Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 1297–1320
Vision Research
www.elsevier.com/locate/visres

Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays

Philip L. Smith [a,*], Roger Ratcliff [b], Bradley J. Wolfgang [a]

[a] Department of Psychology, University of Melbourne, Victoria 3010, Australia
[b] The Ohio State University, USA

Received 9 June 2003; received in revised form 30 December 2003

Abstract

Mask-dependent cuing effects, like those previously found in yes–no detection, were found in a task in which observers judged the orientations of orthogonally-oriented Gabor patches presented at cued or uncued locations. Attentional cues enhanced sensitivity for masked, but not unmasked, stimuli. Responses were faster to cued than to uncued stimuli, irrespective of masking. The distributions of response times and accuracy were well described by a diffusion process model of decision making. Mask-dependent cuing was explained by an orienting model in which: (a) decisions are based on stable stimulus representations in visual short term memory that determine the rate of evidence accumulation in the diffusion process; (b) inattention delays the entry of stimuli into short term memory, and (c) masks limit the visual persistence of stimuli.

1. Introduction

When an observer makes a psychophysical decision about a weak, briefly-presented stimulus, the time required to make the decision, as measured by the observer's response time (RT), may exceed the time for which the stimulus is present by a factor of 10 or more. For example, RTs to detect or identify stimuli exposed for a few tens of milliseconds may be of the order of several hundred milliseconds or longer, depending on the difficulty of the task (Luce, 1986; Teichner & Krebs, 1972). During the time taken to make the decision, the observer has access to a decaying stimulus trace, or "icon," whose duration, as a function of the time since stimulus offset, depends on such factors as the background luminance of the display (Averbach & Sperling, 1960) and the spatial frequency composition of the stimulus (Di Lollo & Woods, 1981). Under normal viewing conditions, the icon may retain vestiges of usable stimulus information for several hundred milliseconds or more, but if the display is masked by a backward, pattern mask, the erasure of the icon is almost immediate (Turvey, 1973).

In this article, we ask the question: What is the relationship between the physical stimulus trace and the observer's decision and how is this relationship affected by spatial attention? The fact that observers can render accurate judgments about stimuli that are masked after a few tens of milliseconds seems to imply that stimulus information is preserved in some relatively durable form that survives the destruction of the icon. Indeed, the existence of such a durable, posticonic, visual short term memory (VSTM) is well established (Phillips, 1974), as is the role of spatial attention in the selective transfer of stimulus information from the icon to VSTM (Sperling, 1960). Here we seek to investigate the relationship between attention and VSTM in simple psychophysical decisions.

The models that give the most natural account of accuracy and RT in simple decision tasks are sequential sampling models. These models, which have been developed during the last 40 years in mathematical psychology, assume the representation of stimuli in the visual system is noisy, as assumed in signal detection theory (Green & Swets, 1966). As in signal detection theory, the noise may be inherent in the stimulus itself, or may reflect processes that are internal to the observer. To make a decision about the presence or the identity of a stimulus, successive samples of the noisy stimulus trace are accumulated until a criterion amount of information

* Corresponding author. Tel.: +61-3-8344-6343; fax: +61-3-9347-6618.
E-mail address: philipls@unimelb.edu.au (P.L. Smith).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.01.002

needed for a response is obtained. The time taken to accumulate this information is identified with the decision component of RT. Sequential sampling models differ from signal detection theory in that they assume that the decision time depends on the information content of the sequence of noisy samples, whereas in signal detection theory it does not.

This structure allows sequential sampling models to predict two basic, but important, features of simple decisions. First, mean RT and accuracy, considered as functions of the stimulus condition, are negatively correlated (Pachella, 1974). As the difficulty of the decision increases, accuracy decreases and RTs get longer. The models predict negatively correlated speed and accuracy because the rate of information accumulation depends on the signal strength of the stimuli. As the signal strength decreases, so too does the rate of information accumulation, producing a decrease in accuracy and an increase in RT. Second, observers are able to trade off speed and accuracy and respond rapidly and inaccurately, or slowly and accurately, as the task requires. The models predict speed–accuracy tradeoffs because decision criteria are assumed to be under the observer's control. By adopting low criteria, observers can respond rapidly, on the basis of a small amount of information, but at the expense of sacrificing accuracy. By adopting high criteria, they increase the amount of information needed for a decision, resulting in higher accuracy, but at the expense of longer RTs. Introductions to these models may be found in Townsend and Ashby (1983) and Luce (1986). A recent comparative analysis may be found in Ratcliff and Smith (in press).

1.1. Stationary and nonstationary decision models

Signal detection theory assumes that a decision is based on a single sample of noisy stimulus information; sequential sampling models assume it is based on a sequence of noisy samples. This sequence is modelled mathematically as a stochastic process—that is, as a probabilistic process extended in time—whose statistics depend on the stimulus encoding mechanisms and moment-by-moment fluctuations in noise within the observer. If this process is identified theoretically with the decaying iconic trace, it is natural to model it as a *nonstationary* process, that is, as a process whose mean, and possibly also whose variance, change as a function of time. The nonstationarity of the process reflects the fact that the information extractable from the stimulus decreases progressively with the decay of the icon.

