Psychonomic Bulletin & Review
2004, 11 (2), 197-231

THEORETICAL AND REVIEW ARTICLES

Why is it difficult to see in the fog? How stimulus contrast affects visual perception and visual memory

ERIN M. HARLEY
University of California, Los Angeles, California

and

ALLYSS M. DILLON and GEOFFREY R. LOFTUS
University of Washington, Seattle, Washington

Processing visually degraded stimuli is a common experience. We struggle to find house keys on dim front porches, to decipher slides projected in overly bright seminar rooms, and to read 10th-generation photocopies. In this research, we focus specifically on stimuli that are degraded via reduction of stimulus *contrast* and address two questions. First, why is it difficult to process low-contrast, as compared with high-contrast, stimuli? Second, is the effect of contrast *fundamental* in that its effect is independent of the stimulus being processed and the reason for processing the stimulus? We formally address and answer these questions within the context of a series of nested theories, each providing a successively stronger definition of what it means for contrast to affect perception and memory. To evaluate the theories, we carried out six experiments. Experiments 1 and 2 involved simple stimuli (randomly generated forms and digit strings), whereas Experiments 3–6 involved naturalistic pictures (faces, houses, and cityscapes). The stimuli were presented at two contrast levels and at varying exposure durations. The data from all the experiments allow the conclusion that some function of stimulus contrast combines multiplicatively with stimulus duration at a stage prior to that at which the nature of the stimulus and the reason for processing it are determined, and it is the result of this multiplicative combination that determines eventual memory performance. We describe a stronger version of this theory—the sensory response, information acquisition theory—which has at its core, the strong Bloch's-law-like assumption of a fundamental visual system response that is proportional to the product of stimulus contrast and stimulus duration. This theory was, as it has been in the past, highly successful in accounting for memory for simple stimuli shown at short (i.e., shorter than an eye fixation) durations. However, it was less successful in accounting for data from short-duration naturalistic pictures and was entirely unsuccessful in accounting for data from naturalistic pictures shown at longer durations. We discuss (1) processing differences between short- and long-duration stimuli, (2) processing differences between simple stimuli, such as digits, and complex stimuli, such as pictures, (3) processing differences between biluminant stimuli (such as line drawings with only two luminance levels) and multiluminant stimuli (such as grayscale pictures with multiple luminance levels), and (4) Bloch's law and a proposed generalization of the concept of *metamers*.

This article has two purposes. The first is to investigate the effects of stimulus contrast on perception of, and memory for, visually presented material. The second is to demonstrate the effectiveness of testing a series of progressively stronger—that is, nested—theories. We will discuss these two goals in turn.

This research was supported by NIMH Grant MH41637 to G.R.L. We thank Tom Busey for providing the hooded faces used in Experiment 3, Martin Oberg and Jessica Cooke for assistance in running the subjects in Experiments 3–6, and John Palmer, Trammel Neill, and two anonymous reviewers for very useful comments on an earlier version of this article.
Correspondence concerning this article should be sent to G. R. Loftus, Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195-1525 (e-mail: gloftus@u.washington.edu).

Stimulus Contrast

In the course of everyday visual behavior, a person must frequently detect and identify visual stimuli that are somehow degraded. For example, deciphering lecture slides in a brightly lit room, reading traffic signs through a rain-smeared windshield, and navigating a ski slope while wearing foggy goggles are all situations in which the visual signal is degraded to one degree or another. It seems, subjectively, that perceiving degraded stimuli is more difficult than perceiving undegraded, or *clean*, stimuli.

Although there are many ways in which visual stimuli may be degraded, a large body of research indicates that stimulus *contrast*, defined as the ratio of foreground to background luminance, is critical in determining the fundamental response of the visual system. This effect can be seen in the cat's visual system, wherein retinal ganglion cells are considerably more sensitive to stimulus contrast than to absolute light levels (see, e.g., Wandell, 1995, p. 139). The effect continues up through experiments in which human sensitivity has been investigated, wherein contrast sensitivity has varied over a range of approximately 20:1 as absolute light level has varied over a range of more than 1,000,000:1 (e.g., van Nes & Bouman, 1967).

Stimulus contrast is often the key experimental variable in modern vision science (Wandell, 1995). This is largely because, unlike stimulus intensity, where linearity fails, a high degree of linearity is observed in neuronal responses when intensity is fixed and stimulus contrast is treated as the input variable. This linearity extends beyond physiology to perception. For example, Ginsburg, Cannon, and Nelson (1980) demonstrated that perceived contrast is a linear function of stimulus contrast for sine-wave gratings, and Olds and Engel (1998) showed that object identification is predicted well by a theory within the context of which responses to different spatial frequency components of independently varying contrasts are simply summed to determine the overall response.

Whereas a great deal of research has been conducted to investigate contrast effects on low-level sensory processes, there has been considerably less research in which contrast effects on higher level cognitive processes have been investigated. We ask whether some of the fundamental laws that have been found by vision scientists to characterize simple stimuli in simple situations (e.g., detection of a monochromatic light patch) may be extended to more complex stimuli in more complex situations (e.g., face recognition). In this article, we investigate contrast effects on visual information acquisition and visual memory. In our experiments, we presented stimuli for varying durations at one of two contrast levels, followed by various kinds of memory tests. The main result from such an experiment is referred to as a *performance curve*, which relates memory performance to stimulus duration. By comparing performance curves issuing from stimuli shown at different contrasts, the effects of contrast can be precisely evaluated, as we will describe in detail below.

We report six experiments. In the first, memory for random visual forms was tested in a two-alternative forced-choice (2AFC) recognition test. Each form was originally viewed under different combinations of duration and contrast. In addition, the 2AFC test was either easy or difficult, and our question could be roughly framed, Does the effect of contrast on memory performance depend on test difficulty? In the second experiment, we combined variation in duration and contrast with two very different *stimulus–task combinations* (STCs). In the first STC, random forms were again used as stimuli and were tested in a 2AFC procedure. In the second STC, digit strings were used as stimuli and were tested by serial recall. Here, our main question could be roughly framed as follows: Does the effect of contrast depend on STC? In Experiments 3–6, we incorporated and compared two types of naturalistic stimuli within each experiment. In addition, we tested memory for the stimuli both immediately, via a prospective confidence rating, and in a later delayed recognition test. In each of these experiments, we (again roughly) asked, Is the effect of contrast the same for the two stimulus types and for the two performance measures?

Nested Theories

What does it mean to claim that the effect of contrast on memory performance depends or does not depend on some variable, such as task difficulty, STC, stimulus type, or test delay? This question is necessarily framed and answered within the context of some *theory*. Accordingly, we now turn to the second purpose of this article, which is to demonstrate the usefulness of postulating a series of nested theories of some effect. In particular consider a series of theories, Theory 1 . . . Theory k such that each Theory $k + 1$ is nested within, or *stronger than*, Theory k. By this we mean that Theory $k + 1$ is a special case of Theory k. For instance, Theory $k + 1$ might be identical to Theory k, except that whereas some particular function in Theory k is assumed to be monotonic, the corresponding function in Theory $k + 1$ is assumed to be linear. Such a theoretical strategy is effective for two complementary reasons. First, *rejection* of a theory implies rejection of all nested (i.e., stronger) theories. In our example, rejection of the monotonic function theory would also imply rejection of the linear function theory. Second, *confirmation* of a theory implies confirmation of all weaker theories. To use the same example, confirmation of a linear function implies confirmation of a monotonic function.

We now will describe a series of four nested theories in order to investigate the question, How does stimulus contrast affect perception and memory? Following our report of Experiment 1, we will describe in detail how these four theories, plus an additional theory, are tested.

Basic Assumptions

We first will establish the basic assumptions and a notation within which the theories are couched. We assume, in particular, that processing a visual stimulus in anticipation of a later memory test can be viewed as the accumulation over time of *information* about that stimulus—that is, that

$$I(t) = A(t, C). \tag{1}$$

Here, $I(t)$ is acquired information about a stimulus. By Equation 1, $I(t)$ is some function, A, of t, the time following the stimulus onset, and C, the stimulus contrast. When the stimulus is shown for some particular stimulus duration d at contrast level C, the total acquired information about the stimulus is

$$I = A(d, C). \tag{2}$$

Performance in any task that measures perception and memory for the stimulus, then, depends on—that is, is a monotonic function m_T of—acquired information. Thus,

$$P_T = m_T(I), \tag{3}$$

where P_T is performance in Task T. With respect to the present experiments, variation in T corresponds to the influence of different test difficulties (Experiment 1), different STCs (Experiment 2), or different stimulus types and test delays (Experiments 3–6).

Within the context of the general theory defined by Equations 1–3, some candidate answers to the question of how contrast affects perception and memory present themselves, such as "different kinds of information are acquired from low-contrast stimuli than from high-contrast stimuli," or "the same information is acquired, but at a slower rate, from low-contrast stimuli as from high-contrast stimuli." To formalize such answers, we consider four specific nested theories of how stimulus contrast affects accumulation of information over time.

Multidimensional Theory

Multidimensional theory, the weakest theory, consists of the following assumptions: (1) Information, I, is represented by J different dimensions, and thus, I is a J-element vector; and (2) the different dimensions of I may be affected in different ways by stimulus contrast; in particular,

$$I_j = A[d, g_j(C)], \tag{4}$$

where I_j is the value of the jth informational dimension, g_j is a function such that $g_j(C)$ is a scalar, and A is monotonically increasing in all arguments. Performance in Task T is then obtained by

$$P_T = m_T[I] = m_T[I_1, I_2, \ldots, I_J], \tag{5}$$

where m_T is a function that is monotonic in all arguments.

Numerous sorts of multidimensional theories have appeared in the memory literature over the past 4 decades, where dimensions have consisted, for example, of short-term versus long-term memory (Atkinson & Shiffrin, 1968), levels of processing (Craik & Lockhart, 1972), multiple memory "attributes" (Underwood, 1969), convolution theories (Hintzman, 1984; Murdock, 1982, 1993), verbal versus imaginal codes (Paivio, 1969, 1971), strength-fragility theory (Wickelgren, 1972, 1974), implicit versus explicit memory (Schacter & Tulving, 1994), and, related to the present work, strength-certainty theory (Busey, Tunnicliff, Loftus, & Loftus, 2000), which we will describe in more detail below.

Multidimensional theory is our weakest theory. Is it valid? The assumptions of multidimensional theory seem sufficiently general to be almost self-evident. To confirm them, one need only find that memory performance increases with both stimulus duration and stimulus contrast. In the world of visual psychophysics, virtually any performance measure in any experiment improves with contrast (e.g., Graham, 1989; Olzak & Thomas, 1986), and much the same has been found with visual memory tasks (Loftus, 1985c; Loftus, Kaufman, Nishimoto, & Ruthruff, 1992; Loftus & Ruthruff, 1994). Likewise, performance almost invariably increases with increasing duration (e.g., Kaswan & Young, 1963; Laughery, Alexander, & Lane, 1971; Loftus, Busey, & Senders, 1993; Rumelhart, 1970; Shibuya & Bundesen, 1988).

Nevertheless, these findings are not absolutely universal, and circumstances can be concocted in which even these assumptions can be shown to be false. An excellent example has been provided by Nairne (1988), who demonstrated that interfering with encoding processes via a reduction in stimulus duration can lead to an increase, not a decrease, in memory performance. In a simple naming task, Nairne presented masked words at two durations: a short duration that led to 50% naming accuracy and a long duration that led to 100% naming accuracy. A large retention advantage was found for the short-duration words on a surprise recognition test, whereas no such advantage was found for recall. Nairne compared these results with the *generation effect*, in which people show better memory performance for items that are self-generated during study than for those that are only read (e.g., Slamecka & Graf, 1978). He hypothesized that as a result of insufficient processing time for the short-duration words, participants engaged in a data-driven generation process that led to greater performance on a test that is sensitive to data-driven processing—for example, recognition. If the effect is due to an increase in data-driven processing, interfering with encoding should lead to increased performance on any test that is sensitive to such processing. In an extension of Nairne's original study, Hirshman and Mulligan (1991) replicated the naming task with short- and long-duration words and replaced the follow-up recognition test with a perceptual identification task. Counter to the prediction of Nairne's hypothesis, no advantage was found for the short-duration words in the perceptual identification task, suggesting that the recognition advantage was not due to data-driven processes but, rather, to conceptually driven processes. The conclusion appears to be that in some circumstances, the vigor of observers' postperceptual stimulus processing can be made inversely dependent on the physical quality of the stimulus—a kind of *overcompensation* effect. It appears that this phenomenon does not happen ordinarily; nevertheless, that it does happen *sometimes* places boundaries on even this weakest of our theories.

Unidimensional Theory

Unidimensional theory is a special case of multidimensional theory in which there is only a single dimension. Thus, information, I, is a one-element vector—that is, information can be represented by a single number, or

$$I = A[d, g(C)], \tag{6}$$

and performance in Task T is obtained by

$$P_T = m_T(I) \tag{7}$$

Numerous theories make the implicit assumption of unidimensionality. Examples are strength theories (e.g., Norman, 1966) and logogen theory (Morton, 1969). Basic signal detection theory entails an underlying unidimensional scale of *strength* or *familiarity*; however, because two numbers, strength and criterion, are necessary to generate performance, the complete theory is two-dimensional, not unidimensional.

Multiplicative Theory

The remaining two theories retain the unidimensionality assumption and entail successively more specific instantiations of the function g in Equation 6.

Multiplicative theory assumes that the function A combines duration d and some function of contrast $g(C)$ multiplicatively to produce the single value of information—that is,

$$I = A[d \times g(C)], \tag{8}$$

with performance on Task T as described in Equation 7.

Multiplicative theory has been confirmed in several studies reported by Loftus (1985c; see also Sperling, 1986) and by Loftus et al. (1992, Experiment 3), wherein complex naturalistic color pictures were used as stimuli and were tested using both a prospective confidence rating acquired after each study trial and a long-term yes–no recognition procedure. In a related experiment in which degradation was accomplished by imposition of random noise, rather than by contrast reduction, multiplicative theory was again confirmed in a recognition task (Loftus et al., 1992, Experiment 4).

Bloch's Law

Multiplicative theory assumes that information is a function of the product of duration d and some function of contrast $g(C)$. A more specific version of this theory assumes that information is a function simply of the product $(d \times C)$—that is, that $g(C) = C$. Thus, acquired information is obtained by

$$I = A(d \times C), \tag{9}$$

and performance is described in Equation 7. This theory is akin to Bloch's law, wherein below a critical duration, perception, as measured in any task, is determined solely by the product of stimulus luminance and stimulus duration—that is, by total stimulus energy.

The Notion of a *Fundamental Effect* of Contrast

In perception and memory tasks, variables differ in terms of where in the cognitive system they exert their effects. Suppose, for example, that an observer were asked to encode a stimulus, either by using rote rehearsal or by generating a mental image. This *processing type strategy* would be construed as having its effect at a relatively high level in the cognitive system; the variable's effect would be influenced, for example, by the nature of the stimulus and the observer's intent in processing the stimulus, among other things.

The last three theories we have discussed, from unidimensional to Bloch's law, all place the contrast effect at a low level in the system. That is, if contrast and duration simply combine to generate a single number, the value of contrast itself is lost at an early stage. Informally, it can be construed as a "dumb" variable whose effect is independent of the nature of the stimulus or the reason for processing the stimulus (e.g., the eventual task for which the stimulus will be relevant).

EXPERIMENTS

To evaluate these theories, we carried out six experiments, all using a memory task for visually presented material. In Experiments 1 and 2, we used simple stimuli (digits and random forms), which were what we define to be *biluminant*: They were composed of only two luminance values, a foreground and a background; hence, a given stimulus could be characterized by a single contrast value, which is essentially the ratio of the two luminances. In Experiments 3–6, we used naturalistic grayscale images (faces, houses, and cityscapes) as stimuli. Each of these *multiluminant* stimuli used the full range of 256 grayscale values; hence, there was no single contrast value, and definition of contrast was somewhat more complex, as will be described below. In all six experiments, we addressed the question, Is the effect of contrast on memory performance the same for both levels of a third independent variable? The answer to this question will, as we shall see, allow us to identify which of the nested theories described earlier are confirmed and which are disconfirmed.

Experiment 1
Task Difficulty

In Experiment 1, computer-generated random forms were shown for different durations and contrasts and were then tested in a 2AFC recognition test. Task difficulty was manipulated by varying the similarity between a target and its associated distractor.

Method

Observers. The observers were 6 University of Washington undergraduates and graduates. Each reported normal or corrected-to-normal vision and was paid $60.

Stimuli. The stimuli were random computer-generated line drawings (hereafter, *forms*) created in MATLAB as follows. For each form, seven sets of x–y coordinates within a 50 (columns) $\times$ 58 (rows) pixel matrix were chosen at random and connected, in the order they were chosen, by six straight lines. The seventh point was not connected back to the first. For each target form, two corre-

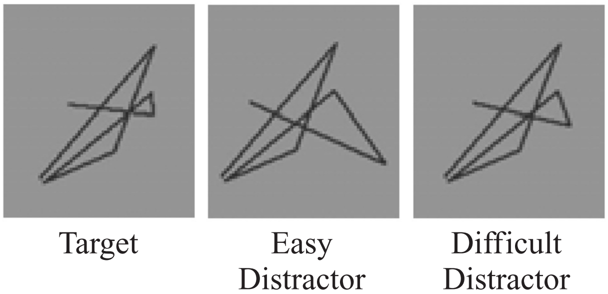

Figure 1. Example of a form stimulus triad used in Experiment 1.

sponding distractor forms were created—one for the *easy* test and one for the *difficult* test. Thus, there were three versions of each form—the target and its two corresponding distractors. We will refer to these as stimulus *triads*, an example of which is shown in Figure 1. Distractors were created by altering one set of target coordinates by some number of pixels, which we term *delta*. For the *easy* distractors, delta was set to 11 pixels, whereas for the *difficult* distractors, delta was set to 7 pixels. The coordinate to be altered was chosen at random for each pair of distractors, excluding the two end coordinates, 1 and 7. For half the distractor pairs, delta was added to both the *x*- and the *y*-values of the altered coordinates, whereas for the other half, the delta value was subtracted from both the *x*- and the *y*-values of the altered coordinates.

Apparatus. The stimuli were displayed on a 17-in. Macintosh ColorSync monitor with a screen refresh rate of 13.5 msec, driven by a Power Macintosh G3 computer. The experiment was written and executed in MATLAB, using the libraries provided in the Psychophysics Toolbox (Brainard, 1997) and the Video Toolbox (Pelli, 1997). The experimental lab room was dimly illuminated during the experiment. All the stimuli were shown against a uniform gray background whose luminance was 8.6 cd/m^2.

Design. There were 20 conditions in the experiment: 5 stimulus durations × 2 contrasts × 2 difficulty levels. Because we wanted to achieve roughly equal performance ranges for the two contrast levels, the duration ranges were different for the high and the low contrasts: For the low-contrast conditions, durations ranged from 40.5 to 148.5 msec in equal increments, whereas for the high-contrast conditions, durations ranged from 13.5 to 121.5 msec in equal increments. Stimulus contrast was defined to be $(L_{\mathrm{Max}} - L_{\mathrm{Min}})/(L_{\mathrm{Max}} + L_{\mathrm{Min}})$ where L_{Max} is the maximum luminance and L_{Min} is the minimum luminance. Specific contrast values varied among observers to accommodate a range of observer ability but were held constant

Table 1
Observer Information for Experiment 1

Observer	Number of Blocks	Low Contrast	High Contrast	Contrast Ratio	*k* (Best Shift)
S.K.C.	160	0.043	0.073	1.698	3.320
E.C.U.	40	0.043	0.073	1.698	3.004
E.M.H.	60	0.058	0.088	1.517	2.460
J.C.M.	70	0.058	0.088	1.517	2.460
C.A.W.	160	0.073	0.104	1.425	2.460
J.S.F.	80	0.073	0.104	1.425	2.460
M		0.056	0.086	1.547	2.694

Note—The "Contrast Ratio" is the ratio of high contrast to low contrast. The "*k* (Best Shift)" column shows the ratio of durations required to achieve equal performance in the low-contrast, as compared with the high-contrast, condition. Note that *k* is constrained to be the same for both the easy and the difficult form recognition tasks.

in the two tasks for each individual observer. Table 1, columns 3 and 4, shows the high- and low-contrast values for each of the 6 observers. The relevance of the two rightmost columns of Table 1 will be discussed in a later section.

Each observer completed at least four sessions of data collection, preceded by at least two practice sessions. Each session involved 10 blocks of form recognition trials. Each block consisted of 40 stimulus trials, 2 in each of the 20 conditions.

Procedure. Each trial began with a fixation point displayed in the center of the screen for 250 msec, accompanied by a warning tone. Following the fixation point, the target stimulus was presented for the exposure duration appropriate to that trial. Five hundred milliseconds after target offset, the target and its corresponding distractor (*easy* or *difficult*) were displayed side by side on the screen, where they remained until a response was made. The observer's task was to identify which of the two test stimuli was the previously presented target. The observer entered one of six response options, corresponding to *definitely left*, *probably left*, *maybe left*, *maybe right*, *probably right*, and *definitely right*. Feedback was given in the form of a tone—2000 Hz if the observer was correct and 500 Hz if the observer was incorrect.

