

An Integrated Theory of Attention and Decision Making in Visual Signal Detection

Philip L. Smith
The University of Melbourne

Roger Ratcliff
The Ohio State University

The simplest attentional task, detecting a cued stimulus in an otherwise empty visual field, produces complex patterns of performance. Attentional cues interact with backward masks and with spatial uncertainty, and there is a dissociation in the effects of these variables on accuracy and on response time. A computational theory of performance in this task is described. The theory links visual encoding, masking, spatial attention, visual short-term memory (VSTM), and perceptual decision making in an integrated dynamic framework. The theory assumes that decisions are made by a diffusion process driven by a neurally plausible, shunting VSTM. The VSTM trace encodes the transient outputs of early visual filters in a durable form that is preserved for the time needed to make a decision. Attention increases the efficiency of VSTM encoding, either by increasing the rate of trace formation or by reducing the delay before trace formation begins. The theory provides a detailed, quantitative account of attentional effects in spatial cuing tasks at the level of response accuracy and the response time distributions.

Keywords: attention, detection, spatial cuing, diffusion process, response time

In 1980, Posner, Snyder, and Davidson reported that detection of a bright spot of light in an otherwise empty visual field was faster if it occurred at a location to which attention had been drawn by a cue. In the same year, Bashinski and Bacharach (1980) reported a corresponding result for detecting near-threshold stimuli: Sensitivity to a brief, backwardly masked, luminance stimulus was higher if it occurred at a cued location. Together, these findings challenged a theoretical distinction that, in one form or another, had underpinned much of the research on attention since the 1950s. This was the idea that the processes involved in identifying a stimulus can be divided into two classes that differ in their attentional demands. One is a class of preattentive processes, which are parallel and unlimited in capacity; the other is a class of focal attention processes, which are selective and tightly capacity limited, possibly serial (Broadbent, 1958; Egeth, 1977; Neisser, 1967). According to this distinction, detection—that is, determining whether a simple, featurally defined stimulus is present in, or absent from, the visual field—is carried out preattentively and neither benefits from nor requires attention. Focal attention is needed only for more complex judgments, such as discrimination or recognition of form. Clearly, if this were so, there should be

neither a response time (RT) advantage nor a sensitivity advantage for cued stimuli in detection tasks.

Together, the studies of Posner et al. (1980) and Bashinski and Bacharach (1980) ushered in a period of research on attention and visual signal detection that, over the next 25 years, progressively identified the conditions under which detection does and does not benefit from attention. The upshot is that we now know that no simple characterization of the effects of attention in detection is possible or can be expected. Rather, the effects that are found depend on the way in which detectability is manipulated and on the choice of dependent variable: sensitivity (accuracy) or RT. Attention interacts with a number of variables that affect detectability, including the use of visual masks (Smith, 2000a), the presence of external noise in the display (Doshier & Lu, 2000a, 2000b; Smith & Wolfgang, 2007), and spatial uncertainty (Gould, Wolfgang, & Smith, 2007). Under some conditions, attention affects detection sensitivity and RT in equivalent ways; under others, it affects RT but has no effect on sensitivity (Smith, Ratcliff, & Wolfgang, 2004). These results are inconsistent with any simple form of attention–preattention dichotomy, but they nevertheless emphasize the need to distinguish between detection and other kinds of perceptual judgment, because the effects found in detection may have no direct counterpart in other tasks.

In this article, we describe a theory of attention and perceptual decision making that accounts for the pattern of sensitivity and RT effects found in cued signal-detection tasks. The theory gives rise to a class of models that make detailed, quantitative predictions about how attentional cues and other variables jointly affect sensitivity and the distributions of RT. Our theory differs from other computational models of attention, such as those of Bundesen (1990), Reeves and Sperling (1986), and Sperling and Weichselgartner (1995), in its emphasis on performance in the simplest possible visual task, namely, detection of a single stimulus in an otherwise empty display. It also differs from attentional models

Philip L. Smith, Department of Psychology, The University of Melbourne, Melbourne, Victoria, Australia; Roger Ratcliff, Department of Psychology, The Ohio State University.

The research in this article was supported by Australian Research Council Discovery Grants DP0209249 to Philip Smith and DP0558761 and DP0880080 to Philip Smith and Roger Ratcliff. We thank Claus Bundesen, John Palmer, Erik Reichle, and Jim Townsend for comments on an earlier version of the article. Matlab and C code to fit the models described in this article can be downloaded from <http://web.psych.unimelb.edu.au/philip/SmithRatcliff09.zip>

Correspondence concerning this article should be addressed to Philip L. Smith, Department of Psychology, University of Melbourne, Victoria 3010, Australia. E-mail: philip@unimelb.edu.au

based on signal-detection theory, such as those of Foley and Schwarz (1998); Lu and Doshier (1998); Shimozaki, Eckstein, and Abbey (2003); Palmer, Ames, and Lindsey (1993); and many others, in that it predicts both RT and accuracy in cued detection tasks. Despite the simplicity of the experimental task, the patterns of data it produces are complex and have not yet been captured satisfactorily in theoretical models. Before describing the theory, we first provide a description of two key sets of experimental data to give readers an idea of the phenomena to be explained.

Attentional Cuing Effects in Signal Detection

The data we consider were obtained from experiments on the detection of Gabor patch stimuli (Gaussian-vignetted sinusoidal gratings), as shown in Figure 1. Similar results would likely be obtained using other kinds of stimuli, such as spots of light (e.g., Smith, 1995), but we used Gabor patches for the same reasons as they are favored by many vision researchers, namely, that they are band-limited and their intensity profiles correspond roughly to the receptive fields of cells in primary visual cortex (Webster & De Valois, 1985). This means that the set of visual mechanisms they stimulate is localized in space and frequency. Although our theoretical focus is on detection, the data we describe were actually obtained from a very easy discrimination task, in which observers discriminated between horizontally and vertically oriented Gabor patches in a two-alternative forced-choice design. We follow Lee, Koch, and Braun (1997) and use this task as a proxy for detection. The main advantages of the orthogonal discrimination task over the traditional yes/no task of signal-detection theory are that it is relatively unbiased, which minimizes criterion effects, and that it can be represented satisfactorily using an equal-variance signal-detection model, whereas yes/no tasks typically require an

unequal-variance model (Green & Swets, 1966). These properties simplify the task of obtaining reliable estimates of sensitivity for attended and unattended stimuli. A further advantage in the present setting is that, unlike the yes/no task, the distributions of RT for the two responses are very similar, which simplifies the task of fitting our mathematical models.

Treating orthogonal discrimination and yes/no detection as functionally equivalent is justified empirically by the results of Thomas and Gille (1979), who showed that contrast thresholds in the two tasks are indistinguishable: If a stimulus can be detected, its orientation can be discriminated, and vice versa. This finding was used by Lee et al. (1997) and by Cameron, Tai, and Carrasco (2002) as justification for treating the two tasks as equivalent for the purpose of drawing inferences about attention. In support of this approach, we have shown that our main experimental findings on the effects of attention in the orthogonal discrimination task are also found in yes/no and rating scale tasks (Smith, 2000a; Smith & Wolfgang, 2004; Smith, Wolfgang, & Sinclair, 2004). These similarities support the idea that the two tasks can be treated as equivalent and interchangeable in studies of attention.

Our experimental task was the Posner (1980) spatial cuing paradigm, in which a cue is used to draw attention to a particular location in the visual field. The cue is followed, after a brief delay, by the stimulus, which can appear either at the cued location or at some other location, with some probability. In the version of the task shown in Figure 1, the cue consisted of four pairs of perpendicular lines, which marked the corners of a 1.8° square, located on the circumference of an imaginary 6.2° diameter circle, surrounding a central fixation point. The cue was flashed for 60 ms at a stimulus onset asynchrony (SOA) of 140 ms before presentation of the stimulus. The stimulus could appear at any one of three locations on the circumference of the circle, one cued and two uncued. On each trial, the angular position of the cue, α ($0 < \alpha \leq 360^\circ$), was selected randomly; the uncued locations were at $\alpha + 120^\circ$. With this configuration, the two uncued locations were equally distant from the cued location and so, under reasonable assumptions about the spatial distribution of attention, should have received similar processing resources. In our studies, the cue-target SOAs and the stimulus exposures were chosen so that their sum was 200 ms or less. This ensured that the stimulus could not be brought into central vision by making a saccadic eye movement during the course of an experimental trial (Hallett, 1986).

One of the ongoing theoretical challenges in the visual signal-detection literature has been to distinguish between two mechanisms whereby attention can influence performance. These mechanisms have been variously termed *signal* or *stimulus enhancement* and *uncertainty* or *noise reduction* (Lu & Doshier, 1998; Shiu and Pashler, 1994; Van der Heijden & Brouwer, 1999). Signal enhancement is a selective increase in the signal-to-noise ratio for stimuli at attended locations and is usually thought of as a perceptual-level phenomenon. Uncertainty reduction is an increase in performance due to foreknowledge of the likely location of the stimulus in the display, increasing the efficiency with which it can be segregated from its background or distinguished from distractors (Palmer, Verghese, & Pavel, 2000; Tanner, 1961). Uncertainty reduction is usually thought of as a decision-level phenomenon.

Whereas signal enhancement implies some form of selective, limited-capacity processing mechanism, uncertainty reduction de-

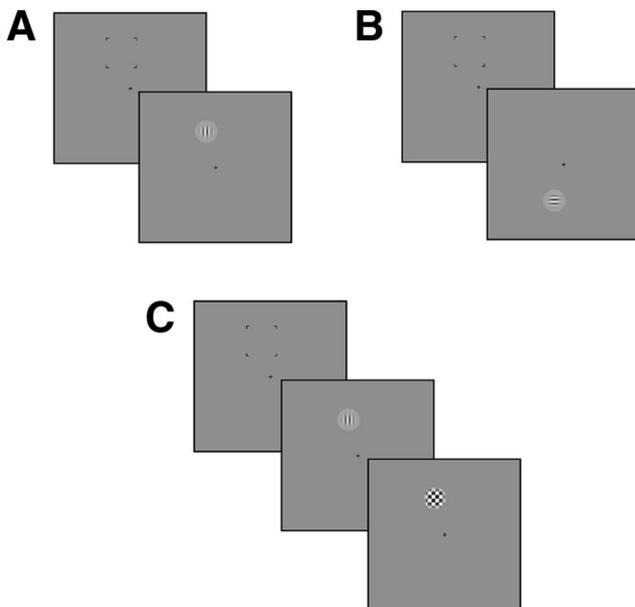


Figure 1. Examples of the stimulus configurations used by Smith, Ratcliff, and Wolfgang (2004). A. Vertically oriented pedestal Gabor patch at a cued location. B. Horizontally oriented pedestal Gabor patch at a miscued location. C. Cued patch followed by a checkerboard backward mask.

scribes changes in the total amount of noise entering the processing system and occurs even in the absence of capacity limitations. Consequently, signal enhancement can be identified experimentally only when the effects of uncertainty are eliminated or otherwise controlled (Shaw, 1984). A number of visual search phenomena once thought to reflect limited-capacity processing can be well described by multichannel signal-detection models that attribute them to a combination of display noise and spatial uncertainty (Baldassi & Burr, 2004; Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer et al., 1993; Shimozaki et al., 2003; Verghese, 2001).

The effects of uncertainty are likely to be particularly significant in cued detection tasks in which near-threshold stimuli are presented against a uniform background at unknown locations. Under such conditions, the background acts as a field of noise against which the stimulus must be distinguished. Because manipulations of attention are usually also manipulations of uncertainty, some way is needed to decouple the effects of the cue on uncertainty from its effects on stimulus signal-to-noise ratio. Figure 1 shows how this can be done experimentally, by ensuring stimuli remain well localized perceptually, even at very low intensities. Rather than presenting stimuli directly on a uniform background, they are presented on top of circular, suprathreshold contrast, luminance pedestals (Smith, 2000a). Detectability is manipulated by varying the contrast of the grating relative to the luminance of the pedestal, but the stimulus itself remains well localized because the pedestal is kept at a constant, suprathreshold contrast.

Effects of Backward Masking

Probably the best predictor of whether attentional effects will be found in detection tasks is whether or not stimuli are backwardly masked. In a review of the literature on visual signal detection, Smith (2000a) pointed out that most of the detection studies reporting signal enhancement limited the information from the stimuli with backward masks (Bashinski & Bacharach, 1980; Brawn & Snowden, 2000; Downing, 1998; Hawkins et al., 1990; Luck et al., 1994; Müller & Humphreys, 1991; Smith, 1998), whereas most of those finding small effects or no effect limited stimulus information using contrast or exposure duration alone (Bonnell & Hafter, 1998; Bonnell, Stein, & Bertucci, 1992; Davis, Kramer, & Graham, 1993; Foley & Schwarz, 1998; Graham, Kramer, & Haber, 1985; Palmer, 1994; Palmer et al., 1993; Shaw, 1984). This conclusion echoes earlier suggestions that the attentional effects found in other tasks may also depend on the use of masks (Cheal & Lyon, 1992; Giesbrecht & Di Lollo, 1998; Morgan, Ward, & Castet, 1998; Shiu & Pashler, 1994). It is also consistent with the attentional dependencies found in masking paradigms such as metacontrast masking (Ramachandran & Cobb, 1995), object-substitution masking (Enns & Di Lollo, 1997; Francis, 2000; Tata, 2002), and radial frequency-pattern masking (Smith, Lee, Wolfgang, & Ratcliff, in press).

In a series of studies comparing masked and unmasked stimuli in the same experimental task, Smith and colleagues found systematic cuing effects in detection with masked stimuli but none with unmasked stimuli. We refer to this as the *mask-dependent cuing effect*. As we noted previously, we have found the same effect using a yes/no task (Smith, 2000a; Smith & Wolfgang, 2004), a rating-scale task (Smith, Wolfgang, & Sinclair, 2004), and the orthogonal discrimination task (Smith, Ratcliff, & Wolfgang,

2004). Because these studies all used pedestals to localize the stimuli, the results could be attributed to some form of signal enhancement, and not to uncertainty reduction. This article presents a theory of how such enhancement arises.

Figure 2 shows a summary of the data of Smith, Ratcliff, and Wolfgang (2004), who used the orthogonal discrimination task to compare the effects of cues for masked and unmasked stimuli. A novel feature of this study was that accuracy and RT data were collected from the same task. In one condition of the experiment, stimuli were flashed for 60 ms and then extinguished; in another, they were flashed for 60 ms and then followed immediately by a high-contrast checkerboard mask, as shown in Figure 1. Stimulus contrast was varied in five steps, using the method of constant stimuli, to yield psychometric functions ranging from near-chance to near-perfect performance for each observer. Roughly twice the stimulus contrast was needed to obtain performance under masked conditions equal to that under unmasked conditions.

The top panels of Figure 2 show detection sensitivity (signal detection d') for masked and unmasked stimuli; the bottom panels show mean RT. The data in the figure are averaged over five observers; the individual observer data may be found in the original article. Each data point in Figure 2 is based on 400 trials per observer (2,000 trials per point; 8,000 trials in total per observer). Perhaps the most striking feature of the data is that there is a dissociation in the effects of cues on sensitivity and on RT. When stimuli were masked, detection sensitivity for cued stimuli exceeded that for miscued stimuli at all levels of contrast. When stimuli were unmasked, sensitivity to cued and miscued stimuli did not differ. For RT, the situation is different. Mean RT shows the usual Piéron's law (power law) reduction with increasing contrast (Luce, 1986) but, more important, cues had an unconditional effect on mean RT. Although the cuing effect in RT was larger for masked stimuli, mean RTs were shorter for cued than for miscued stimuli for both masked and unmasked stimuli.

The pattern of data in Figure 2 highlights what has been one of the longstanding theoretical puzzles in the attention literature. Whereas many detection studies have found no effect of cues on sensitivity,

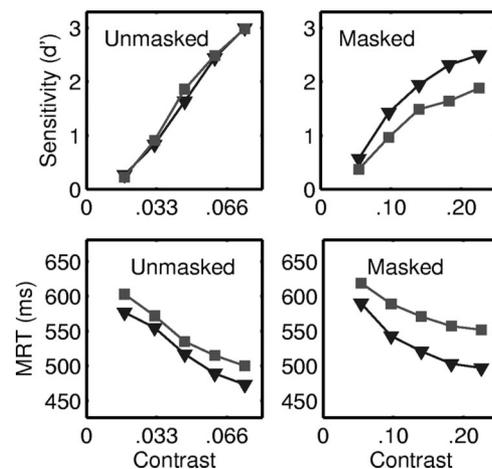


Figure 2. Results from Smith, Ratcliff, and Wolfgang (2004). Triangles are cued stimuli; squares are miscued stimuli. The upper panels show detection sensitivity (d') for masked and unmasked stimuli; the lower panels show mean response time (MRT). Data from five observers.

shorter RTs to cued stimuli—the so-called Posner effect—are routinely found. This has led some researchers to suggest a criterion-setting explanation for the RT effects (Pashler, 1998; Sperling, 1984; Sperling & Doshier, 1986). According to the criterion-setting account, observers in RT tasks set lower decision criteria at attended locations and so base their decisions about attended stimuli on less evidence, resulting in shorter RTs for attended stimuli. This explanation attributes the RT differences wholly to differences in the observers' decision strategies and not to any increase in the quality of stimulus information at attended locations. Clearly, the pattern of data in Figure 2 is more complex than can be accommodated by a simple criterion-setting account, which cannot explain the interaction with backward masking in the sensitivity data. Nevertheless, we agree with Sperling and Pashler that the relationship between accuracy and RT found in tasks of this kind implies the involvement of a sequential-sampling decision mechanism that accumulates noisy information to a response criterion. Such mechanisms play a fundamental role in the theory that follows.

The mask dependencies in d' in Figure 2 were obtained using weakly predictive peripheral cues. Stimuli were presented at the cued location on 50% of trials and at each of the uncued locations on 25% of trials. We have found similar mask dependencies using a number of different cue configurations and a variety of cue validities, including 80% valid cues (Smith, 2000a) and 100% valid cues (Smith, Wolfgang, & Sinclair, 2004). The effect thus does not depend critically on cue validity.

Effects of Spatial Uncertainty

The pattern of sensitivity values shown in Figure 2 is obtained only with perceptually well-localized stimuli. Two recent studies by Carrasco and colleagues (Cameron et al., 2002; Carrasco, Penpeci-Talgar, & Eckstein, 2000) investigated the effects of cues on detecting and discriminating oriented Gabor patch stimuli. Both studies included conditions requiring discrimination between pairs of stimuli of dissimilar orientation ($\pm 45^\circ$ in the Carrasco et al., 2000, study; $\pm 15^\circ$ in the Cameron et al., 2002, study), which should have yielded results comparable to detection. Unlike the studies of Smith and colleagues described previously, these studies presented stimuli directly against a uniform background. Both yielded significant cuing effects with unmasked stimuli, which Carrasco and colleagues attributed to signal enhancement.

We suspected that the reason why we obtained different results with unmasked stimuli to those of Carrasco and colleagues was because of the greater perceptual saliency of our stimuli and the resulting differences in the effects of uncertainty. Whereas our stimuli were localized perceptually by pedestals, the stimuli used by Carrasco and colleagues were not. Consequently, performance in their task may have been influenced by uncertainty.

To test this, Gould et al. (2007) compared performance in an unmasked, orthogonal discrimination task under differing conditions of uncertainty. Half of the stimuli were presented directly on a uniform field, without localizing markers, like the stimuli used by Carrasco and colleagues. The other half were localized perceptually, using fiducial crosses (Eckstein, Pham, & Shimozaki, 2004), as shown in Figure 3. These consisted of four radial lines centered on the stimulus location, whose onset and offset coincided with that of the stimulus. The fiducial crosses served the same role as did the pedestals in our earlier studies, namely, to ensure that stimuli were well

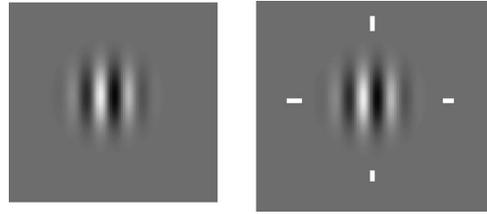


Figure 3. Examples of the stimuli used by Gould, Wolfgang, and Smith (2007). Low-contrast Gabor patches were presented directly on a uniform field without localizing markers (left), or were localized perceptually by fiducial crosses (right). The display configuration was the same as that used in the study of Smith, Ratcliff, and Wolfgang (2004), shown in Figure 1.

localized perceptually, even at very low contrasts. All aspects of the stimulus display, cue contingencies, and the experimental procedure were identical to those in the study of Smith, Ratcliff, and Wolfgang (2004), except that stimuli were presented for 40 ms, rather than 60 ms. As in that study, stimuli were presented on the circumference of an imaginary circle surrounding a fixation point. Cued stimuli were presented at randomly chosen angles (α , $0 < \alpha \leq 360^\circ$); uncued stimuli were presented at $\alpha \pm 120^\circ$. Also as in that study, we collected measures of both accuracy and RT.

Figure 4 summarizes the results of Experiment 1 of Gould et al. (2007). The upper panels show detection sensitivity for stimuli presented with and without fiducial crosses; the lower panels show mean RT. The data in the figure are averages over five observers, as before. Each data point in the figure is based on 360 trials per observer (1,800 trials per point; 7,200 trials total per observer). When no fiducial crosses were used, the results replicated those of Carrasco and colleagues: Sensitivity was uniformly higher for cued than for miscued stimuli, even though no backward masks were used. When stimuli were localized by fiducial crosses, the results replicated those of Smith and colleagues using pedestals: Sensitivity for cued and miscued stimuli did not differ.

A similar dissociation appears in the pattern of sensitivity and mean RT to that found in the masking data of Smith, Ratcliff, and Wolfgang (2004). As in that study, mean RTs were shorter at higher levels of contrast, decreasing roughly according to Piéron's law. Also, as in that study, mean RTs were shorter for cued than for miscued stimuli, irrespective of whether fiducial crosses were used and irrespective of whether a sensitivity effect was obtained. The exception to this was at the very lowest contrasts in the no-fiducial condition, in which cued and miscued RTs did not differ. We discuss possible reasons for a breakdown in the cuing effect in RT at low contrasts subsequently.

The data show that uncertainty is a significant determinant of attentional effects when a single stimulus is presented without localizing information in an otherwise empty display. We refer to the pattern of results in Figure 4 as the *uncertainty-dependent cuing effect*, to emphasize the parallels with the results of Smith, Ratcliff, and Wolfgang (2004). Researchers have characterized the effects of uncertainty mathematically using multichannel signal-detection models (Pelli, 1985; Shaw, 1982; Tanner, 1961). The signal-detection approach to uncertainty assumes the uniform field on which the stimulus is presented is coded by a population of noisy spatial frequency and orientation tuned filters. Detection of a near-threshold target at an unknown location requires that activity in the filter coding the target be distinguished from noise in the filters coding the uniform field.