However, as suggested previously, there is another possibility. If the observer's decision is not based directly on the decaying iconic trace, but on a durable representation of the stimulus, computed during the period of stimulus presentation and retained in VSTM, then the only changes that would be expected during the course of the observer's decision would be those associated with moment-to-moment fluctuations in internal noise. Under these circumstances, it should be possible to model the process as a *stationary* stochastic process, that is, as one whose statistics do not change over time—at least not during the period of a second or so needed to make a decision.

Historically, most sequential sampling models have assumed stationary stimulus representations (e.g., Edwards, 1965; Laming, 1968; Link & Heath, 1975; Pike, 1966; Ratcliff, 1978; Stone, 1960; Vickers, 1970), but recently some researchers have investigated nonstationary sequential sampling models (Diederich, 2003; Heath, 1992; Ratcliff, 1980; Smith, 1995). Heath (1992) proposed that the distributions of RT to briefly-flashed stimuli could be described by a model that combined a linear-filter stimulus encoding stage and a sequential sampling decision stage. Smith (1995) proposed such a model to account for the effects of sustained and transient mechanisms on the distributions of simple RT. The model also accounts for Bloch's law and the differential breakdown of Bloch's law at long durations for high and low spatial frequency stimuli (Smith, 1998a).

The distinction between stationary and nonstationary decision models is illustrated in Fig. 1. In both kinds of

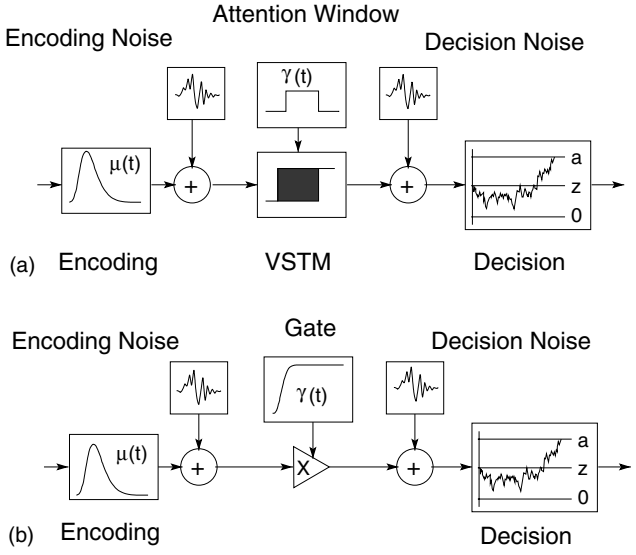

Fig. 1. Stationary and nonstationary models. The noisy output of a visual filter, $\mu(t)$, is accumulated by a sequential sampling decision stage, which is represented as a diffusion process between absorbing boundaries, 0 and a, starting at z at time $t = 0$. (a) Stationary model. The output of the encoding stage is admitted to VSTM by an attention window, $\gamma(t)$. The drift of the diffusion process is proportional to the integral of the filter output over the attention window. (b) Nonstationary model. The drift of the diffusion process is proportional to the product of the filter output, $\mu(t)$, and an attention gate, $\gamma(t)$, that determines the rate of information gain from attended and unattended locations. In (a) the drift is constant over the course of a trial; in (b) it is time-varying.

model, the perceptual effect of a stimulus is represented by a time-varying encoding function, $\mu(t)$. When a stimulus is presented briefly with no mask, $\mu(t)$ is identified with the information in the decaying iconic trace. The information in $\mu(t)$, perturbed by noise, forms the basis for the observer's decision. As shown in the figure, noise may enter the system either while the stimulus is being encoded or while a decision is being made. The key difference between the two kinds of models is that the decision mechanism in the nonstationary model is driven directly by the decaying iconic trace whereas in the stationary model it is driven by the contents of VSTM. In other words, in the nonstationary model the encoding function and the decision mechanism are directly coupled whereas in the stationary model they are separated by an intervening stage of VSTM.

A natural way to represent how a constant VSTM trace is formed from a time-varying encoding function in stationary models is via an *attention window* (e.g., Reeves & Sperling, 1986; Shih & Sperling, 2003). The attention window computes area under the encoding function (i.e., its integral) over a set of preceding time values. The output of the window is proportional to this area and determines the strength of the VSTM trace. The more discriminable the stimulus, the greater the amplitude of the encoding function, and the stronger the VSTM trace. The attention-window formulation also expresses the idea that entry into VSTM is selective. Relative to the capacity of the icon, the capacity of VSTM is sharply limited (Gegenfurtner & Sperling, 1993; Sperling, 1960), so a selection mechanism is needed to resolve competition between stimuli whenever the number of stimuli in the visual field exceeds the capacity of VSTM.