Randomization and counterbalancing. Conditions were counterbalanced across 20-block chunks. For each 20-block chunk, 40 stimuli were rotated through each of the 20 conditions. Condition order for each of the 40 stimuli within a set was initially randomized, with the constraint that each of the 20 conditions occurred twice. Then, for 20 consecutive 40-trial blocks, each form was rotated through all 20 conditions; thus, each of the 40 forms was shown once during each block. The ordering of the forms within each block was randomized. A unique stimulus set of 40 form triads was used for each 20-block chunk completed by an observer. Table 1, column 2 shows the number of blocks run by each observer.

Recall that on each test trial, a pair of stimuli was displayed: the target and its corresponding distractor (*easy* or *difficult*). Of the two stimuli in a pair, the choice of which stimulus was displayed as the target prior to test was determined by a virtual coin flip on each trial. This was done to prevent the observers from learning a particular target in each stimulus triad. The observers were informed at the outset of the experiment that although they would see the same stimulus pairs repeated across blocks, on any particular trial, of the two forms in the pair, the form chosen to be the target would vary.

Results

Proportion correct, corrected for the 50% guessing rate, was computed for each of the 20 conditions and averaged across sessions for each observer. Figures 2A (easy task) and 2B (difficult task) show mean performance as a function of exposure duration (log scale) for the two contrast levels, averaged across the 6 observers. The error bars, which are mostly obscured by the curve symbols, represent standard errors. Numerous aspects of Figure 2—the reason for plotting performance on a log duration scale, the genesis of the theory lines through the data points, the rectangles on panels A and B, each enclosing two data points, and panels C and D in their entirety—will be described below. For the moment, we note that, as was expected, performance increases with increasing duration, is higher for high than for low contrast, and is higher in the easy than in the difficult condition.

Evaluation of Theories

We now will turn to an evaluation of the theories described earlier. In the process, we will describe in detail what each theory predicts about our data.

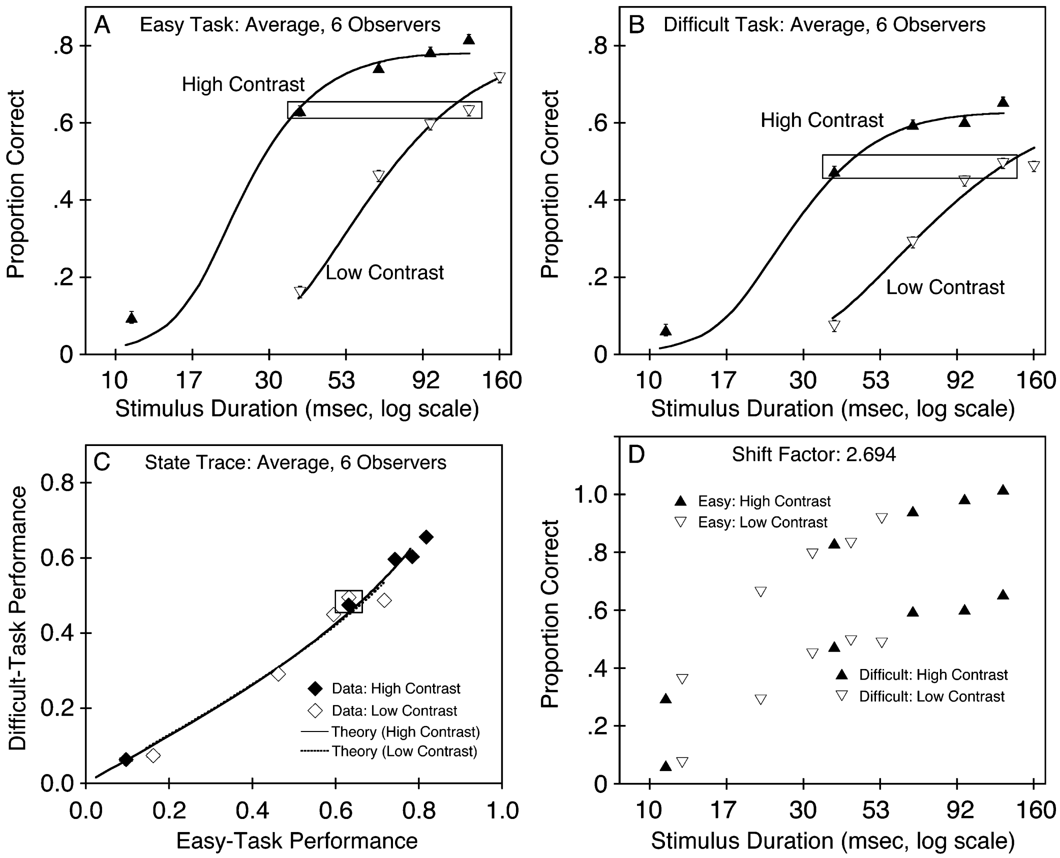

Figure 2. Experiment 1 data. (A and B) Response probability corrected for guessing as a function of stimulus duration (log scale). (C) Difficult performance as a function of easy performance (state-trace plot). (D) The low-contrast curves from the top two panels have been shifted leftward by identical amounts, corresponding to a factor of 2.694 (for visual clarity, the "easy" curves have been shifted upward by 0.30). With this shift, the high-contrast and the low-contrast curves align quite well for both difficulty levels. Solid and dotted lines in panels A–D represent theoretical fits of the sensory response, information acquisition theory described in the text. All error bars (most of them obscured by the plot symbols) represent standard errors.

Multidimensional theory. Multidimensional theory makes only the assumptions that all the elements of information, I, increase with both increasing duration and increasing contrast and that performance increases with increasing information. To confirm multidimensional theory, it is therefore sufficient to note that for both the easy and the difficult tasks, performance increases with increasing duration and luminance.

Unidimensional theory. By unidimensionality theory, stimulus duration and stimulus contrast are combined into a single number representing acquired information, I, and performance is a monotonic function of I. In Experiment 1, there were two performance measures—one for the easy task, P_E, and one for the difficult task, P_D—and there were, correspondingly, two monotonic functions, m_E and m_D, that mapped I to P_E and P_D (see Equation 7 above).

To describe unidimensionality theory's predictions, we use what is referred to as a *state-trace analysis*, described in detail by Bamber (1979; see also Busey et al., 2000; Loftus & Bamber, 1990; Loftus & Irwin, 1998; Palmer, 1986a, 1986b). In the Appendix we provide a brief tutorial describing state-trace analysis. Applied to the present data, the general idea is this. Consider two different duration × contrast conditions, (d_1, C_1) and (d_2, C_2), that lead to *equal performance*, P_E, on the easy task. The prediction is that these same two conditions must also lead to equal performance, P_D, on the difficult task. The logic behind this prediction is that if the two conditions, (d_1, C_1) and (d_2, C_2), produce the same value of P_E, then according to the unidimensional model, these two conditions must have produced the same value of information,

$$I = m_E^{-1}(P_E), \tag{10}$$

where, m_E^{-1} is the inverse of m_E. Because the two conditions produce the same value of information, they must also produce the same performance on the difficult task, $P_D = m_D[m_E^{-1}(P_E)]$. Loftus et al. (1992, Experiment 3) showed this to be the case for picture recognition where the joint effects of luminance and duration were measured on two tasks: prospective confidence and yes–no recognition.

To gain an intuition for this prediction, consider the 40-msec high-contrast and the 121-msec low-contrast conditions for the easy task. Mean performances for these two conditions are enclosed in the rectangle in Figure 2A. Because these two conditions are approximately equal in the easy task condition, they are predicted to also be equal in the difficult task condition. It is evident that they are, as is shown by the rectangle in Figure 2B, which encloses the same two conditions. Thus, the equality, in both difficulty conditions, of these two conditions constitutes a confirmation of unidimensional theory.

The success of state-trace analysis does not require that one be lucky enough to find pairs of duration × contrast conditions that happen to produce identical performances. The formal rationale for this assertion is described in Bamber (1979), and an intuitive treatment of it is provided in our Appendix. Essentially, one assumes that the measured points are samples from an underlying continuous function whose form can be estimated from the data. The unidimensional theory's prediction is evaluated using what is referred to as a *state-trace plot*, which is a scatterplot, over experimental conditions, of one performance measure (in this case, difficult task performance) against another performance measure (in this case, easy task performance). If the continuous function were measured, the scatterplot would, by unidimensional theory, form a continuous monotonic function, with higher valued points corresponding to greater information (i.e., longer duration and higher contrast) conditions. Because the continuous function is not measured completely but is only sampled—the sample corresponding here to the chosen set of stimulus durations for the two contrast levels—we must estimate the underlying function by "connecting the dots" in the scatterplot. Unidimensional theory's prediction remains that the sampled scatterplot points will be monotonic over all 10 points in the scatterplot—that is, across both the high- and the low-contrast conditions.

The state-trace plot for Experiment 1 is shown in Figure 2C. As in the other Figure 2 panels, the filled curve symbols represent high-contrast conditions, whereas the open symbols represent low-contrast conditions. On a state-trace plot, two contrast × duration conditions that are equal for both the easy and the difficult tasks would appear as two points in the scatterplot that fall almost entirely atop one another. The two such conditions that we have just discussed indeed appear as two (almost) overlapping points enclosed by the rectangle in Figure 2C. It is evident that the Figure 2C scatterplot is, for all intents and purposes, monotonic, thereby confirming unidimensional theory.

Just as the data points from panels A and B are transposed to panel C, the theoretical curves (whose genesis is to be explained soon) are likewise transposed. The two theoretical curves corresponding to high and low contrast also fall virtually atop one another, thereby indicating that the theory that generates these curves incorporates the assumption of unidimensionality.

Multiplicative theory. Multiplicative theory assumes that stimulus contrast and stimulus duration combine multiplicatively to produce information—that is, that

$$I = A[d \times g(C)]. \tag{11}$$

Consider, as we did above, two pairs of duration–contrast combinations, (d_1, C_1) and (d_2, C_2), that produce the same easy performance value, P_E. We have just argued that any unidimensional theory implies that (d_1, C_1) and (d_2, C_2) must then have produced equal values of I. Therefore, by Equation 11,

$$A[d_1 \times g(C_1)] = A[d_2 \times g(C_2)]. \tag{12}$$

Applying the inverse of the function A to Equation 12, taking logarithms of each side, and rearranging terms,

$$\log(d_1) - \log(d_2) = \log\left(\frac{g(C_2)}{g(C_1)}\right) = \log(k), \tag{13}$$

where $\log(k)$—and thus, $k = g(C_2)/g(C_1)$—is a constant. That is, the difference between the logarithms of d_1 and d_2 (the durations that produce equal performance levels for contrasts C_1 and C_2) is predicted by the multiplicative theory to be constant. Another way of characterizing this prediction is to note that, when plotted on a log-duration scale, performance curves for the two contrast values must be *horizontally parallel*, separated by a constant, $\log(k)$. Equation 13 includes no task difficulty component, which indicates that the single constant, $\log(k)$, applies to both the easy and the difficult tasks. Note that if lowering stimulus contrast from a higher value, C_2, to a lower value, C_1, confirms the multiplicative theory, a natural intuitive interpretation emerges—that the same perceptual processes are occurring with the lower contrast as with the higher contrast, but at a rate that is slower by a factor of k.

A casual inspection of the pairs of performance curves shown in Figures 2A and 2B indicates that they are at least roughly horizontally parallel and separated by about the same amount. To evaluate the multiplicative theory, we carried out the following procedure. For each of the 6 observers, we iterated through a progression of k values. For each k value, we shifted the low-contrast data points from Figure 2, panels A and B, to the left by $\log(k)$. We then judged which k value produced the best simultaneous alignment of the low- and the high-contrast data points. This part of the procedure yielded the data in the "k (Best Shift)" column of Table 1. We then computed the mean k value across observers (2.694) and used it to left-shift the mean low-contrast data from Figure 2, panels A and B. The results are shown in Figure 2, panel D for both the easy task data and the difficult task data (note that to avoid visual confusion, the two easy condition curves have been shifted upward by 0.30 to make them readily distinguishable from the two difficult condition curves). It is evident that the single shift value produces reasonably, although not perfectly, aligned data points for both the easy and the difficult curve pairs. We conclude that multiplicative theory is weakly confirmed. Following a more detailed analysis below, in conjunction with our discussion of a specific quantitative theory, we will have more to say about the fit of the multiplicative theory.

Bloch's law. Bloch's law is that information and, therefore, performance are determined by (i.e., are a monotonic function of) the product of duration and contrast. This means that if contrast is increased by some factor, k, duration must be decreased by the same factor, k, to achieve equal performance. To see why this is, consider Equation 12 above and make the Bloch's law assumption that $g(C) = C$. Suppose contrasts C_2 and C_1 are such that $C_2/C_1 = k$. Applying A^{-1}, the inverse of function A, and rearranging terms produces $d_1/d_2 = C_2/C_1 = k$.

This prediction is evaluated via the data shown in Table 1, where columns 5 and 6 show the contrast ratio C_2/C_1, and k (Best Shift), which corresponds to d_1/d_2. The prediction clearly fails. For each of the 6 observers, the estimated k value is considerably greater than the physical high- to low-contrast ratio. Averaged across observers, the k value is greater than the contrast ratio by a factor of approximately 1.74 (95% confidence interval = ± 0.13). In short, increasing contrast produces a performance improvement that is considerably greater than would be predicted if stimulus information were based on the simple product of contrast and duration.

The Loftus and Busey sensory response, information acquisition theory. Because Bloch's law holds in some low-level tasks, it provides a starting point—a kind of plausible null hypothesis—for predicting the precise effect of contrast on perception and memory tasks. As we have just seen, it failed. We now consider the possibility that Bloch's law, nevertheless, still governs the effect of contrast on information acquisition, but in a disguised fashion. As an analogy, consider Newton's law of gravitation, which states that a falling object accelerates at a constant rate. In fact, an object falling in an atmosphere will be observed to accelerate at a decreasing rate (eventually ceasing to accelerate when it reaches what is called terminal velocity). However, no one would conclude from this observation that Newton's laws are disconfirmed; rather, one would note that an additional factor, air friction, is preventing the Newton's law prediction from being met.

Perhaps, in similar fashion, Bloch's law truly describes the way in which contrast and duration combine into information, but its workings are obscured by some other process. One such possibility is that there is a contrast *threshold* involved in perception of our stimuli such that the effect of contrast "begins at" some greater-than-zero level. To provide a rough intuition about what we mean by this, suppose that, from the visual system's point of view, a stimulus contrast of 0.03 corresponded to zero—that is, that the system did not respond to contrasts less than 0.03. Now consider two contrast levels, 0.12 and 0.06. From the experimenter's point of view, the ratio of these two contrast levels is 2:1, whereas from the visual system's point of view, the ratio is $(0.12 - 0.03):(0.06 - 0.03) = 3:1$. Therefore, the visual system could still be responding linearly to *its* representation of contrast, but not to *the experimenter's* representation of contrast.

Loftus and his colleagues (Busey & Loftus, 1994, 1998; Loftus et al., 1993; Loftus & McLean, 1999; Loftus & Ruthruff, 1994; Massaro & Loftus, 1996) have developed a theory that incorporates exactly this threshold assumption. The theory, called the sensory response, information acquisition (SRIA) theory, describes perception and memory of relatively simple stimuli shown at relatively low contrast levels. This theory is described in detail in numerous published articles (see Busey & Loftus, 1994, for the most complete description). We describe its essence here, demonstrate how it relates to nested Theories 1–4 described above, and apply it to the present data.

Description of the SRIA theory. The theory begins with a representation of some briefly presented stimulus as a temporal waveform, $f(t)$, relating stimulus contrast to time t since stimulus onset. For stimuli presented on a CRT, as in the present experiments, $f(t)$ would consist of a series of scallops, each scallop corresponding to a single screen refresh. However, as has been shown by Busey and Loftus (1994, Appendix D), given the nature of the human visual system and the timing of CRT refresh rates, a very good approximation can be generated with the simplifying assumption that $f(t)$ is rectangular with a width equal to d (the stimulus duration) and a height equal to C (the stimulus contrast).

The physical stimulus is assumed to initiate a *sensory response function*, $a(t)$, obtained by convolving $f(t)$ with the *impulse response function*, which is the system's theoretical response to an infinitely brief stimulus. The impulse response function is assumed to be a gamma function that is a convolution of n exponentials, each with a decay time of τ msec (Watson, 1986). Given specification of d and C, therefore, the sensory response function has the same two free parameters, n and τ, as the impulse response function. On the basis of past data, we always set n to 9, whereas τ remains a free parameter. Examples of sensory response functions for different contrasts and durations are shown in Figure 3. The top panel shows functions based on a contrast of $C = 0.10$, whereas the bottom panel shows functions based on a contrast of $C = 0.05$. Note that C simply scales $a(t)$; thus, the $C = 0.10$ functions are twice the height of the corresponding $C = 0.05$ functions.

At any given time t since stimulus onset, information is assumed to be acquired from the stimulus at some rate

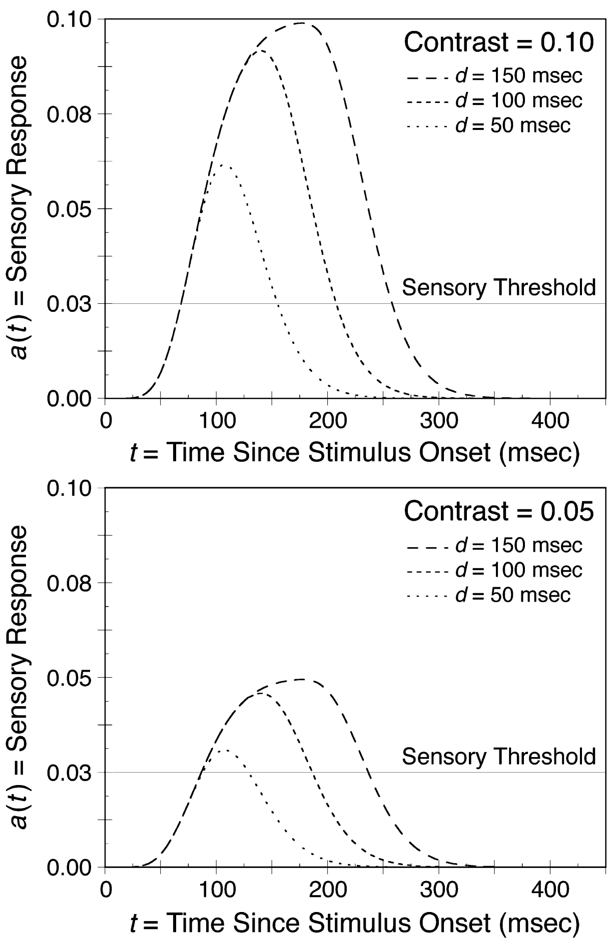

Figure 3. Sensory response, information acquisition theory: sensory response functions for three durations (50, 100, and 150 msec) and two contrast levels (0.10 and 0.05). The curves are obtained by convolving the impulse-response function generated by parameters *n* and *t* with the original, physical, stimulus contrast function *f*(*t*). The sensory threshold *q*, here set at 0.03, allows computation of the above-threshold area $A_\theta(\infty)$ for each of the six curves.

$r(t)$, whose magnitude is based in part on the instantaneous value of $a(t)$. Note that $I(t)$, the integral over time of $r(t)$, is the proportion of stimulus information acquired at time t. It is further assumed that there is a *sensory threshold* such that whenever $a(t)$ is below threshold, $r(t)$ is zero. This threshold, designated θ, is set, for illustrative purposes, to 0.03 and is shown in the two panels of Figure 3; thus, θ is a second free parameter in the theory. For ease of exposition, a new function, $a_Q(t)$ is defined to be $a_\theta(t) = [a(t) - \theta]$ for $a(t) > \theta$ and 0 for $a(t) \leq \theta$.

The information acquisition rate, $r(t)$ is, more specifically, assumed to be proportional to the product of $a_\theta(t)$ and $[1 - I(t)]$, the proportion of stimulus information not acquired by time t. The constant of proportionality is termed $1/c$, where c is a third free parameter in the theory. Given these assumptions, an equation can be written for stimulus information, which is,

$$I(t) = 1.0 - e^{-A_\theta(t)/c}, \tag{14}$$

where $A_\theta(t)$ is the integral under $a_\theta(t)$ from 0 to t. Note that the total amount of stimulus information acquired on a trial is obtained by letting t equal infinity; thus,

$$I(\infty) = 1.0 - e^{-A_\theta(\infty)/c}, \tag{15}$$

where $A_\theta(\infty)$ is the total area under $A_\theta(t)$ (see Loftus & McLean, 1999, pp. 396–397, for formal derivations of Equations 14 and 15).