Information about the probable target location allows perceptual activity at other display locations either to be excluded from the observer's decision, or for its effects to be attenuated. This leads to a reduction in the amount of noise entering the decision process, increasing signal-to-noise ratio, and thus improving sensitivity. Because knowledge of the target location leads to a reduction in decision noise, researchers typically regard "noise reduction" and "uncertainty reduction" as synonyms.¹

Signal-detection theory has proven to be a powerful tool for understanding the relationship between stimulus uncertainty and attentional effects in a variety of perceptual tasks (Baldassi & Burr, 2004; Doshier & Lu, 2000a, 2000b; Eckstein et al., 2000; Foley & Schwarz, 1998; Lu & Doshier, 1998; Palmer et al., 1993; Palmer et al., 2000; Shaw, 1980, 1982, 1984; Shaw, Mulligan, & Stone, 1983; Shimozaki et al., 2003; Smith, 1998; Verghese, 2001). However, one of its limitations is that it is a static (i.e., a random variable) theory, and so offers little insight into the time course of uncertainty reduction. Our theory seeks to extend signal-detection theory to give a characterization of the dynamic nature of uncertainty reduction. Following Pelli (1985), we assume that uncertainty changes the shape of the visual contrast response function that relates the detectability of a stimulus to its contrast. The effect of an increase in uncertainty is to make low contrast stimuli less detectable than they would be otherwise. In our theory, uncertainty and attention jointly determine the efficiency with which a durable representation of the stimulus is formed in visual short-term memory (VSTM). This representation serves as the basis of the observer's decision.

The Elements of the Theory

We claim that an adequate account of the mask-dependent cuing effect in sensitivity, and of the relationship between RT and accuracy in Figures 2 and 4, requires a theory that links visual encoding, masking, attention, VSTM, and decision making in an integrated dynamic framework. Figure 5 shows the elements of such a theory.

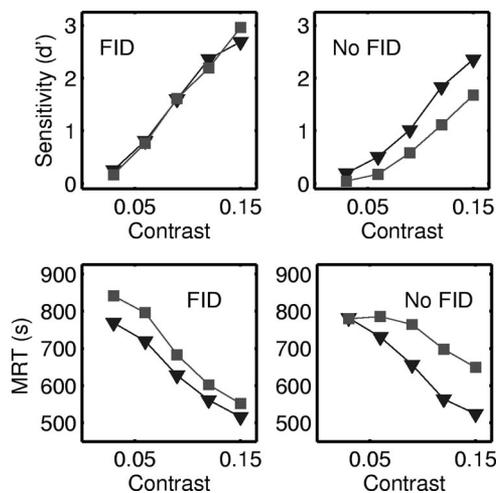


Figure 4. Results from Gould, Wolfgang, and Smith (2007). Triangles are cued stimuli; squares are miscued stimuli. The upper panels show detection sensitivity (d') for stimuli presented with fiducial crosses (FID) or without fiducial crosses (no FID); the lower panels show mean response time (MRT). Data from five observers.

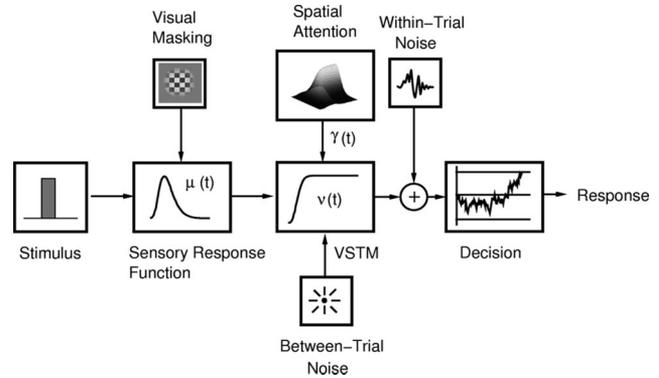


Figure 5. System model of attention and decision making. The sensory response function, $\mu(t)$, describes the visual response to a briefly flashed stimulus. The information in the sensory response function is encoded in a durable form in visual short-term memory (VSTM) under the control of spatial attention, $\gamma(t)$. The VSTM trace is subject to moment-by-moment perturbations by broad-spectrum, Gaussian noise (the noise at the top). Successive samples of the noise-perturbed VSTM trace are accumulated over time by a diffusion process to make a decision. The noise source at the bottom represents the effects of trial-to-trial variation on the quality of the encoded stimulus information. Its effect in the model is to introduce variability in the strength of the VSTM trace.

The theory assumes that decisions are made by a sequential-sampling mechanism that accumulates successive samples of noisy stimulus information to a response criterion. The criterion that is reached determines the response that is made; the time taken to reach it determines the RT. We assume that evidence accumulates continuously in time and that the accumulating evidence state is itself continuously distributed. This leads to a representation of the decision stage as a diffusion process of some kind (Busemeyer & Townsend, 1993; Ratcliff, 1978; Ratcliff & Smith, 2004; Smith, 1995, 2000b). We consider two decision mechanisms based on diffusion processes in more detail subsequently.

The rate at which evidence accumulates in the decision stage—the so-called *drift* of the diffusion process—depends on the quality, or the strength, of a representation of the stimulus in VSTM. Strong VSTM traces lead to a high rate of evidence accumulation, resulting in high accuracy and short RTs. The strength of the VSTM trace depends, in turn, on stimulus contrast and on the properties of the early spatiotemporal filters that encode the stimulus, interacting with any masking stimulus that may be present, under the control of spatial attention. The role ascribed to attention in our theory is that it controls the formation of a representation of the stimulus in VSTM. This is, of course, a classical notion, with a long pedigree in the psychological literature. It was first proposed by Sperling (1960) in his pioneering study of iconic memory and has been developed by him and his coworkers in a number of articles subsequently (e.g., Reeves &

¹ Lu and Doshier (1998) use a similar term, "noise exclusion," in a sense that is unrelated to uncertainty reduction. They use it to describe an attention-dependent mechanism that helps distinguish signal from noise at a single, specified display location. The action of this mechanism is unrelated to changes in an observer's uncertainty that result from knowledge of the target location. Lu and Doshier's theory is discussed in more detail later in this article.

Sperling, 1986; Sperling & Weichselgartner, 1995), as well as by many others. The same idea is instantiated dynamically in our theory.

Perceptual Encoding and Visual Masking

We assume stimuli are encoded by perceptual analyzers that, to a first approximation, act as linear spatiotemporal filters (Graham, 1989). The filters are low-pass in the sense that they remove fast transients, or high temporal frequencies, from the stimuli they encode. The effect of filtering is to transform brief, rectangular pulse stimuli into smooth time-varying functions, like those shown in Figure 6. We model the response of a filter to a step-onset, unit increase in contrast using a gamma function

$$\Gamma(t; \beta, n) = 1 - e^{-\beta t} \sum_{j=0}^{n-1} \frac{(\beta t)^j}{j!}. \quad (1)$$

In probabilistic settings, Equation 1 describes the distribution of a sum of n exponentially distributed random variables. In linear filtering settings, Equation 1 describes the deterministic response of a filter comprising n identical resistance-capacitance stages in cascade to a unit-step function stimulus. (The unit-step function is defined to be zero for all times $t < 0$ and unity thereafter.) The exponential rate constant, β , determines the filter's temporal response characteristics. Smaller values of β lead to a slower temporal response, resulting in greater attenuation of the high temporal frequencies in the filter's output. Models based on linear filters, which generalize and extend Equation 1 in various ways, have been widely used in studies of visual temporal sensitivity (Watson, 1986).

The information an observer extracts from a stimulus depends on the temporal response of the filter that encodes it and on the effects of any masking stimulus that is present. The assumptions we make about visual masking are the simplest ones possible. We assume that the effect of a backward, pattern mask, like the one shown in Figure 1, is to limit the time for which stimulus infor-

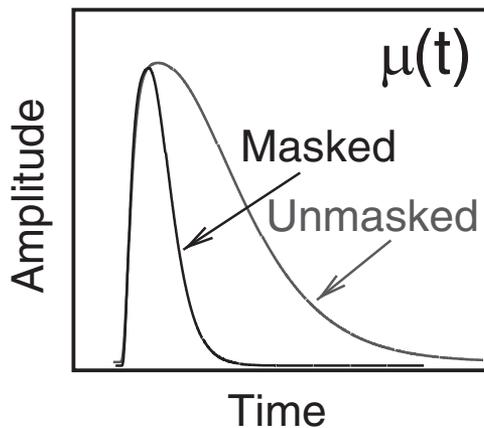


Figure 6. Sensory response functions for masked and unmasked stimuli. When stimuli are unmasked, they are encoded and then undergo slow iconic decay. When stimuli are masked, they are rapidly suppressed by the mask.

mation is available to later processing stages. In Coltheart's (1980) terms, masks limit the *informational persistence* of the stimulus. When stimuli are unmasked, informational persistence is long. They are encoded and are then subject to slow iconic decay, so the information they contain is available to later processing stages for a relatively long time. When they are masked, informational persistence is short. They are encoded and are then rapidly suppressed by the mask, so the information they contain is available for a limited time only. The difference in the informational persistence of masked and unmasked stimuli is fundamental to the theory's ability to predict mask-dependent cuing effects.

In our theory, the time course of stimulus processing is described by a *sensory response function*, $\mu(t)$, as shown in the box on the left of Figure 5. To model the different sensory response characteristics of masked and unmasked stimuli we use an extended linear filter representation of the form

$$\mu(t) = \alpha \Gamma(t; \beta_{\text{on}}, n) [1 - \Gamma(t - d; \beta_{\text{off}}, n)], \quad (2)$$

where the function $\Gamma(t; \beta, n)$ is defined by Equation 1. This equation was introduced by Smith and Wolfgang (2004) in a precursor to the present theory. Equation 2 describes the sensory response of the visual system to a rectangular pulsed stimulus of amplitude α , and duration d . It extends the usual linear filter representation by allowing the filter rise (onset) time and fall (offset) time to be different. The onset and offset times are characterized, respectively, by the rate constants β_{on} and β_{off} . When stimuli are unmasked, we assume that $\beta_{\text{off}} < \beta_{\text{on}}$; when they are masked, we assume the converse.

Examples of the sensory response functions yielded by Equation 2 for masked and unmasked stimuli are shown in Figure 6. When $d > 0$ and $\beta_{\text{off}} < \beta_{\text{on}}$, Equation 2 becomes indistinguishable from the simpler, additive representation

$$\mu(t) = \alpha [\Gamma(t; \beta_{\text{on}}, n) - \Gamma(t - d; \beta_{\text{off}}, n)], \quad (3)$$

at least for values of stimulus duration and onset and offset times of the kind that would be encountered experimentally.² Equation 3 is a straightforward generalization of the symmetrical linear filter model, in which $\beta_{\text{off}} = \beta_{\text{on}}$, that has been used by many authors (e.g., Busey & Loftus, 1994; Sperling & Weichselgartner, 1995). When $\beta_{\text{off}} > \beta_{\text{on}}$, Equation 2 may instead be interpreted as representing rapid, multiplicative suppression of the stimulus by the mask. Equation 2 thus allows us to model the theoretical sensory response characteristics of masked and unmasked stimuli in a flexible way using a single equation.

In adopting Equation 2 as a model of perceptual encoding, we assume that backward, pattern masks act as *interruption masks* (Kahneman, 1968). The distinction between *interruption masks*

² The extended linear filter representation in Equation 2 is well approximated by the additive representation in Equation 3 whenever $n/\beta_{\text{on}} \ll d$, that is, when the mean of the onset term, $\Gamma(t; \beta_{\text{on}}, n)$, is small relative to the stimulus duration. Then $\Gamma(t; \beta_{\text{on}}, n) \approx 1$, for $t > d$. This means that the onset term will be close to asymptote for all times at which the offset term, $\Gamma(t - d; \beta_{\text{off}}, n)$, is nonzero. Under these circumstances, the product $\Gamma(t; \beta_{\text{on}}, n)\Gamma(t - d; \beta_{\text{off}}, n)$ in Equation 2 is approximately equal to $\Gamma(t - d; \beta_{\text{off}}, n)$, which reduces Equation 2 to the additive representation. Note that in these equations, and in Equations 2 and 3 in the text, $\Gamma(t; \beta, n) = 0$ for $t < 0$.

and *integration* masks is one of several distinctions made in the literature between masking mechanisms of different kinds (Breitmeyer, 1984). In interruption masking, a trailing mask terminates stimulus processing prematurely, before it has been completed. In integration masking, the stimulus and mask fuse to form a perceptual composite, whose signal-to-noise ratio is lower than that of the stimulus in isolation. The magnitude of integration masking is the same for forward (leading) and backward (trailing) masks, and it is maximal when target and mask are simultaneous. Interruption masking is found only with backward masks, and its effects are maximal when the mask trails the target after a critical delay, usually of the order 60–100 ms. By comparing the functions obtained with dichoptic and monoptic masks, Turvey (1973) was able to show that integration masking mainly occurs peripherally, in retina, primary afferent pathways, or lateral geniculate body, whereas interruption masking occurs centrally, in visual cortex. Although there is some cortical integration of targets and masks, the effect decreases sharply with increasing temporal separation and is virtually nonexistent by 60 ms (Michaels & Turvey, 1979).

Two pieces of evidence support our claim that the cuing effects in Figure 2 depend on an interaction between attention and an interruption masking mechanism. The first comes from a recent study by Smith and Wolfgang (2007) that compared the magnitude

of the cuing effect in the orthogonal discrimination task with simultaneous masks and with backward masks that trailed the target by 60–90 ms. In all five of the experiments reported in the article, there were large cuing effects with backward masks, whereas the cuing effects with simultaneous masks were weak or absent. This is consistent with the idea that the cuing effect arises when the mask terminates stimulus processing prematurely.

The second piece of evidence comes from a study by Smith and Wolfgang (2004), the results of which are summarized in Figure 7. This study investigated the effects of attentional cues in yes/no detection using pedestal Gabor stimuli under dichoptically masked, monoptically masked, or unmasked conditions. The design compared detection sensitivity for cued and miscued stimuli to sensitivity with a neutral cue, which provided no information about the likely target location. When stimuli were unmasked, there was no evidence of a cuing effect of any kind, a result replicating that of Smith, Ratcliff, and Wolfgang (2004; see our Figure 2). However, there were systematic cuing effects for all observers with both monoptic and dichoptic masks. Indeed, as Figure 7 shows, the dichoptic effect was actually slightly larger than the monoptic effect. This is consistent with the view that cuing effects in detection occur when the mask interrupts stimulus processing, because interruption masking should be maximized by

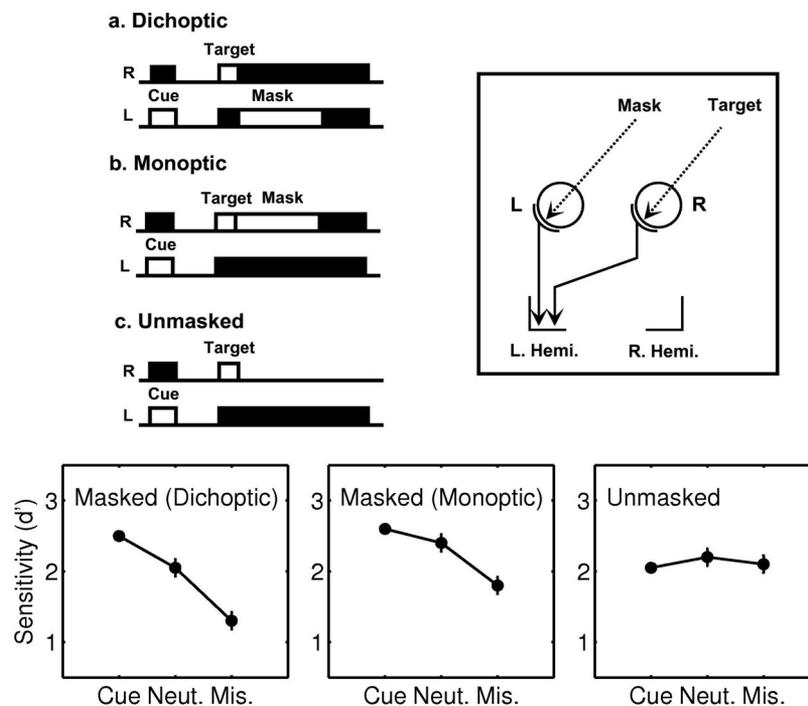


Figure 7. Results of Smith and Wolfgang (2004). Stimuli were dichoptically masked (targets and masks to different eyes), monoptically masked (targets and masks to the same eye), or unmasked. The inset at the upper left shows the sequence of stimulus events in the three conditions. The white rectangles are stimulus events; the black rectangles are occlusions. The inset at the upper right shows the anatomical pathways stimulated by targets and masks in the dichoptic presentation condition. A target and mask are presented at the same spatial location in the visual field to the temporal hemiretina of the left eye and nasal hemiretina of the right eye, respectively. They travel via different afferent pathways to area V1 in the left cerebral hemisphere. The panels at the bottom show detection sensitivity (d') for cued stimuli (Cue), neutrally cued stimuli (Neut.), and miscued stimuli (Mis.) in the three masking conditions. Data from five observers. Stimulus contrasts were set individually to try to ensure average performance of $d' = 2.0$ for each observer in each condition.

dichoptic viewing. Interruption masking is thought to arise when two competing pattern stimuli arrive in close succession in visual cortex. These effects are largest when stimuli have no opportunity to fuse in primary afferent pathways (Turvey, 1973).

Visual Short-Term Memory

In the second stage of processing, the information in the stimulus is encoded in a durable form in VSTM under the control of spatial attention. We denote the strength of the VSTM trace as a function of time by $v(t)$. As shown in Figure 5, following stimulus presentation, the VSTM trace grows smoothly to an asymptote whose value depends on the contrast and duration of the stimulus and on whether it appears at an attended or an unattended location. The idea that there exists a relatively durable, posticonic, precategorical form of VSTM is well established—the classic study being that by Phillips (1974). VSTM is “posticonic” in the sense that it is able to survive visual masking; it is “precategorical” in the sense that the stimulus information it contains has not been categorized or identified. Recently, Vogel, Woodman, and Luck (2006) showed that the VSTM trace is formed within the first 200 or 300 ms of stimulus onset, consistent with the estimates we report here. Our recognition of the need for an explicit VSTM stage was in response to these findings and to the results of Ratcliff and Rouder (2000), who showed that the information in a transient stimulus event is preserved without decaying for the time needed to make a decision.

We modeled the growth of VSTM mathematically using a particular kind of differential equation called a *shunting equation*. The distinguishing feature of shunting equations is that the input to the system—the so-called forcing function, which describes the information in the stimulus—enters into the equation multiplicatively, rather than additively, as occurs in the usual linear system formulation. This gives shunting equations desirable characteristics in models of short-term memory, as we discuss subsequently.

A number of authors have proposed shunting equations as models of biological computation, the most comprehensive theoretical treatment being that of Grossberg and colleagues (e.g., Grossberg, 1987, 1988). The first use of shunting equations in a psychophysical setting was by Sperling and Sondhi (1968) in a model of flicker perception. However, the application closest in spirit to our own is that of Loftus and colleagues (Busey & Loftus, 1994), who, like us, used a shunting equation to model the encoding of stimulus information in VSTM. Unlike ours, their model is a functional model, which uses a shunting equation to describe the growth of the proportion of correct responses over time. Our model is a process model, which uses a shunting equation to describe the psychological processes that underlie the observed performance, rather than the observed performance itself. The advantage of the process model approach over the functional approach is that it provides an account of both RT and accuracy, whereas the functional approach provides an account of accuracy only.

To ensure the VSTM trace grows to an asymptote and does not saturate at long stimulus exposures, we assume it arises as the result of an opponent channels, or excitatory-inhibitory, coding process. Denoting the intensity of the stimulus by Δ_I and the intensity of the background by I_0 , we assume that the growth of VSTM is described by the differential equation

$$\frac{dv}{dt} = \Delta_I \mu(t) [\theta - v(t)] - I_0 \mu(t) v(t). \quad (4)$$

Two features of this equation are noteworthy. First, as mentioned previously, the information in the stimulus, $\mu(t)$, enters into the equation multiplicatively, rather than additively. This means that when the stimulus is removed—that is, when $\mu(t) = 0$ —the derivative dv/dt goes to zero and the trace stops changing. Shunting equations thus provide a natural way to model the formation of a durable VSTM trace in response to a transient stimulus. Second, assuming an initial condition of $v(0) = 0$, the VSTM trace remains bounded in the interval $[0, \theta]$, regardless of the intensity or duration of the stimulus. This allows such equations to, in effect, retune their sensitivity in response to changes in the dynamic range of the input, allowing them to circumvent what Grossberg referred to as the “noise-saturation dilemma” (Grossberg, 1987).

Using standard techniques for solving first-order, linear differential equations (see Appendix A), it is straightforward to show that the solution of Equation 4 is

$$v(t) = \theta \left(\frac{\Delta_I}{\Delta_I + I_0} \right) \left\{ 1 - \exp \left[- (\Delta_I + I_0) \int_0^t \mu(s) ds \right] \right\}. \quad (5)$$

This equation states that, asymptotically, the VSTM trace grows to a value proportional to $\Delta_I / (I_0 + \Delta_I)$, the modified Weber contrast of the stimulus. We use the term “modified” Weber contrast because the stimulus increment is divided by the maximum intensity in the display, $I_0 + \Delta_I$, rather than by the background intensity, I_0 , as in the usual definition of Weber contrast. The modified Weber contrast has been advocated by some authors (e.g., Burr, Ross, & Morrone, 1985) as a measure of stimulus intensity that reflects that action of early gain control mechanisms. We subsequently denote the asymptotic trace strength in Equation 5 as $v(\infty)$. This notation expresses the property that the asymptotic trace strength is the value to which $v(t)$ converges if the stimulus is prolonged indefinitely.

Because shunting equations rescale luminances into contrasts (i.e., absolute intensities into relative intensities), the model of Equations 4 and 5 is indifferent to whether Δ_I is expressed in luminance or contrast units. We have used the neutral term “intensity” to reflect this fact. Physiological evidence suggests that gain control mechanisms transform luminance into contrast fairly early in visual processing and probably prior to the formation of the VSTM trace (Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990). Under these circumstances, we would interpret Δ_I as a contrast increment and assume that $I_0 = 1$. In fitting our models to data, we assume that stimulus intensities are scaled in this way.

The rate of approach to the asymptote $v(\infty)$ in Equation 5 depends on the exponential rate constant, $I_0 + \Delta_I$. For small values of the increment, Δ_I , the rate depends only weakly on the stimulus intensity and is approximately equal to the intensity of the background, I_0 . The approach to asymptote is controlled by the area under the sensory response function, $\mu(t)$. When stimulus persistence is long, the area under $\mu(t)$ is large and VSTM trace strength closely approaches its theoretical maximum. When stimulus persistence is short, the area under $\mu(t)$ is reduced and the maximum is not attained. This property is central to our theory’s ability to

predict the mask-dependent cuing effect. The constant θ in Equations 4 and 5 is a scaling parameter that could be set equal to unity without loss of generality, but we retain it in our equations to represent the mapping of stimulus contrast into VSTM trace strength.

In adopting Equation 5 as a model of VSTM trace formation, we are not assuming that contrast transduction in the visual system is linear. We know, rather, that the contrast response of the visual system is a saturating, nonlinear function of contrast, which is often modeled mathematically using a Naka-Rushton function,

$$r(\Delta_I) = \frac{\Delta_I^\rho}{\Delta_I^\rho + I_{in}}. \quad (6)$$

The constant I_{in} in the denominator of Equation 6 is referred to as a *divisive inhibition* term and determines the horizontal position of the function on a log contrast scale (Boynton, 2005). The exponent ρ determines the shape of the nonlinearity and typically takes a value of around 2.0. For these values of ρ , the Naka-Rushton function has a sigmoid form.³

It is possible to incorporate transduction nonlinearities like that in Equation 6 into the VSTM growth equation by making an appropriate choice of coefficients, writing it as

$$\frac{dv}{dt} = \Delta_I^\rho \mu(t) [\theta - v(t)] - I_{in} \mu(t) v(t). \quad (7)$$

The solution to Equation 7 is

$$v(t) = \left(\frac{\Delta_I^\rho}{\Delta_I^\rho + I_{in}} \right) \left\{ 1 - \exp \left[- (\Delta_I^\rho + I_{in}) \int_0^t \mu(s) ds \right] \right\}. \quad (8)$$

This equation can be viewed as a dynamic counterpart of the divisive inhibition models that have been widely used to model cortical gain control in physiological and psychophysical settings (e.g., Boynton, 2005; Heeger, 1991; Foley, 1994; Wilson & Kim, 1998).