Fig. 1 also depicts an analogous selection mechanism for nonstationary models. In this *attention gating* model, the rate at which information flows from the encoding stage to the decision mechanism is regulated by attention. Information is accumulated rapidly from attended locations and slowly from unattended locations. For a given stimulus, the accumulation rate depends jointly on the gate and the encoding function, both of which may vary with time. Elsewhere we have proposed a gated nonstationary model to explain attention cuing effects in luminance detection (Smith, 2000a; Smith & Wolfgang, in press; Smith, Wolfgang, & Sinclair, in press). Experimental support for the model's main assumption, that the rate of information accumulation is affected by attention, was provided by Carrasco and McElree (2001).

The attention gate and attention window can be contrasted with two more general mechanisms of attention control that have been proposed in the literature, namely, attention *orienting* and attention *gain*. The idea of attention orienting goes back to the filter theory of Broadbent (1958) and, in vision, to the "spotlight" account of Posner (1978, 1980) and Posner, Snyder, and Davidson (1980). Orienting models assume that, in order to identify a stimulus, a central mechanism must be shifted to the stimulus location. Benefits occur when attention is directed to the stimulus location before stimulus presentation because the delay associated with shifting the central mechanism is eliminated. Costs occur when attention is allocated to the wrong location because there is an additional delay in disengaging the central mechanism before it can be reengaged elsewhere. Dobkins and Bosworth (2001) recently proposed that the attentional effects in a dot motion task could be explained by a model of this kind.

In contrast to attention-orienting models, gain models assume that attention increases the signal-to-noise ratio for stimuli at attended locations relative to that at unattended locations (e.g., Hawkins et al., 1990; Smith, 1998b). The increase in signal-to-noise ratio may be conceived of as occurring either by a process of *signal enhancement* or of *noise reduction* (Shiu & Pashler, 1994). Signal enhancement is an attention-dependent increase in the quality of stimulus information extracted from attended locations (i.e., in the numerator of the signal-to-noise ratio). Noise reduction is a decrease in the aggregate amount of noise entering into the decision, whether from the stimulus location or from elsewhere in the display (i.e., in the denominator of the signal-to-noise ratio). Evidence for both kinds of mechanism have been reported in the literature (e.g., Cameron, Tai, & Carrasco, 2002; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Dosher & Lu, 2000; Foley & Schwarz, 1998; Lu & Dosher, 1998).

Orienting and gain mechanisms can be combined with either stationary or nonstationary sequential sampling models to produce plausible models for attentional effects in tasks in which both RT and accuracy are measured. Orienting can be represented as a delay in the opening of an attention window or gate and gain can be represented as a change in its maximum amplitude. In this article, we show that the RT distributions and accuracy measures in a luminance discrimination task are well described by a stationary model, both when backwardly-masked and unmasked stimuli are used. We interpret this as evidence that the perceptual basis of observers' psychophysical decisions are stable stimulus representations in VSTM. We then show that the effects of attention in this task can be described by an orienting model, in which the formation of stable VSTM traces for unattended stimuli is delayed relative to those for attended stimuli. When stimuli are masked, this leads to a degradation of the information extracted from the stimulus and a decrement in performance.

2. Mask-dependent attentional cuing in visual signal detection

Evidence for a fundamental relationship between attention and visual masking has come from a number

of recent studies. In metacontrast and objection substitution masking paradigms, stronger masking effects are obtained when the stimulus to be masked is unattended (Enns & Di Lollo, 1997; Ramachandran & Cobb, 1995; Tata, 2002). In detection, orientation discrimination, and character recognition tasks, attentional benefits are typically larger when stimuli are backwardly masked (Giesbrecht & Di Lollo, 1998; Kawahara, Di Lollo, & Enns, 2001; Morgan, Ward, & Castet, 1998).

In visual signal detection, the idea that attention increases the detectability of luminance stimuli only when they are backwardly masked has provided an explanation for the inconsistent results previously obtained in this area. Whereas some studies have found that detection sensitivity is selectively increased for signals at attended locations (Bashinski & Bacharach, 1980; Downing, 1988; Hawkins et al., 1990; Luck et al., 1994; Müller & Humphreys, 1991; Smith, 1998b), others have found either no effect of attention or effects that are attributable to the effects of uncertainty reduction alone (Bonnel, Stein, & Bertucci, 1992; Davis, Kramer, & Graham, 1983; Foley & Schwarz, 1998; Graham, Kramer, & Haber, 1985; Lee, Koch, & Braun, 1997; Müller & Findlay, 1987; Palmer, 1994; Palmer, Ames, & Lindsey, 1993; Shaw, 1984). Smith (2000a) pointed out that with few exceptions (e.g., Carrasco et al., 2000; Lee et al., 1997) studies finding increased detection sensitivity for attended stimuli were performed with backwardly-masked displays, whereas those finding no increase were performed without masks (see Smith, 2000a, Table 1). [1]

Direct confirmation of the hypothesis that masking is the critical variable that distinguishes between these studies was obtained by Smith (2000a), Smith and Wolfgang (in press), and Smith et al. (in press). In these studies the effects of attention on the detectability of masked and unmasked Gabor patch stimuli was investigated using a spatial cuing paradigm. To allow the effects of attention to be distinguished from those of spatial uncertainty, stimuli were presented atop a suprathreshold luminance pedestal rather than directly against a uniform field. Because the location of the pedestal was always clearly visible, irrespective of the cue condition, this manipulation decoupled variations in attention from variations in uncertainty about stimulus location, thereby eliminating the latter as an influence on performance.