Within the theory's context, $I(\infty)$ is determined by stimulus duration and stimulus contrast, plus the constrained parameter $n = 9$ and the free parameters τ, θ, and c. To complete the theory's application to the present data, we need a function mapping $I(\infty)$ to proportion correct for the easy and the difficult tasks. To do so, we must consider more carefully what is signified by the term *information*. Past applications of this theory have been to experiments in which study variables—for example, the shape of $f(t)$—were manipulated and a single memory task (e.g., digit recall) was used. Therefore, the term *information* could be roughly interpreted as *proportion of stimulus features acquired*, and the intuitively reasonable assumption was made that proportion correct on the memory test was equal to acquired information. This assumption served quite well, inasmuch as past fits of the theory to the data have been most excellent.

The present experiment is different in that there are two memory tasks, difficult and easy. Given that these two tasks are designed to and do yield different performance levels, it is obviously inadequate to assume that proportion correct is simply equal to $I(\infty)$; this would incorrectly imply equal performance for the difficult and the easy tasks. To solve this problem, we allow different c values, c_D and c_E, that correspond to the two memory tasks. This means that Equation 15 must be modified to index task difficulty—that is,

$$I_T(\infty) = 1.0 - e^{-A_\theta(\infty)/c_T}, \tag{16}$$

where the index T refers to task difficulty and can be either D or E.

This modification violates a simplistic intuition about what is meant by *acquired information*, for the following reason. The experimental design is such that an observer seeing a stimulus during the study phase of a trial does not even know whether the trial is a difficult or an easy trial; this knowledge cannot become apparent until the test portion of the trial. So how can *acquired information* be different depending on whether the test will eventually be difficult or easy? (It sounds like precognition!) The answer is that acquired information in memory is not a static entity, like acquired money in a bank account; rather, information must be viewed as *task relevant*. One might maintain a presumed-to-be-static theoretical construct, such as *proportion of acquired features*; however, the value of proportion of acquired features differs depending on the task—and it is the *value* of the acquired features that $I(\infty)$ refers to in Equation 16. To use a simple example, suppose that either low-contrast digits or a blank field had been briefly presented. Presentation of a digit would yield some information; however this information would have a different value if the task was one of simply *detecting* whether a digit versus a blank was presented than it would if the task was *identifying* which digit was presented.

Asymptotic performance. Ideally, we could simply equate observed performances P_E and P_D in the easy and the difficult tasks, to $I_E(\infty)$ and $I_D(\infty)$, respectively. Alas, we cannot, because by Equation 16, $I_T(\infty)$ asymptotes at 1.0. That is, $A_\theta(\infty)$ can become arbitrarily large—for example, as stimulus duration increases—and as it does so, $I_T(\infty)$ approaches 1.0. Form recognition data, however, do not asymptote at 1.0. Instead, for reasons that are beyond the scope of this article, they invariably asymptote at some lower value. Therefore, at the very least, we must include an asymptote in the theory in order that the theory adequately predicts performance. Indeed, again for reasons that are beyond the scope of this article, the easy and the difficult tasks may asymptote at different levels, which would imply two separate asymptotes. The issue of whether one or two asymptotes are used is irrelevant to the main point of this article; somewhat arbitrarily, we chose to use two asymptotes in the data fits shown in Figure 2.

Theory fit. There are four basic free parameters in the theory: τ, a parameter of the impulse response function; θ, the sensory threshold; and c_E and c_D, just described. In addition, there can be either one asymptote (which we term Y) or two asymptotes corresponding to the two tasks (which we term Y_E and Y_D); thus, the theory has either five or six free parameters. We fit the theory to the 6 observers individually, using a minimum root-mean square error (RMSE) fit criterion. Each fit was carried out twice, once using a single asymptote and again using two asymptotes. The results of the data fits are shown in Table 2. The predicted theoretical curves were averaged across observers to provide the mean predictions shown as curves through the data points in Figure 2. The RMSEs between the mean data and the mean predictions were 0.048 and 0.033 for the one- and the two-asymptote fits, respectively.

How well does the SRIA theory fit? The average mean square error of the data is 0.012, which means that the F values corresponding to the theory null hypothesis are 3.928 and 2.727 for the one- and the two-asymptote fits, respectively. This allows us to statistically reject the SRIA theory as not being sufficient to describe the data points. However, a glance at the data fits in Figures 2A and 2B, along with the barely visible standard errors, makes it evident that we can reject the theory only because the data have enormous statistical power. Another way to evaluate the theory–data fit is to calculate the over-conditions Pearson r^2 between the data and the theoretical predictions, which are .970 and .987 for the one- and the two-asymptote fits, respectively. Thus, the theory captures a very large proportion of between-conditions variance.

Which fit measure should we take most seriously? It is axiomatic that any theory, no matter how closely it approximates reality, cannot fit an empirical data set perfectly, which means that if enough statistical power is brought to bear, the theory will always be disconfirmed via an F or a related statistic. A correlation between theory and data, on the other hand, behaves quite differently: To the degree that the theory correctly describes reality, the underlying (i.e., population) correlation will be high, and with greater statistical power, the obtained correlation will be a better estimate of the population correlation. Thus, a high correlation in conjunction with high power provides a reasonable basis for the claim that

Table 2
Stimulus Response, Information Acquisition Theory Parameter Fits to the Experiment 1 Data

Observer	Parameter Value							RMSE
	τ	θ	c_E	c_D	Y	Y_E	Y_D	
	One-Asymptote Fit							
S.K.C.	3.759	0.035	0.603	1.081	0.830	–	–	0.060
E.C.U.	5.820	0.034	0.243	0.503	0.839	–	–	0.106
E.M.H.	4.803	0.040	0.698	1.140	0.812	–	–	0.084
J.C.M.	4.227	0.044	0.375	1.314	0.669	–	–	0.088
C.A.W.	3.040	0.048	0.886	2.663	0.649	–	–	0.064
J.S.F.	1.286	0.050	1.382	3.059	0.703	–	–	0.075
	Two-Asymptote Fit							
S.K.C.	3.763	0.035	0.618	0.956	–	0.850	0.790	0.061
E.C.U.	5.823	0.034	0.265	0.451	–	0.861	0.803	0.108
E.M.H.	4.720	0.042	0.741	0.580	–	0.884	0.675	0.062
J.C.M.	3.944	0.048	0.293	0.549	–	0.698	0.533	0.077
C.A.W.	2.910	0.054	0.718	0.885	–	0.675	0.467	0.042
J.S.F.	1.762	0.059	0.870	0.873	–	0.717	0.497	0.058

Note—RMSE, root-mean square error.

"the theory fits" even if a high *F* ratio demonstrates what would be agreed to by all a priori—that no theory, no matter how good it is, can be perfect. On these bases, we assert that the correlations are better measures, and we conclude that the theory fits the data reasonably well.

Discussion

There are two important take-home messages from Experiment 1. First, the data allow us to confirm a unidimensional theory with respect to the way in which contrast affects memory for visually presented material. It is reasonable to conclude that duration and contrast combine at an early information-processing stage to produce a single number (*information*) upon which immediate 2AFC recognition memory for random forms is based. Confirmation of a unidimensional theory allows us to assert, on the basis of a firm definitional foundation, that contrast affects the easy and the difficult tasks in the same way: That is, the effect of contrast can be construed to be *only* on the generation of a information value—which is irrelevant to task difficulty—and it is this information value that then determines, in different ways, eventual performance on the two tasks. A simple way of saying this is that contrast affects the easy and the difficult tasks in the same way.

Second, we have, with reasonable success, fit a specific quantitative theory to the data. This theory incorporates unidimensionality. Because of its property of allowing a nonzero threshold, it does not generally predict Bloch's law; nor, as we shall see in a later section, does it even predict a multiplicative effect, which is why we tempered our conclusions about the *confirmation* of the multiplicative theory indicated in Figure 2D. It does, however, incorporate some very Bloch's-law-like and, accordingly, some very multiplicative-theory-like properties, which we will discuss in more detail following our report of Experiment 2.

Experiment 2
Form Recognition Versus Digit Recall

Experiment 1 showed that reducing contrast affects performance equally on both an easy and a difficult form recognition task. However, the design of the experiment was such that the nature of the easy–difficult manipulation could not become apparent until the 2AFC test. That is, the easy–difficult manipulation was not relevant at the time that the form was originally viewed, and for this reason, it is perhaps not surprising that the stimulus contrast effect was the same. In Experiment 2, we implemented a manipulation that was relevant at study as well as at test. In particular, instead of having two test difficulty levels, we had two entirely different STCs. One STC consisted of random forms tested in a 2AFC recognition procedure, as in Experiment 1, whereas the other STC consisted of digit strings tested by serial recall. Thus, Experiment 2 was conceptually similar to Experiment 1, where STC was substituted for task difficulty. Our basic question was analogous to that posed in Experiment 1: Is the contrast effect the same for the two STCs, where "the same" is as defined by the various theories.

Method

The method was similar to that used in Experiment 1. The principal change was that instead of using two form recognition difficulty levels, we used two STCs: form recognition and digit recall.

Observers

Observers. The observers were 7 University of Washington undergraduates and graduates. Each reported normal or corrected-to-normal vision and was paid $180.

Stimuli. Two sets of stimuli were used in Experiment 2: random forms and four-digit number strings. A total of 336 forms were created in Experiment 2, using the same stimulus creation methods as those described in Experiment 1. The delta value for Experiment 2 varied among observers (10–14 pixels), to accommodate a range of observer ability, but was held constant for each individual observer. The second stimulus set consisted of four-digit strings. Each string was created randomly, selecting digits with replacement, just prior to the start of the trial on which it was to be used. Digit strings were displayed on the screen at a 20-pixel font height.

Apparatus. The computer, monitor, and display software were the same as those in Experiment 1.

Design. There were 28 conditions: 7 stimulus durations × 2 contrasts × 2 STCs. As in Experiment 1, because we wanted to achieve roughly equal performance ranges for the two contrast levels, the duration ranges were different for the high and the low contrasts: Again, for the low-contrast conditions, durations ranged from 40.5 to 148.5 msec in equal increments, whereas for the high-contrast conditions, durations ranged from 13.5 to 121.5 msec in equal increments. Again, specific contrast values varied between observers, to accommodate a range of observer ability, but were held constant in the two tasks for each individual observer. Table 3, columns 2 and 3, shows the high- and low-contrast values for each of the 7 observers.

Each observer first participated in two practice sessions, one for digits and the other for forms, and then participated in 16 data collection sessions. During each session, each observer completed seven blocks of form recognition trials, followed by five blocks of digit recall trials, or vice versa. The form-trials/digit-trials order alternated over sessions. Each block consisted of 42 stimulus trials, 2 in each of the 28 conditions. Thus, over the course of the experiment, each observer completed 4,704 form recognition trials and 3,360 digit recall trials. The 4,704 form recognition trials consisted of the

Table 3
Contrast Information for Experiment 2

Observer	Low Contrast	High Contrast	Contrast Ratio	*k* (Best Shift)
T.M.B.	0.043	0.073	1.698	2.305
S.K.C.	0.043	0.073	1.698	2.535
J.R.S.	0.043	0.073	1.698	2.192
L.E.P.	0.043	0.073	1.698	2.872
M.M.R.	0.043	0.073	1.698	2.472
K.D.L.	0.058	0.088	1.517	2.086
C.A.W.	0.058	0.088	1.517	1.984
M	0.047	0.077	1.646	2.349

Note—The "Contrast Ratio" column indicates the ratio of high contrast to low contrast. The "*k* (Best Shift)" column shows the ratio of durations required to achieve equal performance in the low-contrast, as compared with the high-contrast, condition. Note that *k* is constrained to be the same for forms and digits.

rotation of all 336 target forms through the 14 contrast/duration conditions. The 3,360 digit recall trials consisted of 240 replications of each of the 14 conditions.

Procedure. The procedure for the form recognition task was the same as that in Experiment 1. For the digit recall task, each trial began with a 250-msec display of a fixation point accompanied by a warning tone, followed 500 msec later by the target stimulus. The observers then used the number pad on the keyboard to type in four digits, guessing if necessary. Immediately following each response, visual feedback was given in the form of four Xs and Os corresponding to incorrect and correct digits in the string. To be coded as correct, both digit identity and position in the string had to be correct. For example, if the stimulus was 9323 and the observer entered 9228, feedback would read OXOX.

In the form recognition part of the experiment, the conditions were counterbalanced as in Experiment 1. In the digit recall part of the experiment, each of the 14 conditions occurred three times during each 42-trial block. Because digit stimuli were created randomly prior to each trial, there was no need to counterbalance them.

Results

Digit recall was scored as number of digits correctly reported in their correct serial positions. For both the form recognition and the digit recall tasks, proportion correct, corrected for the guessing rates of 0.5 for forms and 0.1 for digits, was computed for each of the 14 conditions and was averaged across sessions for each observer. Figures 4A and 4B show mean performance as a function of exposure duration for the two contrast levels, averaged across the 7 observers for form recognition and digit recall. As was expected, performance in both tasks increased with stimulus duration and was greater for high contrast than for low contrast.

Evaluation of Theories

To test our series of nested theories, we carried out the same analyses as that in Experiment 1.

Unidimensional theory. Figure 4C shows the state-trace plot: Form recognition is plotted as a function of digit recall. Again, low- and high-contrast conditions are represented by open and closed symbols. For all intents and purposes, the state-trace plot is monotonic, thereby confirming a unidimensional theory.

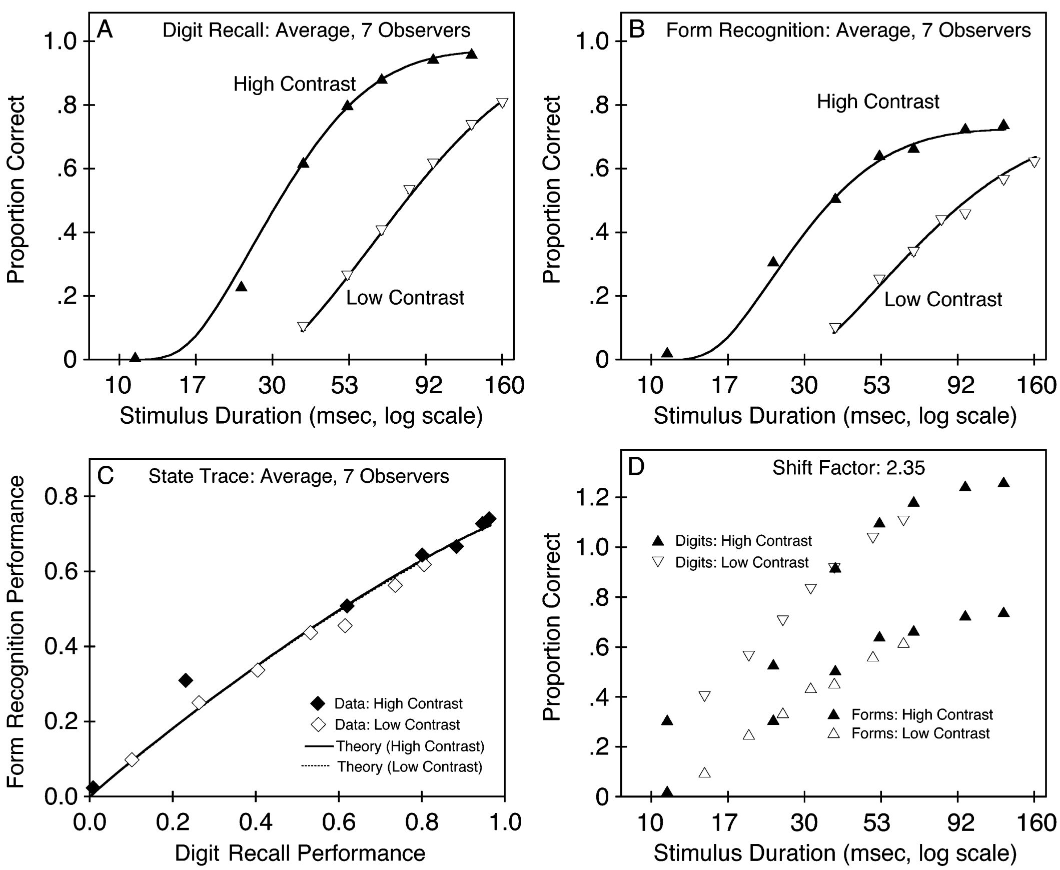

Figure 4. Experiment 2 data. (A and B) Response probability corrected for guessing as a function of stimulus duration (log scale). (C) Difficult performance as a function of easy performance (state-trace plot). (D) The low-contrast curves from the top two panels have been shifted leftward by an identical amount, corresponding to a factor of 2.35 (for visual clarity; the "digits" curves have been shifted upward by 0.30). With this shift, the high-contrast and the low-contrast curves align essentially perfectly for both digits and forms. Solid and dotted lines in panels A–D represent theoretical fits of the sensory response, information acquisition theory described in the text. All error bars (most of them obscured by the plot symbols) represent standard errors.

Multiplicative theory. As in Experiment 1, we carried out a *shifted-curve analysis* to test the multiplicative theory. The results are shown in Table 3, rightmost column. For each observer, we shifted both the form recognition and the digit recall low-contrast curves horizontally, on a log-duration scale, and determined the shift magnitude that afforded the optimal overlap of the low- and the high-contrast curves. We then calculated the mean shift, which corresponded to a linear duration scaling factor of 2.349. Figure 4D shows the low-contrast curves left-shifted by this amount (with the two digit curves shifted upward for visual clarity). It is apparent that they overlap reasonably well, although by no means perfectly, with their high-contrast counterparts, thereby again weakly confirming a multiplicative theory. Again, we will qualify this conclusion in conjunction with our discussion of the SRIA theory.

Bloch's law theory. The Bloch's law prediction was evaluated as it was in Experiment 1, now via the data shown in the two rightmost columns of Table 3, and again failed. For each of the 7 observers, the estimated $d_1/d_2 = k$ value was greater than the physical high- to low-contrast ratio. Averaged across observers, the k value was greater than the contrast ratio by a factor of approximately 1.43 (95% confidence interval = ± 0.15). Again, increasing contrast produced a performance improvement that was greater than would be predicted if stimulus information had been based on the product of contrast and duration.

SRIA theory. To apply the SRIA theory to the present data, we again fixed n, the number-of-stages impulse response function parameter, at 9. We let τ, the other impulse response function parameter, and θ, the sensory response threshold, vary as free parameters. We allowed separate values of c and the asymptote Y for forms and digits. The individual-fit data are shown in Table 4 for the individual observers, and the predicted performance is included with the data in Figure 4, panels A–C. For the average data, the theory–data RMSE is 0.024.

How well does the SRIA theory fit? The average mean square error of the data is 0.007, yielding an F value of 3.38, which allows us to statistically reject the SRIA theory as being insufficient to describe the data. Nevertheless, the Pearson r^2 between the data and the theoretical predictions is .996: As in Experiment 1, the theory–data correspondence is extremely close, and again we can reject the theory only because of the enormous statistical power.

Discussion

The results of Experiment 2 are, frankly, surprising. We began with two very different kinds of stimuli, random forms and digit strings, and subjected them to two very different memory tasks: 2AFC recognition versus recall. The only common element was the identical duration–contrast design structure. One might suppose on intuitive grounds that the effect of *anything*—and of contrast in particular—would be somehow different for the two different STCs. And yet to a very high degree of precision, the contrast effect was identical. This identity was demonstrated at a general level in the state-trace plot (Figure 4C), which confirmed a unidimensional theory, and at a more specific level in the SRIA theory fits, just discussed.

Placement of the SRIA Theory Within Nested Theories 1–4

The SRIA theory has been remarkably successful in accounting for data from Experiments 1 and 2, as well as those from numerous previously published data sets (Busey & Loftus, 1994, 1998; Loftus & Irwin, 1998; Loftus & McLean, 1999; Loftus & Ruthruff, 1994; Massaro & Loftus, 1996). We now consider where the SRIA theory fits within the context of the four nested theories that we described above: multidimensional theories, unidimensional theories, multiplicative theories, and Bloch's law theories. We first will show why the SRIA theory is an example of a unidimensional theory, and then we will address the somewhat more complicated issue of whether it may be construed as a Bloch's law theory. This context within which the SRIA theory may be viewed will be helpful in interpreting its eventual application to the complex-stimuli data of Experiments 3–6. In what follows, we shall, somewhat arbitrarily, illustrate our points by using Experiment 1 as an example; it should be understood that analogous arguments would apply to its application to Experiment 2.

The SRIA theory is a unidimensional theory. By the SRIA theory, contrast and duration are combined at

Table 4
Stimulus Response, Information Acquisition Theory Parameter Fits to the Experiment 2 Data

Observer	Parameter Value τ	θ	c_F	c_D	Y_F	Y_D	RMSE
T.M.B.	5.262	0.027	0.873	1.037	0.740	0.973	0.044
S.K.C.	5.599	0.032	0.804	0.862	0.823	0.978	0.034
J.R.S.	4.991	0.030	0.364	0.568	0.841	0.960	0.036
L.E.P.	4.841	0.034	0.797	0.906	0.713	0.955	0.039
M.M.R.	5.321	0.031	0.586	0.685	0.649	1.000	0.039
K.D.L.	4.815	0.043	0.649	0.568	0.647	0.960	0.036
C.A.W.	5.898	0.039	0.784	1.006	0.664	1.001	0.050

Note—RMSE, root-mean square error.

an early stage via the parameters n, τ, and θ, to produce a single number, $A_\theta(\infty)$ (see Figure 3). Then computation of *information* depends only on $A_\theta(\infty)$. Thus, it is perfectly possible for two conditions—a short-duration high-contrast condition and a long-duration low-contrast condition—to lead to the same $A_\theta(\infty)$ value. Two such conditions would be indistinguishable from the perspective of any process, such as response generation, that depended on $A_\theta(\infty)$ without access to the original values of duration and contrast. In short, the SRIA theory is a unidimensional theory, with $A_\theta(\infty)$ constituting the single value that is produced by combining duration and contrast. This unidimensional property is demonstrated by the overlapping low- and high-contrast theoretical state-trace curves in Figures 2C and 4C.

