogel asked in their paper is whether verbal memory mediated the results. To be able to rule out this possibility, they added an additional verbal task that presumably occupied verbal memory. Specifically, at the beginning of each trial subjects had to hold two digits in memory through the end of the trial and report them. Alec was not satisfied with the number of digits that they used. He thought this may not completely prevent people from using verbal memory. John said this is a common procedure. People use tasks such as counting backwards by 3 to prevent the use of short-term memory. More generally, even if this didn't prevent the use of short-term memory, one would expect an effect of this verbal task if verbal memory was used for the visual task. In fact, there was no apparent effect of adding the memory load. Nevertheless, it would be good to look more closely at studies that have investigated the degree of independence between visual and verbal short term memories. In particular, they would have been more convincing about their verbal memory control if they had showed an effect of the verbal load on a additional verbal task (see next paragraph for an analogous argument).

Role of perceptual limits. Serap Yigit said that she was not satisfied with the way Luck and Vogel ruled out perceptual constraints as a possible contributor to their observed set-size effects. They increased the display duration of sample stimuli from 100ms to 500ms. They argued that such a manipulation would affect the observed set-size effects if perception limited the number of objects that could be reported. John said that he also wasn't satisfied with this control condition. To be able to rule out a perceptual contribution, they need to show they have an effective manipulation of perception. Specifically, one would have to demonstrate an effect of increasing duration on performance in these tasks. Duration is a poor choice because with these highly discriminable feature judgments, duration would have no effect. Indeed, duration often has effects only if conditions are relatively special such as by adding dynamic external noise. Thus, manipulating duration is a particularly poor choice for this control. They must show an effect of the perceptual manipulation.

Serap was concerned that subjects might not be judging stimuli individually but rather judging the configuration of multiple stimuli as a whole. In other words, might there be effects of grouping stimuli? John agreed that such grouping effects are possible. However, Luck and Vogel did do some things to minimize this possibility such as randomizing the locations of the individual stimuli. By presenting stimuli in different configurations for every trial, systematic effects of grouping are probably reduced. The use of identical displays and precues to vary set size would be a further way to investigate this possibility.

Is location special? Alec asked if location is another feature. There is an ongoing debate in the literature (e.g. Nissen, 1985). Many tasks depend on knowing where the stimulus is as well as another attribute. Thus, they assume good knowledge of feature–location conjunctions. This is made explicit by the idea of feature maps in feature integration theory. Location cuing experiments also emphasize the coding of location along with other features. This would be good question to return to in the future.

Encoding, storage or retrieval? Luck and Vogel interpret their effects in terms of a limit on storage. Essentially, they propose a slot model of memory. In contrast, Palmer interprets his limits on performance in terms of the encoding into durable memory. Another possibility not considered explicitly by either author is a limit on retrieval. Perhaps the information is stored but interference from the many trials makes it hard to retrieve the appropriate memory trace. To address this retrieval hypothesis, one can manipulate the retention interval, attempt to manipulate interference (e.g. release from proactive interference) or try to manipulate the quality of the retrieval cues (e.g. recall versus recognition). This appears to be untouched territory for this topic but we need to read the relevant memory literature more closely.

Generalizations. John cited studies by Mary Hayhoe (Hayhoe, et al., 1998; 2003) where she measured eye movements while subjects compared complex objects that were made out of many interchangeable parts (Lego structures). The subjects were presented with two stimuli side by side and they made eye movements back and forth between the two stimuli comparing corresponding locations. This is consistent with serial processing and very limited verbal memory.

Seminar on the Empirical Foundations of Visual Attention
Questions for week 8: Final synthesis
John Palmer and Cathleen Moore
24 May 2007

Goals

For our last meeting, we consider issues that reach across the previous sessions. What are the themes revealed by our readings and discussions?

Please write or outline brief answers to the 5 questions below. We will compare our answers in the seminar.

Readings

No additional readings

Discussion Questions

1. What is selective attention? How can one define the phenomena of selective attention? What is selection? What are the possible mechanisms? What are the major theories of which mechanisms apply to which stimuli and which tasks?

2. What is divided attention? How can one define the phenomena of divided attention? What is capacity? What are the possible mechanism of capacity? What are the major theories of which mechanisms apply to which stimuli and which tasks?

3. How can attentional and nonattentional phenomena be distinguished?

4. How can attentional effects that are mediated by perceptual processes be distinguished from attentional effects mediated by memory processes?