In addition, the studies used a display that contained only a single stimulus that appeared at either a cued or a miscued location. This allowed the effects of the cue on stimulus detectability to be distinguished from its effects on target selection. When a target stimulus is presented among a background of distractors, cuing the target location may benefit performance by helping the observer select the target from among the distractors by a process of *distractor exclusion* (Lu & Dosher, 1998; Shiu & Pashler, 1994). These effects are often well described by a multichannel signal detection model in which each distractor or distractor feature acts as an independent source of noise (Eckstein, Thomas, Palmer, Shimozaki, & Steven, 2000; Foley & Schwarz, 1998; Palmer et al., 1993). By using a display containing only a single stimulus, the studies eliminated the possibility of a cuing effect caused by distractor exclusion.

In all three studies, *mask-dependent cuing effects* were found. When stimuli were masked by backward, pattern masks, detection sensitivity (as measured by signal detection d' or a related statistic) was greater for attended than unattended stimuli. When stimuli were presented without masks, detection sensitivity for attended and unattended stimuli did not differ. These effects have been obtained with a number of different cue configurations and stimulus manipulations.

Smith (2000a) and Smith and Wolfgang (in press) explained the mask-dependent cuing effect in single-element displays using a nonstationary diffusion process model like that shown in Fig. 1. Mask-dependent cuing arises in this model from an interaction of two factors: differences in the rates of information accumulation from attended and unattended stimuli, and differences in the visual persistence of masked and unmasked stimuli. The model provides an account of the main psychophysical phenomena (mask-dependent cuing as a function of stimulus contrast and exposure duration), but it relies on the assumed nonstationarity of the decision process, that is, on the assumption that the decision mechanism samples directly from a decaying perceptual trace.

Contrary to the assumptions made by Smith (2000a) and Smith and Wolfgang (in press), Ratcliff and Rouder (2000) found evidence against nonstationary decision models in a letter discrimination task with backwardly-masked stimuli. They pointed out that nonstationary models predict an increase in RT distribution skewness as stimulus discriminability is reduced that exceeds that typically found in empirical data. This increase in skewness arises because the decision mechanism is driven by the information in the decaying perceptual trace. As the trace decays, the rate of information accumulation progressively slows. If the trace decays completely before the evidence needed for a response has been accumulated, the decision mechanism samples noise until a criterion amount of evidence has been attained.

[1] Lee et al. (1997) found no effect of attentional load on the ability to discriminate vertical from horizontal gratings that were masked after 120 ms. However, their data showed no differences in the contrast thresholds for the masked and unmasked versions of the task, suggesting that, at the stimulus-mask asynchronies used, the masks had no effect on performance. The results of Carrasco et al. (2000) are considered in Section 5.

Such responses are made more slowly than they would be otherwise because the decision mechanism is driven only by noise, rather than by stimulus information, at least for part of the decision time.

In this article, we investigate the effects of attention on the speed and accuracy of judgments about grating patch stimuli, like those used in previous studies of mask-dependent cuing by Smith (2000a), Smith and Wolfgang (in press), and Smith et al. (in press). Our study differs from previous studies of this kind in that we measured both RT and accuracy for decisions about near-threshold stimuli. Our aim in collecting data of this kind was to obtain more detailed information about the properties of the underlying attentional mechanisms than can be obtained from accuracy measures alone.

Most studies of RT typically use suprathreshold stimuli, whereas studies that use near-threshold stimuli typically consider only accuracy data. Often, especially when attention is manipulated, accuracy data are interpreted using signal detection models. The data we report show systematic dependencies of both RT and accuracy on stimulus contrast, which signal detection models cannot explain. Such data have a natural interpretation within a sequential sampling framework, as we show subsequently.

3. Discrimination of orthogonal orientations: effects of cues and backward masks

Smith (2000a), Smith and Wolfgang (in press) and Smith et al. (in press) found mask-dependent cuing effects using a yes–no (''signal present''/''signal absent'') detection task. Here we used an easy discrimination task, in which observers discriminated between horizontal and vertical grating patches. Because contrast thresholds in this task are indistinguishable from those for yes–no judgments (Thomas & Gille, 1979), a number of investigators have argued that it can be used as a proxy for detection in studies of attention (Carrasco et al., 2000; Lee et al., 1997). If so, it should yield similar mask-dependent cuing effects to those previously obtained in yes–no tasks. We used orthogonal discrimination instead of yes–no detection in this study in order to ascertain whether the two tasks were in fact functionally equivalent, in the sense of producing similar patterns of mask dependencies in a cuing paradigm. We also wished to avoid the asymmetries in the distributions of RT for the two responses that are often found in yes–no tasks.

3.1. Method

Observers. Six paid undergraduate volunteers, all with normal or corrected-to-normal vision served as observers. Each observer served in a total of 20 experimental sessions, each of around one hour duration, preceded by 7–8 practice and calibration sessions.