Is the SRIA theory a Bloch's law theory? A key property of sensory response functions generated by the SRIA theory is that they are linear with respect to both contrast and duration. As a consequence, the *total* area under a sensory response curve, $A(\infty)$, is proportional to (and actually, in the present treatment, equal to) the area under the original $f(t)$ function $F(\infty)$. Because $f(t)$ is rectangular, with a width equal to d and a height equal to C, $F(\infty)$ is simply the product of duration and contrast.

The zero-threshold special case: The SRIA theory is a Bloch's law theory. An implication of this property is that if the sensory threshold θ were zero, the theory would be a Bloch's law theory. This is illustrated in Figure 5. Consider first the solid and the dashed-dotted $a(t)$ functions that correspond to two contrast–duration conditions: $C = 0.10$, $d = 50$ msec and $C = 0.05$, $d = 100$ msec. These two conditions have the same contrast × duration product (5.00) and, accordingly, the same $F(\infty)$ values. Therefore, as is indicated in Figure 5, the total areas under these two $a(t)$ curves are equal as well: They also are both equal to 5.00. This example illustrates Bloch's law: Doubling the duration requires halving the contrast for equal values of $A(\infty)$ and, hence, equal performance.

Greater-than-zero threshold: The SRIA theory is neither a Bloch's law nor a multiplicative theory. However, if the threshold is greater than zero—in the Figure 5 example, $\theta = 0.04$—the situation is different. The above-threshold area under the high-contrast short-duration curve is 2.14, whereas the above-threshold area under the low-contrast long-duration curve is only 0.69. This illustrates a general principle that holds for any greater-than-zero threshold: For two different stimuli whose contrast × duration product is the same, the above-threshold area is greater for the shorter high-contrast stimulus than for the longer low-contrast stimulus. In the Figure 5 example, to achieve 0.05-contrast performance equal to that in the 0.10-contrast condition, duration must be more than doubled: It must increase to about 245 msec—that is, it must be raised by a factor of almost 5:1 to compensate for the 2:1 contrast decrease.

Indeed, with a greater-than-zero threshold, the SRIA theory is not even a multiplicative theory. This is illustrated in Figure 6. We first consider a continuum of high-contrast ($C = 0.10$) *base durations* ranging from 50 to 400 msec and calculate by what ratio this base duration must be increased in order for a $C = 0.05$ stimulus to match above-threshold area and, thus, performance for var-

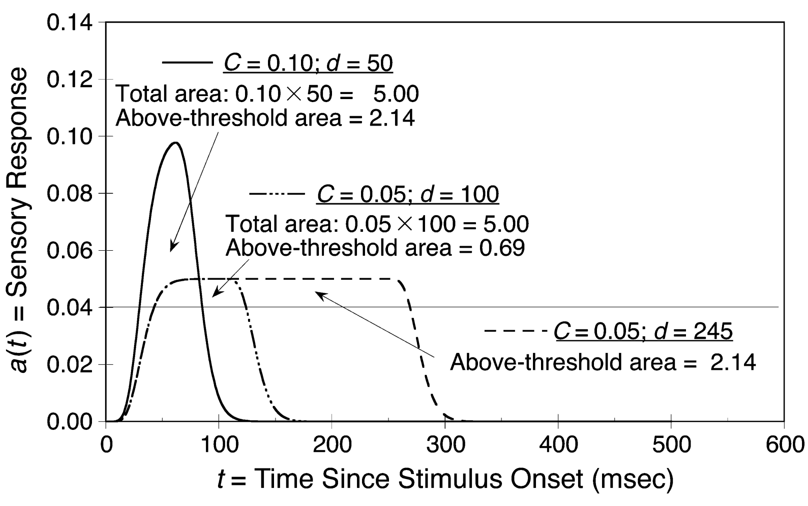

Figure 5. Illustration of why the sensory response, information acquisition theory is a Bloch's law theory for zero threshold, but not a Bloch's law theory for greater-than-zero thresholds. For a threshold of $q = 0$, the $C = 0.10$, $d = 50$ and the $C = 0.05$, $d = 100$ conditions yield the same area under the sensory response curve (5.00) and, therefore, lead to the same performance; in general, to maintain equal performance, duration must be increased by the same factor as that by which contrast is decreased. For a threshold greater than zero—$q = .04$, in this example—the low-contrast duration needs to be increased from 50 to 245 msec in order to maintain the same above-threshold area and, thus, the same performance.

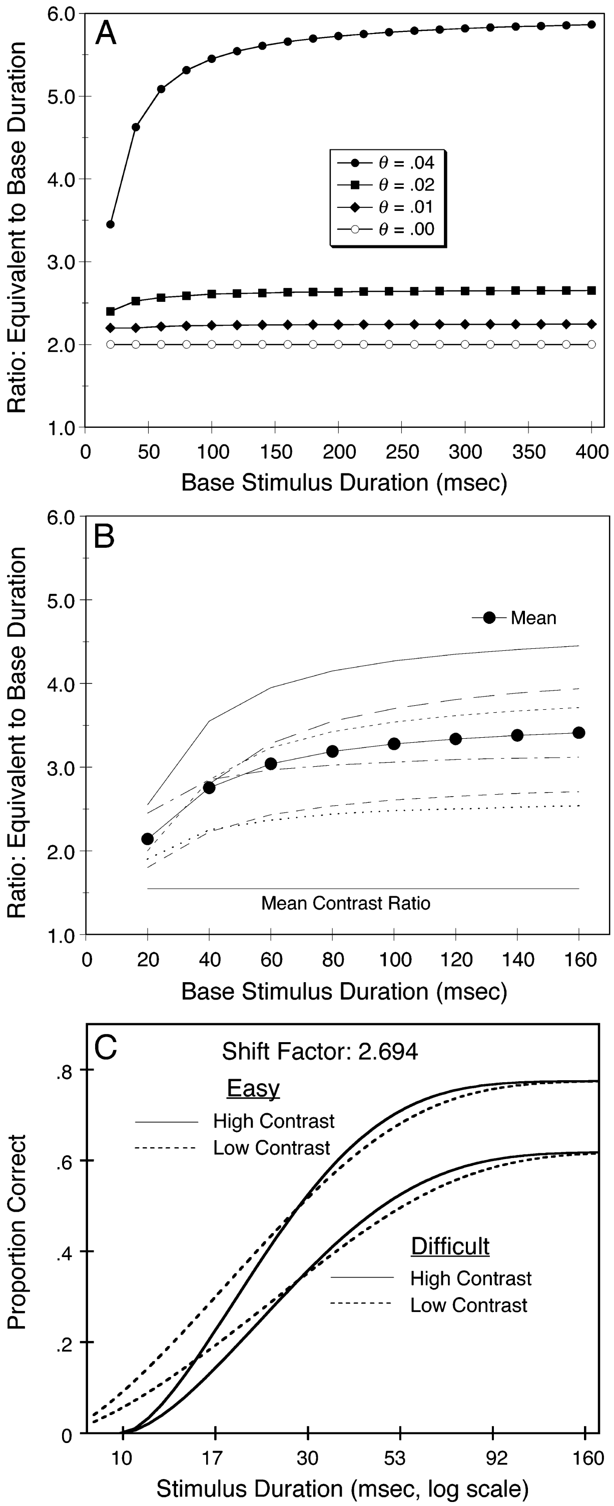

Figure 6. Illustration of why the sensory response, information acquisition (SRIA) theory with greater-than-zero thresholds is not a multiplicative theory. (A) Theoretical illustration with four *q* values. (B) Data from individual observers based on estimated parameters. In panels A and B, the bottom, flat curve shows the Bloch's law prediction. (C) SRIA theory predictions of best-shifted data of Figure 2D.

ious threshold values. Figure 6A shows the ratio of matching low-contrast to high-contrast durations as a function of base duration for four θ values ranging from zero to 0.04. With $\theta = 0$, the theory is, as has just been described, a Bloch's law theory that yields a constant ratio equal to the ratio of high to low contrast—in this case, 0.10/0.05 = 2.0. A θ of 0.04 is, in a sense, at the other extreme, in that $\theta = 0.04$ is close to the actual low-contrast value of 0.05. This is relevant because the contrast value, C, represents the maximum possible value of $a(t)$. Therefore, if θ exceeds C, $a(t)$ can never exceed threshold, and performance can never exceed chance. When C is close to θ, the theory becomes highly nonmultiplicative, in that the matching-durations ratio is still rising notably even with base durations as high as 150 msec.

Figure 6B shows similar calculations for the 6 observers in Experiment 1. For each observer, we used that observer's estimated τ and θ parameters, along with their high- and low-contrast values, to generate equivalent-performance low-contrast durations for the high-contrast base durations in the general range of the durations used in Experiment 1 (20–160 msec). Data are shown for the individual observers, as well as for the mean over observers. It is evident that the curves all rise over the durations that we used, implying that the theory predicts departures from multiplicativity. This prediction is at odds with the seemingly well-behaved shifted curves shown in Figures 1D and 3D. As it turns out, the SRIA theory predicts curves that align reasonably well, but not perfectly. This is shown in Figure 6C, where the mean predicted performance curves are shown shifted by the amount that the empirical curves were shifted in Figure 2D: As can be seen, the overlap between the high-contrast and the low-contrast curves, although reasonable, is not perfect, particularly at short durations.

It is clear, in short, that in its predictions about response probability, the SRIA theory, although unidimensional, is neither a Bloch's law nor a multiplicative theory when $\theta > 0$. However, at the heart of the theory is a very simple assumption, which is that contrast and duration do combine multiplicatively and, indeed, in a Bloch's law fashion to produce the area under the $a(t)$ function. At the risk of some hubris, we return to our analogy with physical laws. A complete theory designed to predict the rate at which an object falls in the atmosphere would require assumptions about the effects of air friction and would predict a somewhat complex function relating the object's velocity to time since the object began to fall; however, at the heart of this function would be a simple law: Objects in a vacuum fall at a constantly increasing rate.

Experiments 3–6
Complex Stimuli

Experiments 3–6 were similar in concept and in design to Experiments 1 and 2 in that they entailed semifactorial combinations of stimulus duration and stimulus contrast, along with some third variable akin to difficulty in Experiment 1 and STC in Experiment 2. There were two main changes in Experiments 3–6. First, the stimuli were grayscale naturalistic pictures. Second, memory was tested in a delayed yes–no recognition procedure, in addition to being tested immediately following each stimulus presentation via a prospective confidence judgment.

In Experiments 1 and 2, the main question was, essentially, What is the effect of contrast, and how is it similar over task difficulty (Experiment 1) or STC (Experiment 2)? In each of Experiments 3–6, we asked two analogous questions. First in each of Experiments 3–6, the same contrast and duration levels were applied to two different stimulus types (e.g., faces and cityscapes); thus, we can ask, To what degree is the effect of contrast similar over the two stimulus types? The second question addressed in Experiments 3–6 was, Is the effect of contrast the same for the two performance measures, immediate prospective confidence and delayed recognition performance? More formally, the unidimensional theories described earlier assume that as part of the process of perception, duration and contrast combine into a single value called *information*, which then, via monotonic functions, determines immediate performance. To extend these theories to longer term memory performance, it must further be assumed that a long-term memory representation of a picture is formed, that the *quality* (however defined; see, e.g., Gillund & Shiffrin, 1984, Hintzman, 1987, and Murdock, 1993, for examples) of this long-term representation is monotonically related to original information, and that long-term recognition performance is monotonically related to the quality of the long-term representation. As was discussed earlier, the latter assumption is not always met: It is possible to structure an experiment in which shorter durations actually lead to higher recognition performance in a later, surprise test (Nairne, 1988). However, as also was discussed, this assumption is usually met: In the vast majority of picture recognition experiments, longer durations and/or higher contrasts lead to better performance.

Busey et al. (2000) reported an experiment addressing these issues that was similar to Experiments 3–6. They used a face recognition paradigm in which study duration was varied and three dependent variables—prospective confidence, recognition accuracy, and retrospective confidence—were measured. The difference between Busey et al.'s experiment and the present Experiments 3–6 is that Busey et al. substituted *rehearsal* for *contrast*: For 15 sec following the offset of each studied face, observers were either required to rehearse or were prevented from rehearsing the appearance of the face that they had just seen. Although rehearsal increased both prospective confidence and eventual recognition performance, the two measures did not issue from the same unidimensional structure: Roughly speaking, prospective confidence increased following rehearsal considerably more than was warranted given its eventual effect on performance. Busey et al. offered a theory of these results wherein rehearsal raised the value of not one, but two internal constructs, which they labeled "strength" and "certainty." Prospective confidence was then a joint monotonic function of strength and certainty, whereas recognition performance and retrospective confidence depended on strength only.

If the contrast effect for naturalistic scenes is the same as that for simple stimuli (Experiments 1 and 2), we would expect a different result from that obtained by Busey et al. (2000): In particular, we expect the contrast effect to be the same for prospective confidence and recognition performance, and as well, we would expect the contrast effect to be the same for different stimulus types. In short, we would expect contrast, a low-level sensory variable, to behave differently from rehearsal, a high-level cognitive variable.

Method

The experimental designs were identical for Experiments 3–6; the only thing that changed among them was the nature of the stimuli.

Observers. A total of 662 observers participated in Experiments 3–6, distributed across the four experiments as shown in Table 5, column 2. The observers were recruited from the University of Washington observer pool and received course credit for participating in a single session that lasted approximately 40 min. No observer served in more than one of the experiments. In each experiment, the observers were run in 24 groups, each of which ranged in size from 6 to 8 observers.

Stimuli. Five sets of naturalistic stimuli, obtained from various sources, were digitized as grayscale images and were normalized with respect to luminance in such a way that the pixel values for each image spanned the full intensity range from 0 to 255. All the stimuli were 450 pixels high and, except for members of one of the stimulus sets (celebrity faces; see below), 400 pixels wide. The stimulus sets were as follows.

1. Computer-generated (CG) faces. The FACES "Identikit" application was used to generate a set of 144 faces, 73 males and 71 females, examples of which are shown in Figure 7, row 1. The faces were quite realistic and included wide variations in hairstyle, face shape, facial hair, and all basic facial features—for example, eyes, nose, or mouth.
2. Hooded faces. These were 144 photos of Indiana University students, 54 males and 90 females, examples of which are shown in Figure 7, row 2. No face included any facial hair. All the faces were hooded so that there was no discernable hairstyle variation among the faces.
3. Houses. These were 144 photos of similar houses in a middle-class Seattle neighborhood, examples of which are shown in Figure 7, row 3.
4. Cityscapes. These were 144 heterogeneous cityscapes scanned from various books and magazines, examples of which are shown in Figure 7, row 4. They included skylines, individual buildings, street scenes, sports stadiums, and bridge/water scenes.

Table 5
Experiments 3–6 Design Information

Experiment	No. of Observers	Subsession 1 Stimuli	Subsession 2 Stimuli
3	171	CG faces	hooded faces
4	167	CG faces	houses
5	154	CG faces	cityscapes
6	170	CG faces	celebrity faces

Note—CG, computer generated.

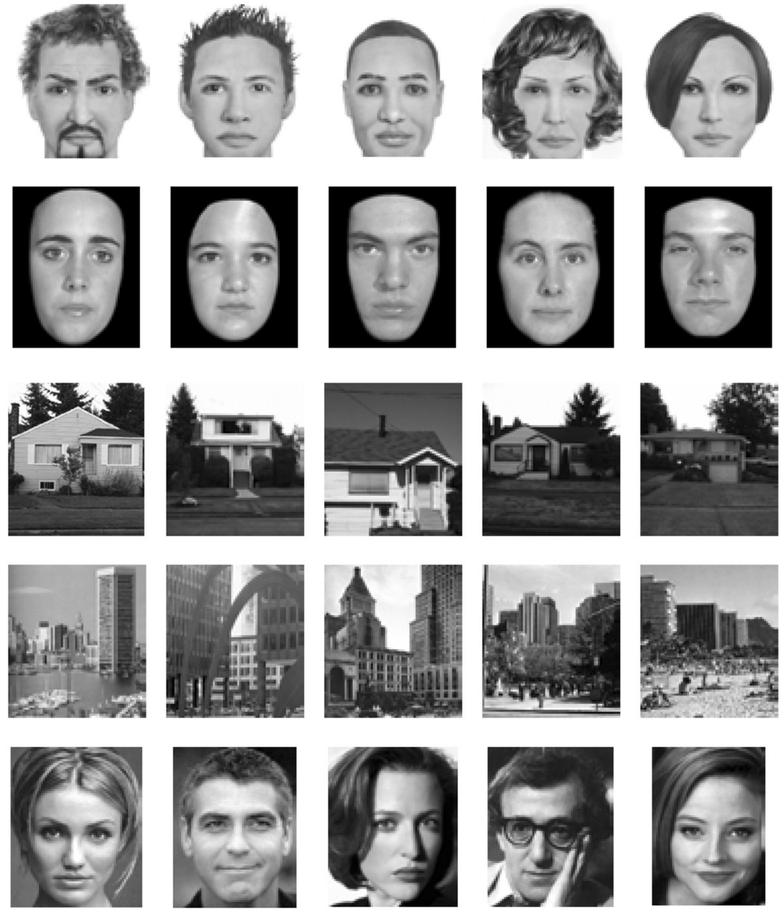

Figure 7. Examples of the stimuli used in Experiments 3–6.

5. Celebrity faces. These were 144 photos of reasonably well known celebrities, 90 males and 54 females, examples of which are shown in Figure 7, row 5. They were mostly obtained from the Internet, although some were scanned from books and magazines. Like all the other stimuli, the celebrity faces were 450 pixels high; unlike the other stimuli, however, their widths varied from 294 pixels (John Travolta) to 466 pixels (Marilyn Monroe).

Apparatus. The stimuli were displayed on a flat-white wall via a 60-Hz LCD projector interfaced to a Macintosh G4 running under MATLAB. The observers sat in eight seats arranged in two rows of four. The stimuli subtended visual angles of 26º or 44º vertically depending on whether the observer sat in the back or the front row, and except for the celebrity faces, their horizontal extent varied 34º or 57º over pictures. Picture luminance ranged from 0.2 cd/m^2 (at a zero grayscale value) to 14.8 cd/m^2 (at a 255 grayscale value). When no picture was displayed (a *blank screen*) the screen was at a 128 grayscale value, which corresponded to a luminance of 3.7 cd/m^2. Each observer was equipped with a keypad for responding.

Design. Each experimental session incorporated two subsessions. The CG faces were always used as the stimuli in the first subsession, whereas another stimulus set was used in the second subsession, as is indicated in Table 5, columns 3–4. Except for the different stimuli, the two subsessions were identical. The basic unit of each subsession is called a *tray*, honoring its physical slide-tray ancestry. A tray contained 24 pictures (virtual "slides"), and each subsession involved six trays; hence, there were 144 pictures in each subsession, or 288 pictures in all. Each tray constituted an independent replication that had nothing to do with counterbalancing considerations. The selection of six as the number of trays was dictated by (1) the amount of time in an experimental session and (2) the limited number of hooded face pictures that we had available.

For each tray within each subsession, 12 target pictures were shown in a study phase, followed by all 24 of the tray pictures (i.e., the 12 just-viewed targets plus 12 never-seen distractors) in a test phase. The targets were shown in the same order for all the groups in a given experiment. Likewise, all the test pictures were shown in the same order for all the groups in a given experiment; also at test, the target pictures, although of course intermingled with distractors, appeared in the same order as they had at study.

Each target stimulus fell into one of 12 study conditions. Each study condition was defined by a combination of one of two stimulus contrasts and one of six stimulus durations.

As has been described by Peli (1990), the definition of contrast is somewhat arbitrary in grayscale pictures. There are numerous such definitions. *Contrast energy* is the average squared deviation between individual pixels and the mean luminance. Peli has offered a definition of *band-limited contrast*, computed at successive nonoverlapping spatial frequency bands. We computed both of these, plus several others, for all our stimuli. A problem arises, which is that the ordering of mean contrast across stimulus sets differs slightly depending on which definition of contrast is used. Accordingly we used a simple, generic definition: Each picture was first scaled so that its grayscale value ranged from 0.0 to 1.0—that is, from minimum to maximum values. Then, to attain a desired contrast *C*, every pixel was adjusted by $i' = (i \times C) + [1 - (C/2)]$, where *i* and *i′* are the original and the new grayscale intensity levels. The two contrasts used in Experiments 3–6 were 0.03 (low contrast) and 0.20 (high contrast). Within each contrast level were six durations. For low contrast, the durations were 50, 99, 150, 250, 450, and 750 msec, whereas for high contrast, the durations were 17, 33, 50, 83, 150, and 250 msec.