Equation 8 is similar to Equation 5, apart from the obvious addition of the Naka-Rushton exponent, ρ . However, a more consequential difference is the difference in the size of the divisive inhibition terms in the two equations, namely, I_0 versus I_{in} . Whereas I_0 is equal to the background luminance of the display and is typically large relative to the increment Δ_I , the inhibition term I_{in} and the increment Δ_I are typically of similar magnitudes. The upshot of this is that the rate of growth of VSTM in Equation 5 is relatively independent of stimulus contrast, but the rate in Equation 8 depends on contrast strongly. As we see subsequently, the data of Gould et al. (2007) are consistent with intensity-dependent growth, but the data of Smith, Ratcliff, and Wolfgang (2004) are not.

Attention

The transient information in the stimulus is encoded in a durable form in VSTM under the control of spatial attention. In our theory, attention translates the information in the sensory response function, $\mu(t)$, into the VSTM trace, $v(t)$. This idea is expressed graphically in Figure 5, in which the icon in the box depicting attention is intended to suggest that it is a dynamic process that possesses both a spatial and a temporal extent. In previous articles,

we investigated two different models of the translation process, both with long pedigrees in the attention literature. One is a *gain* model (Smith & Wolfgang, 2004); the other is an *orienting* model (Smith, Ratcliff, and Wolfgang, 2004). Our current theory is sufficiently flexible to allow us to represent both gain and orienting models within a unified theoretical framework and to compare their predictions.

Gain models—in the sense in which we use the term—are parallel processing models, in which attention affects the rate or efficiency of stimulus processing. In our theory, attention affects the rate of VSTM trace formation. Smith (2000a) proposed that attention affects the rate at which stimulus information becomes available to later processing mechanisms as an explanation for the mask-dependent cuing effect. Carrasco and McElree (2001) provided experimental support for this idea at about the same time. However, the idea underlying gain models can be traced back to the capacity theory of Kahneman (1973). Kahneman proposed that the efficiency of stimulus processing is governed by the availability of a unitary resource, or “capacity,” that is needed to activate neural structure. Under focused attention conditions, capacity is concentrated at a single spatial location, leading to efficient stimulus processing at the attended location and inefficient processing elsewhere. Under divided attention conditions, capacity is distributed evenly over a number of locations, leading to intermediate efficiency everywhere. Subsequently, Townsend et al. showed how the idea of capacity limitations in a parallel processing system could be expressed in a precise and mathematically rigorous way, by assuming that capacity determines the rates in the underlying processing channels (Townsend & Ashby, 1978; Townsend & Wenger, 2004).

We can represent gain models very simply in our theory by introducing an attentional gain constant, γ , into the VSTM equation, Equation 7:

$$\frac{dv}{dt} = \gamma_i \Delta_I^\rho \mu(t) [\theta - v(t)] - I_{in} \mu(t) v(t), \quad \gamma_i \in \{\gamma_A, \gamma_U\}. \quad (9)$$

We write the gain constant with a subscript, i , to express the fact that gain differs as a function of position in the visual field. Minimally, we assume that gain has one value, γ_A , at attended locations and another value, γ_U , at unattended locations, with $\gamma_A > \gamma_U$. The solution to Equation 9 is

$$v(t) = \theta \left(\frac{\Delta_I^\rho}{\Delta_I^\rho + I_{in}} \right) \left\{ 1 - \exp \left[- \gamma_i (\Delta_I^\rho + I_{in}) \int_0^t \mu(s) ds \right] \right\}. \quad (10)$$

³ In applications, I_{in} is often written as a function of the Naka-Rushton exponent, as $I_{0.5}^\rho$. When written in this way, the constant $I_{0.5}$ is known as the semisaturation constant—so designated because it is the value at which the function equals 0.5, which is half its theoretical maximum of 1.0. In our work, we prefer the representation of Equation 6 because it makes fewer assumptions about how the exponent and the inhibition term covary across experimental conditions. To model the data of Gould et al. (2007) we have assumed that ρ varies with uncertainty, whereas the inhibitory term remains fixed. Allowing the inhibition term to vary as $I_{0.5}^\rho$ leads to a family of functions with a fixed point at $I_{0.5}$ (i.e., $r(I_{0.5}) = 0.5$ for all ρ). Under these circumstances, increasing ρ makes low-contrast stimuli less detectable and high-contrast stimuli more detectable, in a way that does not fit the data.

The effect of differences in gain on the formation of the VSTM trace is shown in Figure 8a. When stimuli are at attended locations the VSTM trace grows rapidly to its asymptote, $v(\infty)$; when they are at unattended locations, the trace grows more slowly. Providing stimulus duration is long enough for the trace formation process to run to completion, however, the asymptotes for attended and unattended stimuli are the same.

Orienting models are, in essence, switching models. The defining characteristic of orienting models is that they assume that a limited-capacity central mechanism must first be switched to, or aligned with, the location of the stimulus before stimulus processing can be completed. The most influential expression of this idea has been Posner's attentional spotlight formulation (Posner, 1980), but the idea has its origins in Broadbent's (1958) filter theory. Like gain models, orienting models have been extremely influential in the literature and have been discussed by many authors in a variety of settings.

We represent orienting models in a similar way to the way we modeled gain, by introducing the idea of an *attention gate* or

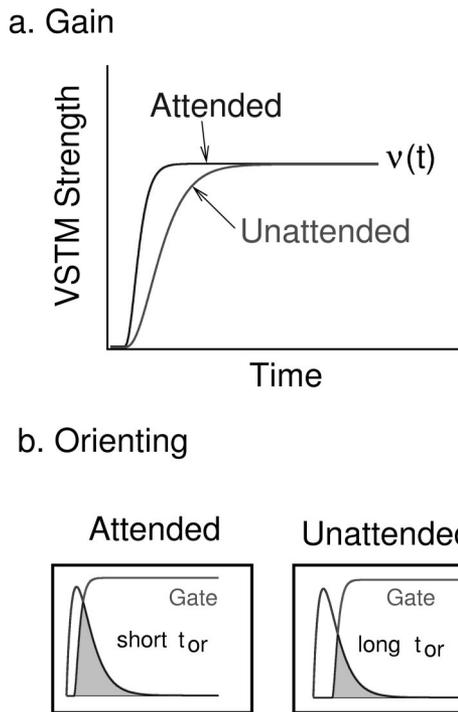


Figure 8. Gain and orienting models. **a.** Gain model. The visual short-term memory (VSTM) trace forms rapidly for stimuli at attended locations and slowly for stimuli at unattended locations. If stimulus persistence is sufficient for the trace-formation process to run to completion, the traces for attended and unattended stimuli grow to the same asymptote. If stimulus persistence is short, attended stimuli have an advantage because of their higher rate of trace growth. **b.** Orienting model. VSTM trace formation begins after the opening of an attention gate or window. The effective sensory response function equals the proportion of the sensory response falling within the window. When orienting time (t_{or}) is short, most of the sensory response falls within the window; when orienting time is long, much of the sensory response falls outside the window. The effect of missing the beginning of the stimulus is greater when stimuli are masked and stimulus persistence is short.

attention window that opens and transfers the stimulus into VSTM (Reeves & Sperling, 1986). We again denote the attention gate by $\gamma(t)$, but we now write it as a function of time. The defining property of orienting models is that the process of VSTM trace formation begins only once the attention gate opens. The time at which this occurs is the *orienting time*, denoted t_i . Minimally, again, we assume that the orienting time has one value, t_A , for attended stimuli and another value, t_U , for unattended stimuli, with $t_U > t_A$. That is, relative to attended stimuli, the opening of the attentional gate for unattended stimuli is delayed. The VSTM growth equation for an orienting model is

$$\frac{dv}{dt} = \gamma(t - t_i) \{ \Delta_f^p \mu(t) [\theta - v(t)] - I_{in} \mu(t) v(t) \}, t_i \in \{t_A, t_U\}. \quad (11)$$

The solution of this equation is

$$v(t) = \theta \left(\frac{\Delta_f^p}{\Delta_f^p + I_{in}} \right) \left\{ 1 - \exp \left[- (\Delta_f^p + I_{in}) \int_0^t \gamma(s - t_i) \mu(s) ds \right] \right\}. \quad (12)$$

The effect of differences in orienting time for attended and unattended stimuli is depicted in Figure 8b. In orienting models, the effective stimulus information depends on the product of functions $\gamma(s - t_i) \mu(s)$. Because the VSTM trace begins to form only when a stimulus is present and when the gate is open, the area under the product determines how closely the VSTM trace approaches its theoretical maximum. As Figure 8b shows, a delay in opening the attention gate results in the onset of the stimulus being missed, reducing the effective stimulus information driving the VSTM trace. The ultimate effect of this on performance depends on the persistence of the stimulus, which, in turn, depends on whether or not the stimulus is backwardly masked.

Stimulus energy and stimulus information. There is one further elaboration needed to complete the VSTM model. Our previous discussion of the studies of Smith, Ratcliff, and Wolfgang (2004) and Gould et al. (2007) stressed the similarities between the two data sets. However, as we noted previously and discuss in detail later, there is one critical dimension on which they differ. This is in the shapes of the RT distributions—in particular, in the way the leading edges of the RT distributions change with stimulus contrast. In most RT tasks, most of the change in RT that occurs with changes in stimulus intensity or discriminability occurs in the tails, or the upper quantiles, of the distributions (Ratcliff & Smith, 2004; Ratcliff, Thapar, Smith, & McKoon, 2005). The tails of the distributions are determined by the slowest responses in each experimental condition. The leading edges of the distributions, which are determined by the fastest responses in each condition, show comparatively little change. This pattern is found in a wide variety of tasks, ranging from low-level perceptual tasks like brightness discrimination to higher-level cognitive tasks like recognition memory. This same pattern is also found in the data of Smith, Ratcliff, and Wolfgang (2004), where the change in the leading edge between the lowest and highest stimulus contrasts is less than 50 ms. In contrast, the leading edges of the distribution in the data of Gould et al. (2007) change by around 100 ms. A significant

theoretical challenge for us has been to provide a model that can simultaneously accommodate both of these patterns of data.

In our theory, the leading edge of the RT distribution depends on the rate of VSTM trace formation. Delaying the entry of stimuli into VSTM delays responding and shifts the RT distribution to the right. If this assumption is correct, it implies that the rate of VSTM growth in the study of Smith, Ratcliff, and Wolfgang (2004) was largely independent of stimulus contrast, whereas in the study of Gould et al. (2007), VSTM growth rate was highly dependent on it. What is it about the two experimental tasks that would lead to these different patterns of growth? The answer, clearly, is that the stimuli used by Smith, Ratcliff, and Wolfgang were presented on top of suprathreshold contrast luminance pedestals, whereas the stimuli of Gould et al. were presented directly against a uniform field. A natural hypothesis, then, is that the rate of VSTM growth depends on the total power or energy in the stimulus compound. By “stimulus compound,” we mean the combination of grating and pedestal. (We remind the reader of the standard engineering definitions of power and energy at this point [McGillem & Cooper, 1991, p. 25]. The power in a waveform is equal to the square of its amplitude; the energy in a waveform is equal to the sum or integral of its power over time.)

For a zero-mean stimulus like a grating patch, the power in the compound equals the sum of the power in the pedestal and the power in the patch. Denoting the power in the patch and the power in the pedestal by Δ_I^2 and I_p^2 , respectively, the power in the compound equals $I_p^2 + \Delta_I^2$. Expressed in amplitude rather than power units, the intensities of the patch and the compound equal the square roots of their powers, that is, Δ_I and $\sqrt{\Delta_I^2 + I_p^2}$, respectively. The latter quantity is closely related to the root-mean-square measure of stimulus contrast, which has been advocated as a measure of contrast in natural scenes and random dot images by Moulden, Kingdom, and Gatley (1990) and others.

To represent the way in which the rate of VSTM growth depends on the power in the pedestal and the patch, we use a somewhat more general formulation of stimulus power, in which the intensities in the pedestal and patch are subject to separate, nonlinear transduction before being combined. We denote the square root of the summed powers of the transduced components in the compound by I_C :

$$I_C = \sqrt{\Delta_I^{2\rho} + I_p^{2\rho}}.$$

The content of this equation is as follows. Psychophysically, the pedestal is a small region of uniform luminance, which acts as a background for the patch in exactly the same way as the surrounding uniform field does for the pedestal. The equation states that the transduction of the patch relative to the pedestal is described by the same power law, with exponent ρ , as is the transduction of the pedestal relative to the uniform field. The quantity I_C is the square root of the summed power of the transduced contrasts of the patch and pedestal, expressed in amplitude units. When no pedestal is present and $I_p = 0$, I_C reduces to Δ_I^ρ , as it should. There is an obvious, alternative model, which assumes a single, nonlinear transduction operation is applied to the power in the compound, but the resulting representation of the power in the compound, $\sqrt{(\Delta_I^2 + I_p^2)^\rho}$, does not fit our data well. Further discussion of the relationship between different measures of stimulus intensity can be found in Appendix B.

In the study of Smith, Ratcliff, and Wolfgang (2004), the power in the stimulus compound was dominated by the pedestal. Direct calculation (cf. Appendix B) shows that at the highest stimulus contrast the power in the pedestal was around nine times the power in the patch. Consequently, the rate of VSTM growth should have been largely independent of stimulus contrast. This differs from the study by Gould et al. (2007), in which no pedestal was used. In this situation, the stimulus power was contained entirely in the patch, so the rate of VSTM growth should have increased systematically with stimulus contrast. (We assume that the fiducial cross is not part of the stimulus compound and does not affect the rate of VSTM formation in the same way as does the pedestal.) Assuming a gain model, the shunting equation for a pedestal Gabor is

$$\frac{dv}{dt} = \gamma_i \{ I_C \mu(t) [\theta - v(t)] - I_{in} \mu(t) v(t) \}, \quad (13)$$

the solution of which is

$$v(t) = \theta \left(\frac{I_C}{I_C + I_{in}} \right) \left\{ 1 - \exp \left[- \gamma_i (I_C + I_{in}) \int_0^t \mu(s) ds \right] \right\}. \quad (14)$$

The equation for an orienting model is obtained in a similar way, by an obvious modification of Equation 11.

Equation 14 describes the growth of a representation of the stimulus compound in VSTM. Unlike Equation 8, the trace in Equation 14 comprises the power in both the pedestal and the patch, but only the latter carries information used to make a perceptual decision. If we wish to interpret Equation 13 as describing the growth of discriminative information in VSTM over time, we need to modify the asymptote in Equation 14 so that it depends only on the power in the patch and not on the power in the compound. An equation that possesses this property is

$$v(t) = \theta \left(\frac{\Delta_I^\rho}{\Delta_I^\rho + I_{in}} \right) \left\{ 1 - \exp \left[- \gamma_i (I_C + I_{in}) \int_0^t \mu(s) ds \right] \right\}. \quad (15)$$

In this equation, asymptotic VSTM trace strength has a Naka-Rushton-like functional dependency on the power in the patch, whereas the rate of trace growth depends on the power in the compound. When a pedestal is present, the rate of growth depends predominantly on the power in the pedestal; when no pedestal is present, it depends solely on the power in the patch. Equation 15 can be obtained as the solution of the shunting equation

$$\frac{dv}{dt} = \gamma_i (I_C + I_{in}) \{ r(\Delta_I) \mu(t) [\theta - v(t)] - [1 - r(\Delta_I)] \mu(t) v(t) \}, \quad (16)$$

where $r(\Delta_I)$ is the transduced stimulus increment of Equation 6. This equation can be viewed as expressing a form of fixed-activation property, in which the sum of the excitatory and inhibitory effects produced by a stimulus is constant. Further discussion of such equations can be found in Appendix A.

The decoupling of asymptotic trace strength from the rate of trace growth in Equation 15 is needed to account for the differences in the shapes of the RT distributions in Smith, Ratcliff, and Wolfgang (2004) and in Gould et al. (2007). However, this decoupling comes at the cost of introducing an additional, contrast-

dependent rate term into the differential equation, Equation 16. To obtain the additional generality in a theoretically principled way, we need to consider a process of VSTM trace formation that is somewhat more complex than the simple, single-channel process of Equation 16. We describe a neurally plausible computational model of trace formation that can provide this generality later in the article.

Decision Making

We assume that $v(t)$, the VSTM trace, is subject to moment-by-moment perturbation by noise. The noise may be inherent in the trace itself or may be injected from elsewhere in the processing system. Successive samples of the noisy VSTM trace are accumulated over time to a response criterion to make a decision. We assume that the noise is broad-spectrum Gaussian noise, or white noise. This leads to a representation of the accumulating information as a diffusion process, which we denote by $X(t)$. The process of information accumulation can be described by a stochastic differential equation (SDE) of the form

$$dX(t) = [v(t) - \lambda X(t)]dt + \sigma(t)dW(t), \quad (17)$$

(Brown, Ratcliff, & Smith, 2006; Smith, 1995, 2000b). In this equation, the term $W(t)$ is a white noise process, which is the (formal) derivative of the Brownian motion, or Wiener, diffusion process. SDEs are usually written in differential form, as in Equation 17, rather than in the more familiar form involving derivatives (e.g., Equation 4), because the highly irregular sample paths of diffusion processes mean that they do not possess derivatives in the same way as smooth functions do.

The quantity $dX(t)$ on the left of Equation 17 can be interpreted as the change in decision stage activation occurring during a small time interval, dt . This change is the sum of two parts, a deterministic part and a stochastic part. The deterministic part, or drift, is equal to $[v(t) - \lambda X(t)]$; the stochastic part is equal to $\sigma(t)dW(t)$. In any interval, dt , the stochastic part is normally distributed with mean zero and variance $\sigma^2(t)dt$ and is independent of the change in any other, nonoverlapping interval. The drift itself is also a sum of two terms: a time-varying function, $v(t)$, which depends on the strength of the VSTM trace, and a passive, state-dependent decay term, $\lambda X(t)$. When $\lambda = 0$, the process $X(t)$ is a Brownian motion, or Wiener, diffusion process (Ratcliff, 1978, 1980). When $\lambda > 0$, the process $X(t)$ is an Ornstein-Uhlenbeck (OU) diffusion process (Busemeyer & Townsend, 1993; Diederich, 1997; Heath, 1992; Smith, 1995, 2000b; Usher & McClelland, 2001). The Wiener process can be viewed as a perfect stochastic integrator, which accumulates noisy information over time without loss or decay. The OU process can, in contrast, be viewed as an imperfect, or “leaky” integrator, which loses information at a rate that depends on the decay constant, λ .

We consider two decision models in this article, both based on diffusion processes. One is the (Wiener) diffusion model of Ratcliff (1978, 1988). The other is a model that Ratcliff and Smith (2004) called the “leaky accumulator” and Ratcliff, Hasegawa, Hasegawa, Smith, and Segraves (2007) subsequently termed the “dual diffusion” model. Smith (2000b) proposed this model as a way to combine the properties of accumulator models—like that of Vickers (1970, 1979; Smith & Vickers, 1988) and others—and random walk or diffusion models in a simple way. Ratcliff

and Smith (2004) subsequently developed the model theoretically and Ratcliff et al. (2007) applied it to neural firing data recorded from the superior colliculi of monkeys performing an eye-movement decision task. In that article, we showed the model could simultaneously account for the observed choice probabilities and distributions of RT and for the patterns of neural firing in build-up cells in the colliculus in the interval preceding the response.

We chose to investigate two decision models in this article to ensure that our inferences about attention and VSTM were robust to changes in our assumptions about the decision mechanism. As we show, the single and dual diffusion models produced only very minor differences in fit, consistent with this expectation. The two decision models we considered, the single, Wiener, diffusion model and the dual diffusion model, were chosen because they have been shown to provide a satisfactory account of performance in a wide variety of simple decision tasks. We did not consider models that have been shown to be unsatisfactory on other grounds. The Usher and McClelland (2001) model would likely provide a similarly good account of performance on the tasks we consider here, but it is relatively less tractable analytically.

The Wiener diffusion model. Figure 9 depicts the main elements of the Wiener diffusion model. The model in the figure is the same as the diffusion model of Ratcliff (1978), except that both the drift and diffusion coefficient depend on time. Diffusion processes with time-varying coefficients are called *time inhomogeneous* processes in the literature. The time inhomogeneity in the model arises because the decision process is driven by the VSTM trace, $v(t)$, which depends on time. We obtained predicted RT distributions and choice probabilities for the model using numerical integral equation methods described by Smith (2000b). These methods are sufficiently general to allow one to obtain predictions for a wide class of diffusions with time-varying coefficients. Ratcliff and Smith (2004, Appendix) give a summary of these methods.

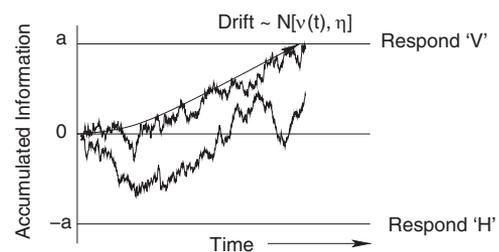


Figure 9. Wiener diffusion model. Decision making is modeled as diffusion on a line between absorbing boundaries located at a and $-a$. The boundaries represent the criteria for vertical (V) and horizontal (H) responses. The rate of information accumulation, or drift, is proportional to the visual short-term memory trace, which makes the process time inhomogeneous. The drift is normally distributed, with a mean proportional to the visual short-term memory trace strength, $v(t)$, and an asymptotic standard deviation, η , that depends on the trial-to-trial quality of the stimulus encoding. The irregular paths represent the accumulating stimulus information on two different experimental trials. The response that is made depends on the first boundary (upper or lower) crossed by the process; the response time depends on the time at which the first boundary crossing occurs.

The model represents decision making as diffusion between a pair of absorbing boundaries located at a_1 and $-a_2$. The boundaries represent the evidence criteria used to make the two responses. A process starting at $X(0) = 0$ drifts in either an upward or a downward direction, depending on the stimulus, until it reaches one or other of the absorbing boundaries. The first boundary crossed by the process determines the response. If the boundary is the upper boundary the observer responds (say) “vertical”; if it is the lower boundary the observer responds “horizontal.”

As shown in Figure 5, there are two distinct sources of noise in the model. The first is diffusion noise, whose cumulative effects are represented by the irregular sample paths in Figure 9. Such noise reflects moment-by-moment perturbations in the strength of the VSTM trace, as we discussed previously. The second is trial-to-trial variation in the mean of the VSTM trace. We model these effects as variations in the drift term in Equation 17. In our presentation of the VSTM model, we described the process of VSTM trace formation in purely deterministic terms, using ordinary differential equations. In practice, however, we assume that the encoded stimulus representation is a random variable, just as in signal-detection theory. Psychologically, this variability represents trial-to-trial differences in the quality or efficiency of the encoding by the visual system of nominally equivalent stimuli. Following Ratcliff (1978) we assume that drift is normally distributed with a mean equal to the VSTM trace strength, $\nu(t)$, and a standard deviation of η . Specifically, we define η to be the standard deviation of the asymptotic trace strength $\nu(\infty)$.