Apparatus. Stimuli were presented on a 17 in. Sony 200PS monitor driven at a frame rate of 100 Hz by a Cambridge Research Systems VSG 2/4 15-bit framestore housed in a Pentium computer. The display response was linearized (gamma corrected) from measurements made with a Pritchard PR8800 photometer. Stimulus presentation and response recording were controlled by software written in C++. The computer's clock chip was reprogrammed to tick in milliseconds to allow accurate measurement of RT. Observers performed the task in a dimly-lit laboratory at a viewing distance of 50 cm. Viewing position was stabilized with a chin rest.

Stimuli. The stimuli were circularly symmetrical, sine-phase Gabor patches presented atop circular, 15% contrast luminance pedestals, which were displayed on a 25° square, 30 cd/m^2, uniform field. The mathematical form of the Gabor patch stimuli was as given by Graham (1989, p. 53). The sinusoid had a period of 6 pixels; the Gaussian envelope had a space constant (full width at half height) of 10 pixels, giving a bandwidth of 0.78 octaves. Viewed from 50 cm, the spatial frequency of the sinusoid was 3.5 cycles/deg and the width of the Gabor patch (half height) was 28′. The diameter of the luminance pedestal was 1.47° (30 pixels). The backward mask was a Gaussian vignetted checkerboard, consisting of alternating 17′ (6 pixels) squares whose contrasts were modulated by a circularly symmetrical Gaussian vignette with a space constant of 42′ (15 pixels) and a peak contrast, relative to the luminance of the pedestal of 95%. Examples of the stimuli and mask are shown in Fig. 2.

The attentional cues consisted of four black corner markers that identified the corners of a 1.8° square, centered on the target location. The cues were flashed for 60 ms, at a cue–target stimulus onset asynchrony (S0A) of 140 ms, and then extinguished. A pure exogenous (50% validity) cuing manipulation was used: Stimuli appeared at the cued location and at uncued locations with equal frequency.

On any trial there were three potential stimulus locations, one cued and two uncued, located at an angular separation of 120° on the circumference of an imaginary 3.2° radius circle, centered on a fixation cross. On each trial, a randomly chosen angle α ($0 < \alpha \leqslant 360°$) determined the position of the cued location. The possible uncued locations were at $\alpha \pm 120°$, these locations being chosen on uncued trials with equal frequency. This display configuration has the property that the two possible uncued locations are equidistant from the cued location and thus, according to a symmetrical model of attentional gradient effects, should receive equal processing resources.

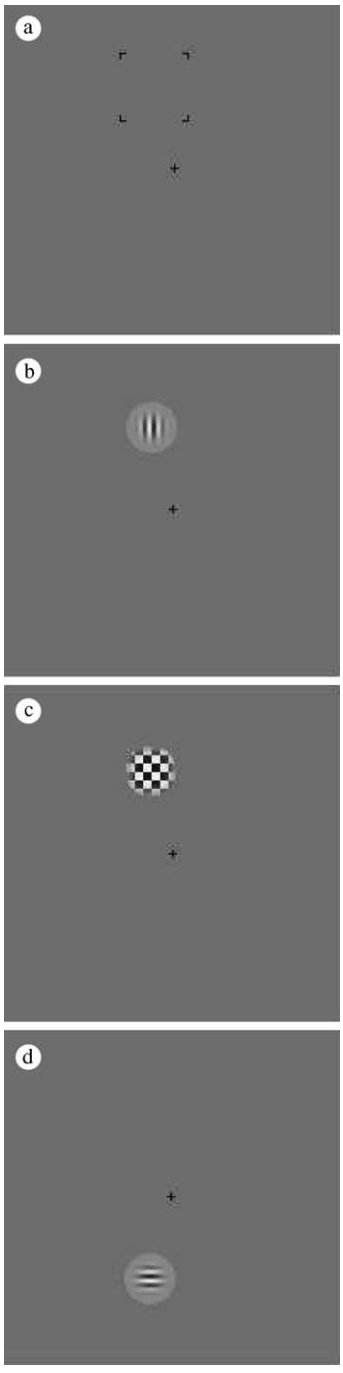

Fig. 2. Example stimuli. (Top to bottom) (a) Cue stimulus, (b) vertical pedestal Gabor at cued location, (c) checkerboard backward mask, (d) horizontal pedestal Gabor at uncued location. Cues were flashed for 60 ms at a stimulus onset asynchrony of 140 ms. Stimuli were presented for 60 ms and followed immediately by the mask on masked trials or erased on unmasked trials.

Procedure. The experiment was run in two phases, each consisting of 10 experimental sessions. In one phase, observers discriminated the orientation of masked, pedestal Gabors; in the other they performed the same task without masks. Each session consisted of 400 experimental trials, divided into 20 blocks of 20 trials. Stimuli were presented for 60 ms at five equally-spaced contrasts, the values of which were chosen individually for each observer during practice to produce a range of performance that varied from near chance to near perfect. Equal numbers of horizontal and vertical stimuli, and cued and miscued stimuli, were presented at each of the five levels of contrast, in random order, in each session. On cued trials the stimulus was presented at the previously cued location (α); on miscued trials it was presented at one of the other two possible stimulus locations ($\alpha \pm 120°$). Because the time from cue onset to stimulus offset (200 ms) was too short to refixate the display (Robinson, 1965), eye movements were not monitored.