During the study phase of a given tray, each of the 12 study conditions was assigned to one of 12 target stimuli. At test, all the stimuli were shown at full contrast.

Procedure. An experimental session began with the display of four practice CG faces shown in a randomly selected 4 of the 12 study conditions, along with detailed study instructions and general test instructions. Following the first study phase, detailed test instructions were provided. Following the six trays in the first experimental subsession, the observers had a brief break while the second stimulus set was loaded. They then saw four practice examples of the stimulus set to be used in the second subsession. Following a brief reminder of study and test instructions, they then proceeded through the six trays of the second subsession.

Each study trial consisted of a 1,000-msec fixation cross that appeared at the center of the screen, accompanied by a 1000-Hz warning tone. After the fixation-cross/warning-tone, there was a 300-msec blank screen and, then, the target picture at its assigned contrast level for its assigned duration, followed by the blank screen. After a poststimulus offset delay of 500 msec, signaled by a beep, the observers rated their confidence that they would be able to subsequently recognize the picture on a 4-point scale, where 1 was *definitely no*, 2 was *probably no*, 3 was *probably yes*, and 4 was *definitely yes*. The scale, with its verbal labels, was displayed perpetually at the top of the screen during the study phase. The observers responded using their keypads. For ease of responding, the actual keypad keys used were 4, 5, 6, and +, corresponding to the four verbal labels. These four nominal responses were converted into a scale of 1–4 when they were recorded. If anyone typed in an illegal response, a noxious sound issued forth from the computer, and all the observers were asked to respond again. After everyone had responded, the test picture vanished, and there was a half-second, blank screen pause prior to the onset of the next study trial.

Each test trial began with the display of a test picture. Following a postonset delay of 800 msec, signaled by a beep, the observers were asked to provide a confidence rating of 1–4 that the test picture was one that they had seen during the previous study phase. The test picture remained visible until everyone had provided a valid response. The scale labels and the response keys were identical to those for the study confidence rating and, as at study, were perpetually in view. The procedures for dealing with illegal responses and for moving along to the next test trial were the same as those at study. Note that the main recognition measure was, therefore, retrospective confidence. Given the results of Busey et al. (2000), it may be assumed that retrospective confidence and any other reasonable measure of recognition accuracy—for example, hit rate or d'—measured the same thing; that is, they would be monotonic functions of one another. For clarity of discourse, we will henceforth use the term *recognition performance*, instead of *retrospective confidence*, to denote performance at test.

Randomization and counterbalancing. Prior to the start of each experiment, the 144 pictures for each of the two stimulus sets used in that experiment were randomly assigned to the 6 trays for the two subsessions. Counterbalancing was carried out independently for each of the total 12 trays and for each tray, as follows. Prior to the start of the experiment, the 24-tray pictures were randomly divided into two sets (A and B). Recall that each experiment involved 24 groups of observers. For the 12 odd groups, the A pictures were used as targets and the B pictures as distractors, and vice versa for the even groups. The 12 study slides were shown in the same order over the 12 groups that saw them as targets. The 12 conditions occurred in random order across the 12 study trials. This condition order was rotated over the 12 groups that saw it as a target, so that over the 12 groups, each target picture was shown once in each of the 12 conditions.

Results and Discussion

Because Experiments 3–6 were identical in structure, the results for each are provided in identical formats in Figures 8–11. We will describe Figure 8 in detail and then Figures 9–11 more briefly.

Experiment 3: Results. In Experiment 3, CG faces were compared with hooded faces. Our goal in Experiment 3 was to compare the contrast effects for two sets of conceptually similar naturalistic pictures that differed in recognition difficulty—much as, in Experiment 1, we compared two form recognition tests that differed in difficulty. Here, we had anticipated that hooded faces would be more difficult to recognize because the hooded faces contained considerably less distinguishing information than did the CG faces (see Figure 7, top two rows).

Figure 8 shows the results of Experiment 3. The six panels provide the following information. The top two panels show performance curves: mean prospective confidence (top left) and recognition performance (top right) as functions of exposure duration. There are four performance curves within each panel, representing the four combinations of two contrasts and two stimulus types. Note that upward-facing solid triangles represent high contrast, whereas downward-facing open triangles represent low contrast. Solid lines in this and other, analogous figures for Experiments 4–6 represent CG faces, whereas dashed lines represent the other stimulus type used in the experiment—here, hooded faces. As was anticipated, for both performance measures, high-contrast pictures are better than low-contrast pictures, and CG faces are better than hooded faces. Prospective confidence increases monotonically with exposure duration for both stimulus types. Recognition performance increases monotonically, except for low-contrast CG faces, the performance curve for which asymptotes after about 300 msec (an asymptote that was replicated in Experiments 4–6). Because of the shorter durations we used for the high-contrast conditions, we cannot tell whether recognition performance would similarly asymptote after about 300 msec. Nevertheless, the low-contrast asymptote that we observe, subtle though it is, is important for inferring the general effects of contrast on naturalistic pictures, and we shall have more to say about it in a later section.

The middle two panels of Figure 8, analogous to panel C of Figures 2 and 4, show state-trace plots that are designed to answer the question, Are the two stimulus types affected in the same way by contrast? Here, hooded face performance is plotted against CG face performance for prospective confidence (left middle panel) and recogni-

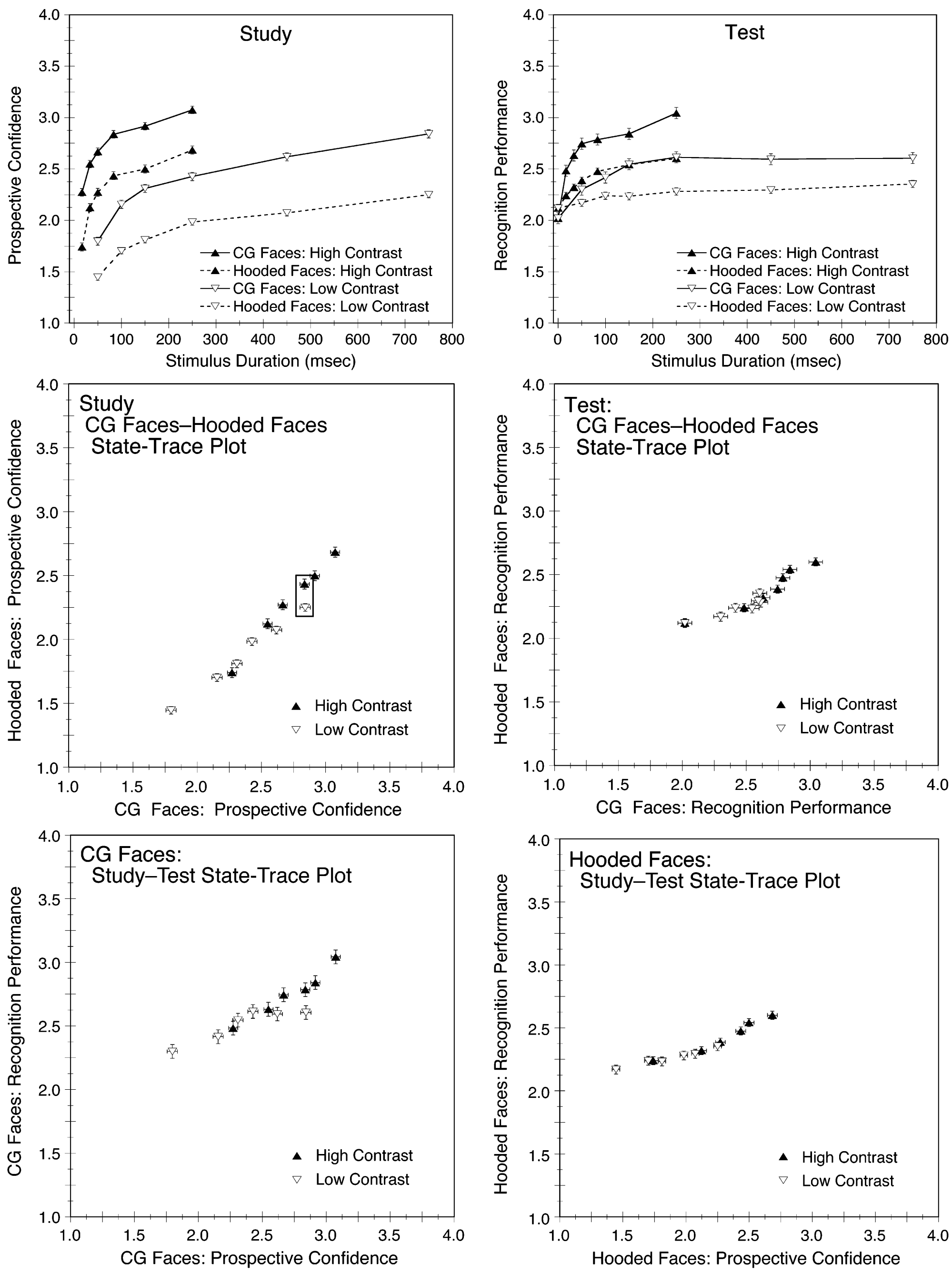

Figure 8. Experiment 3 data: computer-generated (CG) faces and hooded faces. Error bars represent standard errors.

tion performance (right middle panel). For prospective confidence, there is some separation of the low- and the high-contrast curves, which may be interpreted as follows. Consider two separate duration–contrast conditions that are equivalent for CG faces (i.e., vertically aligned, as exemplified by the vertical rectangle). For these two conditions, the high-contrast condition produces higher prospective confidence for the hooded faces, which indicates that at least for longer exposure durations, contrast affects hooded faces more than it affects CG faces. For recognition performance (right middle panel), however, there is no curve separation whatsoever: The function is essentially completely monotonic over both low- and high-contrast conditions. This replicates the results of Experiments 1 and 2 and is consistent with the proposition that the CG faces and the hooded faces, although quite different in appearance and physical structure and quite different in terms of recognition performance level, are, as measured by recognition performance, affected in the same way by stimulus contrast.

The bottom two panels of Figure 8 show state-trace plots designed to answer the question, Are the two types of memory measures, prospective confidence and recognition performance, affected in the same way by contrast? These panels show recognition performance plotted against prospective confidence for CG faces (bottom left panel) and for hooded faces (bottom right panel). For hooded faces, the function is perfectly monotonic: The two performance measures can be construed as being affected in the same way by contrast. For the CG faces, there is one clear departure from monotonicity: the longest duration low-contrast point. This departure reflects an already discussed property of the curves in the top panel: For prospective confidence, the low-contrast CG face curve increases monotonically with exposure duration, whereas for recognition performance, this same curve asymptotes after about 300 msec.

Experiment 3: Discussion. Experiment 3 replicates Experiments 1 and 2: As is shown in the right middle panel, the state-trace plot of the two stimulus types is monotonic, thereby confirming that contrast affects two quite different kinds of naturalistic faces in the same way.

Of our two performance measures, prospective confidence and recognition performance, only recognition performance was a genuine test, in the sense that performance depended entirely on information stored in memory. Prospective confidence, on the other hand, although depending in part on short-term information about the just-seen target picture, can also depend in part on metacognitive judgments about the relation between the nature of the stimulus and/or presentation conditions and the accuracy of future memory performance. In particular, there is evidence for two such metacognitive judgments in the Figure 8 data. The first is seen in the separation of the two contrast curves for CG faces and hooded faces in the left middle panel: Essentially, the observers gave more weight to high contrast for hooded faces than for CG faces. Here, it is likely that the observers had so little distinguishing information for each hooded face that stimulus contrast played a greater role than it did for CG faces, which were considerably more heterogeneous. The second apparent contribution of a metacognitive judgment is seen in the separation of the two contrast curves for the prospective confidence and recognition performance in the bottom left panel. Here, it is likely that the observers believed that increased duration always leads to better recognition performance, and this belief was reflected by their monotonically increasing prospective confidence functions for CG faces (top left panel). As it turns out, however, the observers were incorrect; their actual recognition performance did not increase beyond an exposure duration of 300 msec (upper right panel). That this was true for CG faces and, as we shall see, for houses, cityscapes, and celebrity faces as well, was unexpected for us, as well as for the observers, and the reasons for it are unknown.

Experiment 4: Results. In Experiment 4, CG faces were compared with houses. Our goal in Experiment 4 was to use two sets of conceptually dissimilar naturalistic pictures that were approximately the same in heterogeneity and, hence, in recognition difficulty (see Figure 7, rows 1 and 3).

Figure 9 shows the results of Experiment 4. The results are largely consistent with those in Experiments 1–3. As is shown in the middle right panel, contrast appears to affect the two stimulus types in the same way. As in Experiment 3, prospective confidence for CG faces increased monotonically with exposure duration, whereas recognition performance asymptoted at about 300 msec; this same effect occurred with houses (compare the upper two Figure 9 panels, with the corresponding systematic lacks of monotonicity in the bottom two panels).

Experiment 4: Discussion. Experiment 4 replicated Experiments 1–3, indicating that even with two conceptually different stimulus types—faces and houses—contrast affects recognition performance in the same way. Experiment 4 also demonstrates another stimulus type, homogenous houses, for which recognition-relevant memory performance does not appear to be acquired after about 300 msec of viewing time.

Experiment 5: Results. In Experiment 5, CG faces were compared with cityscapes. Our goal in Experiment 5 was to compare two sets of conceptually dissimilar naturalistic pictures, one of which (CG faces) was quite homogenous, whereas the other of which (cityscapes) was quite heterogeneous (see Figure 7, rows 1 and 4).

Figure 10 shows the results of Experiment 5. The results are largely consistent with those in Experiments 1–4. As is shown in the middle right panel, contrast appears to affect the two stimulus types in the same way. As in Experiments 3 and 4, prospective confidence for CG faces increased monotonically with exposure duration, whereas low-contrast recognition performance asymptoted at about 300 msec. In Experiment 5, there was also a recognition performance asymptote, albeit at a higher duration, for cityscapes.

Experiment 5: Discussion. Experiment 5 replicated Experiments 1–4, indicating that again with two conceptually different stimulus types—CG faces and cityscapes—contrast affects recognition performance in the same way. Experiment 5 also demonstrates another stimulus type, heterogeneous cityscapes, for which recognition-relevant memory performance does not appear to be acquired after some asymptotic viewing duration—in this case, around 500 msec.

Experiment 6: Results. Experiment 6 departed from Experiments 1–5 in that it used a stimulus set—celebrity faces—each member of which was encodable with a verbal label (e.g., at test, a photo of Jennifer Lopez could be recognized as having been seen at study either by the vi-

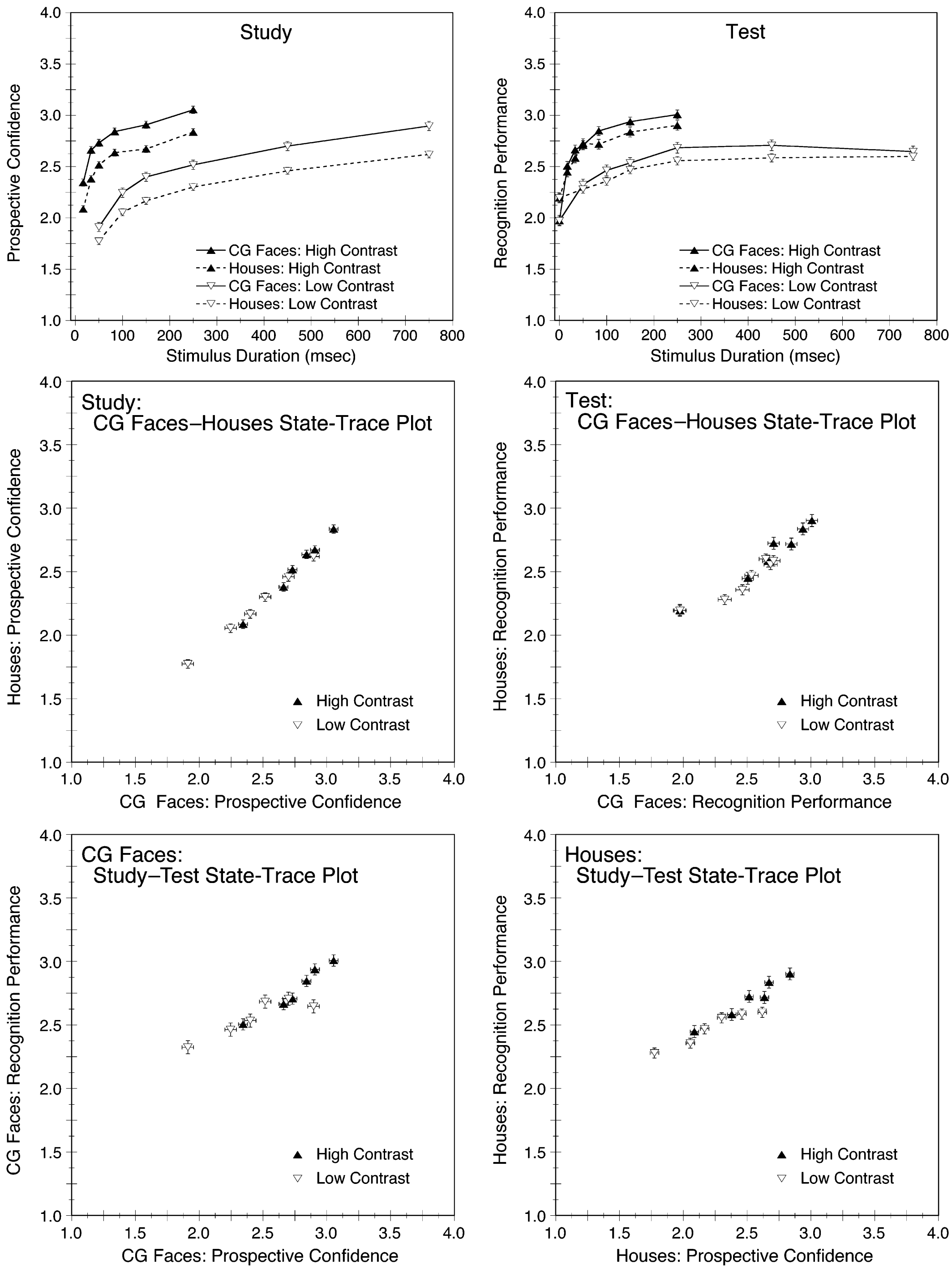

Figure 9. Experiment 4 data: computer-generated (CG) faces and houses. Error bars represent standard errors.

sual representation of her face or by the verbal label corresponding to her name). Because the potential encoding of celebrity faces is different in this respect from the potential encoding of CG faces, we anticipated that we might detect a difference in contrast effects on the two stimuli.

Nevertheless, we did not. Celebrity faces were, unsurprisingly, predicted by the observers to be more recognizable than CG faces, and equally unsurprisingly, they were. As is shown in Figure 11 top panels, both prospective confidence and recognition performance were considerably higher for celebrity faces than for CG faces.