5. Are there any paradigms not discussed in the seminar that should be included in an introduction to attention? There are many other paradigms and we don't want an exhaustive list. Rather, which paradigms are most useful to introduce the basic vocabulary of effects and mechanisms upon which to build a more compete description? Additional candidates include:

a. Task selection (e.g. Stroop and Simon effects)

b. Time course of task selection

c. Attentional control (e.g. capture by luminance transients)

Seminar on the Empirical Foundations of Visual Attention

Summary of discussion for Week 8: Finale
Alec Scharff, John Palmer and Cathleen Moore
8 June 2007

Phenomena, Mechanisms, and Theories

John Palmer introduced three questions to organize the discussion. What defines the phenomena of attention? What mechanisms mediate the basic effects of attention? What theories describe which mechanisms apply to what phenomena?

Phenomena. Defining phenomena is not a trivial task in attention research because the border between attentional phenomena and other kinds of phenomena is not clearly delineated. There are non-attentional factors which can vary with experience and instruction. For example, transient cues used to direct attention might also have perceptual effects (e.g. Schneider, 2006). Other things being equal, central cues are preferable, because they do not produce any obvious perceptual effects.

Mechanisms. John recalled that some people at Michigan used to talk about mathematical models as a "bag of tricks." For any situation, you can dig through your bag of tricks and see if see if any of the tools you have are right for the job. John thought this was a good analogy for attentional mechanisms. A mechanism is a specific theory of an attentional phenomena that might be applied to a variety of stimuli and situations. For example, the 'all-or-none selection' mechanism might apply to some situations but not others.

Theories. Theories address which mechanisms apply to what stimuli for what situations. In this seminar, we have focused on mechanisms rather than the larger theories in attention. Such larger theories include feature integration theory or early and late selection theories.

Selective and Divided Attention

Selective Attention. Serap Yigit summarized her view of selective attention. The key phenomenon in selection is the consequence of choosing to direct special processing to one subset of the available stimuli and not others. What is the effect of this selection on the attended stimuli? What is the fate of the unattended stimuli?

In addition to these questions, Serap also considered attentional control to be a fundamental issue. For example, in her foil experiments, a high-contrast foil can wrest attention away from the task-relevant locations against the observer's volition. For another example of involuntary selection, Serap described trying to read on the ferry while people are having a loud conversation nearby.

John agreed with Serap's account of selection. He began a discussion of the choice of words in describing attentional phenomena. He cautioned against using the term attention when defining attentional phenomena because these can result in a circular definition. He also stressed the importance of using language carefully to distinguish between phenomena (e.g., the cuing paradigm) and mechanisms (e.g., attentional spotlight).

Alec Scharff added his own thoughts about selection. Generally, selecting a subset of available stimuli has the effect of improving performance for the selected stimuli and decreasing performance for other stimuli. Alec thought that the question of where in the processing stream the facilitation takes place was of key importance. Selection can affect sensory, perceptual, or decision processes. One might term this the "locus" question. John added that the nature of facilitation is also an integral part of this issue. Attenuation and all-or-none selection are two possible mechanisms that can occur at any stage of processing.

Divided Attention. Divided attention phenomena are found in search paradigms, dual tasks, psychological refractory period experiments, and simultaneous-successive display experiments. All of these paradigms involve multiple stimuli or multiple tasks. The key theoretical idea is one of capacity. Performance suffers when an attentional capacity limit is exceeded.

Pashler describes capacity in terms of monitoring multiple channels, avoiding the use of the 'divided attention'. This may be an unconventional terminology, but it may have merits in using terms that are less associated with theory. We need to think about what terms to use as neutral and which to specifically relate to either phenomena or internal mechanisms. In this note, I generally use "divided attention" as the neutral or phenomena term and capacity as the mechanistic term.

The key mechanisms for this domain are bottlenecks due to serial processing or capacity sharing among parallel processes. Limited-capacity parallel models are often characterized as sharing or deployment of attentional resources, while an all-or-none processing mechanism reflects a serial model. Of the capacity sharing models, the sample size model is perhaps the best characterized (e.g. Shaw, 1980). Many other compromise positions and hybrid models are possible, of course. John pointed out that the capacity sharing can start to sound a lot like selection if you're not careful with your words.

Serap brought up the issue of why different mechanisms seem to apply to different stimuli. John agreed that the goal of capacity theories was to characterize how and why various stimuli show different processing properties. Full-blown theories of capacity are central to characterizing cognition in general.