Observers were instructed to try to perform the task accurately but not to spend too much time on each decision. They were told that their decision times were being measured but that they should not regard the task as an RT task. They were instructed to maintain central fixation during the course of each trial but to use the cue to direct their attention to the cued location.

Each trial began, after a 3 s intertrial interval, with presentation of the fixation cross, 1 s before the cue. This served both as a warning signal and as an instruction to the observer to maintain fixation. In the masked condition, the mask was presented at the end of the 60 ms stimulus exposure period and remained on until the observer responded; in the unmasked condition the stimulus was erased at the end of the exposure period. The RT on each trial was measured as the time from stimulus onset until the time at which the observer pressed one of two microswitched response buttons to indicate a decision. After each response, accuracy feedback was provided auditorily using distinctive tones. In addition, summary accuracy feedback was provided at the end of each block on the visual display. No feedback about RT was given during the course of a session.

3.2. Results

The proportions of correct responses to horizontal and vertical stimuli, $P_H(C)$ and $P_V(C)$, for each observer in each stimulus condition were converted to d' sensitivity measures using the formula

$$d' = \frac{z[P_H(C)] + z[P_V(C)]}{\sqrt{2}},$$

where $z[\cdot]$ denotes the inverse normal (z-score) transformation. The factor of $\sqrt{2}$ in the denominator of this expression puts d' values obtained from a two-alternative forced-choice task onto the same scale as those

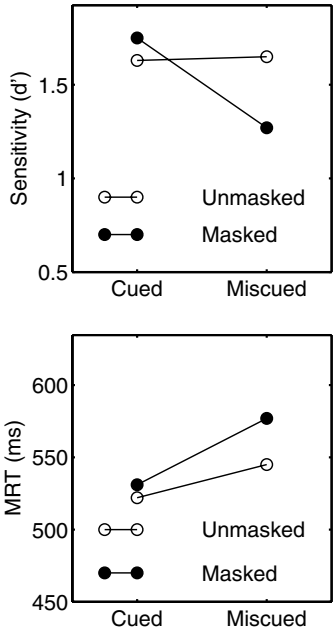

Fig. 3. MRT and sensitivity (d') averaged across contrasts and observers. Miscuing depresses sensitivity for masked but not unmasked stimuli. MRT is faster for cued stimuli, irrespective of masking.

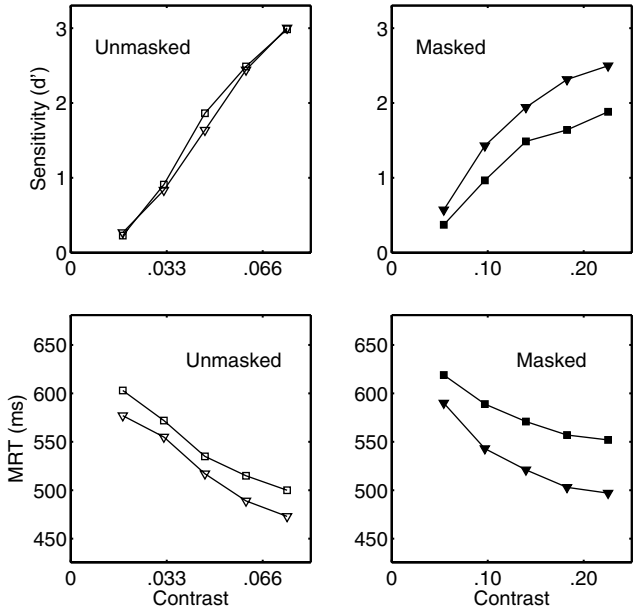

Fig. 4. Sensitivity (d') and MRT as a function of contrast for masked and unmasked stimuli. Triangles are cued stimuli; squares are miscued stimuli. The abscissa values are average contrasts across observers (cf. Figs. 5 and 6).

obtained from a yes–no task (Macmillan & Creelman, 1991). [2] The d' values and the associated mean response times (MRT) were analyzed using a three-way Cue (cued, miscued) × Mask (masked, unmasked) × Contrast (five levels) repeated-measures analysis of variance. The significant results from this analysis are shown in Figs. 3 and 4. The abscissa values in Fig. 4 are averages across observers of the contrast values used in the task.

As shown in Fig. 3, there was a significant mask-dependent cuing effect for d', as reflected in the Cue × Mask interaction, $F(1,5) = 18.05$, $p < 0.001$. Attentional cues had no effect on sensitivity for unmasked stimuli, but significantly increased sensitivity for masked stimuli. The corresponding interaction for MRT did not attain significance, $F(1,5) = 5.53$, n.s., although there is some suggestion of an interaction in the plot. However, there was a significant main effect of cue for MRT, $F(1,5) = 11.37$, $p < 0.05$, whereas the corresponding main effect for d' was not significant, $F(1,5) = 1.82$, n.s. Taken together, these results show that attentional cues have different effects on sensitivity and RT: Cuing results in shorter MRTs, regardless of whether stimuli are masked, but it increases sensitivity only for masked stimuli.