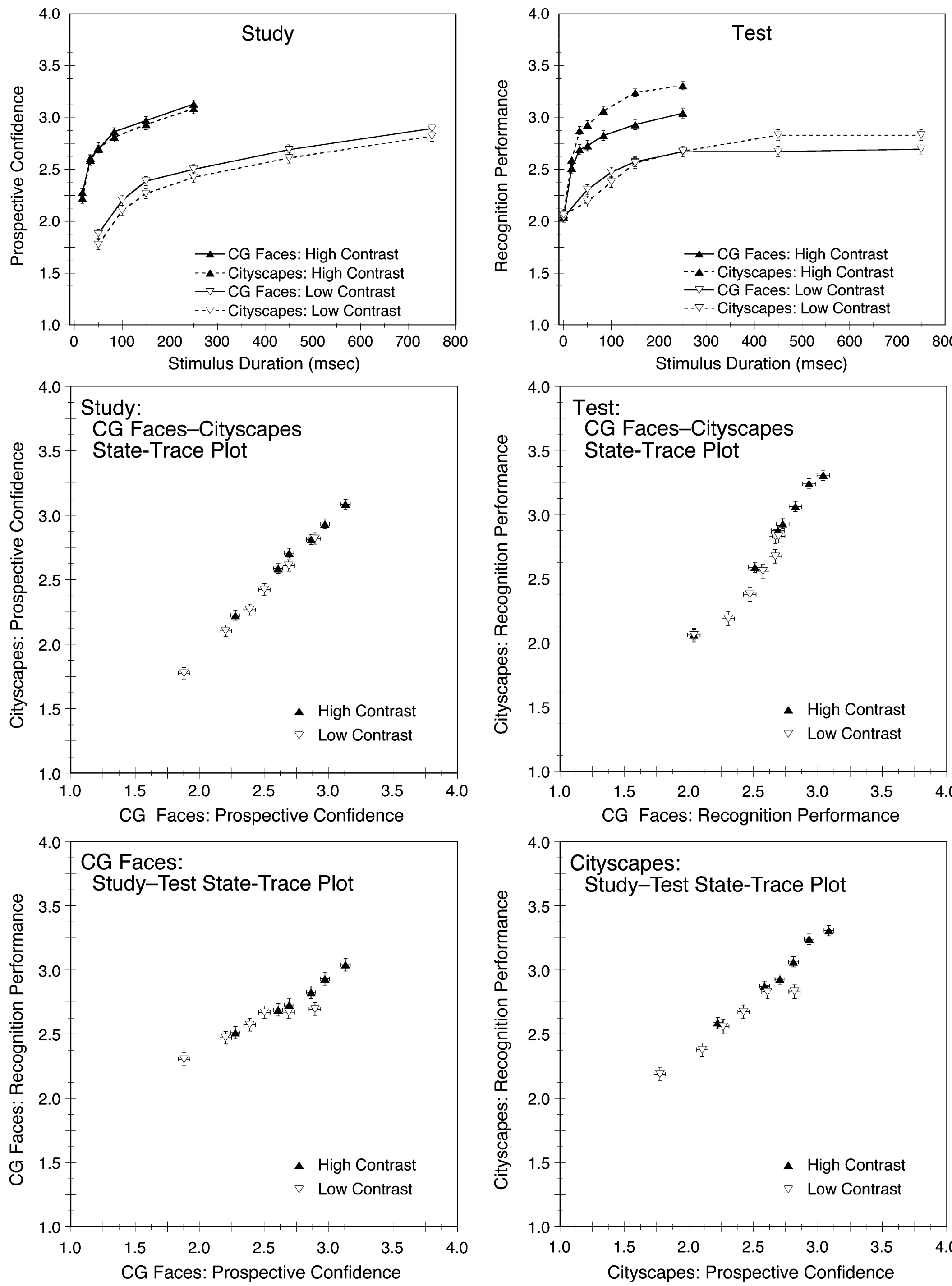

Figure 10. Experiment 5 data: computer-generated (CG) faces and cityscapes. Error bars represent standard errors.

Nevertheless, as is shown in the middle right panel, the contrast effect was identical for the two kinds of faces. Indeed, Experiment 6 showed very clean and regular data: With the exception of the single nonmonotonic point in each of the bottom two panels, characteristic of all of Experiments 3–6, the data indicate that (1) contrast affected CG faces and celebrity faces in the same way at both study and test (middle two panels) and (2) contrast affected study and test performance in the same way for both stimulus types (bottom two panels).

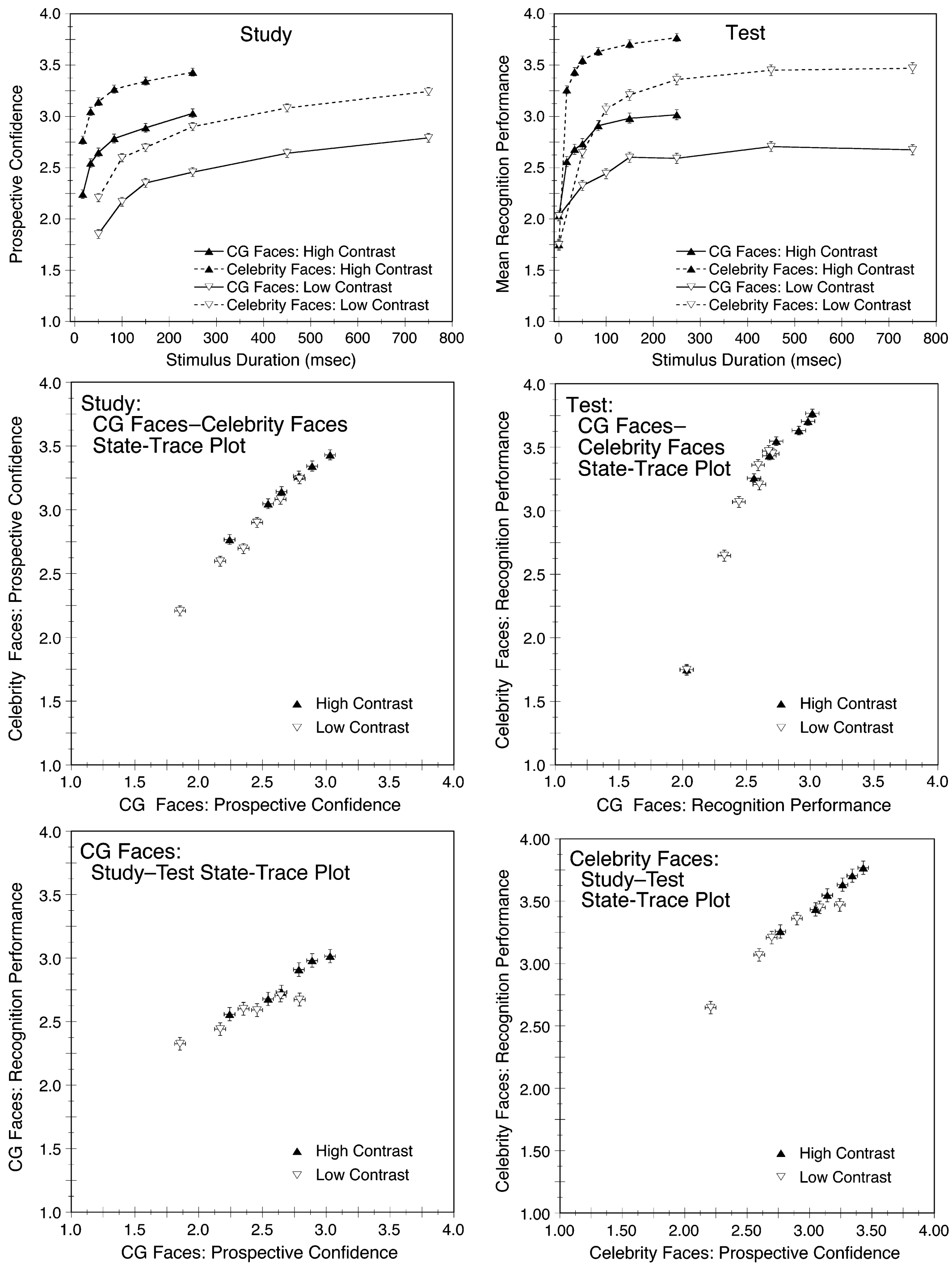

Figure 11. Experiment 6 data: computer-generated (CG) faces and celebrity faces. Error bars represent standard errors.

Experiment 6: Discussion

Experiment 6 replicated Experiments 1–5, indicating that with two kinds of faces differing considerably in their potential study strategies and, commensurately, their recognizability—CG faces and celebrity faces—contrast affects recognition performance in the same way. Experiment 6 also demonstrates another stimulus type, celebrity faces, for which recognition-relevant memory performance does not appear to be acquired after some asymptotic viewing duration—in this case, around 500 msec.

Experiments 3–6: Discussion

For each of Experiments 3–6, we addressed two interrelated questions about the effect of contrast on naturalistic scenes: Is the contrast effect the same for immediate prospective confidence ratings and delayed recognition performance, and is the contrast effect the same for prospective confidence ratings for CG faces and each of the four other stimulus types?

Conceptual differences between prospective confidence and recognition performance. It is worthwhile to reiterate that the two variables used in Experiments 3–6—prospective confidence and recognition performance—differ in an important respect. Immediately after study presentation, at the time a prospective confidence rating is provided, the observer knows the contrast level and the approximate duration of the just-seen picture. This means that prospective confidence can, in principle anyway, be influenced by metacognitive judgments, such as "I saw this for a long time, so I'll give it a high rating." Recognition performance, on the other hand, is a genuine memory test, which cannot depend directly on the nature of the original display conditions; if, for instance, an observer realized that a test picture was originally displayed at a long exposure duration, the observer would then know that the test picture must have been a target to begin with and could respond accordingly. In other words, recognition performance must depend only on information stored in memory about the test picture. In this sense, recognition performance is akin to the 2AFC form recognition test and the digit recall test of Experiments 1 and 2. The latter tests were also genuine tests of what was stored in memory, so that knowledge of display conditions could not influence performance.

Prospective confidence. Because prospective confidence is not a performance measure in the same sense as is recognition performance, we will not depend on it strongly for theory assessment. We do, however, have a number of observations about the prospective confidence data.

Each of Figures 8–11 provides information about prospective confidence in the left middle panel and in the two bottom panels. The left middle panel allows one to assess whether prospective confidence for the two stimulus types may be construed as issuing from a unidimensional theory—one in which contrast and duration combine into a single number that is passed to subsequent stages of the perceptual system and, thence, determines the magnitude of the prospective confidence. The two bottom panels allow one to assess whether the two performance measures, prospective confidence and recognition performance, may be construed as being based on the same unidimensional trace for the two stimulus types. In all cases, a unidimensional theory would imply monotonic functions.

In 8 of the 12 relevant panels over the four figures, there are departures from monotonicity. We have discussed them and have noted that they probably issue from metacognitive judgments of various sorts. Thus,

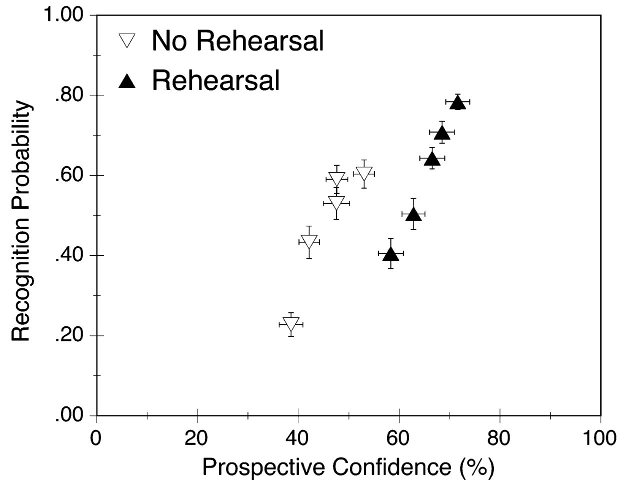

Figure 12. State-trace plot from Busey, Tunnicliff, Loftus, and Loftus (2000): recognition performance as a function of prospective confidence.

these nonmonotonicities inform us that prospective confidence may be construed as being based on more than one dimension—information acquired from the stimulus plus metacognitive judgments.

Although these occasional departures from monotonicity are undeniable, they are small. It is instructive to compare the prospective-confidence–recognition-performance scatterplots (Figures 8–11, bottom panels) with analogous ones reported by Busey et al. (2000, Experiment 1). This experiment, described earlier, used a face recognition procedure much like that in our Experiments 3–6, except that rehearsal was varied in place of stimulus contrast. Figure 12 shows prospective confidence as a function of recognition performance. It is obvious that the plot over all 10 data points is highly nonmonotonic. The nonmonotonicity is systematic: The rehearsal data points are displaced to the right from the no-rehearsal data points. That is, if one considers two duration × rehearsal/no-rehearsal points that are equal for recognition performance, the shorter rehearsal condition is given a higher prospective confidence rating. This finding implies that, when making a prospective confidence judgment, observers ascribe more importance to rehearsal than is warranted in terms of rehearsal's eventual effect on recognition performance. Clearly, more than one dimension is required to describe these data. Busey et al. postulated, as an example explanatory model, that prospective confidence was determined by two dimensions, which they termed "strength" and "certainty," whereas recognition performance was determined only by a single dimension, strength. In their treatment, strength was influenced by both duration and rehearsal, whereas certainty was influenced only by rehearsal.

For our data, the prospective-confidence–recognition-performance scatterplots are nothing like this; in comparison with the Figure 12 data, they are almost completely monotonic. Apparently, rehearsal strongly influences metacognitive judgments, whereas contrast may be construed as a more basic perceptual variable whose effect is confined almost exclusively to basic sensory processes.

Recognition performance: Assessment of nested theories. At the outset of this article, we described a set of four nested theories—multidimensional, unidimensional, multiplicative, and Bloch's law—that could potentially describe the effect of contrast. For Experiments 1 and 2, we clearly confirmed the unidimensional theory, found weak confirmation of multiplicative theory, and disconfirmed Bloch's law theory. To what degree can we assess these theories with respect to Experiments 3–6? To answer this question, we will confine ourselves to recognition performance, which, as just was described, can be construed as a "pure" test of memory, unsullied by possible metacognitive judgments.

Unidimensional theory. For each of Figures 8–11, we consider the right middle panels, which show the state-trace plots analogous to those in Figures 2C and 4C for Experiments 1 and 2. As with Experiments 1 and 2, in which simple stimuli were used, Experiments 3–6, with complex stimuli, confirm unidimensional theory: For Figures 2C and 4C and for Figures 8–11, middle right panels, the low-contrast and the high-contrast data points fall almost entirely atop one another.

Multiplicative theory. Recall that for Experiments 1 and 2, we assessed the degree to which shifting low-contrast performance curves to the left allowed them to align with the corresponding high-contrast performance curves. This enterprise yielded reasonable correspondence (see Figures 2D and 4D), thereby providing some support for multiplicative theory. Although we could, in principle, do the same thing for Experiments 3–6, there is another aspect of the data that allows us to unambiguously reject multiplicative theory, which is that the high-contrast and the low-contrast test performance curves almost certainly asymptote at different levels, as is indicated in Figures 8–11, top right panels.

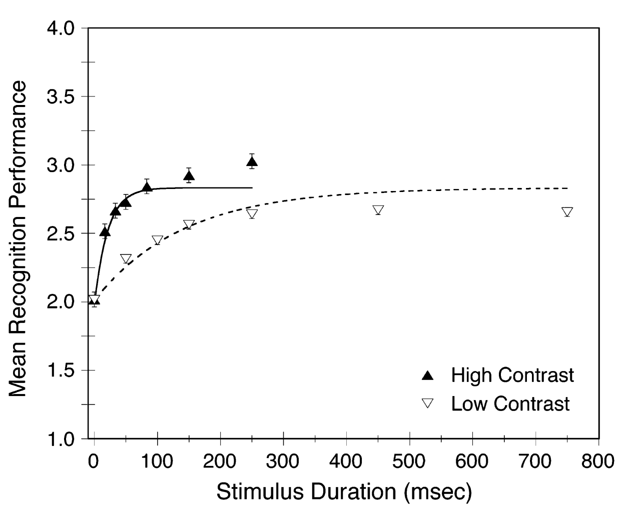

Figure 13. Sensory response, information acquisition (SRIA) theory fit to computer-generated test data (mean over Experiments 3–6). Best SRIA theory fit is shown by the solid line (for high-contrast conditions) and the dashed line (for low-contrast conditions).

That this finding disconfirms multiplicative theory is simple to demonstrate. To do so, we begin by averaging the CG faces data over Experiments 3–6. Note that such averaging is reasonable because CG faces were shown first and under identical circumstances over Experiments 3–6. These averaged data are shown in Figure 13 (the theoretical curves will be described below). Now consider any high-contrast condition for which recognition performance exceeds low-contrast asymptotic performance (e.g., the highest duration high-contrast condition). According to a multiplicative model, such performance would be

$$P_H = m[d_1 \times g(C_H)], \tag{17}$$

where d_1 and C_H are the longest high-contrast duration and the highest contrast, respectively, and m and g are monotonic functions. Now consider the low-contrast value C_L and define the ratio $r = g(C_H)/g(C_L)$; thus, $g(C_H) = g(C_L) \times r$. Accordingly,

$$d_1 \times g(C_H) = (d_1 \times r) \times g(C_L) = d_2 \times g(C_L), \tag{18}$$

where $d_2 = d_1 \times r$. Applying the function m to the left and right sides of Equation 18,

$$m[d_1 \times g(C_H)] = m[d_2 \times g(C_L)], \tag{19}$$

which, by Equation 17, are equal to P_H. In short, by a multiplicative theory, one can always find a sufficiently high duration, d_2, that, when combined with any low contrast, will produce performance equal to any shorter duration combined with any arbitrary high contrast. Different asymptotes imply that two such matching durations cannot be found, thereby disconfirming a multiplicative model.

In Experiments 1 and 2, we confirmed something close to a multiplicative theory, whereas in Experiments 3–6, we have soundly disconfirmed a multiplicative theory. Why is there this difference between the simple stimuli of Experiments 1 and 2 and the naturalistic stimuli of Experiments 3–6? Past data provide some clues. Loftus (1985c) confirmed multiplicative theory with heterogeneous naturalistic scenes that varied in contrast. Loftus et al. (1992) found much the same result, also with naturalistic scenes. However, these confirmations occurred with short durations (less than about 300 msec, the duration of an eye fixation); at longer exposure durations, the predictions of the multiplicative theory began to fail. The suggestion here is that for complex scenes, multiplicative theory holds only within the duration of an eye fixation. This makes sense. Both in the present and in past experiments, the first eye fixation was at a controlled part of the scene (the part indicated by the prestimulus fixation point). This first fixation was used both for acquiring information from the scene and for planning the next fixation's location. If low-contrast visual information provided a poorer basis for deciding where the next fixation should be than did high-contrast information, it would follow that subsequent fixations would be on systematically less informative areas of the picture in low-contrast, as compared with high-contrast, pictures. Therefore, in addition to information's being acquired more slowly in the low-contrast pictures, the information itself was less valuable. This is a form of failure of a multiplicative theory, a central tenet of which is that slower information acquisition is the *only* consequence of low contrast.

In Experiments 1 and 2, the maximum stimulus duration was 160 msec, so all stimulus presentations were within the duration of an eye fixation. Perhaps in Experiments 3–6, a multiplicative theory would describe the short-duration stimuli. Accordingly, we consider short durations only; that is, for the moment, we ignore the low-contrast 450- and 750-msec data points, leaving only data based on durations ranging from 50 to 250 msec for the low-contrast data and from 17 to 250 msec for the high-contrast data.

We evaluated a multiplicative model as we did for Experiments 1 and 2; that is, we investigated the degree to which the low-contrast curve, when scaled leftward, could be made to align with the high-contrast curve. We began by considering the CG face recognition performance data averaged over the four experiments and determined the factor that produced the best shift; this factor was 8.5. We then scaled all five recognition performance low-contrast curves—those for CG faces, hooded faces, houses, cityscapes, and celebrity faces—leftward by this same factor of 8.5. The results are shown in Figure 14. Figure 14 has multiple uses; the theory curves and the small box in the lower right of each panel involve SRIA theory fits and will be described below. For the moment, however, they should be ignored while we focus on the data points and make two conclusions. First, with the minor exception of cityscapes, the alignments are quite good for all five stimulus types. This is remarkable given that the same scaling factor was used for all of them. Second, when the alignments are slightly off, as they are for hooded faces and celebrity faces, the two sets of data points are horizontally parallel, so that a slight additional horizontal shift would bring them into alignment; that is, a multiplicative theory would still describe them, but with a shift factor slightly greater than 8.5. Thus, for short durations, the data for these five stimulus types are reasonably well described by a multiplicative theory.

Bloch's law theory. As was indicated, the factor by which the short-duration curves were shifted to bring them into alignment was 8.5 (this factor would be slightly more if the best alignment were sought for the hooded faces, cityscape, and celebrity faces data). The ratio of the two actual stimulus contrasts was $0.20/0.03 = 6.67$. Therefore, as with Experiments 1 and 2, the shift factor exceeds the contrast ratio, thereby disconfirming a Bloch's law account of the data.

The SRIA theory. As was discussed earlier, the SRIA theory is not generally a multiplicative theory, although a multiplicative assumption (indeed, a Bloch's law assumption) is central to it. Is it possible to adequately fit the SRIA theory to the full data from Experiments 3–6? We began with the CG data, which, averaged as it is over all four of Experiments 3–6, have considerable statistical power. To create a performance measure compatible with SRIA theory predictions, we converted the recognition performance confidence ratings, which, recall, were on a 1–4 scale, to proportions (both *hits* for the targets and *false alarms* for the distractors) using the equation H or FA $= (\text{Conf} - 1)/3$, where Conf is the confidence rating and H or FA is the resulting probability (a hit or a false alarm). We then corrected the hits for the false alarm rate, using $p = (\text{H} - \text{FA})/(1 - \text{FA})$, as in Loftus and McLean (1999). The data fit is shown in Figure 13, and the details of the fit are provided in Table 6, row 1.

For this data set, the SRIA theory clearly fails. The RMSE is 0.115, which, when compared with the standard error of 0.025, produces an $F(8,920)$ of 21.15. As with Experiments 1 and 2, this F value is not particularly meaningful, because of the very high experimental power; more meaningful is the fact that the r^2 between the data and the theory points is only .766. Because the estimated value of θ is 0, this version of the SRIA theory is, for reasons described earlier, a multiplicative theory (indeed a Bloch's law theory), and the reason for its failure is the same as that provided by Equations 18–20: According to this SRIA theory with $\theta = 0$, high-contrast and low-contrast performance curves must asymptote at the same level. It is evident from Figure 13 that this is the reason that the fit is poor.