To obtain a well-behaved model with time-varying drift, we need to assume the diffusion coefficient, $\sigma(t)$ in Equation 17, also varies with time. This contrasts with Ratcliff’s (1978) model, in which the diffusion coefficient (denoted s) is constant. Had we assumed a constant diffusion coefficient, the early part of the decision process would have been dominated by noise. This would have resulted in a large number of fast errors and substantially degraded performance. To avoid this, we assumed that the diffusion coefficient grows in proportion to the drift, that is, $\sigma(t) \propto \nu(t)$. To obtain well-behaved performance across a range of stimulus contrasts, we needed to assume that the diffusion coefficient grew to a fixed, constant value while the drift grew to the asymptotic trace strength, $\nu(\infty)$, whose value depended on stimulus contrast. We also assumed the diffusion coefficient continued to grow after stimulus offset until it reached asymptote. This ensured that the asymptotic value of the diffusion coefficient was unaffected by the use of masks. In Ratcliff’s model, the diffusion coefficient acts as a scaling constant for the other parameters of the model and is arbitrarily set to $s = 0.1$. We followed this convention here and assumed that $\sigma(t)$ grew to an asymptote of 0.1. Physiologically, the growth of $\sigma(t)$ to a fixed constant might reflect a stimulus-dependent release from inhibition, as Ratcliff et al. (2007) have argued.

To complete the decision model, we assume that the predicted RT is the sum of the decision time and the time required for other processes not specified in the model. We follow Ratcliff (1978) and assume that this time is rectangularly distributed with mean T_{er} and range s_r . In Ratcliff’s model, T_{er} represents all components of RT, apart from the decision time itself. This includes the central and peripheral components of stimulus encoding and response execution. In our model, T_{er} includes only components of RT not

otherwise identified in the sensory encoding, VSTM, and decision models.

Ratcliff’s (1978) diffusion model includes one additional source of variability not included in the models we consider here. This is variability in the starting point of the diffusion process, denoted z in Ratcliff’s model. Variability in starting point allows the model to predict fast errors, which are a feature of many experimental tasks, as we have just discussed. In the present setting, however, in which stimuli were near threshold and accuracy of responding was stressed, errors were uniformly slower than correct responses. Slow errors arise in our model, as in Ratcliff’s, as the result of variability in drift across experimental trials (see Ratcliff & Smith, 2004, for a discussion of this property). Because there is little role for starting-point variability in the tasks we consider here, we omitted it from the model for the sake of computational efficiency.

The dual diffusion model. Figure 10 depicts the main elements of the dual diffusion model. Like the single, Wiener, diffusion model, the dual diffusion model we considered here was time inhomogeneous. The dual diffusion model represents the decision process as a race between a pair of independent, parallel diffusions. One process, $X_V(t)$, represents evidence for a vertical response and the other, $X_H(t)$, represents evidence for a horizontal response. The diffusions take place on a pair of positive half-lines, $X_V(t) \geq 0$ and $X_H(t) \geq 0$, between an upper, absorbing boundary and a lower, reflecting boundary. The absorbing boundaries are decision criteria, just as in the single diffusion model. The reflecting boundaries can be thought of physiologically as hyperpolariz-

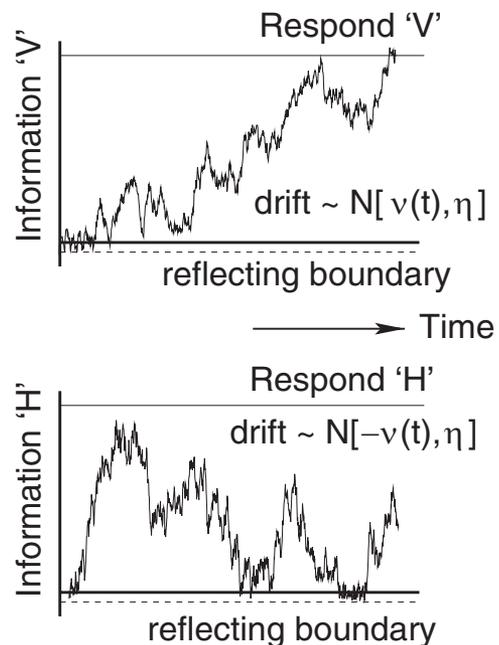


Figure 10. Dual diffusion model. Decision making is modeled as a race between two diffusing evidence totals on the positive half line, between reflecting and absorbing boundaries. The absorbing boundaries represent evidence criteria for the two responses. The response that is made depends on the first process to reach its criterion, and the response time depends on the time it takes for this to occur. The drifts in the two processes are mirror images (additive inverses) of each other.

ing limits, below which the processes cannot be driven (cf. Usher & McClelland, 2001). Diederich (1995) previously used a single diffusion process between an absorbing and a reflecting boundary to model simple RT. In the dual diffusion model, the response that is made depends on the first process to reach its criterion, and the decision time is the time taken to reach it. We obtained predicted choice probabilities and RT distributions for the model using the matrix method of Diederich and Busemeyer (2003). This method approximates the continuous-time, continuous-state diffusion process by a finite-state Markov chain on a discrete state space, much like the models proposed by Pike (1966). The method is sufficiently flexible that it can be used with time-inhomogeneous models with both absorbing and reflecting boundaries.

Unlike the single process model, the accumulation processes in the dual diffusion model are OU processes. As we noted previously, the Wiener diffusion is a perfect stochastic integrator, whereas the OU process is a stochastic integrator with decay. The OU process can also be viewed as a bounded accumulation process. Whereas the average accumulated information in the Wiener process grows linearly with time, in the OU process, it grows to a bound that depends on the ratio of the drift and the decay (Smith, 1995). Several authors have argued that the OU process should be preferred on the grounds of biological plausibility, because biological accumulation processes must be inherently bounded (Smith, 1995; Usher & McClelland, 2001).

As shown in Figure 10, we assume the drifts in the two processes are mirror images of one another. That is, if the drift in the vertical accumulator is ξ , the drift in the horizontal accumulator is $-\xi$. This differs from what was assumed by Ratcliff and Smith (2004) who followed Usher and McClelland (2001) in assuming the drifts in the two accumulators sum to a constant. For the data sets we considered here, we found that mirror-image drifts provided better fits.

We can justify this assumption theoretically by considering the computations underlying a decision. We assume that the information encoded in the developing VSTM trace, $\nu(t)$, is simultaneously and continuously matched against mental representations of the two stimulus alternatives for the task. We denote the instantaneous values of the goodness of match of the trace to the representations of the vertical and horizontal stimuli by $\xi_v(t)$ and $\xi_H(t)$, respectively. Positive values of the difference $\xi_v(t) - \xi_H(t)$ are evidence for a vertical response, whereas positive values of the complementary difference $\xi_H(t) - \xi_v(t)$ are evidence for a horizontal response. We assume it is these two differences that provide the drifts to the two accumulators.

We also assume that both accumulators start at a value of $X(0) = 0$, at, or just above, the lower reflecting boundary. This means that the process in one accumulator drifts toward the upper boundary, whereas the other tends to remain in the vicinity of its starting point because it is prevented from drifting in the negative direction by the reflecting boundary. Although the drift in the second accumulator is negative, the process does not remain at the lower boundary uniformly; rather, because of diffusive variability, it makes random excursions into the positive half-space, $X(t) > 0$. Some of these excursions tend to take it away from the lower boundary, whereas the drift tends to take it back to the lower boundary. The sample path of a typical process bumps along the lower boundary, making random excursions into the positive half-space and then returning to the lower boundary. On some trials,

however, the cumulative effects of diffusive variability are sufficient to take the process above its upper boundary, and an error results.

The model also has a second mechanism for producing errors. Because drift is normally distributed, on some trials on which a vertical stimulus is presented, the drift in the vertical accumulator is negative and drift in the horizontal accumulator is positive; that is, the drifts in the accumulators are in the wrong directions.⁴ On such trials, the most probable outcome is an incorrect response. This mechanism is the counterpart of the between-trial variability in drift in the single diffusion model. We also assume, as in that model, that the time for other processes is rectangularly distributed with a mean of T_{cr} and a range s_r . The two decision models are therefore exact analogues of one another in the components of processing they assume and the parameters used to describe them. The only difference is that the dual diffusion model has one additional parameter, the OU decay constant, λ , that has no counterpart in the single diffusion model.

Model Evaluation

Our approach to model evaluation closely follows that of Ratcliff and Smith (2004). The data for which we wish to account are the choice probabilities (response accuracy) and the distributions of RT for correct responses and errors. We summarize the information in RT distributions using the distribution quantiles. In our evaluation, we used five quantiles: the .1, .3, .5, .7, and .9 quantiles. The .1 quantile describes the distribution's leading edge, the .5 quantile describes its central tendency (the median), and the .9 quantile describes its tail. We chose to use five quantiles because this number suffices to characterize the shape of the distribution while being relatively insensitive to outliers.

To evaluate our models we fitted them to quantile-averaged data. We averaged the five distribution quantiles across observers to obtain the quantiles of the group RT distribution. We did this for the distributions of correct responses and errors for each of the five levels of stimulus contrast in each experimental condition. We also averaged the choice probabilities across observers. For the data sets of Smith, Ratcliff, and Wolfgang (2004) and Gould et al. (2007), this yielded a total of 40 RT distributions (20 distributions of correct responses and 20 distributions of errors) and 20 choice probabilities. Ratcliff and colleagues have shown repeatedly that the parameter estimates obtained by fitting the diffusion model to quantile-averaged data agree closely with the averages of parameter estimates obtained by fitting the model to individual subject data (e.g., Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2003, 2004; Thapar, Ratcliff, & McKoon, 2003). Smith, Ratcliff, and Wolfgang (2004) found the same thing in a precursor to the present theory. There is thus no evidence that

⁴ We abuse terminology here for ease of exposition. Strictly, the drift in the diffusion process in Equation 17 is the sum of a stimulus-dependent, excitatory term, $\nu(t)$, and a state-dependent decay term, $-\lambda X(t)$. When we refer to "trial-to-trial variability in drift," we actually mean trial-to-trial variability in $\nu(t)$. For the Wiener process, the drift and the excitatory component of drift are the same; for the OU process, the drift comprises both excitation and decay. For the OU process, we assume that the excitatory component varies normally from trial to trial, just as in the Wiener process, whereas the decay remains constant.

the picture is materially altered by fitting models to group data. In the present setting, in which we fitted several hundred different model variants to our data, this restriction was necessary to make the overall task of model evaluation manageable.

Our theory is sufficiently flexible that it allowed us to compare two different attentional models and two different decision models. We thus have a total of four model classes in total: single diffusion, gain; dual diffusion, gain; single diffusion, orienting; and dual diffusion, orienting. We fitted these models to the quantile-averaged data by minimizing the likelihood-ratio chi-square statistic (G^2),

$$G^2 = 2 \sum_{i=1}^{20} n_i \sum_{j=1}^{12} p_{ij} \log \frac{p_{ij}}{\pi_{ij}},$$

using the Matlab implementation of the Nelder-Mead Simplex algorithm (fminsearch). In this equation, p_{ij} and π_{ij} are, respectively, the predicted and observed probabilities (proportions) in the bins bounded by the quantiles. The inner summation over j extends over the 12 bins formed by each pair of joint distributions of correct responses and errors. (There are five quantiles per distribution, resulting in six bins per distribution, or 12 bins in total for each distribution pair.) The outer summation i extends over the five stimulus contrasts in each of the four experimental conditions (20 distribution pairs in all). The quantity n_i is the number of experimental trials in each condition. For the data of Smith, Ratcliff, and Wolfgang (2004) we set $n_i = 400$ and for the data of Gould et al. (2007) we set $n_i = 360$. This is consistent with our interpretation of the quantile-averaged distributions as the performance of an “average observer.” Because G^2 computed on the joint distributions depends on the relative proportions of correct responses and errors, it characterizes goodness-of-fit to the distribution shapes and the choice probabilities simultaneously.

Effect of Backward Masks (Smith, Ratcliff, & Wolfgang, 2004)

Before describing the fits of the models to experimental data, we try to give the reader an intuition of how they predict mask-dependent cuing effects like those in Figures 2 and 7. Figure 11 shows how gain models predict these effects. In our theory, backward masks reduce the area under the sensory response function, $\mu(t)$. This area determines how closely the final VSTM trace approaches its theoretical maximum, $v(\infty)$. When stimuli are unmasked, the area under $\mu(t)$ is large and the final trace is close to $v(\infty)$. When stimuli are masked, the area is small and the maximum is not attained.

Gain models assume that attention determines the rate at which the VSTM trace is formed. When stimuli are attended, the trace grows at rate γ_A ; when they are unattended the trace grows at rate γ_U , with $\gamma_A > \gamma_U$. Providing stimulus persistence is sufficiently long, however—that is, providing the area under $\mu(t)$ is sufficiently large—the two traces grow to the same asymptotic value, $v(\infty)$. The rate at which information accumulates in the decision stage, which is equal to the drift of the diffusion process, is proportional to the VSTM trace strength, $v(t)$. Because the final trace strength is the same for attended and unattended stimuli, the asymptotic rates of information accumulation for attended and

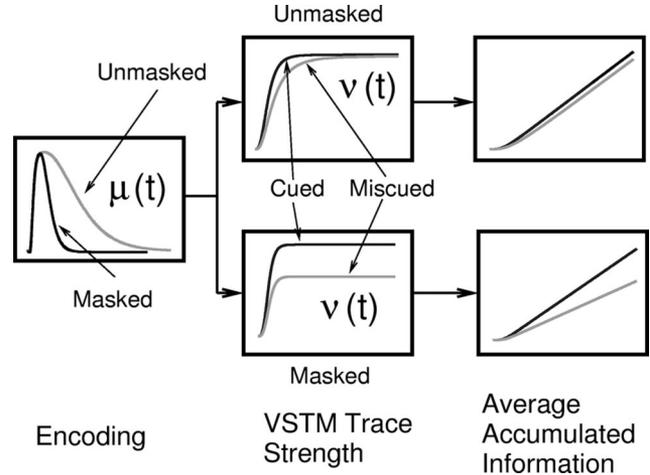


Figure 11. The mask-dependent cuing effect for gain models. When stimuli are unmasked and stimulus persistence is long, the visual short-term memory (VSTM) traces for cued and miscued stimuli grow at different rates to the same asymptote. The average accumulated stimulus information in the decision stage increases asymptotically at the same rate for cued and miscued stimuli. This results in response time difference but no accuracy difference. When stimuli are masked and stimulus persistence is short, cued stimuli have an advantage because they produce faster VSTM trace growth. This means that more of the trace has formed by the time the sensory response decays, resulting in different asymptotic trace strengths for cued and miscued stimuli. The average accumulated stimulus information in the decision stage increases asymptotically more rapidly for cued than for miscued stimuli. This results in a large response time difference and an accuracy difference.

unattended stimuli are equal. Under these circumstances, the curves that describe the average accumulated information as a function of time are roughly parallel to one another, as shown on the right side of Figure 11. This results in an RT difference for attended and unattended stimuli but little or no sensitivity or accuracy difference.

For masked stimuli the situation is different. The smaller area under the sensory response function means the VSTM trace has not reached its maximum, $v(\infty)$, when the trace stops growing. In this situation, attended stimuli have an advantage, because they produce higher rates of VSTM growth. This means that more of the trace has formed before the stimulus is suppressed by the mask. The final trace strength for attended stimuli exceeds that for unattended stimuli and, consequently, the decision process accumulates information more rapidly for attended than for unattended stimuli. This means that the curve that describes average accumulated information as a function of time has a steeper slope for attended than for unattended stimuli, as shown in Figure 11. This results in both an RT and an accuracy difference. It is this property that allows the model simultaneously to predict the mask-dependent cuing effect in accuracy and the unconditional Posner effect in RT shown in Figure 2.

Orienting models, like the one in Equation 11, also have a mechanism for producing the mask-dependent cuing effect. Orienting models assume that, relative to attended stimuli, the opening of the attentional gate for unattended stimuli is delayed. Specifically, we assume that the opening time of the gate is t_A for

attended stimuli and t_U for unattended stimuli, with $t_U > t_A$. A delay in opening the gate causes the system to miss the beginning of the stimulus, reducing the effective stimulus information, as shown in Figure 8b. The effect of a delay is greater when stimuli are masked, because the proportional reduction in effective stimulus information is greater when stimuli are masked, as Figure 8b shows. Thus in orienting models, differences in the time at which the attention gate opens interact with the differential persistence of masked and unmasked stimuli to predict the mask-dependent cuing effect.

Figure 12 shows the fit of one of the four models, the single-diffusion, gain model, to the data from the masking study of Smith, Ratcliff, and Wolfgang (2004). The data and fitted values are shown as a quantile probability plot (Ratcliff & Smith, 2004; Ratcliff & Tuerlinckx, 2002). In such plots, the quantiles of the RT distributions are plotted against the choice probabilities for correct responses and errors in each stimulus condition. These plots provide a compact way of representing how the RT distributions, the ordering of mean RTs for correct responses and errors, and response accuracy all vary with the intensity or discriminability of the stimulus.

The five lines in each panel of the figure are, in ascending order, the .1, .3, .5, .7, and .9 quantiles of the predicted RT distributions; the symbols are the quantiles of the empirical distributions. For a

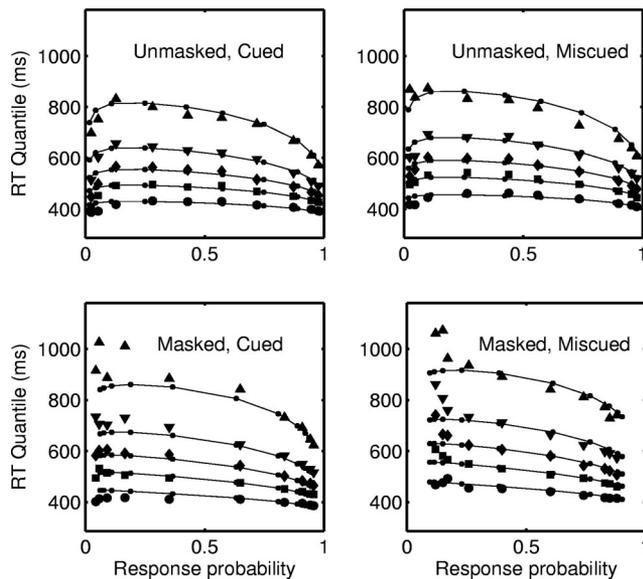


Figure 12. Fit of the single-diffusion, attention-gain model to the response time (RT) distributions and choice probabilities for the Smith, Ratcliff, and Wolfgang (2004) study. The large symbols are the experimental data, and the continuous curves and small circles are the fitted values. The five lines in each panel are, in ascending order, the .1, .3, .5, .7, and .9 quantiles of the predicted RT distributions. The corresponding empirical quantiles are shown as circles, squares, diamonds, inverted triangles, and upright triangles, respectively. The plot is parameterized by stimulus contrast. The outermost pair of distributions in each plot (response probabilities closest to 0 and 1.0) are distributions of errors and correct responses for the highest contrast stimulus; the innermost pair of distributions (response probabilities closest to 0.5) are distributions of correct responses and errors for the lowest contrast stimulus.

given level of stimulus contrast, i , the quantiles of the distribution of RTs for correct responses are plotted against the probability of a correct response (say q_i), and the quantiles of the distribution of RTs for errors are plotted against the probability of an error response ($1 - q_i$). The five distributions on the right of the .5 point on the x axis are distributions of correct responses and the five on the left of the .5 point are the corresponding distributions for errors. The two outermost distributions on the plot are distributions of correct responses and errors for the easiest stimulus condition (the highest level of contrast), and the two innermost distributions are for the most difficult condition.

All the main features of performance on the task are represented in this plot. First, the RT distributions show the same unimodal, positively skewed form that is found with RTs to suprathreshold stimuli. This is shown by the spacing of the quantiles: The tail quantiles are spaced more widely than are the quantiles at the leading edge. Second, both the mean and standard deviation of the RT distributions increase as stimulus contrast is reduced. (Indeed, the relationship between them is approximately linear, as discussed by Wagenmakers & Brown, 2007.) Third, most of the change in RT with changing contrast occurs in the distribution tails (the .7 and .9 quantiles); the leading edges show comparatively little change. Fourth, error RTs are slower than correct RTs. This appears as an asymmetry of the plot across its vertical midline. If distributions of correct and error RTs were the same, the plot would be left-right symmetrical. In Figure 12, the quantile RTs for errors are longer than the corresponding RTs for correct responses at each level of contrast. This ordering is typically found when discrimination is difficult and accuracy is stressed (Luce, 1986).

Figure 12 shows that the model captures all of the main features of the data. Within each condition, it correctly characterizes the way the RT distributions and choice probabilities change with stimulus contrast and the ordering of RTs for correct responses and errors. Between conditions, it correctly characterizes the joint effects of attentional cuing and backward masking. When stimuli are unmasked, cues affect RT but have no effect on accuracy. This appears in the quantile probability plot as a systematic upward shift in the quantiles for miscued stimuli relative to cued stimuli, with no change in the horizontal extent of the plot. When stimuli are masked, cues affect both RT and accuracy: There is a (large) upward shift in the quantiles for miscued stimuli and a reduction in the horizontal extent of the plot.

The largest discrepancies in the predicted and observed RTs are in the tail quantiles of the distributions of errors to high-contrast stimuli (i.e., the cluster of points at the upper left in each panel). This is not due to any shortcoming of the model but is simply a reflection of the fact that these quantiles are estimated empirically with very low accuracy. This is because of the sparseness of points in the tail and because distributions of error RTs for high-contrast stimuli are based on only a small number of trials. Consequently, confidence intervals for a quantile probability plot increase substantially at the upper left of the plot, as Ratcliff and Tuerlinckx (2002, Figure 3) have shown.

The fits of the four models are summarized in Table 1, which shows both G^2 values and values of the Bayesian Information Criterion (BIC) model-selection statistic. The latter is defined as

$$\text{BIC} = G^2 + m \log N,$$

Table 1
Fit Statistics (Smith, Ratcliff, & Wolfgang, 2004)

Model	G^2	df	BIC
Diffusion gain	166.2	207	283.0
Diffusion orienting	173.7	207	290.6
Dual diffusion gain	173.9	206	299.5
Dual diffusion orienting	171.8	206	297.6

Note. BIC = Bayesian Information Criterion.

where m is the number of free parameters in the model and N is the number of observations used in calculating G^2 . The BIC is a sample-size-dependent penalty statistic that penalizes models according to their number of free parameters. The best model from a competing set of candidate models is the one with the smallest BIC. As Table 1 shows, the quality of the fits of the four models was very similar. The best model, in both a G^2 and BIC sense, was the single diffusion/gain model. Manifestly, however, the similarities between models far outweigh the differences. The largest and smallest G^2 values are within 4% of each other, and the quantile probability plots for the four models are virtually indistinguishable.

We have avoided assigning p values to the goodness-of-fit statistics in Table 1 because G^2 statistics computed from averaged quantile probability data do not satisfy multinomial sampling assumptions and also because the interpretation of p values from chi-square-type tests is inherently problematic.⁵ When multinomial sampling assumptions are satisfied, the expected value of G^2 when the null hypothesis is true is equal to its degrees of freedom (the number of degrees of freedom in the data minus the number of free parameters in the model), which, for these data, is equal to 206 or 207. The fact that the G^2 values in the table are all substantially less than this is because quantile averaging reduces variability, leading to data which, in a statistical sense, are “underdispersed.” As our primary focus is on the models’ ability to capture the main qualitative features of the data, the presence of underdispersion does not concern us.