[2] The $\sqrt{2}$ scaling relationship between yes–no performance and two-alternative forced-choice performance is predicted by the Gaussian, equal-variance signal detection model. We use it here only as a conventional, summary measure of forced-choice performance, without implying any commitment to the detailed assumptions of this model.

Fig. 4 shows the corresponding results for stimulus contrast. There was a significant Cue × Mask × Contrast interaction for d', $F(4,20) = 4.11$, $p < 0.05$. This reflects a tendency for d' to grow more slowly with contrast for masked than for unmasked stimuli, which, combined with the differential effect of cuing for masked and unmasked stimuli, results in a three-way interaction. The cuing effect for masked stimuli in Fig. 4 represents an average sensitivity advantage for cued stimuli of around 3–4 dB across the range of the psychometric function. This agrees with the estimates for yes–no detection found by Smith et al. (in press).

For MRT there was a significant main effect of contrast, $F(4,20) = 18.01$, $p < 0.001$, and a significant Mask × Contrast interaction, $F(4,20) = 4.72$, $p < 0.01$. As shown in Fig. 4, MRTs decreased systematically from around 600 ms to around 500 ms with increasing contrast, but this decrease was more pronounced with unmasked stimuli.

3.3. Individual observers

We also analyzed sensitivity and MRT on an observer-by-observer basis to characterize the individual differences in cuing effects. These data are shown in Figs. 5 and 6. To quantify the cuing effects for sensitivity, Weibull functions were fitted to the psychometric functions for sensitivity for each observer (see Appendix A for details). Two models were compared: a single-function model, in which the same Weibull function was

fitted to the data for cued and miscued stimuli, and a two-function model, in which the cued and miscued data were fitted separately. The smooth curves in Fig. 5 are fits of the latter model. The difference between the fits of the two-function and single-function models is a measure of the magnitude of the cuing effect for each observer. The results of these fits are summarized in Table 1.

Under masked conditions, four observers (MS, FM, LB and ST) showed significant cuing benefits; a fifth (EH) showed a marginally significant effect and the remaining observer (IG) showed no effect of any kind. For unmasked stimuli, no observer showed any evidence of a cuing benefit. The unmasked data for one observer (EH) were better fitted by a two-function model than a single-function model and the difference in the fits for two others (IG and ST) were near significance. However, these effects are due to small sensitivity reversals, in which sensitivity for cued stimuli was slightly lower than for miscued stimuli. These reversals are similar to those found previously in yes–no detection (Smith, 2000a; Smith & Wolfgang, in press; Smith et al., in press), which we attributed to forward masking of the target by the cue (although see Smith and Wolfgang (in press), for another possibility). The range of individual differences in the cuing effect is comparable to that found for yes–no detection by Smith et al. (in press).

To quantify the effects for MRT, a two-parameter Piéron's law function (Piéron, 1920; Pins & Bonnet, 2000) was fitted to the MRTs for cued and miscued stimuli for each observer (see Appendix A). As shown in Fig. 6, Piéron's law describes a power law decrease in MRT with increasing stimulus intensity. As with sensitivity, the cuing effect was evaluated by comparing the fits of a one-function and a two-function model. These fits are summarized in Table 1.

For the masked data, MRT was well described for all observers by a two-function model and in all cases this model was significantly better than a single-function model. For the unmasked data, there is some evidence of systematic departures from Piéron's law although, qualitatively, the model again captures the main features of the data. (The difference in the fit statistics for masked and unmasked stimuli appears to be due mainly to dif-

Table 1
Cuing effects for individual observers

Observer	One-function model		Two-function model		Difference	
	$\chi^2(7)$	p	$\chi^2(6)$	p	$\Delta\chi^2(1)$	p
Detection sensitivity (d'), Weibull function fits						
Masked						
MS	15.95	0.03	8.16	0.23	7.79	0.01
FM	271.76	0	3.73	0.71	268.03	0
LB	140.67	0	12.41	0.05	128.26	0
IG	5.93	0.55	5.69	0.46	0.23	0.63
EH	13.85	0.05	10.44	0.11	3.41	0.06
ST	50.17	0	1.62	0.95	48.55	0.00
Unmasked						
MS	9.07	0.25	6.71	0.35	2.36	0.12
FM	24.33	0	23.24	0	1.09	0.30
LB	9.50	0.29	8.89	0.18	0.60	0.44
IG	5.45	0.60	2.13	0.91	3.32	0.07
EH	16.53	0.02	10.56	0.10	5.96	0.01
ST	8.27	0.31	5.25	0.51	3.02	0.08
MRT, Piéron's law fits						
	$\chi^2(8)$	p	$\chi^2(6)$	p	$\Delta\chi^2(2)$	p
Masked						
MS	30.72	0	6.02	0.42	24.70	0
FM	95.25	0	3.26	0.77	91.99	0
LB	365.49	0	5.91	0.43	359.53	0
IG	236.51	0	10.12	0.12	226.40	0
EH	206.80	0	4.67	0.59	202.13	0
ST	275.29	0	7.78	0.25	267.51	0
Unmasked						
MS	111.16	0	103.32	0	7.84	0.02
FM	61.88	0	15.49	0.02	46.38	0
LB	181.76	0	37.84	0	143.92	0
IG	66.79	0	22.35	0	44.44	0
EH	249.32	0	47.82	0	201.50	0
ST	111.47	0	20.75	0	90.72	0