By the arguments we made earlier, it is reasonable to believe that the SRIA theory may fit the short-duration data even if it fails for the complete data set. Accordingly, we fit it to the short-duration data for all five stimulus types. The details of the fits are provided in Table 7, rows 2–6, and the fits themselves are shown in Figure 14 in two ways: First, the theoretical predictions are represented by the solid (high contrast) and dashed (low contrast) lines through the data points. Second, the scatterplot relating theoretical predictions to the data points over the 10 short-duration conditions is shown at the bottom right of each panel, along with the associated data theory Pearson r^2.

It is apparent in three ways that the SRIA theory does not fit the data particularly well. First and second, except for hooded faces and houses, the F values for the fits are quite large, and/or the r^2 values are low. Third, for all the stimulus types, the small embedded scatterplots show systematic deviations from the 1.0 linear slope that represents perfect theory–data correspondence.

We considered one more possibility why the SRIA theory may not fit the data. As was discussed earlier, a definition of *contrast* is problematic with the present multiluminant stimuli composed of the full grayscale range, rather than of only two luminance values. Perhaps the definition we selected is incorrect. Accordingly, we

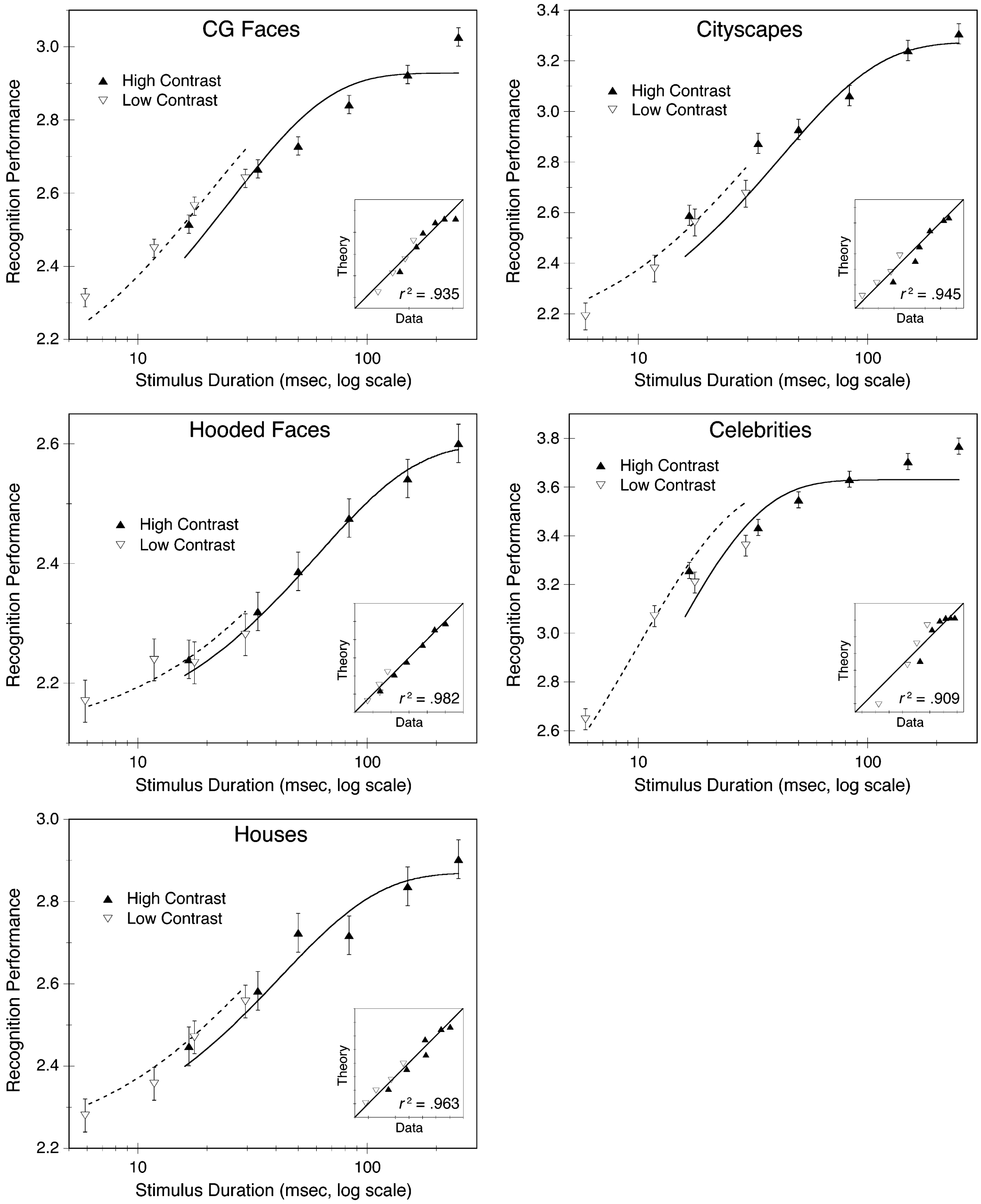

Figure 14. Theory evaluation for Experiments 3–6 short-duration recognition performance. Each panel shows low-contrast and high-contrast performance as a function of duration (log scale). For the multiplicative theory evaluation, each low-contrast curve has been scaled to the left by the same factor (8.5). For the sensory response, information acquisition (SRIA) theory evaluation, theory lines (dashed for low contrast, solid for high contrast) show the best fit, and the small inserted figures show the scatterplots relating SRIA theory predictions to data points over the 10 short-duration conditions (within the small scatterplots, the r^2 values are the data–theory correlations, and the diagonal lines indicate perfect theory–data correspondence).

refit the data from each of the five short-duration complex stimuli sets, with the addition of the high-contrast to low-contrast ratio as a free parameter. The resulting fits, expressed as data-theory r^2s are shown in Table 6, rightmost column. Although the fits improve (as they must with the additional free parameter), they are still not uniformly good.

Our conclusion is that these complex stimuli data cannot be fit by the SRIA theory with anything like the precision that the simple stimuli data from Experiments 1

Table 6
Stimulus Response, Information Acquisition Theory Parameter Fits to Experiments 3–6 Along With Fit Parameters

Condition	n	τ	θ	c	Y	RMSE	F	r^2	Adj. r^2
CG faces: all data	9	9.987	0.000	4.917	0.411	0.115	21.15	.766	–
CG faces: short durations	9	9.991	0.000	6.083	0.46	0.075	9.12	.935	.935
Hooded faces: short durations	9	9.226	0.004	14.272	0.255	0.025	0.57	.982	.990
Houses: short durations	9	9.998	0.000	9.998	0.374	0.049	1.25	.963	.974
Cityscapes: short durations	9	9.995	0.000	9.98	0.625	0.106	5.23	.945	.975
Celebrity faces: short durations	9	9.842	0.000	2.977	0.836	0.128	11.31	.909	.925

Note—The r^2 value represents the theory–data correlation with the low-contrast to high-contrast ratio fixed at its physical value (6.667). The "Adj. r^2" value is the theory–data correlation when the low-contrast to high-contrast ratio is a free parameter. RMSE, root-mean square error.

and 2 could be fit (see Figures 2 and 4). Despite this glass-half-empty view, however, there is also a glass-half-full view, according to which the theory is not totally without merit in its application to these data. This can be seen in two ways. First, as is indicated in Figure 14, the fits, although not perfect, are at least reasonable; in particular, the data-theory r^2s are all above .90. Second, as is indicated in Table 6, the parameter values are well behaved: The values of sensory parameters t and θ, which we would expect to remain relatively invariant over stimulus sets, are indeed extremely close; variation across the five stimulus sets is, instead, reflected in different values of c and the asymptote.

GENERAL DISCUSSION

We assessed the effects of stimulus contrast on perception and memory for two quite different classes of visual stimuli: simple biluminant stimuli (digits and random forms) and complex multiluminant stimuli (naturalistic photos of different sorts). Table 7 summarizes the degree to which the set of nested theories described in our introduction, along with the SRIA theory, can account for the data of both the simple stimuli (Experiments 1 and 2) and the complex stimuli (Experiments 3–6). For multiplicative theory and Bloch's law theory, Table 7 is vertically divided into data involving only durations shorter

Table 7
Summary of the Degree to Which Various Theories Are Supported

Theory	Simple Stimuli	Complex Stimuli
Unidimensional theory	*Strong support*. Monotonic state-trace plots are shown in Figures 2C and 4C.	*Strong support*. Monotonic state-trace plots are shown in Figures 8–11, right middle panels.
Multiplicative theory: shorter-than-an-eye fixation durations	*Weak support*. See Figures 2D and 4D. In each case, scaling the low-contrast data points to the left brings them into reasonable alignment with the high-contrast data points.	*Strong support*. See Figure 13. For each of the five stimulus types, scaling the low-contrast data points to the left by a common factor of 8.5 brings them into close-to-perfect alignment with the high-contrast data points.
Bloch's law theory	*Failure*. See Tables 1 and 3, columns 4 and 5. The mean best-shift factors exceed the mean physical contrast ratios: For Experiment 1, 2.604 versus 1.547; for Experiment 2, 2.349 versus 1.639.	*Failure*. The best-shift factor (8.5) exceeds the physical contrast ratio (6.667).
SRIA theory: shorter-than-an-eye fixation durations	*Strong support*. See Figures 2 and 4, panels A and B. Theory fits are extremely good, as is indicated by the data–theory correlations: .987 for Experiment 1 and .996 for Experiment 2.	*Weak support*. See Figure 13. Fits are poor; data–theory scatterplots are systematically nonlinear; data–theory correlations range from .909 to .982; fit quality decreases with increased data range. However, theoretical parameter values were well behaved across the five stimulus sets.
Multiplicative theory: longer-than-an-eye-fixation durations	*Technically unknown*. Longer durations were not used. However, performance was close to asymptote for the longest durations, so if longer durations had been used, the degree of support for multiplicative would likely not have been affected.	*Failure*. See Equations 18–20 and associated text. Multiplicative theory cannot account for different low-contrast and high-contrast asymptotes.
SRIA theory: longer-than-an-eye-fixation durations	*Technically unknown*. Longer durations were not used. However, performance was close to asymptote for the longest durations, so if longer durations had been used, the theory would likely have fit.	*Failure*. See Figure 12. The SRIA theory cannot account for different low-contrast and high-contrast asymptotes.

than 250 msec (the approximate duration of an eye fixation) versus durations longer than 260 msec that likely involve multiple eye fixations.

Table 7 is pretty much self explanatory and can be summarized as follows.

1. For all the data, unidimensional theory is unambiguously confirmed.

2. Multiplicative theory is weakly confirmed in the case of simple stimuli and strongly confirmed in the case of short-duration complex stimuli. We use the term *strongly confirmed* for multiplicative theory applied to short-duration complex stimuli, for two reasons. First as is shown in Figure 14, a single scaling factor ($k = 8.5$) brought the low-contrast and the high-contrast curves into almost perfect alignment for all five complex stimulus sets. Second, the small remaining discrepancies could have been essentially eliminated with slightly different scaling factors.

3. For all the data, Bloch's law theory is unambiguously disconfirmed.

4. The SRIA theory is strongly confirmed for simple stimuli. In the case of short-duration complex stimuli, the SRIA theory fits are marginal, although the parameter values are well behaved.

5. Both multiplicative theory and the SRIA theory are unambiguously disconfirmed for long-duration complex stimuli.

Implications of a Unidimensional Process

As was just noted, for both simple and complex stimuli, our results allow precise confirmation of a unidimensional theory: In all cases, the data were consistent with the proposition that stimulus duration and stimulus contrast combine at an early perceptual stage and that subsequent memory performance is then based on this combination; that is, neither duration nor contrast affects performance in isolation. The interpretation of these results is that contrast is a low-level variable that operates at a stage prior to that at which the system "knows" what stimulus is being analyzed or why such analysis is taking place. We note that this does not necessarily mean that the contrast level of some stimulus is not stored as part of the memory representation; rather, it simply means that it does not play a direct role in recognition performance.

Simple Versus Complex Stimuli

There were several differences between the simple and the complex stimuli both in terms of the observed pattern of results and in terms of the nature of the theories describing perception of them.

Asymptotic Performance

With simple stimuli, stimulus contrast did not appear to affect asymptotic performance level. There are three bases for this conclusion. First, an inspection of Figures 2 and 4 indicate that, with the longest exposure durations, performance in both contrast conditions is close to what appear to be equal asymptotes. Second, the digit recall task was such that the low-contrast, as well as the high-contrast, stimuli were always completely perceivable given enough time and, therefore, any less-than-1.0 asymptote results from careless or motor errors on the observer's part. On an a priori basis, there is no reason to expect that such errors would be influenced by stimulus contrast. Third, the SRIA theory assumes a common asymptote for low-contrast and high-contrast stimuli; thus, the almost-perfect confirmation of the theory for Experiments 1 and 2 lends credence to this assumption.

With complex stimuli, asymptotic performance is strongly dependent on stimulus contrast. The basis for this conclusion comes from an inspection of the upper right panels of Figures 8–11. For all five stimulus sets, low-contrast recognition performance asymptoted or almost asymptoted after a duration of approximately 250 msec, at a level that was considerably less than that of the highest high-contrast performance.

Fit of the SRIA Theory

There was another clear and consistent difference between the simple stimuli and the short-duration complex stimuli: The SRIA theory fit data issuing from simple stimuli—random forms and digits—essentially perfectly, and as was indicated earlier, it has done so consistently over a wide variety of circumstances (Busey & Loftus, 1994, 1998; Loftus et al., 1993; Loftus & Ruthruff, 1994). Why, then, did the SRIA theory fit the short-duration complex stimuli data so much less well. We will consider and reject three simple possibilities.

A poor definition of "contrast." First, the simple stimuli both in Experiment 1 and 2 and in all the earlier referenced studies contained only two luminance levels, and the definition of *contrast* was, therefore, quite simple and presumably valid. The definition of *contrast* was considerably less straightforward for our multiluminant pictures, whose luminance levels spanned the full 0–255 range. However, even if our definition of *contrast* was not correct, and the high-contrast to low-contrast ratio is allowed to vary as a free parameter, the fit of the SRIA theory to the short-duration complex-stimuli data is not greatly improved (compare Table 6, two rightmost columns). So, a poor definition of contrast is not a sufficient explanation.

An inappropriate dependent variable. Second, perhaps we used an inappropriate dependent variable. Recall that the SRIA theory produces a value of *information* that then must be mapped to whatever dependent variable is measured in an experiment. With digit recall, it is reasonable to assume simply that there is an identity match: If X% of the information is acquired from the stimulus display, then X% of the digits are reported. However for the random-form data of Experiments 1 and 2, a simple corrected-for-guessing hit rate was used as the dependent variable, and the fit was still essentially perfect.

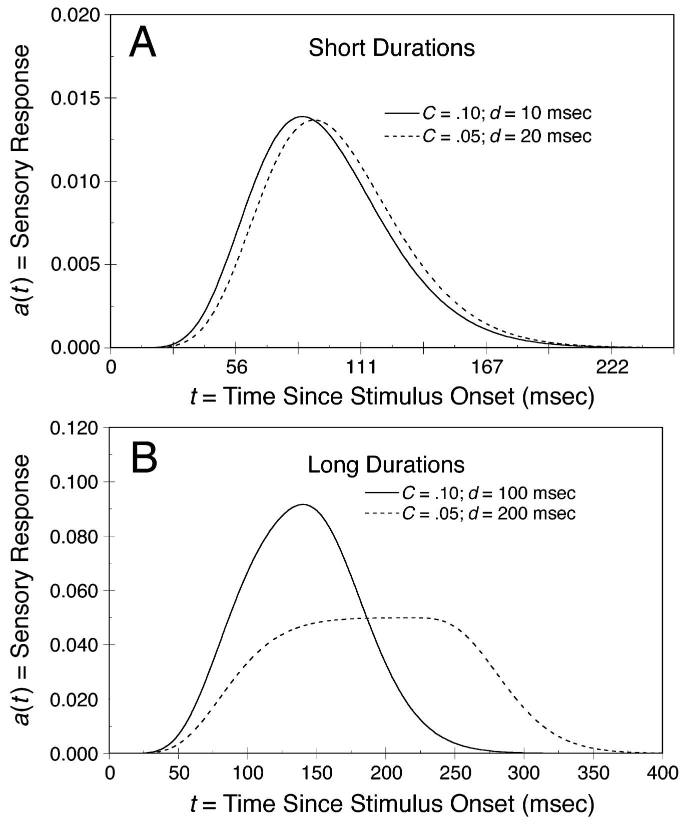

Figure 15. Two pairs of equal-product stimuli. Panel A shows short-duration stimuli that may be considered literal metamers; they are (virtually) identical in all respects. The two stimuli in panel B have identical areas but are not identical in any other respects. Nevertheless, they may be considered metamers with respect to any measure (such as memory performance in the sensory response, information acquisition theory) that depends only on the area under the function.

Simple stimuli versus pictures. Third, perhaps encoding and remembering pictures is sufficiently different from encoding and remembering simple stimuli that the same theory cannot be used to describe both. However, Loftus and McLean (1999) fit the SRIA theory simultaneously to a set of four picture recognition data sets issuing from four experiments whose designs were very different. Loftus and McLean also assumed *information* from the theory to equal hit rate corrected for false alarm rate, using the same formula that we used for fitting the recognition performance data here. Similarly, Olds and Engel (1998) confirmed the SRIA theory by using corrected-for-guessing identification of black-and-white simple objects. This provides evidence that picture recognition data per se are not necessarily outside the theory's domain and provides additional evidence that a corrected-for-guessing hit rate is not entirely inappropriate as a dependent variable, as far as the theory is concerned.

Biluminant versus multiluminant stimuli. In short, the poor fit of the SRIA theory cannot be due to a poor definition of contrast; it is unlikely that it is due to our choice of dependent variable, and it is unlikely that it occurs because we are using picture recognition. By a process of elimination, it would appear that the SRIA theory is incapable of describing data based on multiluminant stimuli such as the ones we used in Experiments 3–6. This makes sense. The SRIA theory makes the fundamental assumption that, associated with a stimulus is a single value of contrast that, in conjunction with duration, dictates the form of the single sensory response function. With multiluminant stimuli, there is no single value of contrast. Instead, there are multiple edges at varying contrast levels, which define features that have varying degrees of importance and relevance and idiosyncratic roles in encoding the picture for later recognition. Moreover, as contrast is reduced, some features that are low contrast to begin with fall below threshold, whereas other, higher contrast features do not. It is possible that by (1) analyzing each individual picture with respect to its various internal contrast levels, (2) applying the SRIA theory simultaneously to these various levels, (3) assessing the roles of the features defined by the various edges defined by these varying contrast levels, and (4) using the results to generate a measure of *infor-*

mation that can be sensibly related to the dependent variable, the SRIA theory could be successfully applied to multiluminant stimuli. However, that particular daunting enterprise is beyond the scope of this article.

Bloch's Law and Informational Metamers

Earlier, we discussed Bloch's law as the simplest way of describing the effect of contrast on memory performance and showed how, at the core of the SRIA theory beats a Bloch's law heart. Here, we will make some final remarks about the implications of a Bloch's law theory and its place within a more general theory of perception and information acquisition.

Bloch's law, as typically treated, incorporates the notion of a *critical duration*, termed d_c. This duration, typically estimated at around 100 msec, is the upper limit wherein Bloch's law holds for detection; that is the duration at which the product of duration and luminance strictly determines contrast. At durations higher than d_c, Bloch's law begins to break down; that is, high-luminance short stimuli are detected better than equal-product lower luminance longer stimuli.

Although Bloch's law has typically been applied to the relation between duration and stimulus luminance, it has also been applied to the relation between duration and stimulus *contrast*. Although many of the same principles hold, the critical duration has been generally found to be lower—approximately 20–50 msec rather than 100 msec—with contrast than with luminance (Gorea & Tyler, 1986; Musselwhite & Jeffreys, 1982; Spekreijse, Van der Tweel, & Zuidema, 1973).

What causes the Bloch's law breakdown? In Figure 15, we show sensory response functions, generated by the SRIA theory for two pairs of what we term *equal product* stimuli: For each pair, the product of duration and contrast is the same (duration × contrast = 1.00 in panel A and 10.00 in panel B). It is apparent that in panel A, the two physically different stimuli generate (virtually) identical sensory response functions. Therefore, any dependent measure that depends on the sensory response function, be it detection, identification, memory, or anything else, must *in principle* be (virtually) identical for the two stimuli. In particular, Bloch's law would hold. The two stimuli in panel B, however, have very different shapes. Therefore, it is in principle possible for a dependent measure to be different for the two stimuli. For example, if detection depended on the peak value of the curve, detection would be better for the short bright stimulus than for the dimmer longer one—just as is found in Bloch's law studies.

Borrowing from color vision, we may term the two stimuli in panel A *metamers*. Like classical color metamers, the information distinguishing the two stimuli is lost at the beginning of the sensory-perceptual-cognitive system, and no way of tapping the mental representation of them could reconstruct which was which. The panel B stimuli are not metamers in this sense. Nevertheless, any dependent measure that depends only on the area under the curve—such as memory performance in the SRIA theory—will be identical for these two stimuli. Kahneman and Norman (1964) directly compared the critical duration for Bloch's law (using luminance). They found that d_c was approximately 100 msec for a brightness-matching task, which might reasonably be determined by the height of the sensory response function, whereas it was considerably longer, up to 347 msec, for a stimulus identification task that, like performance in the present experiments, might reasonably be determined by the area under the curve.