Table 2 lists the parameters that were used to fit the models and shows their estimated values. The free parameters fall into four groups: sensory encoding parameters, attention/VSTM parameters, decision process parameters, and parameters describing other (i.e., nonsensory and nondecision) processes. The encoding parameters describe the transduction of stimulus contrast and the rate at which stimuli pass through early sensory filters. Contrast transduction is described by three Naka-Rushton parameters: θ , ρ , and I_{in} . The parameter θ describes the mapping between contrast and VSTM trace strength (the latter expressed in units of the diffusion coefficient σ); ρ is the Naka-Rushton exponent, and I_{in} is the inhibition constant. The assumption that contrast transduction follows a Naka-Rushton function allows the effects of the 10 different stimulus contrasts, which would otherwise have had to be treated as free parameters, to be characterized using only three parameters.

The remaining encoding parameters were the three rate constants of the masked and unmasked sensory response functions. These were the onset rate, β_{on} , assumed to be the same for masked and unmasked stimuli, and the offset rates, $\beta_{off, m}$ and $\beta_{off, u}$, for masked and unmasked stimuli, respectively. We arbitrarily set n , the number of cascaded stages in $\mu(t)$, to three. In fitting the models to data we found, like other researchers who have used

linear filter models of sensory encoding, that the quality of the fits was largely insensitive to the number of stages in the filter (Smith, 1995). We therefore treated n as a fixed parameter in the fits reported here. The final sensory encoding parameter was I_p , which characterized the perceptual effect of the pedestal.

The VSTM process was parameterized differently for gain and orienting models. For the gain models, VSTM was parameterized by a pair of gain constants, γ_A and γ_U , which describe the rate of VSTM formation for attended and unattended stimuli. For orienting models, VSTM was again described by two free parameters, a gain constant, γ , assumed to be the same for attended and unattended stimuli, and an orienting parameter, t_{or} , which describes the delay in the opening of the attention gate for unattended stimuli relative to attended stimuli, that is, $t_{or} = t_U - t_A$. This parameterization assumes there is no loss of stimulus information on cued trials, because the gate opens at the cued location prior to stimulus onset. The assumption is reasonable psychologically, because the cue-target SOA of 140 ms was chosen to be optimal or near-optimal to produce large cuing effects with peripheral cues. We modeled the attention gate as a gamma function (Equation 1) composed of $n = 2$ cascaded stages and a large ($\beta = 250$) rate constant, both of which were treated as fixed parameters in the fits. With these values the gate approximated a step function, which opened virtually instantaneously on orienting to the stimulus.

The decision process parameters were the decision criterion, a , and the between-trial drift variability for masked and unmasked conditions, η_m and η_u . Initially, we allowed the criteria for masked and unmasked stimuli to differ because the experiment used a blocked design in which masked and unmasked stimuli were presented in different conditions. However, for none of the four model types was the fit better (according to the BIC) with separate criteria for masked and unmasked stimuli than when the criteria were the same. We have therefore reported fits for the single-criterion model in Table 1. We thought it likely that the trial-to-trial variability in stimulus encoding would differ for masked and unmasked stimuli because of the perceptual effects of the mask, so we allowed the drift variability parameters to differ between conditions. The estimates of drift variability confirmed this expectation, so we retained separate parameters for masked and unmasked stimuli in the models reported in Table 1. The dual diffusion models also had an additional parameter, the OU decay constant, λ . The remaining parameters were T_{er} and s_p , the mean and the range of the nondecisional component of RT, respectively. We

⁵ The philosophical objection to chi-square goodness-of-fit tests is the same as the objection to null hypothesis testing, namely, that they test point hypotheses, which have a probability measure of zero in the space of experimental outcomes. Such hypotheses are false with probability 1.0 and, so, must be rejected by a test of sufficient power. This is sometimes expressed by saying that the chi-square test is “too powerful.” This objection applies equally to the likelihood-ratio test, G^2 , and to the Pearson X^2 test. A further, technical objection to assigning p values to G^2 or X^2 computed on bins formed from quantiles was pointed out by Speckman and Rouder (2004). This is that the resulting sampling distribution is not a true multinomial, as required by these tests, because the boundaries of the bins are determined by the data. A somewhat paradoxical feature of this situation is that estimation performance appears to be better when the (incorrect) multinomial sampling model is used than when the correct sampling model is used, which is based on order statistics.

Table 2
Model Parameters (Smith, Ratcliff, & Wolfgang, 2004)

Parameter	Symbol	Diffusion gain	Diffusion orienting	Dual diffusion gain	Dual diffusion gain
Sensory response function					
Onset rate	β_{on}	57.8	121.1	49.5	45.9
Offset rate (masked)	$\beta_{off, m}$	182.2	127.7	108.4	53.7
Offset rate (unmasked)	$\beta_{off, u}$	10.1	10	10.2	10.2
Number of stages ^a	n	3	3	3	3
Naka-Rushton amplitude	θ	1.05	1.35	1.04	1.23
Naka-Rushton exponent	ρ	1.97	2.10	2.31	2.22
Naka-Rushton inhibition	I_{in}	0.008	0.005	0.002	0.003
Pedestal amplitude	I_p	0.88	0.94	0.99	1.06
Attention/VSTM (gain models)					
Gain (attended)	γ_A	15.0	—	14.6	—
Gain (unattended)	γ_U	9.5	—	10.0	—
Attention/VSTM (orienting models)					
Gain	γ	—	5.7	—	6.5
Orienting time	t_{or}	—	0.041	—	0.061
Decision process					
Decision criterion	a	0.096	0.091	0.05	0.048
Drift variability (masked)	η_m	0.22	0.33	0.29	0.35
Drift variability (unmasked)	η_u	0.35	0.54	0.5	0.56
Ornstein-Uhlenbeck decay	λ	—	—	6.0	6.1
Reflecting boundary ^a	r	—	—	-0.01	-0.01
Nondecision processes					
Mean nondecision time	T_{er}	0.281	0.283	0.301	0.276
Nondecision time range ^a	s_t	0.1	0.1	0.1	0.1

Note. VSTM = visual short-term memory.

^a Denotes a fixed parameter.

found that the quality of the fit was relatively insensitive to changes in s_t across a range of plausible values, so we set it to 0.1 s and treated it as a fixed parameter.

With this parameterization, fits of the single and dual diffusion models were based on 13 and 14 free parameters, respectively. Although the number of parameters is fairly large, the degree of data reduction provided by the models is substantial. There are a total of 220 degrees of freedom in the data (20 distribution pairs, each with 11 degrees of freedom). Empirical models of RT, such as the ex-Gaussian, the shifted Weibull, and the shifted gamma, typically require three free parameters to describe the location, dispersion, and shape of a single distribution (Luce, 1986, pp. 507–511; Ratcliff, 1979). An empirical description of the family of 40 distributions would therefore require a total of 120 free parameters and would provide no account of the choice probabilities. In comparison, our process models require an order of magnitude fewer parameters and provide an account of both the choice probabilities and the distributions of RT.

The estimated values of the parameters in Table 2 seem reasonable psychologically. Visual contrast transduction was described by a Naka-Rushton function with an inhibition term of around 0.01 and an exponent of around 2, consistent with values reported in the literature.⁶ Estimates of the perceptual effect of the pedestal, I_p , ranged from 0.8 to around 1.0. These values are consistent with the interpretation of I_p in the VSTM equation, as a measure of the relative energies of the pedestal and the patch. The data do not allow I_p to be estimated with any great precision, as its role is to determine the rate of VSTM trace formation, which in turn determines the location of the first quantile of the RT distribution. The

RT distributions in Figure 12 are consistent with a model in which the rate of trace formation is independent of stimulus contrast and this can be achieved with any large value of I_p .

Estimates of the sensory response function onset rate parameter, β_{on} , were fairly variable. This was unsurprising, as estimates of rate parameters from noisy data typically tend to be labile. Indeed, we found that fits of reasonable quality could be obtained by fixing the value of this parameter arbitrarily anywhere on the range 50 to 300. Estimates of the masked offset rate, $\beta_{off, m}$, were uniformly greater than β_{on} , whereas estimates of the unmasked offset rate, $\beta_{off, u}$, were much less than β_{on} . This is consistent with the idea that $\beta_{off, m}$ describes rapid, multiplicative suppression of the stimulus by the mask whereas $\beta_{off, u}$ describes slow, iconic decay.

The fits in Table 1 were obtained subject to the constraint that $\beta_{off, u} > 10$. We constrained the fits in this way because the models predict mask-dependent cuing effects when the persistence of unmasked stimuli is long relative to masked stimuli. Consequently, estimates of $\beta_{off, u}$ in unconstrained fits tended to go to zero. We constrained $\beta_{off, u}$ to see whether the observed cuing effects could be predicted using realistic estimates of stimulus persistence. As Table 2 shows, for all four models, the estimated value of $\beta_{off, u}$ was at or near the lower bound of 10. With $n = 3$ cascaded stages and $\beta_{off, u} = 10$, $\mu(t)$ decays to 25% of its peak within 350 ms of stimulus offset.

⁶ With $\rho = 2$, a value of $I_{in} = .01$ corresponds to a semisaturation constant $I_{0.5} = 0.1$ (see Footnote 3). That is, the visual contrast transducer function attains half its maximum value at a contrast of 0.1.

It is difficult to assess whether these estimates are reasonable or a little too long. Visual persistence can be estimated in a variety of different ways (Breitmeyer, 1984, chap. 3) and these methods give differing estimates. Moreover, as Coltheart (1980) argued, visual persistence and informational persistence are not the same, and it is the latter that is relevant in our theory. One way to measure the informational persistence of a stimulus was proposed by Loftus, Johnston, and Shimamura (1985), who compared performance with masked and unmasked stimuli to estimate “the worth of an icon.” They found that the icon was worth an additional 100 ms of stimulus exposure. That is, when stimuli were masked, exposure duration needed to be increased by around 100 ms to obtain a level of performance comparable to that obtained with unmasked stimuli.

Using the logic of Loftus et al. (1985) in Equation 2, we find that a plausible value of $\beta_{\text{off}, u}$ is around 20. We estimate this value by equating the areas of $\mu(t)$ for a masked stimulus with $d = 0.16$ s and an unmasked stimulus of $d = 0.06$ s, using typical values of β_{on} and β_{off} to do so. A decay parameter of $\beta_{\text{off}, m} = 20$ represents a sensory response function that decays to 25% of its maximum value within 200 ms of stimulus offset. This value agrees reasonably well with estimates of the visual persistence of a 3.5 cycles-per-degree grating, like those used by Smith, Ratcliff, and Wolfgang (2004), reported by Bowling and Lovegrove (1981). When we constrained $\beta_{\text{off}, m} > 20$ in the fits, the G^2 values worsened, but only by around 4%. The resulting constrained fits are still highly acceptable.

The estimates of attention gain were around 10 to 15 in the gain models and a little smaller in the orienting models. Gain values of 10–15 imply that the VSTM trace is formed within the first 200–300 ms after stimulus onset, consistent with the results of Vogel et al. (2006). In the gain models, γ_A was around 1.5 times greater than γ_U . These estimates are reasonable if we regard gain as a precursor to, and determinant of, relative trace strength. The relative gains correspond to an attentional effect of around 3.5 dB, which agrees with the mask-dependent cuing effects found experimentally (Smith, Wolfgang, & Sinclair, 2004).⁷ The estimates of orienting time in the orienting models ranged from 41 ms to 61 ms. These values are longer than the 16 ms estimated by Smith, Ratcliff, and Wolfgang (2004) in an earlier version of the theory. Given that the cuing effect with peripheral cues takes around 100 ms to reach its maximum, the longer estimates seem more reasonable. These estimates can be attributed to our use of a more elaborated and psychologically realistic model of the sensory response function than the one used by Smith, Ratcliff, and Wolfgang (2004).

The criterion-setting hypothesis. We noted earlier that a number of authors have suggested a criterion-setting explanation for the Posner effect in detection. According to this explanation, RTs are shorter for cued stimuli because people set lower criteria for stimuli at cued locations, allowing them to respond to cued stimuli more rapidly. Although criterion setting does not explain the mask-dependent cuing effect, it is conceivable that some part of the attentional effect in the Smith, Ratcliff, and Wolfgang (2004) data may be due to criterion differences.

To investigate whether this was so, we refitted the data with each of the four models, allowing the criteria for attended and unattended stimuli, a_A and a_U , to differ. The resulting goodness-of-fit statistics are shown in Table 3. The table shows that there is,

Table 3
Test of Criterion Setting (Smith, Ratcliff, & Wolfgang, 2004)

Model	G^2	df	BIC
Diffusion gain	154.5	206	280.3
Diffusion orienting	163.3	206	289.1
Dual diffusion gain	164.9	205	299.7
Dual diffusion orienting	168.7	205	303.4

Note. BIC = Bayesian Information Criterion.

at best, equivocal support for the criterion setting hypothesis. Allowing a_A and a_U to differ produced a uniform improvement in G^2 for all models, but only in some cases is the BIC for a criterion-varying model better than for a fixed-criterion model, and in those cases, the difference is small. There is thus little evidence that the attentional effects in this task are due to observers adopting different criteria for attended and unattended stimuli. Of course, we do not argue that this finding generalizes beyond these data. The study of Smith, Ratcliff, and Wolfgang (2004) used a psychophysical task, in which accuracy was stressed and in which observers were given trial-by-trial accuracy feedback. There was thus little incentive to try to optimize speed by setting low criteria at cued locations. It remains plausible that criterion setting may be a determinant of performance in tasks in which speed of responding is stressed, as previous authors have argued. In the Smith, Ratcliff, and Wolfgang data, however, the differences in RT produced by cuing appear to be due to other mechanisms.

Interim summary. We have shown that our theory provides a good account of the data from a cued detection task with masked and unmasked stimuli. The pattern of data is a fairly complex one: Cues increased accuracy only when stimuli were backwardly masked, but they reduced RT unconditionally. The theory attributes these effects to an interaction between attention and the informational persistence of stimuli. It assumes that attention increases the efficiency with which stimulus information is transferred to VSTM, either by increasing the rate of transfer or by reducing the delay before the transfer begins. The effect of the increased efficiency is most apparent when stimuli are masked and informational persistence is short. Under these circumstances, attention produces an increase in accuracy and a large reduction in mean RT. When stimuli are unmasked, there is no change in accuracy and a smaller reduction in RT.

Our theory not only accounts for the effects of cues and masks on accuracy and mean RT but it also accounts for the shapes of RT distributions for correct responses and errors and how the shapes change as a function of experimental conditions. It thus provides an extremely rich and detailed characterization of performance on this task. We found that two different decision models, both based on diffusion processes, provided equally good accounts of the RT distributions and choice probabilities. This finding reinforces the

⁷ The cuing effect in decibels is equal to $20\log_{10}(d'_A/d'_U)$, where d'_A and d'_U are the sensitivities to cued and miscued stimuli, respectively. A scaling factor of 20, rather than 10, is used because, by convention, the decibel is a unit of relative power rather than relative amplitude. Arguments in relative amplitude units must therefore be squared before taking logarithms.

findings of Ratcliff and Smith (2004), who showed that the critical ingredient of a successful decision model is not whether the model assumes one evidence total or two but the kind of stochastic process used to model evidence accumulation. As Ratcliff (2001) has shown, models based on diffusion processes naturally predict the kinds of RT distributions that are found experimentally.

Effect of Spatial Uncertainty (Gould et al., 2007)

The data of Gould et al. (2007) in Figure 4 show a similar pattern to that in the data of Smith, Ratcliff, and Wolfgang (2004). When stimuli were localized by fiducial crosses, cues shortened mean RT but had no effect on sensitivity. When stimuli were not localized by crosses, cues shortened RT and increased sensitivity. Because these effects were obtained with unmasked stimuli, they cannot be due to an interaction between attention and informational persistence.

To model the data of Gould et al. (2007), we assumed that fiducial crosses affect the process of selecting stimuli for entry into VSTM. Our initial hypothesis was that cues and crosses act in a similar way, to affect the efficiency of VSTM trace formation. Specifically, we assumed that cues have a top-down effect on efficiency, whereas crosses have a bottom-up effect. This implied that gain (or orienting time) would vary as a function of both cues and crosses. We therefore predicted very low gain or very long orienting times when stimuli were miscued and were presented without fiducial crosses.

This idea led to the right qualitative predictions—namely, reduced accuracy and long RTs in the miscued, no-cross condition—but the resulting models failed quantitatively. There were two reasons for this failure. First, values of gain or orienting time that were sufficient to predict the reduction in accuracy in the miscued, no-cross condition also predicted increases in RT that were much greater than those found experimentally. Second, the magnitude of the predicted cuing effect in accuracy in the no-cross condition was largely independent of stimulus contrast. The cuing effect found experimentally, expressed as a proportional reduction in sensitivity, increases at low contrasts (Gould et al., 2007, Figure 4). This differs from the pedestal task, in which the magnitude of the cuing effect is almost independent of contrast (Smith, Wolfgang, & Sinclair, 2004). We therefore sought a different characterization of the effects of the fiducial cross.

Of the many alternatives we considered, the one that provided the most parsimonious account of the data was the assumption that the fiducial cross affects ρ , the exponent of the sensory transducer function. This assumption derives from the idea that the fiducial cross is primarily a manipulation of observer uncertainty (Gould et al., 2007; Petrov, Verghese, & McKee, 2006) and that uncertainty changes the effective shape of the transducer function (Pelli, 1985), making low-contrast stimuli less detectable than they would be otherwise. Pelli (1985) argued that, for a stimulus to be detectable, the activity in the visual filter coding the stimulus must differ by some threshold amount from noise in filters coding the surrounding display. The task of distinguishing the stimulus from the background becomes more difficult with increasing uncertainty about the stimulus location because it increases the number of filters the observer must monitor for evidence of a signal. These effects are largest at low stimulus contrasts, where the activity produced by signals is similar to the noise in the surrounding

display. The result appears as an increase in the exponent of the visual contrast response function. Pelli's argument echoes an argument made earlier by Green (1960) for auditory detection.

Table 4 lists the parameters that were estimated to fit the resulting models. As in the Smith, Ratcliff, and Wolfgang (2004) experiment, the two attention and two decision models gave us a total of four different models. As Table 4 shows, along with gain (or orienting time), the main parameter that varied between cue and cross conditions was the Naka-Rushton exponent, ρ . Initially, we allowed ρ to vary freely as a function of cue and cross conditions (a total of four values in all). However, we found no evidence that ρ changed with cuing in the cross condition; rather, any changes were confined to the no-cross condition. We therefore constrained cued and miscued ρ values in the cross condition to be equal and allowed ρ to vary with cues only in the no-cross condition. We also allowed η , the drift variability in the decision process, to vary as a function of fiducial condition, as it seemed likely that trial-to-trial encoding efficiency would vary, depending on whether fiducial crosses were used. The results of Gould et al. (2007) suggest that crosses reduce the effects of noise from the surrounding display, leading to more efficient stimulus localization and VSTM trace formation. If so, trace strength variability is likely also to be reduced by crosses, because stimulus localization is less affected by trial-to-trial variations in display noise. We also assumed the decision criteria and the onset and offset rates for the sensory response function, β_{on} and β_{off} , would be the same in all conditions. The assumption of equal criteria is required by the experimental design, which mixed cross and no-cross stimuli randomly within blocks of trials. Equal onset and offset rates were assumed because the same, unmasked stimuli were used in all conditions. The fixed parameters were constrained as in the previous experiment.

The models with Naka-Rushton exponents varying in this way provided a reasonable account of all the mean features of the data, including the choice probabilities, the shapes of the RT distributions, and the changes in the leading edge of the distributions as a function of contrast. Figure 13 shows the fit of the single-diffusion, gain model; the top part of Table 5 shows the goodness-of-fit statistics for all four models. As was the case with the Smith, Ratcliff, and Wolfgang (2004) data, all four models performed similarly. The best model, in both a G^2 and a BIC sense, was the dual diffusion, orienting model. Again, however, the G^2 values for the best and worst fitting models were within a few percent of each other.

Although the patterns of sensitivity and mean RT for the Smith, Ratcliff, and Wolfgang (2004) study and the Gould et al. (2007) study are very similar, the quantile probability plots for the two studies are quite different. In comparison with the Smith, Ratcliff, and Wolfgang (2004) study, the distribution quantiles in the Gould et al. data are considerably more bowed, and there is substantially more change in the .1 quantile across stimulus conditions. The models do reasonably well in capturing the bowing of the .1 quantile, as Figure 13 shows. The critical feature of the models that allow them to capture both patterns of data is the presence or absence of the pedestal intensity term, I_p , in the VSTM equation. When I_p is present, the rate of VSTM growth, and hence the leading edge of the RT distribution, is relative independent of stimulus contrast. When I_p is absent, the leading edge of the RT distribution varies substantially.

Table 4
Model Parameters (Gould et al., 2007)

Parameter	Symbol	Diffusion gain	Diffusion orienting	Dual diffusion gain	Dual diffusion gain
Sensory response function					
Onset rate	β_{on}	128.6	105.8	86.4	92.2
Offset rate	β_{off}	32.2	22.6	20.8	16.9
Number of stages ^a	n	3	3	3	3
Naka-Rushton amplitude	θ	3.44	3.95	2.53	2.21
Naka-Rushton exponent					
Fiducial	ρ_{fid}	1.65	1.67	1.59	1.64
Attended, no fiducial	$\rho_{A, \text{no fid}}$	1.69	1.75	1.65	1.68
Unattended, no fiducial	$\rho_{U, \text{no fid}}$	2.00	2.04	2.01	2.00
Naka-Rushton inhibition	I_{in}	0.054	0.036	0.059	0.051
Attention/VSTM (gain models)					
Gain (attended)	γ_A	15.64	—	38.3	—
Gain (unattended)	γ_U	11.7	—	28.6	—
Attention/VSTM (orienting models)					
Gain	γ	—	14.3	—	40.7
Orienting time	t_{or}	—	0.050	—	0.062
Decision Process					
Criterion	a	0.068	0.065	0.046	0.048
Drift variability, fiducial	η_{fid}	0.53	0.80	0.50	0.44
Drift variability, no-fiducial	$\eta_{\text{no fid}}$	0.67	0.85	0.55	0.51
Ornstein-Uhlenbeck decay	λ	—	—	6.6	6.0
Stimulus independent diffusion	σ_2	0.015	0.014	0.020	0.032
Reflecting boundary ^a	r	—	—	-0.01	-0.01
Nondecision Processes					
Mean nondecision time	T_{er}	0.267	0.260	0.276	0.259
Nondecision time range ^a	s_t	0.1	0.1	0.1	0.1

Note. VSTM = visual short-term memory.

^a Denotes a fixed parameter.

Although the models satisfactorily capture the interaction of cues and uncertainty and their joint effects on the shapes of the RT distributions, the fit statistics in Table 5 are about 40% worse than those in Table 1. For these data, G^2 for a well-fitting model should be around 206–207, with a standard deviation of around 20, so the values in Table 5 suggest some failure to fit. A possible reason for this failure can be identified in the pattern of mean RTs in Figure 4. The figure shows that there was a small cuing effect when stimuli were localized by crosses and a large effect when they were not. Unlike the corresponding pattern in the Smith, Ratcliff, and Wolfgang (2004) study, however, the cuing effect in the no-cross condition disappears at low contrasts. At these contrasts, there was no difference in mean RT for cued and miscued stimuli.

What is happening at low contrasts in the no-cross condition in this experiment? Figure 4 shows that, at the lowest levels of contrast, performance is almost at chance. At these levels of contrast, observers were obtaining virtually no stimulus information from the display. One interpretation of the collapse of the Posner effect in RT at low contrasts is that observers were responding according to an internal deadline. That is, they were aware within a few hundred milliseconds of the cue that they had not registered the presence of a sensory event and were unlikely to do if they continued to wait and so chose to terminate sampling from the display. In other words, they were not using a pure information-controlled decision process but, rather, were using a mixture of time-controlled and information-controlled processing. Responses to miscued, low-contrast, no-cross stimuli would be most affected by a mixed strategy of this kind because these

responses are the slowest. The result would be a collapse of the Posner effect at low contrasts.