Terminology for Divided Attention

Capacity. John provided background on the relevant terminology. The use of the word capacity comes from information theory. Early information theorists tried to characterize cognitive processes in terms of bits. For example, if 7 digits can be held in memory, then one could calculate the number of bits needed to distinguish 7 digits. If bits define the capacity limit, then one can predict how many characters or words can be stored by calculating the bits needed to represent a letter or word.

These predictions didn't pan out. People can generally remember as many characters or words as digits. This revelation led to the "death of information theory" in mainstream cognitive psychology. The notion of storing information in more subjective 'chunks' of meaning became popular. Some cognitive psychologists henceforth tended to avoid the term 'capacity,' with its associations to information theory, in favor of the term 'resource.'

Resources. John has tended to favor the term 'capacity' over resource. Capacity invokes the mechanisms described in information processing, for better or for worse. John believes many people shy away from capacity because they think of information theory as a failed paradigm, not having seen in it work in a more generalized sense. Information theory is still alive and well in signal processing. For example, Adrienne Fairhall of UW uses information theory in the analysis and single cell behavior. It is a useful way of conceiving a communications system, though by no means the only way.

Pashler prefers the term capacity as well. Resource is most often used in dual-task literature, about which Pashler is not enthusiastic.

As someone unaware of the historical connotations of the terms resource and capacity, Alec thought resource seemed to denote a sharp capacity limit, whereas as capacity more often refers to a 'soft' limit. When the term resource is used, generally you start with some amount of something, and when it's all used up, you have no more options. For example, when you run out of money, you simple can't buy anything anymore. On the other hand, when capacities are discussed, you start at zero and approach some limit. Often, capacities can be exceeded, but with some detrimental effect. For example, the maximum capacity in an elevator can be exceeded, but results in crowding, suboptimal performance, or the danger of breaking the elevator. Although water glass has a fixed capacity, it's still logically possible to keep trying to fill it once the capacity has been exceeded, albeit with negative consequences. By communicating carefully, you could use either word to reflect exactly the point you are trying to get across. Everyone agreed that it's best to use language that carries the least connotations in order to avoid confusion.

Cross-talk. We also mentioned another metaphor for divided attention. Given parallel processes, there might be interference analogous to electrical cross-talk between circuits.

Dependency and Independence. At the end of this conversation, we considered discussing these effects in terms of independence or dependence among processes. This everyone agreed was a good superordinate terminology. Perhaps we could adapt the degree of independence as the overarching terminology combined with the alternative metaphors of capacity, resource or cross-talk?

"Selection is everything" – Cathleen Moore

John introduced the 'bumper sticker version' of a proposal from Cathleen Moore. She proposed that all attentional phenomena be thought of in terms of selection. The challenge for this framework is to explain the phenomena of divided attention. The central idea is that "dividing attention" is equivalent to not selecting or diffuse attention. Thus, division is simply the complement of selection, not a distinct idea.

This idea is particularly clear if one adapts a single mechanism theory such as early selection. By such a theory, the effects of both selective attention and divided attention are due to the operation of a single filter. The filter determines what stimuli are processed. Processes before the filter are in parallel and processes after the filter are serial. Thus the filter is the single critical component that defines both selection and the degree of independence between processes.

Can one extend this idea to more complex theories such as Pashler's controlled parallel model? In this model, two things are different. First, one can have parallel processes after the early filter. Thus, the filter alone does not mark where in the processing stream parallel processing ends and serial processing begins. It also invites there to be different selection processes at various stages. For example, one can elaborate the basic theory with selection processes that act on sensory processing, encoding into durable memory or a decision process. By this account, one can predict how a particular stimulus in a particular task is bottlenecked at particular stages of processing by what kind of selection is relevant to that stimulus and task. This way of conceiving of capacity limitations again emphasizes selection as the defining theoretical element.

John then went back to “being John” and considered how to break such a multiple selection theory. The specific prediction of the theory is that it requires each selection point to function in a stereotyped way for all stimuli. For example, consider a point in the sequence of processing where a selection for a certain kind of processing is said to take place. Consider word reading where features and letters are recognized to yield word representations. How might such a theory be tested? Consider, Heather Knapp's result where she used the simultaneous-successive paradigm to evaluate the perception of digits or words (of digits, e.g. "two"). The task followed Pashler and Badgio (1987) in that the subject reported highest value of a set of numbers. In different experiments, these numbers were represented as either digits or words. She found evidence for unlimited-capacity parallel processing for digits but not for words. Thus, one would have to propose that digits are processed at a different point in the processing stream than where the words are processed. For one measurement, this need not be a problem. But it constrains the theory and further measurements might yield contradictions if digits are processed not processed in parallel for other cases.