It is, therefore, appropriate to term these kinds of stimulus pairs *informational metamers*: They lead to different representations in early stages of the system (where the sensory response functions reside) but to the same representation at later parts of the system (where *information* of whatever is the determinant of memory performance resides). Further discussion of such informational metamers is beyond the scope of this article. We would like to close by saying, however, that just as an understanding of color metamers enabled a quantum leap in developing a coherent color vision theory, incorporating the notion of informational metamers into theories of perception and memory could greatly assist us in theoretical development within these disciplines.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 90-197). New York: Academic Press.

Bamber, D. (1979). State trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, **19**, 137-181.

Bogartz, R. S. (1976). On the meaning of statistical interactions. *Journal of Experimental Child Psychology*, **22**, 178-183.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, **101**, 446-469.

Busey, T. A., & Loftus, G. R. (1998). Binocular information acquisition and visual memory. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1188-1214.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, **7**, 26-48.

Craik, F. I. M., & Lockhart, R. F. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Ginsburg, A. P., Cannon, M. W., & Nelson, M. A. (1980). Suprathreshold processing of complex visual stimuli: Evidence for linearity in contrast perception. *Science*, **208**, 619-621.

Gorea, A., & Tyler, C. W. (1986). New look at Bloch's law for contrast. *Journal of the Optical Society of America A*, **3**, 52-61.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, **16**, 96-101.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 507-513.

Kahneman, D., & Norman, J. (1964). The time–intensity relation in visual perception as a function of observer's task. *Journal of Experimental Psychology*, **68**, 215-220.

Kaswan, J., & Young, S. (1963). Stimulus exposure time, brightness, and spatial factors as determinants of visual perception. *Journal of Experimental Psychology*, **65**, 113-123.

Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology*, **55**, 477-483.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, **6**, 312-319.

Loftus, G. R. (1985a). Consistency and confoundings: Reply to Slamecka. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 817-820.

Loftus, G. R. (1985b). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 396-405.

Loftus, G. R. (1985c). Picture perception: Effects of luminance level on available information and information extraction rate. *Journal of Experimental Psychology: General*, **114**, 342-356.

Loftus, G. R., & Bamber, D. (1990). Weak models, strong models, unidimensional models, and psychological time. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 916-926.

Loftus, G. R., Busey, T. A., & Senders, J. W. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics*, **54**, 535-554.

Loftus, G. R., & Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology*, **35**, 135-199.

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation durations, perceptual processing, and long-term visual memory. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 203-226). New York: Springer-Verlag.

Loftus, G. R., & McLean, J. E. (1999). A front end to a theory of picture recognition. *Psychonomic Bulletin & Review*, **6**, 394-411.

Loftus, G. R., & Ruthruff, E. R. (1994). A theory of visual information acquisition and visual memory with special application to intensity–duration tradeoffs. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 33-50.

Massaro, D., & Loftus, G. R. (1996). Sensory storage: Icons and echoes. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition* (Vol. 10, pp. 68-101). New York: Academic Press.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, **76**, 165-178.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial order information. *Psychological Review*, **100**, 183-203.

Musselwhite, M. J., & Jeffreys, D. A. (1982). Pattern-evoked potentials and Bloch's law. *Vision Research*, **22**, 897-903.

Nairne, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 248-255.

Norman, D. A. (1966). Acquisition and retention in short term memory. *Journal of Experimental Psychology*, **72**, 369-381.

Olds, E. S., & Engel, S. A. (1998). Linearity across spatial frequency in object recognition. *Vision Research*, **38**, 2109-2118.

Olzak, L. A., & Thomas, J. P. (1986). Seeing spatial patterns. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 7.1-7.56). New York: Wiley.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, **76**, 241-263.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Palmer, J. C. (1986a). Mechanisms of displacement discrimination with and without perceived movement. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 411-421.

Palmer, J. C. (1986b). Mechanisms of displacement discrimination with a visual reference. *Vision Research*, **26**, 1939-1947.

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America*, **7**, 2032-2040.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, **7**, 191-218.

Schacter, D. L., & Tulving, E. (Eds.) (1994). *Memory systems*. Cambridge, MA: MIT Press.

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 591-600.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

Spekreijse, H., van der Tweel, L. H., & Zuidema, T. (1973). Contrast evoked responses in man. *Vision Research*, **13**, 1577-1601.

Sperling, G. (1986). A signal-to-noise theory of the effects of luminance on picture memory: Comment on Loftus. *Journal of Experimental Psychology: General*, **115**, 189-192.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, **76**, 559-573.

van Nes, F. L., & Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America*, **57**, 401-406.

Wandell, B. A. (1995). *Foundations of vision.* Sunderland, MA: Sinauer Associates.

Watson, A. B. (1986). Temporal sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. I, pp. 6.1-6.43). New York: Wiley.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, **9**, 418-455.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, **2**, 775-780.

APPENDIX

Here, we will provide a short tutorial on the logic and methodology of state-trace analysis. For an excellent, detailed, and readable treatment, the reader is referred to Bamber (1979). Our goals here are to provide specific examples of both a unidimensional theory and a multidimensional theory and to show how a single-dimension theory is tested using a data set similar to those in Experiments 1–6. We provide explicit equations corresponding to the two theories so the interested reader can follow along numerically with the help of Excel, MATLAB, Mathematica, S+, or any other application capable of numerical manipulation. We hasten to point out that these equations are not meant to be taken seriously as actual models of psychological events; rather, they were chosen mainly for expositional simplicity.

An Example Experiment

We base this tutorial on an experiment reported by Busey et al. (2000, Experiment 1) concerning the relation between confidence and accuracy in face recognition. Methodologically, their experiment was very much like our Experiments 3–6. A series of target faces were shown in a study phase. In a subse-

APPENDIX (Continued)

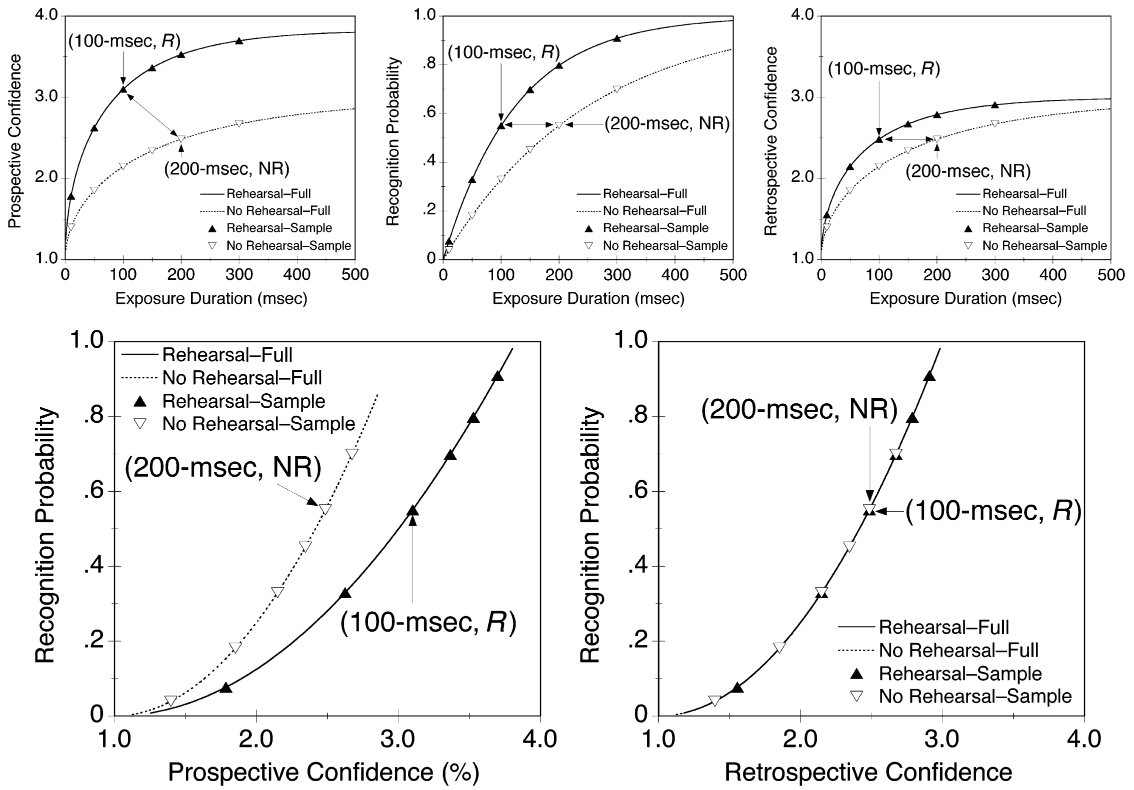

Figure A1. Hypothetical data from the exposure duration × rehearsal/no-rehearsal experiment. Top panels: main data in the form of three dependent variables as functions of exposure duration. Bottom panels: state-trace plots in the form of recognition performance plotted, over conditions, as a function of prospective confidence (bottom left) and retrospective confidence (bottom right). The bottom right state-trace plot depicts a confirmation of single-dimension theory, whereas the bottom left state-trace plot depicts a disconfirmation of single-dimension theory.

quent test phase, the targets were randomly intermingled with distractors, and the resulting set of test pictures was shown one by one in an old–new recognition procedure. At study, two independent variables were manipulated: exposure duration and rehearsal. By the rehearsal manipulation, observers either were required to visually rehearse or were prevented from rehearsing each just-seen target picture for 15 sec following target offset.

There were three performance measures: *prospective confidence*, for which, following each study trial, the observer provided a rating of 1–4 as to whether he or she would subsequently recognize the face; *recognition*, for which, on each test trial, the observer provided an old–new recognition response; and *retrospective confidence*, for which, following each test trial, the observer provided a rating of 1–4 as to whether he or she believed the old–new recognition response to have been correct.

In what follows, we will provide idealized versions of the experiment and the results. We construct two hypothetical scenarios. In the first, or *full*, scenario, there is an extremely large number of exposure durations: Duration ranges from 1 to 500 msec in 1-msec intervals. Thus, it is possible to map out what is essentially the continuous underlying curve relating the performance measures to exposure duration. The second, or *sample*, scenario constitutes a more realistic approximation to this or any other experiment, in which six exposure durations (10, 50, 100, 150, 200, and 300 msec) are used.

The top panels of Figure A1 show the hypothetical results of both scenarios: The three dependent variables are plotted as functions of duration, with separate curves for rehearsed and nonrehearsed conditions. The full-scenario data (with 500 exposure duration conditions) are depicted as smooth curves, whereas the sample scenario data (with 6 exposure duration conditions) are depicted by curve symbols. Qualitatively, there is nothing surprising about the data. All three performance measures behave similarly: They rise with exposure duration and are better for rehearsed than for nonrehearsed conditions.

The major question was, Can confidence (both prospective and retrospective) and recognition performance be construed as measuring the same underlying internal state; that is, can a single unidimensional measure be construed as determining all three measures? The answer obtained by Busey et al. (2000) was that retrospective confidence and recognition performance could be so construed but prospective confidence and recognition performance could not: The relation between retrospective confidence and recognition performance could be described with a single-dimension theory, whereas the relation between prospective confidence and recognition performance required

APPENDIX (Continued)

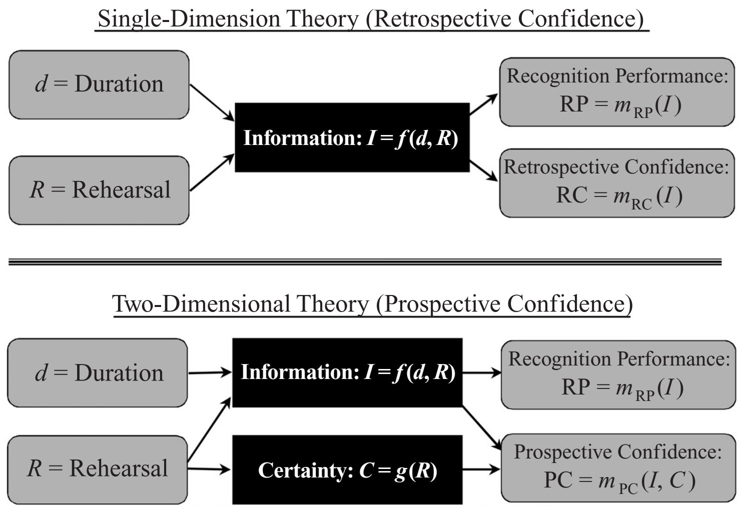

Figure A2. Schematic representations of a single-dimension theory and a two-dimensional theory.

a two-dimensional theory. We now describe these two theories and their predictions.

Retrospective Confidence: Confirmation of a Single-Dimension Theory

We begin with retrospective confidence. The relation between retrospective confidence and recognition performance is adequately described by a single-dimension theory, which is shown in the top panel of Figure A2. Here, the left and right gray boxes represent observables (duration, *d*, and rehearsal, *R*, are on the left, whereas recognition performance, RP, and retrospective confidence, RC, are on the right). The "black box" in the center depicts the unobservable manner in which the independent variables combine and then determine performance. Here, *d* and *R* are combined into a unidimensional construct, information [*I*], by a function, *f*, which is monotonic in both arguments. Both recognition performance and retrospective confidence are then monotonic functions, m_{RP} and m_{RC} of *I*. The specific equation *f* for combining rehearsal and duration to obtain information is

$$I = f(d, R) = \begin{cases} 2d & \text{for rehearsal} \\ d & \text{for no rehearsal}, \end{cases} \tag{A1}$$

where *d* is exposure duration in milliseconds. The monotonic equations *m* for obtaining recognition performance and retrospective confidence from information are

$$\text{RP} = m_{\text{RP}}(I) = 1 - e^{-I/250} \tag{A2}$$

and

$$\text{RC} = m_{\text{RC}}(I) = 2\sqrt{1 - e^{-I/250}} + 1 . \tag{A3}$$

An idealized prediction: The *full* scenario. To understand the prediction of the unidimensional theory, it is useful to consider the *full* scenario. Consider recognition performance—the middle curve at the top of Figure A1. Choose any point on the rehearsal curve—say, the 100-msec point. We call it the (100-msec, *R*) point. Recognition performance at the (100-msec, *R*) point is approximately 0.551. Now we ask, How much exposure duration is required to achieve the same performance level in the no-rehearsal condition? The answer is 200 msec. We call this the (200-msec, NR) condition. Both of these points, along with a depiction of their equality, are indicated on the upper middle panel in Figure A1. Because the (100-msec, *R*) and the (200-msec, NR) conditions produce the same recognition performance, 0.551, they must, by the unidimensional theory, have produced the same information, $I = m_{\text{RP}}^{-1}(0.551) = 200$ (where m_{RP}^{-1} is the inverse of m_{RP}). Now—and this is the crux of the single-dimension theory's prediction—*because the (100-msec, R) and the (200-msec, NR) conditions produce equal information, I, they must also produce equal values of retrospective confidence.* And indeed, the value of retrospective confidence for both these conditions is 2.484, as is indicated in the upper right panel in Figure A1.

This finding—that two duration × rehearsal/no rehearsal conditions producing equal values of recognition performance must also produce equal values of retrospective confidence—is encapsulated in the *state-trace plot* shown at the bottom right of Figure A1. In general, a state-trace plot is a scatterplot of one dependent variable against a second dependent variable, with one point for each experimental condition. There are actually two curves in this plot: one relating recognition performance to retrospective confidence for the rehearsal conditions and the other relating recognition performance to retrospective confidence for the no-rehearsal conditions. However, there appears to be only one curve, because the two curves are exactly superimposed, for reasons that we now will explain.

Note that any two conditions—a shorter duration rehearsal condition and a longer duration no-rehearsal condition—that produce the same value of retrospective confidence must be *vertically aligned*. Likewise, any two such conditions that produce the same value of recognition performance must be *horizontally aligned*. If, as is implied by the unidimensional theory,

APPENDIX (Continued)

any two conditions producing equal recognition performance *also* produce equal retrospective confidence, these points must be *both* vertically and horizontally aligned; that is, they must coincide. That they do is shown in the bottom right panel in Figure A1 for our two example conditions. Because the prediction holds over the entire range of conditions, the two curves must entirely overlap, as they are depicted to do.

A real-life prediction: The *sample* scenario. This, then, is the prediction of the single-dimension theory: The two curves in the state-trace plot must entirely overlap. It is easy to evaluate this prediction with the hypothetical *full* experiment in which there is essentially a continuous range of exposure durations. In a more realistic experiment, exemplified by our 6 *sample* points, evaluating the prediction is not quite so simple. For the single-dimension theory to be confirmed, it is necessary that all data points (12 in this example: 6 rehearsal data points, and 6 no-rehearsal data points) jointly form a monotonic function. Needless to say, however, it is possible for sampled data points to form a monotonic function even if they issue from two underlying nonoverlapping functions. Thus, even if one observes that the sampled points form a monotonic function, one must be cautious in concluding that a single-dimension theory has produced them.

The experimental design of the present experiments is much the same as that in this example, where instead of combining two values of rehearsal with exposure duration, we combine two levels of contrast with exposure duration. The logic is the same, and the general prediction is the following: If all points on the state-trace plot—the low-contrast points and the high-contrast points—form a single monotonic function, the prediction of the single-dimension theory is confirmed.

Prospective Confidence: Disconfirmation of a Single-Dimension Theory

As is shown in the top panels of Figure A1, prospective confidence appears superficially to behave similarly to both recognition performance and retrospective confidence. As we have demonstrated, recognition performance and retrospective confidence are both functions solely of a single underlying variable, information, which led to the coinciding state-trace curves shown in the bottom right panel in Figure A1.

The state-trace plot relating recognition performance to prospective confidence, however, was not monotonic: As is shown in the bottom left of Figure A1, the rehearsal curve is displaced to the right of the no-rehearsal curve. Thus, if two conditions, a shorter duration rehearsal condition and a longer duration no-rehearsal condition, are equal for recognition performance, the shorter duration rehearsal condition produces a larger value of prospective confidence. This is shown with our two example conditions, (100-msec, R) and (200-msec, NR), in both the top left and the bottom left panels in Figure A1. This data pattern has a straightforward interpretation: At the time of study, observers overestimate the value of rehearsal in terms of how much it boosts eventual recognition performance.

Thus, the nonmonotonic state-trace plot disconfirms a single-dimension theory: More than one dimension is required to account for the data. The bottom panel of Figure A2 provides an illustrative two-dimensional theory. In our incarnation of it, information is computed by Equation A1, and recognition performance is computed by Equation A2. However, there is second dimension, called *certainty*, which is affected by rehearsal duration, but not by exposure duration. In particular,

$$C = g(R) = \begin{cases} 2.0 \text{ for rehearsal} \\ 1.0 \text{ for no rehearsal}. \end{cases} \tag{A4}$$

Although recognition performance is determined solely by information, prospective confidence is jointly determined by information and certainty. In particular,

$$\mathrm{PC} = m_{\mathrm{PC}}(I,C) = 2C\sqrt{1-e^{-I/250}} + 1. \tag{A5}$$

Thus, when two conditions, for example the (100-msec, R) and the (200-msec, NR) conditions, produce equal recognition performance, they are still inferred to have produced the same value of information. However, when two conditions produce the same value of information, the rehearsal condition produces a higher value of prospective confidence than does the no-rehearsal condition by virtue of the higher certainty contribution to prospective confidence in the rehearsal than in the no-rehearsal condition.

Summary

The important points of this tutorial are the following.

1. A single-dimension theory stipulates that two or more dependent variables in an experiment are determined solely by a single, internal, single-dimensional variable whose value is, in turn, determined by the values of the independent variables.
2. A state-trace plot is a scatterplot of one dependent variable plotted against another dependent variable, with each point corresponding to a single experimental condition.
3. The prediction of a single-dimension theory is that all the points of the state-trace plot form a monotonic function, as in the bottom right panel in Figure A1. If the points of a state-trace plot are observed to form a nonmonotonic function, as in the bottom left panel in Figure A1, a single-dimension theory is insufficient to describe the data.
4. If a single-dimension theory is disconfirmed, the disconfirming data pattern typically suggests the form of multidimensional theory that would be appropriate for describing the data. This is exemplified in the two-dimensional information certainty model that we have described.
5. Finally, it is worth noting that the predictions of a unidimensional theory are not affected by scaling considerations (as has been described, e.g., by Bogartz, 1976; Loftus, 1978, 1985a, 1985b; Loftus & Bamber, 1990), which, for example, imply that a conclusion about the presence or absence of an interaction with respect to one dependent variable may not be valid for a second dependent variable that is nonlinearly related to the first. That is, the predictions described for a single-dimension theory apply not just to the dependent variables measured in the experiment, but to all dependent variables that are monotonically related to one another. For example, a prediction that is confirmed or disconfirmed with respect to probability correct would ipso facto be confirmed or disconfirmed with respect to d'.

(Manuscript received November 15, 2002;
revision accepted for publication May 12, 2003.)