Superficially, the quantile probability functions in Figure 13 appear consistent with a mixed-strategy account of this kind. In comparison with the other three panels, the bowing of the distribution tails in the miscued, no-cross condition is reduced, as would be expected with deadline responding. However, closer inspection reveals that the bowing of the quantile probability functions is reduced for the other quantiles, as well. This is inconsistent with a simple deadline model of responding—of either a fixed or variable kind—because deadline responding should only affect the longest RTs and should leave the remainder of the distribution unaltered. We therefore considered an alternative hypothesis to try to capture what is happening in the no-cross condition.

Both of our decision models assume that drift (or, more precisely, the excitatory component of drift) and the diffusion coefficient grow in proportion to one another. Psychologically, the time at which the diffusion coefficient changes from zero marks the point at which the diffusion process begins to accumulate evidence. The idea that drift and diffusion coefficient grow proportionally presupposes that evidence accumulation is tightly bound to the spatiotemporal properties of the stimulus. Although this is plausible when stimuli are localized by pedestals or fiducial crosses, it is also plausible that this binding may break down to some extent in the absence of localizing markers. This accords with the suggestion of Laming (1968), who argued that information accumulation can be initiated even in the absence of a stim-

ulus by the expectation of its occurrence at a particular time.⁸ We investigated a very simple form of this idea, in which the SDE describing evidence accumulation in the decision stage had an additional component of diffusive variability that was independent of drift. We modeled evidence accumulation by an SDE of the form

$$dX(t) = [v(t) - \lambda X(t)]dt + [\sigma_1(t) + \sigma_2]dW(t). \quad (18)$$

In this equation, $\sigma_1(t)$ is the component of the diffusion coefficient that grows with drift and σ_2 is a small component independent of drift (cf. Equation 17). Psychologically, σ_2 may be thought of as a nonspecific component of decision noise that is independent of the stimulus.

Consistent with this interpretation, we found the extra decision noise improved fit only in the no-cross condition; it had no effect in the fiducial condition, and its effect in the no-cross condition was most apparent with miscued stimuli. The effect of additional decision noise is to reduce accuracy and shorten RT. Its effects are most pronounced at low stimulus contrasts where the rate of VSTM growth is slowest. Unlike deadline responding, however, its effects extend across the distribution. The net result is a flattening of RT quantiles, as is found in the Gould et al. (2007) data.

The lower part of Table 5 shows the goodness of fit of the four models when the σ_2 term is added. Although the addition of σ_2 improves G^2 by only around 5–7%, it leads to a better qualitative

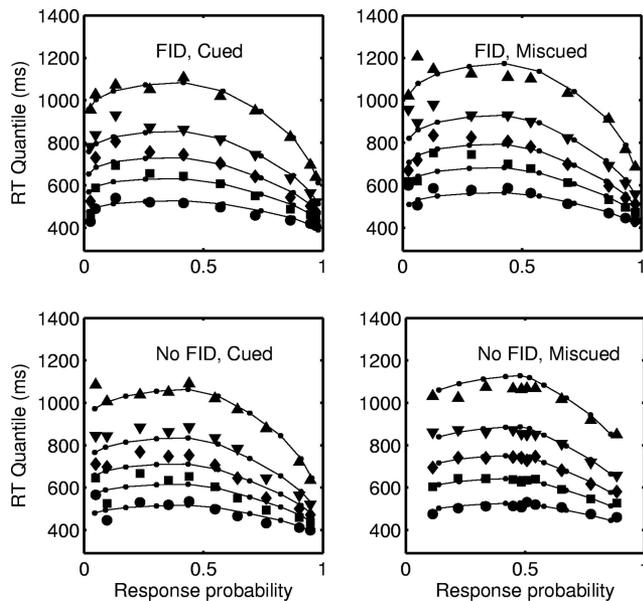


Figure 13. Fit of the single-diffusion, attention gain model to the response time (RT) distributions and choice probabilities for the Gould, Wolfgang, and Smith (2007) study. The large symbols are the experimental data and the continuous curves and small circles are the fitted values. The five lines in each panel are, in ascending order, the .1, .3, .5, .7, and .9 quantiles of the predicted RT distributions. The corresponding empirical quantiles are shown as circles, squares, diamonds, inverted triangles, and upright triangles, respectively. The plot is parameterized by stimulus contrast, as described in the caption to Figure 12. The model had a σ_2 term in the diffusion coefficient in the no FID, miscued condition (see text for details).

Table 5
Fit Statistics (Gould et al., 2007)

Model	G^2	df	BIC
Diffusion gain	247.9	207	363.5
Diffusion orienting	247.6	207	363.1
Dual diffusion gain	246.6	206	371.0
Dual diffusion orienting	234.8	206	359.2
Models with diffusion term, σ_2			
Diffusion gain	232.2	206	356.9
Diffusion orienting	235.9	206	360.2
Dual diffusion gain	234.0	205	353.3
Dual diffusion orienting	218.7	205	352.0

Note. BIC = Bayesian Information Criterion.

match between models and data. As further support for the idea that the flattening of the RT quantiles in the miscued, no-cross condition of Gould et al. (2007) was due to an uncertainty-dependent increase in diffusion noise, we went back and added a σ_2 term to the fits of the Smith, Ratcliff, and Wolfgang (2004) data. Unlike in Gould et al., this produced no improvement in fit, consistent with the idea that the extra decision noise and associated change in distribution shape only occur under conditions of spatial uncertainty, when stimuli are not well localized perceptually.

Table 4 shows the estimated parameters for the four models. In general, the estimated values were similar to those in Table 2, although there are some differences, both between experiments and among models. Some of this variability can be attributed to the cascaded stage structure of the models, which leads to trade-offs between parameters. The estimated values of the $\mu(t)$ offset parameter, β_{off} , were larger than the corresponding estimates for the Smith, Ratcliff, and Wolfgang (2004) study, whereas the differences between the attentional gain constants, γ_A and γ_U , were smaller. The average estimated value of β_{off} was around 20, which agrees with the figure based on Loftus's "worth of an icon" idea, estimated in the previous experiment.

The important parameters from the point of view of the cuing effect are the values of the Naka-Rushton exponent, ρ , which characterizes nonlinearity in stimulus transduction. To fit the models we assumed a transducer function, $r(\Delta_I)$, of the form in Equation 6, in which the inhibitory coefficient, I_{in} , was held fixed while the exponent varied across conditions. The estimated values of ρ in Table 4 are a little smaller than the corresponding values in Table

⁸ Laming (1968) used the idea of top-down initiated information accumulation to explain the fast errors that are often found under conditions of speed stress. Laming argued that the expectation of stimulus onset may cause the decision process to begin sampling prematurely, from the pre-stimulus field. As a result, it would initially accumulate noise, rather than stimulus information, leading to an increased probability of error. We do not invoke expectation or time uncertainty here, but only the idea that the initiation of information accumulation may be, in part, under top-down control. Smith and Wolfgang (2004, p. 133) used Laming's idea more explicitly, in a precursor to the current theory. The model of Equation 18 could be extended in an obvious and plausible way, by introducing temporal uncertainty in the σ_2 term and allowing the onset of the second source of noise to lag behind the stimulus. We have not pursued this possibility further here, as it is incidental to the main purpose of the article.

2. They were smallest in the fiducial cross condition, a little larger in the cued, no-cross condition, and largest in the miscued, no-cross condition. This is consistent with Pelli's (1985) idea that increasing uncertainty increases the exponent of the contrast response functions. Our fits showed that, when stimuli were localized by fiducial crosses, ρ values were the same for cued and miscued stimuli. Because stimuli were unmasked, asymptotic VSTM trace strengths were also the same. Consequently, differences in gain or orienting time produced differences in RT between conditions but none in accuracy. When stimuli were not localized with crosses, there were large differences in ρ for cued and miscued stimuli. The effect of these differences was to make miscued stimuli less detectable than cued stimuli, especially at low contrasts, as Figure 14 shows. This resulted in both RT differences and accuracy differences. The latter were most pronounced at low contrasts, as Gould et al. (2007) found.

The estimates of ρ were slightly larger for cued, cross stimuli than for cued, no-cross stimuli. Although the difference is small numerically, it reflects a significant difference in performance. Gould et al. (2007) found that sensitivity to cued, cross stimuli was slightly but consistently higher than to cued, no-cross stimuli for all observers. This suggests that the cue and the cross both acted to reduce uncertainty by localizing the stimulus but that the cross was the more effective of the two. This is unsurprising because, unlike the cue, the cross was 100% predictive of the stimulus location and localized the stimulus more precisely in space and in time.

We also tested whether some part of the cuing effect in the Gould et al. (2007) study could be attributed to differences in the criteria used to make judgments about attended and unattended stimuli. The results of these model fits are shown in Table 6. As in the Smith, Ratcliff, and Wolfgang (2004) study, allowing the criteria for attended and unattended stimuli to differ produced a reduction in G^2 for all models, but the BIC improved for only one of them, and this improvement was small. There is thus little evidence that the Posner effect in RT in the Gould et al. study was due to criterion setting.

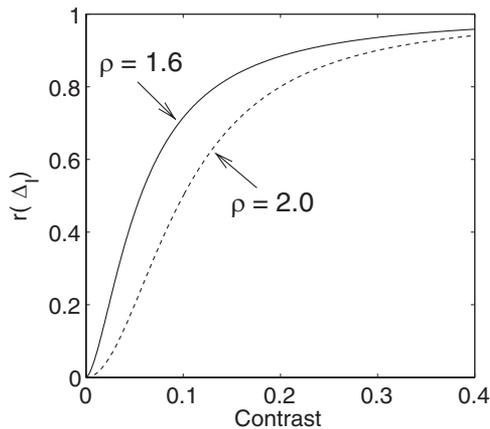


Figure 14. Naka-Rushton sensory transducer functions, $r(\Delta_I)$, as a function of contrast for stimuli presented under conditions of low ($\rho = 1.6$) and high ($\rho = 2.0$) uncertainty. The effect of uncertainty is to make stimuli less detectable, because the activity in the visual filter coding the stimulus must be distinguished from noise in the surrounding display. Proportionally, the impact of uncertainty is largest at low-stimulus contrasts.

Table 6
Test of Criterion Setting (Gould et al., 2007)

Model	G^2	df	BIC
Diffusion gain	229.6	205	362.7
Diffusion orienting	225.3	205	358.5
Dual diffusion gain	228.8	204	362.0
Dual diffusion orienting	214.5	204	352.6

Note. BIC = Bayesian Information Criterion.

A Neurally Plausible Model of VSTM

The heart of our theory's ability to predict the experimental data in Figures 12 and 13 is the VSTM growth equation, Equation 15. In this equation, the pedestal intensity affects the rate of VSTM growth but not asymptotic trace strength. This decoupling of rate and asymptote is needed to account for the data of Smith, Ratcliff, and Wolfgang (2004) and of Gould et al. (2007) with the same VSTM equation. To obtain this equation, we needed to assume that the rate of VSTM growth depends on the energy in the stimulus compound, whereas the asymptotic trace strength depends only on the part of the compound that carries information about stimulus identity. If we think of the VSTM trace as a cognitive representation of the information used in the observer's decision, this property is evidently the right one. How might a computation like the one in Equation 15 be realized in the visual system?

We propose that the VSTM trace is formed by the integration of stimulus activity carried in two parallel pathways. The activity in one of the pathways depends on the total energy in the stimulus compound and is insensitive to the presence of any form information or visual structure the compound may contain. We denote the activity in this pathway by $v_E(t)$ (for "energy"). The activity in the second pathway depends on the presence of visual structure that carries information about stimulus identity. We denote the activity in this pathway by $v_F(t)$ (for "form"). As shown in Figure 15, the relationship between the two pathways is a simple one. Activity in the energy pathway selects a location in the visual field and gates the form information at that location into VSTM. The drift of the diffusion process is proportional to the strength of the resulting VSTM trace.

The computations depicted in Figure 15 are implemented by a pair of coupled shunting equations:

$$\frac{dv_E}{dt} = \gamma_i \{ (I_C + I_{in}) \mu(t) [1 - v_E(t)] - [1 - (I_C + I_{in})] \mu(t) v_E(t) \} \quad (19)$$

and

$$\frac{dv_F}{dt} = \gamma_i v_E(t) \{ r(\Delta_I) \mu(t) [\theta - v_F(t)] - [1 - r(\Delta_I)] \mu(t) v_F(t) \}, \quad (20)$$

where $\gamma_i \in \{\gamma_A, \gamma_U\}$. As defined previously,

$$I_C = \sqrt{\Delta_I^{2\rho} + I_p^{2\rho}}$$

is the square root of the power in the transduced stimulus compound, and $r(\Delta_I)$ is the Naka-Rushton transduced stimulus incre-

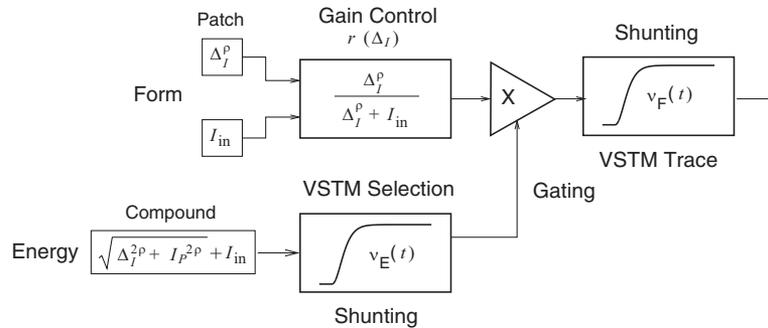


Figure 15. Dual-channel shunting visual short-term memory (VSTM) model. The power in the stimulus compound, $I_C = \sqrt{\Delta_I^{2\rho} + I_P^{2\rho}}$ gates form information in the grating patch, Δ_I^ρ into VSTM. The contrast of the patch is subject to nonlinear transduction and early gain control before gating. The activity in the energy pathway depends on the square root of the total power in the pedestal and patch. The contrast of the patch, Δ_I , and of the pedestal, I_P , are subject to separate nonlinear transduction with exponent ρ before power is computed.

ment of Equation 6. The critical feature of these equations is that the rate of growth of the VSTM trace, $v_F(t)$, is equal to the product of the attention gain, γ_i , and the instantaneous activity in the energy pathway, $v_E(t)$. The activity in this pathway depends on the energy in the stimulus compound. It is large and relatively independent of stimulus contrast when the pedestal is present and small and strongly dependent on contrast when the pedestal is absent.

To obtain the right kinds of properties in the composite VSTM model, we have again used shunting equations that decouple the rates and asymptotic trace strengths from one another.⁹ Unlike the equations described earlier in this article, the rate of VSTM growth in the dual-pathway model depends on stimulus contrast solely via the coupling between the pathways. The model assumes that contrast information in the form pathway is subject to early gain control (Naka-Rushton-like nonlinear transduction) prior to entering the memory system, as depicted in Figure 15. As noted earlier, this assumption is consistent with the known physiology of the visual system (Walraven et al., 1990). The asymptotic trace strength in the form pathway is $\theta r(\Delta_I)$, and the rate of growth is $\gamma_i v_E(t)$. The asymptotic trace strength in the energy pathway is $I_C + I_{in}$, and the rate of growth is γ_r . The composite VSTM model in Equations 19 and 20 assumes a gain model of attention; an orienting model can be obtained by an obvious generalization of Equation 11.¹⁰

To evaluate the dual-pathway model, we fitted it to the data of Smith, Ratcliff, and Wolfgang (2004) and Gould et al. (2007) in the way described previously. Because we previously found no significant differences in fit among the four model classes, in this evaluation, we focused exclusively on the single-diffusion, attention gain model. We found that the multiple-pathway model fit the data just as well and, indeed, slightly better than the model based on Equation 15. For the data of Smith, Ratcliff, and Wolfgang, the goodness-of-fit statistic was $G^2(207) = 160.4$, $BIC = 277.3$, and for the data of Gould et al., it was $G^2(205) = 229.8$, $BIC = 363.0$. These fit statistics indicate that the dual-pathway model nicely captures the differences in the shapes of the RT distributions and the changes in their leading edges in the two data sets.

As a further test of the dual-pathway model, we investigated whether the same set of model parameters could describe the RT distributions and choice probabilities in the two data sets simulta-

neously. The model in Figure 5 involves interacting subprocesses whose parameters can trade off when it is fitted to data. To ensure a consistent characterization of the VSTM process in the pedestal and the no-pedestal task, we fitted the two data sets with a single composite model. Because the two data sets were based on different groups of observers, we expected that any such composite model would misfit to some degree. However, quantile averaging should have reduced the effect of individual differences enough for this comparison to still be meaningful.

Table 7 shows the parameters that were estimated to fit the composite model to the two sets of data, and Figure 16 shows the associated fits. We assumed that the two sets of data could be described by the same decision criteria, sensory response functions, VSTM processes, attention gains, and nondecision times. The parameters that differed between experiments were the Naka-Rushton parameters, which describe the early, nonlinear transduction of contrast, and the between-trials variability in VSTM trace strength. The latter determines the trial-to-trial variability in the

⁹ Unlike the inhibitory coefficient in Equation 20, the inhibitory coefficient in Equation 19 is not guaranteed to remain positive for all values of I_C and I_{in} . In practice, however, it does so for the kinds of values that arise in fitting data. If we adopt a more general formulation, writing the inhibitory coefficient as $K - (I_C + I_{in})$, where K is a positive constant, the asymptotic trace strength in the energy pathway becomes $(I_C + I_{in})/K$. We investigated this more general model, allowing K to vary as a free parameter. When we did so, we obtained a small improvement in G^2 ($G^2 = 460.9$ vs. $G^2 = 466.3$). According to the BIC, this improvement in fit was not sufficient to offset the gain in model freedom associated with the introduction of an extra free parameter ($BIC = 773.5$ vs. $BIC = 760.8$).

¹⁰ Smith (in press) described another way to decouple the rate and asymptote in a shunting equation, which leads to a predicted asymptotic trace strength that is proportional to the Michelson contrast of a stimulus. (Michelson contrast is defined as $(I_2 - I_1)/(I_2 + I_1)$, where I_2 and I_1 are the maximum and minimum luminances in the stimulus, respectively.) In the equation described by Smith, the rate of VSTM growth is equal to $I_2 + I_1$, which, for a grating stimulus, is proportional to the mean luminance of the display. Although this approach has some appealing features, it does not extend readily to the case in which VSTM growth can depend on stimulus contrast, as we have argued occurs in the data of Gould et al. (2007).

Table 7
Two-Channel VSTM Model, Joint Fit (Single-Diffusion, Gain)

Parameter	Value
a	0.087
β_{on}	155
$\beta_{\text{off}, m}$	346
$\beta_{\text{off}, u}$	24
γ_A	103
γ_U	75
η_1	0.274
η_2	0.427
η_3	0.278
η_4	0.359
T_{er}	0.229
θ	1.13
$I_{\text{in}, 1}$	0.001
$I_{\text{in}, 2}$	0.021
ρ_1	2.47
ρ_2	1.84
ρ_3	1.89
ρ_4	2.24
I_P	0.30
σ_2	0.022

Note. VSTM = visual short-term memory; $s_r = 0.10$ fixed.

drift of the diffusion process. We expected both Naka-Rushton shape and between-trial variability to differ between experiments because they are features of stimulus encoding likely to vary with uncertainty and the presence or absence of masks.

The notation used to designate the parameters in Table 7 is the same as that used in Tables 2 and 4, with the following exceptions. The parameters η_1 through η_4 are, in order, the drift variabilities in the unmasked and masked conditions of Smith, Ratcliff, and Wolfgang (2004) and in the fiducial and the no-fiducial conditions of Gould et al. (2007). The parameters $I_{\text{in}, 1}$ and $I_{\text{in}, 2}$ are the Naka-Rushton inhibition constants for Smith, Ratcliff, and Wolfgang and Gould et al., respectively. The parameter ρ_1 is the Naka-Rushton exponent for all four conditions of Smith, Ratcliff, and Wolfgang; ρ_2 is the exponent for the cued and miscued

fiducial conditions of Gould et al.; and ρ_3 and ρ_4 are, respectively, the exponents for the cued and miscued no-fiducial conditions of Gould et al. This choice of parameters was guided by the fits of the models based on Equation 15.

The parameter I_P is a component of the energy pathway response that depends on the pedestal. It is nonzero for the Smith, Ratcliff, and Wolfgang (2004) study and zero for the Gould et al. (2007) study. The amplitude of the energy pathway response in Equation 19 is the sum of two terms: I_C , which depends on stimulus contrast, and I_{in} , which does not. In the single-pathway model of Equation 14, I_{in} is the inhibitory coefficient in an excitatory-inhibitory shunting equation. In Equation 19, however, it has no such interpretation. Rather, it is simply a component of the energy pathway response—and, hence, of the rate of VSTM growth—that is independent of contrast. We retained it in the model after finding that models with an energy pathway response proportional to a power of Δ_r produced quantile probability functions that were far more bowed than those in the data. We infer that the rate of VSTM growth increases with stimulus contrast but not as a simple power function of it.

Finally, the parameter σ_2 denotes the drift-independent component of the diffusion coefficient for the miscued uncertainty condition of Gould et al. (2007). As before, our rationale for this parameter is that the accumulation of stimulus information by the diffusion process is unlikely to be tightly bound to the stimulus when stimuli are not localized by either pedestals or fiducial crosses. This assumption was supported by our finding that the model fits improved (according to the BIC) when this parameter was included, but there was no improvement with a corresponding parameter for any of the conditions of Smith, Ratcliff, and Wolfgang (2004), or for the fiducial conditions of Gould et al.

The joint fit of the dual-pathway model in Figure 16 shows it provides a fairly good description of the RT distributions and the choice probabilities in the two studies, $G^2(420) = 466.3$. The most obvious misfit is in the miscued, uncertainty (no FID) condition of Gould et al. (2007), where the model underestimates the range of choice probabilities (i.e., overestimates the magnitude of the cuing effect). Apart from this, most of the main features of the data,

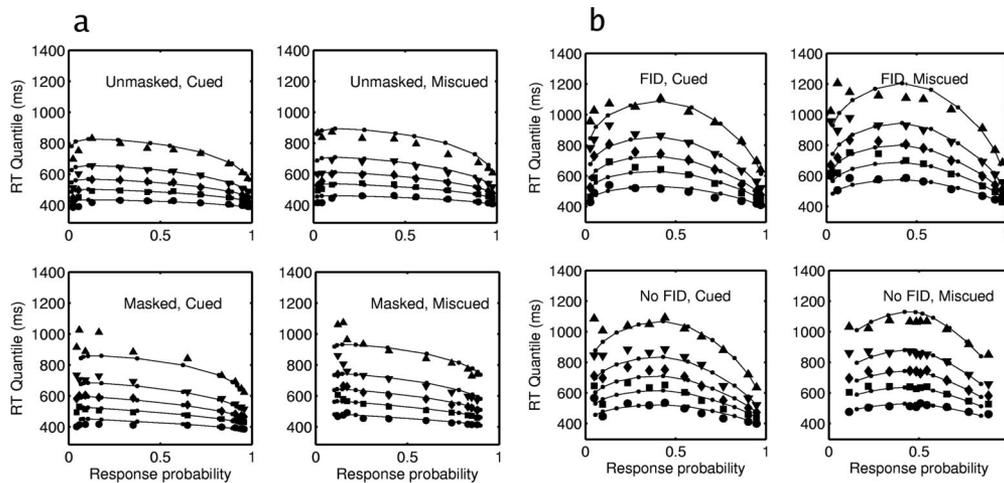


Figure 16. Simultaneous fit of the attention-gain, single-diffusion model to the data of (a) Smith, Ratcliff, and Wolfgang (2004) and (b) Gould, Wolfgang, and Smith (2007).

including the differences in the shapes of the RT distributions and the changes in their leading edges are captured by the dual-pathway model. Although the dual-pathway model uses 20 free parameters to fit the two data sets, it affords a high degree of data reduction, as these parameters jointly describe 80 RT distributions and 40 choice probabilities. This translates to one free parameter per four distributions and pair of choice probabilities.

The dual-pathway model provides a dynamic account of how a VSTM process like the one in Equation 15 could be computed in the visual system. It assumes that the patch and compound are each subject to excitatory-inhibitory shunting interactions and that the final VSTM trace is the confluence of the activity carried by the two pathways. Although the dual-pathway model is more complex than the simple, single-channel shunting VSTM model with which we began, the model is plausible physiologically. There is now considerable evidence that visual information is processed in parallel, through anatomically distinct dorsal “where” and ventral “what” pathways (Mishkin, Ungerleider, & Macko, 1983). According to one version of the dual-pathway account, the dorsal pathway locates a stimulus in space and segregates it from its background while the ventral pathway processes form and identity information. Indeed, Vidyasagar and colleagues (Vidyasagar, 1999; Saalmann, Pigarev, & Vidyasagar, 2007) have proposed a physiologically motivated model of attentional selection that is based on dorsal and ventral pathways interacting in this way.

In Vidyasagar’s (1999) model, stimulus attributes processed through the dorsal stream allow task-relevant stimuli to be located spatially and segregated from their surroundings. Processing of form information from selected locations via the ventral stream then allows the stimulus to be identified. In a similar way, the lower pathway in Figure 15 selects stimuli into VSTM on the basis of their overall energy properties. The more energy a stimulus contains, the more rapidly and efficiently it is encoded in VSTM. However, this pathway carries no information about stimulus identity. Identity information is processed via the upper pathway; the activity in this pathway is gated by the activity in the lower pathway to determine the overall strength of the VSTM trace. The lower and upper pathways in Figure 15 therefore exhibit, respectively, dorsal-like and ventral-like properties similar to those envisaged by Vidyasagar.

Although our dual-pathway model is consistent with visual physiology, our aim in developing it was not primarily to account for the effects of dorsal and ventral pathways in attentional selection; rather, it was to provide an account of the dynamics of VSTM formation within a shunting equation framework. The form of the dual-channel VSTM model was guided purely by mathematical and empirical considerations, rather than by any considerations of physiology or anatomy. Nevertheless, we find it striking that the minimum set of computational principles needed to account for our data should map so naturally onto structures and processes that have been identified physiologically.

Episodic Dynamics of Visual Attention

Our account of the mask-dependent cuing effect assumes that attention affects the efficiency of selection into VSTM and that VSTM selection interacts with the differential persistence of masked and unmasked stimuli to produce the dependency on masking. We showed that two different kinds of attentional model

could produce this interaction: a gain model and an orienting model. The goodness-of-fit statistics in Tables 1 and 5 show that these two models provide equally good accounts of the experimental data. We have therefore shown a form of model mimicry, in which two models with different process structures make predictions that are experimentally indistinguishable. Model mimicry is, of course, not uncommon in cognitive psychology. There are a variety of examples in the literature, the best known being Townsend’s (1972) demonstration of mimicry between serial and parallel search models. Smith (1998, in press) discussed another example of mimicry in the attentional domain, between maximum-outputs and independent-detectors signal-detection models.

From one viewpoint, the mimicry of gain and orienting models might be seen as disappointing, because we could not distinguish between alternative models using the available data. From another viewpoint, however, the distinction between gain and orienting models—or, more generally, between capacity and switching models—although it is well entrenched in the literature, can be viewed as an artificial and perhaps ultimately spurious one. Rather than viewing gain and orienting as mutually exclusive, they can instead be viewed as the spatial and temporal aspects of a single, dynamic attentional control system. The theory that best captures this spatiotemporal view of attention is the episodic theory of Sperling and Weichselgartner (1995). As shown in Figure 17, episodic theory assumes that the movements of an attentional spotlight in the visual field can be parsed into a series of discrete episodes, each with its own spatial and temporal extent.

In the version of the theory shown in Figure 17, attention is viewed as a time-dependent gradient of resources that can be flexibly allocated across space according to the demands of the task. This is depicted in Figure 17 for a single spatial dimension, x . The function in the figure, $\gamma(x, t)$, shows the resources allocated to location x at time t . The theory assumes the observer begins each experimental trial, before the cue has been processed, in a diffuse or divided attention state. In this initial state, resources are distrib-

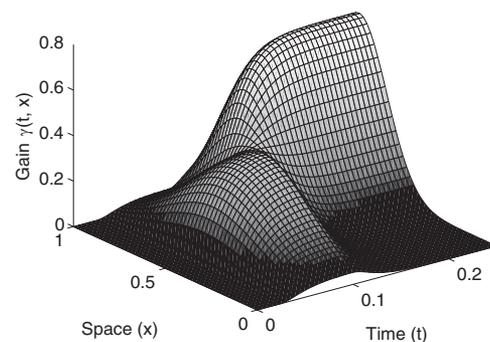


Figure 17. The attention-gating function in episodic theory. The left-hand axis, x , represents a spatial coordinate of the visual field (e.g., horizontal displacement); the right-hand coordinate, t , represents time. Two attentional episodes are shown: an initial, diffuse, episode, in which resources are spread broadly across the visual field, and a second, focused, episode, in which resources are withdrawn from the miscued location and concentrated at the cued location. Gain models assume that the rate of visual short-term memory growth, $\gamma(x, t)$, is proportional to the resources allocated to location x at time t . The theory implies that the rate of trace growth may change dynamically during the course of an experimental trial.

uted equally across the possible stimulus locations. After the cue has been processed, the observer enters a focused attentional state in which resources are withdrawn from uncued locations and concentrated on the cued location. These two episodes are depicted in Figure 17. As Smith and Wolfgang (2004) argued, Sperling and Weichselgartner's (1995) episodic theory combines in a very natural way with diffusion-process decision models if we assume that the function $\gamma(x, t)$ represents the instantaneous value of attentional gain at time t and location x .

Under this interpretation, prior to cue processing, when the observer is in a divided attention state, gain is intermediate at all display locations. After cue processing is complete, when the observer moves into a focused attention state, gain is high at cued locations and low at uncued locations. The orienting time is the time at which the second attentional episode is initiated, when the observer moves from a divided to a focused attention state. Viewed in this way, gain and orienting are not mutually exclusive mechanisms. Rather, gain refers to the spatial dimension of attention and orienting to its temporal dimension. Gain describes the rate of acquisition of information as a function of space at a specific, fixed time. Orienting describes changes in the rate of acquisition as a function of time at a specific, fixed location.

The breakdown of the strict dichotomy between gain and orienting in Figure 17 makes our failure to find quantitative differences between them less surprising than it might otherwise have been. We noted earlier that the effect of a change in either gain or orienting time is to change the efficiency of VSTM selection. Differences in selection efficiency lead to predicted differences in RT and accuracy as a function of spatial cuing. Our theoretical preference is to view attentional selection within the Sperling and Weichselgartner (1995) episodic framework, because of the natural way it marries with our decision models, and because it appears to us to best capture the flexible and dynamic nature of attentional control. Indeed, the manner in which we implemented gain and orienting models in our theory assumed a spatiotemporal attentional control system similar to the one proposed by Sperling and Weichselgartner.

There is, however, an important difference between the view of attentional orienting implied by Figure 17 and the strict view of orienting implied by Posner's (1980) spotlight account and expressed in the model of Equation 11. According to a strict orienting account, no information about stimulus identity is extracted until after a central decision mechanism has been switched to, or aligned with, the stimulus location. This is what our Equation 11 expresses. According to the composite gain-orienting model of Figure 17, however, observers begin to extract information about the identity of unattended stimuli as soon as they appear, but at a slower rate before orienting than afterward. The distinction between a strict view of orienting and the view implied by Figure 17 is thus between a model in which no identity information is extracted prior to orienting and one in which information extraction prior to orienting is slow and inefficient. The distinction parallels that originally made in the classical auditory literature between Broadbent's (1958) filter model, which was an all-or-none switching model, and Treisman's (1960) attenuation model, which assumed that the efficiency of information acquisition from unattended stimuli is reduced.

Potentially, one might test between these alternatives by using a paradigm in which the stimulus event and the discriminative

information it carries are decoupled. The manipulation of interest would be one in which the visual structure that carries information about stimulus identity is delayed for a few tens of milliseconds after the energy change that signals the appearance of a new stimulus. According to a strict interpretation of orienting, delaying the availability of discriminative information should have no effect on performance if the delay is shorter than the orienting time, because no information about stimulus identity is extracted until after orienting has been completed. According to the composite gain-orienting account, however, any delay of discriminative information should degrade performance, because extraction of identity information begins as soon as the stimulus appears. We have not pursued the distinction between strict orienting and composite gain-orienting models any further here, as our purpose in this article was not to test experimentally between these two kinds of model. Rather, it was to show how different kinds of attentional mechanisms can provide detailed quantitative accounts of the RT distribution and accuracy data from cued signal-detection paradigms.

Extensions of the Theory

Selection From Multielement Displays

The theory developed in this article characterizes performance in the simplest attentional task: detecting a single stimulus in an otherwise empty visual field. We have focused on this task because it is well suited to distinguishing between the low-level processes involved in signal enhancement and uncertainty reduction (Luck et al., 1994; Smith, 2000a). When attempting to test between these accounts, the presence of distractor stimuli in the visual field adds an additional layer of complexity, because the processes involved in selecting target stimuli from among distractors are also then engaged. This makes it more difficult to draw inferences about the effects of attention on the processing of stimuli at cued and miscued locations, because these effects must be distinguished from target selection effects.

Nevertheless, for many researchers, the selection and processing of a target stimulus from among distractors is the hallmark of attention. For such researchers, these selective properties best characterize attention's presumed evolutionary role in processing stimuli in natural scenes. For this reason, it is important to know whether the theory presented here can generalize to tasks of this kind. We do not attempt to provide a detailed treatment of these issues in this article but, rather, limit ourselves to sketching how the theory can be developed in this way.

During the last 25 years or so, a number of researchers have investigated performance in near-threshold versions of the visual search task, in which observers are required to detect or identify weak visual targets presented among a background of distractor stimuli (e.g., Eckstein et al., 2000; Palmer et al., 1993; Shaw, 1982). Unlike the standard visual search task, in which the dependent variable is RT, the dependent variable in these studies was accuracy or contrast threshold, that is, the level of stimulus contrast needed to achieve a specified level of accuracy. One of the most highly replicated findings is that performance in such tasks is well described by some version of a maximum-outputs signal-detection model (Baldassi & Burr, 2004; Palmer et al., 2000). These models assume that preattentive processes identify the max-

imum, or most target-like, stimulus in the visual field. The value of this maximum is then compared to a criterion in order to make a decision. Formally, the decision rule in such models is

$$P(\text{Yes}) = P[X_i \geq c, \text{ where } X_i = \max(X_1 \dots X_n)].$$

This equation states that the probability the observer responds “yes” or “target present” is equal to the probability that the largest of a set of random variables, X_1, \dots, X_n , exceeds a criterion, c . The random variables code the strength of evidence that a target is present at each of n display locations. These models give a good account of how performance varies as a function of the number of stimuli in the display. Smith (in press) and Palmer et al. (2000) have provided detailed analyses of these models.

Although maximum-output models provide a good account of performance on search tasks, they provide no account of how the maximum of the set of random variables is actually computed by the visual system. Grossberg (1987, 1988) has shown how this problem is solved by systems of competitively interacting shunting equations. Specifically, he has shown that systems of shunting equations augmented with faster-than-linear (e.g., quadratic) feedback compute a maximum function on their inputs. The combination of feedback and competitive interactions among the stimuli leads to a VSTM representation in which the strongest, or most target-like, input is driven to its maximum, whereas all other stimuli are suppressed. Although Grossberg’s analysis has focused on noise-free, deterministic systems, his results carry over in a natural way to systems in which the inputs are subject to trial-to-trial perturbations by noise. There is thus a natural correspondence between systems of shunting equations and maximum-outputs, multichannel signal-detection models.

More generally, Grossberg has shown that systems of competitively interacting shunting equations augmented with sigmoidal feedback (faster-than-linear changing to slower-than-linear) exhibit what he terms a “quenching threshold.” This endows such systems with a form of limited capacity property in which the strongest inputs in the display are represented in VSTM and the remainder are suppressed. Smith (in press) argued that this is consistent with the results of cued visual search tasks, in which observers appear to form a VSTM representation of a subset of the most target-like stimuli in the display, but not of all of them. The repeated success of maximum-outputs models in predicting performance on such tasks, and their correspondence with systems of competitively interacting shunting equations, makes us believe that the shunting equation formalism is the correct one.

There is another way in which our theory can be extended to account for selection from multielement displays, which does not require the assumption of competitive interactions among stimuli for entry to VSTM. Bundesen (1990) has modeled selection from multielement arrays using a race model, in which display items are processed in parallel and in which items that finish processing first have priority of entry to VSTM. The model assumes that pertinent stimuli are processed more rapidly, increasing their likelihood of being selected into a limited-capacity memory system. Bundesen has shown that this model can predict accuracy as a function of display size and exposure duration in selective report paradigms and mean RTs in visual search and divided-attention tasks. In its assumption that attentional selection affects processing rates, Bundesen’s theory is similar to the gain models we have proposed here. In summary, then, although our theory was developed to

account for the detection of a single, cued stimulus in an empty display, its principles extend in a very natural way to tasks in which targets are presented among a background of distractors.

Detection, Discrimination, and Other Cued Perceptual Tasks

Earlier we noted that one of the most influential ideas in the classical selective attention literature was that the processes involved in identifying a stimulus could be divided into two classes: a class of focal attention processes that require access to a limited-capacity system and a class of preattentive processes that do not. According to this distinction, stimulus detection can be carried out preattentively; focal attention is required only for complex perceptual judgments like difficult discrimination or recognition of form. Although the data presented here are inconsistent with any simple form of attention–preattention dichotomy, it is nevertheless true that larger attentional effects are typically found for more difficult perceptual judgments. Indeed, a number of the studies reviewed earlier that found little or no attentional effect in detection also found large and systematic effects for more complex perceptual judgments (Bonnell & Hafter, 1998; Bonnell et al., 1992; Brawn & Snowden, 2000; Lee et al., 1997; Müller & Findlay, 1987; Palmer, 1994; Palmer et al., 1993; Shaw, 1984). Many studies that have found large attentional effects for discrimination, recognition, or acuity judgments have done so without backward masks (e.g., Bonnell & Hafter, 1998; Carrasco, Williams, & Yeshurun, 2002; Palmer et al., 1993; Shaw, 1984). Taken together, the results of these studies imply a Cue \times Mask \times Task interaction: When stimuli are well localized perceptually, backward masks are needed for cuing effects in detection but are not needed (or may not be needed) for more complex judgments.

The Cue \times Mask \times Task interaction follows straightforwardly from our VSTM model under the assumption that the information required to make complex judgments becomes available comparatively slowly. The idea that different parts of a stimulus may be processed at different rates is a fundamental property of multiresolution filter models of the visual system. Complex discrimination and recognition tasks often rely on the high spatial frequency content or fine spatial structure of stimuli, and this information is extracted comparatively slowly and becomes fully available only late in perceptual processing (Lupp, Hauske, & Wolf, 1976; Watt, 1987). These effects can be represented very simply in the VSTM model, as changes in the value of β_{on} , the sensory response function onset rate.

Figure 18 shows the effects on the VSTM trace of reducing the value of β_{on} for a briefly presented (50-ms) stimulus. To generate these predictions, the offset rate, β_{off} , was held fixed, and γ_A , the attention gain for cued stimuli, was set equal to twice γ_U , the gain for miscued stimuli. The other parameters of the model were chosen so the asymptotic VSTM trace strength was 1.0. As Figure 18a shows, for a stimulus of fixed duration, the effect of reducing β_{on} is to reduce the area under the sensory response function, $\mu(t)$. How closely the VSTM trace approaches its asymptote depends on the product of this area and the attention gain.

As shown in Figure 18b, when β_{on} is large, the asymptotic VSTM trace is independent of γ . The trace grows rapidly for cued stimuli and slowly for miscued stimuli, but the asymptotic strength for cued and miscued stimuli is the same. Under these circumstances, there is an RT effect but little or no effect on accuracy.

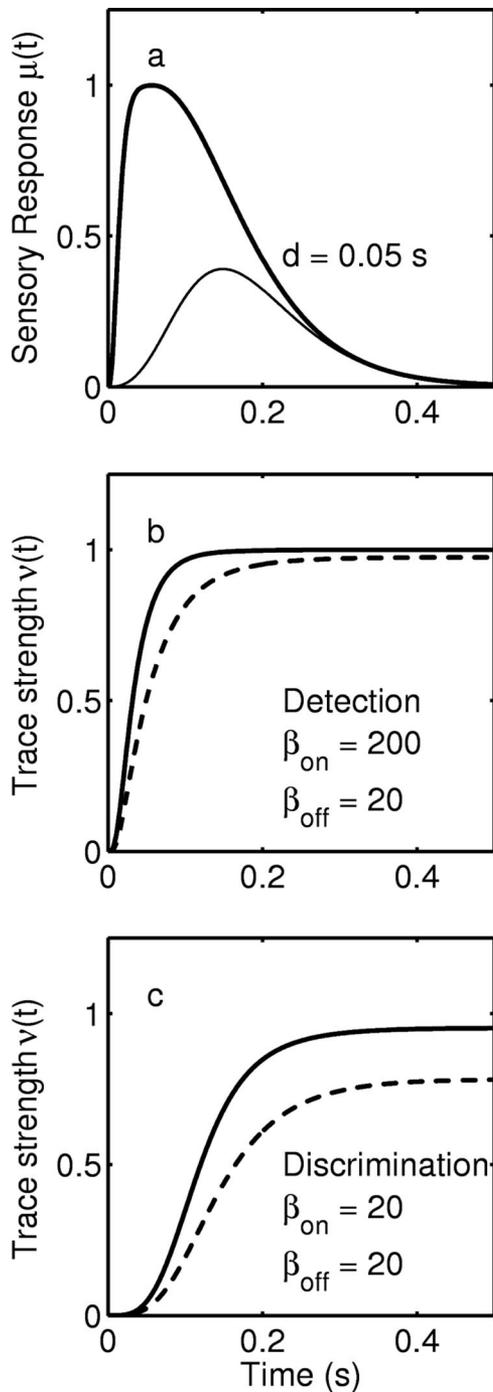


Figure 18. Predicted visual short-term memory (VSTM) traces for detection and discrimination. The model assumes that the information needed for discrimination becomes available more slowly than for detection. This is represented in the model by a reduction of β_{on} from 200 to 20. a. Sensory response functions for detection and discrimination of a 50-ms ($d = 0.05$ s) stimulus. Heavy line, $\beta_{on} = 200$; light line, $\beta_{on} = 20$. b and c. VSTM traces for cued (continuous line) and miscued (dashed line) stimuli for detection (b) and discrimination (c), respectively. Asymptotic VSTM trace strength for cued and miscued stimuli is different for discrimination but not for detection. These predictions are for unmasked stimuli.

This is what is found in unmasked detection tasks with perceptually well-localized stimuli. Figure 18c shows what happens when β_{on} is reduced and all other model parameters are held constant. The reduction in rate means that the area under the sensory response function is insufficient to drive the VSTM trace to asymptote. Cued stimuli then have an advantage because of their higher gain, because more of the VSTM trace will have formed before the stimulus decays to the point at which it cannot support further trace growth. Under these circumstances, there is both an RT effect and an accuracy effect, as is found empirically in difficult discrimination and recognition tasks.

The preceding analysis assumed a gain model of attention, but orienting models predict similar performance because of the interaction between β_{on} and the orienting time, t_i . Of course, we do not claim that all attentional cuing effects can be understood in such a simple way, as an interaction between rate of perceptual processing and efficiency of VSTM transfer. Nevertheless, the fact that a three-way interaction between task type, cue condition, and backward masking emerges in a natural way from our theory is one of its attractive features and is something we are investigating further.

External Noise Exclusion and the Perceptual Template Model

In a series of studies over the past decade, Doshier, Lu, and colleagues have investigated attentional cuing effects in low-level visual tasks within the framework of an extended signal-detection model they call the perceptual template model (Doshier & Lu, 2000a, 2000b; Lu & Doshier, 1998). One of their most replicated findings is that large and systematic cuing effects are obtained when stimuli are embedded in a background of external noise (Doshier & Lu, 2000b; Lu, Lesmes, & Doshier, 2002). In comparison, the cuing effects found in noiseless displays are somewhat smaller and less systematic and are found only with certain types of cues. Lu and Doshier have attributed these effects to the actions of an attention-dependent external noise-exclusion mechanism, which allows observers to efficiently filter out noise at the display location containing the target.

The study by Smith and Wolfgang (2007), described earlier, was carried out to test whether the mask-dependent cuing effect reported by Smith and colleagues could be viewed as a manifestation of Lu and Doshier's external noise mechanism. As a mask may be viewed as a source of noise in the display, the identification of the mask-dependent cuing effect with external noise exclusion is plausible, as Lu et al. (2002) suggested. Smith and Wolfgang argued that if mask-dependent cuing is mediated by an external noise-exclusion mechanism, then the cuing effect should be maximal when the target and mask are simultaneous and should decrease with increasing temporal separation between them. However, if the cuing effect is mediated by an interruption-masking mechanism, it should be maximal at target-mask SOAs that are optimal for producing interruption masking, typically around 80–100 ms. The latter is what they found. Backward masks produced large and systematic cuing effects; simultaneous masks produced much smaller effects, which were significant for only a minority of observers. Exactly the same results were obtained with checkerboard pattern masks and with noise masks. On the basis of these findings, Smith and Wolfgang argued that the mechanism under-

lying the mask-dependent cuing effect was not the same as Lu and Doshier's external noise-exclusion mechanism.

At the same time, however, the finding of weak but significant cuing effects when stimuli are presented with simultaneous masks contrasts with the results of our detection experiments with well-localized, unmasked stimuli, none of which has shown a significant cuing effect. Examples of this are the unmasked condition of Smith, Ratcliff, and Wolfgang (2004) in Figure 2, the fiducial condition of Gould et al. (2007) in Figure 4, and the unmasked condition of Smith and Wolfgang (2004) in Figure 7. Because the cuing effect with simultaneous masks could not have been due to interruption masking, Smith and Wolfgang (2007) argued that it could plausibly be attributed to an external noise-exclusion mechanism of the kind identified by Doshier and Lu. The difference in the magnitude of the external noise effects found by Smith and Wolfgang, which are small and unsystematic, and those of Lu and Doshier, which are large and systematic, can be attributed to differences in the difficulty of the perceptual judgments. Whereas Smith and Wolfgang used the orthogonal discrimination task, Lu and Doshier have typically used more difficult discrimination tasks and have often presented their targets in arrays of distractors to increase the magnitude of the cuing effect.

Given the evidence that the cuing effects with simultaneous masks and trailing masks are produced by different mechanisms, it is important to know whether both kinds of effects can be predicted by our theory. The theory predicts mask-dependent cuing effects via an interaction between attentional dependencies in the efficiency of VSTM transfer and the differential persistence of masked and unmasked stimuli. What about the cuing effects with simultaneous masks? Smith et al. (in press) argued that a VSTM model like the one presented here can predict increased cuing effects when stimuli are embedded in external noise, if one assumes that noise slows the rate at which a perceptual representation of the stimulus is formed. The argument exactly parallels that made in the preceding section for the relationship between detection and discrimination. If noise slows the formation of a perceptual representation of the stimulus, then its effects can be represented in the model as a reduction in β_{on} , the onset rate for the sensory response function. Reducing β_{on} reduces the area under $\mu(t)$, and as shown in the preceding section, reducing this area increases the magnitude of the cuing effect. Consequently, larger cuing effects are obtained with higher levels of external noise.

As the theory attributes the effects of external noise and judgment complexity (i.e., detection vs. discrimination) to the same cause, namely, a reduction in β_{on} , it predicts an interaction between noise and judgment complexity. The increase in the cuing effect produced by adding noise to the stimulus is small in detection and easy discrimination tasks and large in difficult discrimination, recognition, and acuity tasks. A comparison of the results of Smith and Wolfgang (2007) and the studies of Lu and Doshier show that this is indeed the case.

Although this account of the external noise effect is a speculative one, it is supported by unpublished data from Ratcliff's laboratory on the discrimination of pairs of letters embedded in external, dynamic noise (Smith, 2007, December). The RT distribution data from this task show a pronounced bowing of the first quantile in the quantile probability plot, similar to that in the data of Gould et al. (2007) in Figure 13. The bowing of the first quantile means that the leading edge of the RT distribution is shifted

progressively to the right as stimuli become more noisy. These effects are well described quantitatively by a version of the VSTM model in which the rate of VSTM formation decreases with decreasing frame-to-frame correlation in the stimulus. This correlation decreases with increasing stimulus noise. The results of Ratcliff's study are therefore consistent with the idea that VSTM formation is slowed when noise is added to the display. Our theory predicts that this will also lead to an increase in the cuing effect, as the studies of Doshier and Lu have shown.

Conclusions

The literature on attention and visual signal detection from the last 25 years is an inconsistent and confusing one. Studies using accuracy and RT as dependent variables have led to different conclusions about cuing effects in detection, and among accuracy studies, the results have differed, depending on the paradigm. The experimental data we have presented show why this confusion has arisen. Attentional cues do not produce consistent effects on detection accuracy; rather, the effects depend jointly on attention and on the other variables manipulated in the task. The mask-dependent cuing effect is one manifestation of this; the uncertainty-dependent cuing effect is another. Moreover, there is a dissociation in the effects of these variables on accuracy and on RT. Unlike accuracy, cuing effects in RT are found regardless of whether stimuli are backwardly masked or are well localized perceptually. The complexity of these relationships is only fully revealed in studies that jointly manipulate attention and other variables and that measure both accuracy and RT.

Uncertainty can be controlled experimentally by the use of pedestals or fiducial crosses. Theoretically, these manipulations have the same effect, of localizing the decision to a single region of the display. Consistent with this expectation, they abolish the cuing effect in sensitivity but preserve it in mean RT. However, their effects on the shapes of RT distributions are very different. When stimuli are localized with pedestals, most of the change in RT with changes in stimulus contrast occurs in the tail quantiles of the RT distribution; the leading edge of the distribution changes only slightly. When stimuli are localized with fiducial crosses, there is a large change in the leading edge.

We believe that these relationships can only be understood through a detailed analysis of how attention affects the processes of forming a perceptual representation of a stimulus and of making a decision about its identity. This requires a theory that links perceptual encoding, visual masking, attention, VSTM, and decision making in an integrated dynamic framework. The theory presented in this article provides an account of the mask-dependent cuing effect, the uncertainty-dependent cuing effect, the dissociation of accuracy and RT, and of the differences in the shapes of RT distributions when stimuli are localized with pedestals and with fiducial crosses. It also extends to account for the selection of stimuli from multielement displays, the Cue \times Mask \times Task interaction, and the effects of embedding stimuli in external noise.

The heart of the theory is a neurally plausible, shunting model of VSTM. The model assumes that attention increases the efficiency with which a VSTM trace is formed, either by increasing the rate at which stimulus information is transferred to VSTM or by reducing the delay before VSTM trace formation begins. Differences in the shapes of the RT distributions as a function of how stimuli are localized

perceptually are predicted by a dual-pathway model of VSTM. In this model, energy in the stimulus compound gates the form information that is used to make a decision into VSTM. Changes in the leading edge of the RT distribution reflect differences in the rate of VSTM formation that, in turn, depend on the contrast energy in the stimulus. The resulting theory offers a rich, detailed, and quantitatively precise account of fundamental attentional processes and of their effects on the speed and accuracy of performance.

References

- Baldassi, S., & Burr, D. C. (2004). "Pop-out" of targets modulated in luminance or colour: The effects of intrinsic and extrinsic uncertainty. *Vision Research*, *44*, 1227–1233.
- Bashinski, H. S., & Bacharach, V. R. (1980). Enhancements of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, *28*, 241–248.
- Bonnell, A.-M., & Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics*, *60*, 179–190.
- Bonnell, A.-M., Stein, J.-F., & Bertucci, P. (1992). Does attention modulate the perception of luminance changes? *Quarterly Journal of Experimental Psychology A*, *44*, 601–626.
- Bowling, A., & Lovegrove, W. (1981). Two components to visible persistence: Effects of orientation and contrast. *Vision Research*, *21*, 1241–1251.
- Boynton, G. (2005). Attention and visual perception. *Current Opinion in Neurobiology*, *15*, 465–469.
- Brawn, P. M., & Snowden, R. J. (2000). Attention to overlapping objects: Detection and discrimination of luminance changes. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 342–358.
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. Oxford, England: Clarendon Press.
- Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon Press.
- Brown, S. D., Ratcliff, R., & Smith, P. L. (2006). Evaluating methods for approximating stochastic differential equations. *Journal of Mathematical Psychology*, *50*, 402–410.
- Bundesden, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.
- Burr, D. C., Ross, J., & Morrone, M. C. (1985). Local regulation of luminance gain. *Vision Research*, *25*, 717–727.
- Busemeyer, J., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.
- Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, *101*, 446–469.
- Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research*, *42*, 949–967.
- Carrasco, M., & McElree, B. (2001). Covert attention speeds the accrual of visual information. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5363–5367.
- Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, *40*, 1203–1215.
- Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, *2*(6), 467–479.
- Cheal, M., & Lyon, D. R. (1992). Benefits from attention depend on the target type in location-precued discrimination. *Acta Psychologica*, *81*, 243–267.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*, 183–228.
- Davis, E. T., Kramer, P., & Graham, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception & Psychophysics*, *33*, 20–28.
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*, 197–215.
- Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, *41*, 260–274.
- Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322.
- Doshier, B. A., & Lu, Z.-L. (2000a). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*, 1269–1292.
- Doshier, B. A., & Lu, Z.-L. (2000b). Noise exclusion in spatial attention. *Psychological Science*, *11*, 139–146.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 188–202.
- Eckstein, M. P., Pham, B. T., & Shimozaki, S. S. (2004). The footprints of visual attention during search with 100% valid and 100% invalid cues. *Vision Research*, *44*, 1193–1207.
- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, *62*, 425–451.
- Egeth, H. (1977). Attention and preattention. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 277–320). New York: Academic Press.
- Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science*, *8*, 135–139.
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, *11*, 1710–1719.
- Foley, J. M., & Schwarz, W. (1998). Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold-versus-contrast function for contrast discrimination. *Journal of the Optical Society of America A*, *15*, 1036–1047.
- Francis, G. (2000). Quantitative theories of metacontrast masking. *Psychological Review*, *107*, 768–785.
- Giesbrecht, B., & Di Lollo, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1454–1466.
- Gould, I. C., Wolfgang, B. J., & Smith, P. L. (2007). Spatial uncertainty explains exogenous and endogenous attentional cuing effects in visual signal detection. *Journal of Vision*, *7*(13), 4.1–17.
- Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Wiley.
- Graham, N., Kramer, P., & Haber, N. (1985). Attending to the spatial frequency and spatial position of near-threshold visual patterns. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance* (Vol. 11, pp. 269–284). Hillsdale, NJ: Erlbaum.
- Green, D. M. (1960). Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, *32*, 1189–1203.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Grossberg, S. (Ed.). (1987). *The adaptive brain* (Vol. 1). Amsterdam: Elsevier.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, *1*, 17–61.
- Hallett, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York: Wiley.
- Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 802–811.
- Heath, R. A. (1992). A general nonstationary diffusion model for two choice decision making. *Mathematical Social Sciences*, *23*, 283–309.
- Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual

- cortex. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 119–134). Cambridge, MA: MIT Press.
- Kahneman, D. (1968). Methods, findings, and theory in studies of visual masking. *Psychological Bulletin*, *70*, 404–425.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.
- Lee, D. K., Koch, C., & Braun, J. (1997). Spatial vision thresholds in the near absence of attention. *Vision Research*, *37*, 2409–2418.
- Loftus, G. R., & Johnston, C. A., & Shimamura, A. P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 1–13.
- Lu, Z.-L., & Doshier, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*, 1183–1198.
- Lu, Z.-L., Lesmes, L. A., & Doshier, B. A. (2002). Spatial attention excludes noise at the target location. *Journal of Vision*, *2*, 312–323.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 887–904.
- Lupp, U., Hauske, G., & Wolf, W. (1976). Perceptual latencies to sinusoidal gratings. *Vision Research*, *16*, 969–972.
- McGillem, C. D., & Cooper, G. R. (1991). *Continuous and discrete signal and system analysis* (3rd ed.). Orlando, FL: Holt, Rinehart & Winston.
- Michaels, C. F., & Turvey, M. T. (1979). Central sources of visual masking: Indexing structures supporting seeing at a single, brief glance. *Psychological Research*, *41*, 1–61.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6*, 414–417.
- Morgan, M. J., Ward, R. M., & Castet, E. (1998). Visual search for a tilted target: Tests of spatial uncertainty models. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology A*, *51*, 347–370.
- Moulden, B., Kingdom, F., & Gatley, L. F. (1990). The standard deviation of contrast in random dot images. *Perception*, *19*, 79–101.
- Müller, H. J., & Findlay, J. M. (1987). Sensitivity and criterion effects in the spatial cuing of visual attention. *Perception & Psychophysics*, *42*, 383–399.
- Müller, H. J., & Humphreys, G. W. (1991). Luminance-increment detection: Capacity-limited or not? *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 107–124.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*, 1703–1721.
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 108–130.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*, 1227–1268.
- Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*, 1508–1532.
- Petrov, Y., Verghese, P., & McKee, S. P. (2006). Collinear facilitation is largely uncertainty reduction. *Journal of Vision*, *23*, 170–178.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.
- Pike, A. R. (1966). Stochastic models of choice behaviour: Response probabilities and latencies of finite Markov chain systems. *British Journal of Mathematical and Statistical Psychology*, *21*, 161–182.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.
- Ramachandran, V. S., & Cobb, S. (1995, January 5). Visual attention modulates metacontrast masking. *Nature*, *373*, 66–68.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461.
- Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, *21*, 161–182.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, *95*, 238–255.
- Ratcliff, R. (2001). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, *9*, 278–291.
- Ratcliff, R., Hasegawa, Y., Hasegawa, R., Smith, P. L., & Segraves, M. (2007). A dual diffusion model for single cell recording data from the superior colliculus in brightness discrimination task. *Journal of Neurophysiology*, *97*, 1756–1797.
- Ratcliff, R., & Rouder, J. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 127–140.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *19*, 278–289.
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, *65*, 523–535.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*, 408–424.
- Ratcliff, R., Thapar, A., Smith, P. L., & McKoon, G. (2005). Aging and response times. In J. Duncan, L. Phillips, & P. McLeod (Eds.), *Measuring the mind* (pp. 3–31). Oxford, England: Oxford University Press.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 431–481.
- Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*, *93*, 180–206.
- Saalman, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007, June 15). Neural mechanisms of visual attention: How top-down feedback highlights relevant locations. *Science*, *316*, 1612–1615.
- Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance* (Vol. 8, pp. 277–296). Hillsdale, NJ: Erlbaum.
- Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, *14*, 353–409.
- Shaw, M. L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance* (Vol. 10, pp. 109–121). Hillsdale, NJ: Erlbaum.
- Shaw, M. L., Mulligan, R., & Stone, L. (1983). A test between two-state and continuous-state attention models. *Perception & Psychophysics*, *33*, 338–354.

- Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2003). Comparison of two weighted integration models for the cueing task: Linear and likelihood. *Journal of Vision*, *3*, 209–229.
- Shiu, L., & Pashler, H. (1994). Negligible effects of spatial precuing on identification of single digits. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1037–1054.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*, 567–591.
- Smith, P. L. (1998). Attention and luminance detection: A quantitative analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 105–133.
- Smith, P. L. (2000a). Attention and luminance detection: Effects of cues, masks, and pedestals. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1401–1420.
- Smith, P. L. (2000b). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.
- Smith, P. L. (2007, December). *The computational dynamics of visual attention*. Paper presented at the 13th Australian Mathematical Psychology Conference, The Australian National University, Canberra, Australia.
- Smith, P. L. (in press). Attention and the detection of weak visual signals. In V. Coltheart (Ed.), *Tutorials in visual cognition*. Hove, Sussex, England: Psychology Press.
- Smith, P. L., Lee, Y.-E., Wolfgang, B. J., & Ratcliff, R. (in press). Attention and the detection of masked radial frequency patterns: Data and model. *Vision Research*.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, *44*, 1297–1320.
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.
- Smith, P. L., & Wolfgang, B. J. (2004). The attentional dynamics of masked detection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 119–136.
- Smith, P. L., & Wolfgang, B. J. (2007). Attentional mechanisms in visual signal detection: The effects of simultaneous and delayed noise and pattern masks. *Perception & Psychophysics*, *69*, 1093–1104.
- Smith, P. L., Wolfgang, B. J., & Sinclair, A. (2004). Mask-dependent attentional cuing effects in visual signal detection: The psychometric function for contrast. *Perception & Psychophysics*, *66*, 1056–1075.
- Speckman, P. L., & Rouder, J. M. (2004). A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review*, *11*, 574–576.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1–29.
- Sperling, G. (1984). A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 183–242). New York: Academic Press.
- Sperling, G., & Doshier, B. A. (1986). Strategy and optimization in human information processing. In K. R., Boff, L. Kaufman, J. P. Thomas (Eds.), *Handbook of perception and performance* (Vol. 1, pp. 2.1–2.65). New York: Wiley.
- Sperling, G., & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America*, *58*, 1133–1145.
- Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review*, *102*, 503–532.
- Tanner, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, *89*, 752–765.
- Tata, M. S. (2002). Attend now or lose it forever: Selective attention, metacontrast masking, and object substitution. *Perception & Psychophysics*, *64*, 1028–1038.
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, *18*, 415–429.
- Thomas, J. P., & Gille, J. (1979). Bandwidths of orientation channels in human vision. *Journal of the Optical Society of America*, *69*, 652–660.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, *25*, 168–199.
- Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 2, pp. 199–238). Hillsdale, NJ: Erlbaum.
- Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1033.
- Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*, 242–248.
- Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*, 1–52.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.
- Van der Heijden, A. H. C., & Brouwer, R. F. T. (1999). The effect of noise in a single-item localization and identification task. *Acta Psychologica*, *103*, 91–102.
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*, 523–535.
- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*, 37–58.
- Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
- Vidyasagar, T. R. (1999). A neuronal model of attentional spotlight: Parietal guiding the temporal. *Brain Research Reviews*, *30*, 66–76.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1436–1451.
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relationship between the mean and standard deviation of a response time distribution. *Psychological Review*, *114*, 830–841.
- Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A., Schnapf, J. L. (1990). The control of visual sensitivity. In L. Spillman & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 53–102). San Diego, CA: Academic Press.
- Watson, A. B. (1986). Temporal sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and performance* (Vol. 1, pp. 6.1–6.85). New York: Wiley.
- Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express*, *6*, 12–32.
- Watt, R. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, *4*, 2006–2021.
- Webster, M. A., & De Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate cortex cells. *Journal of the Optical Society of America A*, *2*, 1124–1132.
- Wilson, H. R., & Kim, J. (1998). Dynamics of a divisive gain control in human vision. *Vision Research*, *38*, 2735–2741.

(Appendixes follow)

Appendix A

Solution of Shunting Equations

This appendix shows the solution of the shunting equation, Equation 4,

$$\frac{dv}{dt} = \Delta_I \mu(t) [\theta - v(t)] - I_0 \mu(t) v(t).$$

Equations of this form can be solved either by separation of variables or by introducing an integrating factor that renders the equation exact. Here we use separation of variables. Smith, Lee, Wolfgang, and Ratcliff (in press) described the use of the integrating factor method. We write the equation as

$$\frac{dv}{dt} = \mu(t) [\theta \Delta_I - (\Delta_I + I_0) v(t)].$$

and then separate the variables, to obtain

$$\frac{dv}{\theta \Delta_I - (\Delta_I + I_0) v} = \mu(t) dt.$$

The left hand side can be made into an exact differential by writing the equation as

$$- \left[\frac{1}{\Delta_I + I_0} \right] \left[\frac{-(\Delta_I + I_0) dv}{\theta \Delta_I - (\Delta_I + I_0) v} \right] = \mu(t) dt.$$

Multiplying through by $-(\Delta_I + I_0)$ and then integrating both sides of the equation yields

$$\log[\theta \Delta_I - (\Delta_I + I_0) v(t)] = -(\Delta_I + I_0) \int_0^t \mu(s) ds + K,$$

where K is the constant of integration. We assume the initial condition $v(0) = 0$, which implies that $K = \log(\theta \Delta_I)$. Substituting this value in the preceding equation and exponentiating gives

$$\theta \Delta_I - (\Delta_I + I_0) v(t) = \theta \Delta_I \exp\left[-(\Delta_I + I_0) \int_0^t \mu(s) ds\right],$$

which, after rearrangement, yields

$$v(t) = \theta \left(\frac{\Delta_I}{\Delta_I + I_0} \right) \left\{ 1 - \exp\left[-(\Delta_I + I_0) \int_0^t \mu(s) ds\right] \right\},$$

which is Equation 5 in the text. The other equations in the text differ from Equation 4 only in their excitatory and inhibitory coefficients and, so, can be solved in an identical way.

The VSTM model (Equations 16, 19, and 20) uses symmetrical shunting equations in which the inhibitory coefficient is equal to one minus the excitatory coefficient. This allows asymptotic VSTM trace strength to be decoupled from the rate of trace growth. These equations exploit the fact that an excitatory-inhibitory shunting equation of the form

$$\frac{dv}{dt} = b\{a\mu(t)[1 - v(t)] - (1 - a)\mu(t)v(t)\}$$

can be rewritten as a simple, one-term equation:

$$\frac{dv}{dt} = b\mu(t)[a - v(t)],$$

with rate b and asymptote a . Although we could have assumed a one-term equation in developing our VSTM model, we reject one-term models on theoretical grounds because they assume that the asymptote is a property of the memory, rather than the stimulus. This property has no principled psychological interpretation. For the asymptote to be a property of the stimulus, it must be a coefficient of $\mu(t)$, the sensory response function. The symmetrical shunting representation provides a way to transfer stimulus properties "across the brackets," making the asymptote equal to a , the excitatory coefficient of $\mu(t)$. Equations of this form contrast with Equation 4 and Equation 7, in which the asymptote is equal to the excitatory coefficient divided by the sum of the excitatory and inhibitory coefficients. The rate is equal to the sum of the excitatory and inhibitory coefficients multiplied by the gain. Such equations imply a close relationship between rate and asymptote, which is not general enough to account for our data.

Appendix B

Rate of VSTM Growth

The rate of VSTM growth in Equations 13, 16, 19, and 20 depends on I_C , the square root of the power in the stimulus compound. We define

$$I_C = \sqrt{\Delta_I^{2p} + I_p^{2p}},$$

where Δ_I^2 and I_p^2 are, respectively, the power in the patch and the power in the pedestal. This generalizes the simple, amplitude-unit, measure of compound power, $\sqrt{\Delta_I^2 + I_p^2}$, by allowing separate, nonlinear transduction of the components prior to computing

power. The value I_C defined in this way reduces to Δ_I^p when no pedestal is present, as required.

The components of I_C depend on stimulus contrast and area. Let $\Delta_I(x, y)$ be the deviation of the patch intensity from its background (i.e., the pedestal) at display location (x, y) , expressed in Weber contrast units. Then

$$\Delta_I^2 = \iint \Delta_I^2(x, y) dx dy,$$

where the integration is over the spatial extent of the patch. A similar integration for the pedestal gives $I_p^2 = a\Delta_p^2$, where a is the area of the pedestal and Δ_p is the intensity difference between the pedestal and the surrounding uniform field. As noted in the text, I_C is similar to the root-mean-square measure of stimulus contrast that has been used by many authors, except that, unlike RMS contrast, it is not normalized for stimulus area.

Watson (2000), among others, has advocated total contrast energy as a psychophysically appropriate measure of stimulus strength. Total contrast energy is obtained by integrating the square of the local contrast, $\Delta_I^2(x, y)$, over the area and duration of the stimulus. The unit of measurement, termed a Barlow unit, is chosen so that the contrast energy at detection threshold will be about one. Total contrast energy differs from root-mean-square contrast in that it is an energy measure rather than an amplitude measure, and it is not normalized for stimulus area. The measure I_C is similar to the measure proposed by Watson, except that it is based on power rather than energy. Energy dependency is obtained via the integration over time involved in solving the differential equation, that is, in forming the VSTM trace.

An alternative way to define I_C would be to express Δ_I and I_p in units of power rather than amplitude and to assume that the rate of VSTM formation is proportional to the power in the compound

rather than its square root. That is, $\Delta_I = (\Delta'_I)^2$ and $I_p = (I'_p)^2$, where Δ'_I and I'_p are the intensities of the patch and the pedestal in amplitude units. Then I_C takes the simpler form

$$I_C = \Delta_I^p + I_p^p,$$

which also reduces to Δ_I^p when no pedestal is used. The representations in power and amplitude units make virtually indistinguishable predictions, except in the joint fit of the dual-channel model to the Smith, Ratcliff, and Wolfgang (2004) and Gould, Wolfgang, and Smith (2007) data, where the more complex, amplitude-unit representation yielded an 8% better fit.

For a Gabor patch of specified frequency, bandwidth, and duration, measures of stimulus strength based on root or root-mean contrast power or contrast energy are proportional to peak envelope contrast (i.e., the Weber contrast of the peak of the Gabor patch's Gaussian envelope). Because our stimuli were specified using peak envelope contrast, we assumed peak envelope contrast in fitting our models and allowed any difference between units to be absorbed by the VSTM asymptote parameter, θ .

Received July 9, 2008

Revision received November 2, 2008

Accepted November 5, 2008 ■

Call for Nominations: *Health Psychology*

Division 38 (Health Psychology) is currently accepting nominations for the editorship of *Health Psychology* for the years 2011-2016. Robert M. Kaplan is the incumbent Editor.

Candidates should be members of Division 38 and of APA, and should be available to start receiving manuscripts in 2010 to prepare issues to be published in 2011. Division 38 encourages participation by members of underrepresented groups and would welcome such nominees. Self-nominations are also encouraged.

Kevin D. McCaul, Ph.D., has been appointed as Chair for this search.

To nominate candidates, prepare a statement of two pages or less in support of each candidate, and provide a current CV. Submit all materials electronically to: apadiv38@verizon.net.

The deadline for receipt of nominations is April 15, 2009.