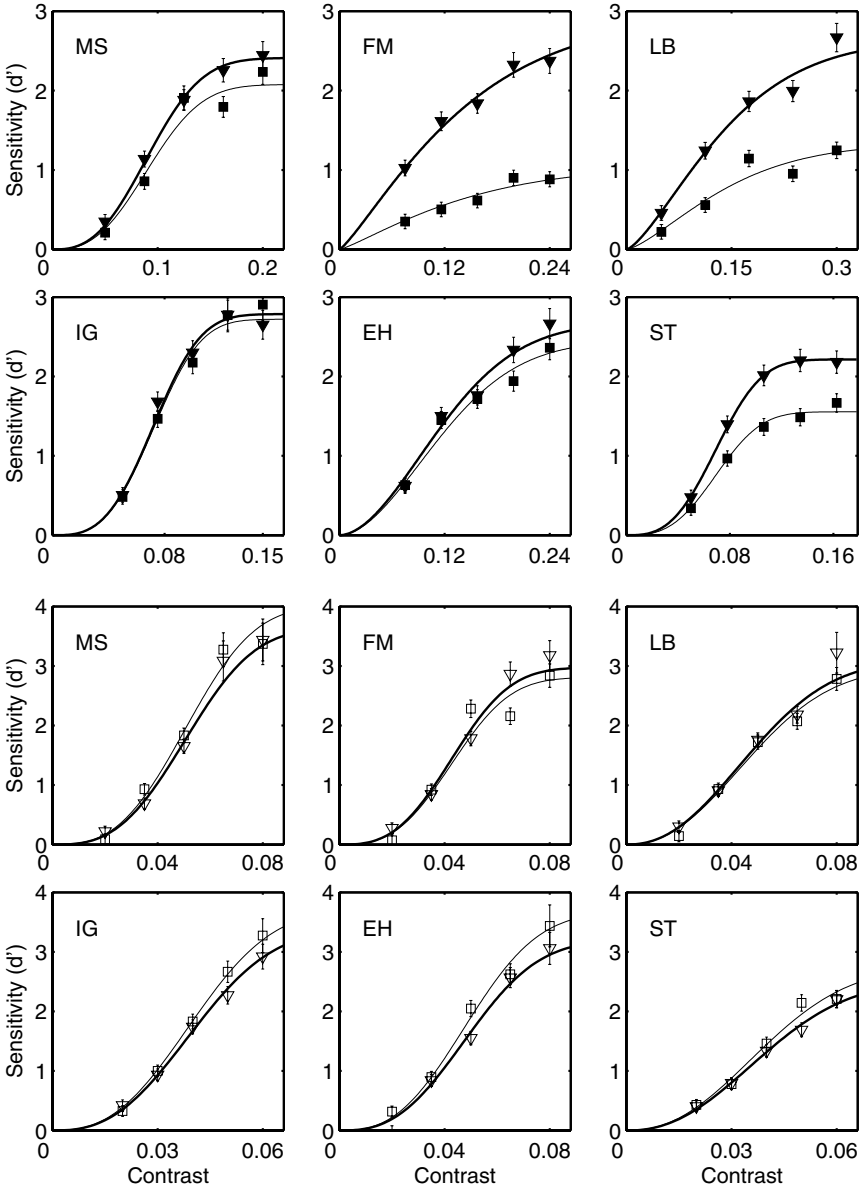

Fig. 5. Effects of cues on sensitivity (d') for individual observers. Filled symbols masked stimuli; open symbols are unmasked stimuli. Triangles are cued stimuli; squares are miscued stimuli. The continuous curves are Weibull function fits (see Appendix A for details).

ferences in the standard errors in the two conditions rather than to any obvious difference in functional form.) In all cases, however, a two-function model provided a better account of the data than a single-function model for all observers, although for observer MS the difference is small. These results show that MRT was faster for cued than for miscued stimuli for both masked and unmasked stimuli, in agreement with the previous analysis of group effects.

3.4. Discussion

The mask-dependent cuing effects found in this experiment replicated those found by Smith (2000a), Smith and Wolfgang (in press), and Smith et al. (in press) using a yes–no task with similar masks and stimuli. The similar results for detection and discrimination of orthogonal orientations is consistent with the assumption made by Lee et al. (1997) and Carrasco et al. (2000) that this latter task may be used as a proxy for detection. The results from both of these tasks are consistent with other reports in the literature that suggest a fundamental connection between spatial attention and masking mechanisms. Smith (2000a) and Smith and Wolfgang (in press) proposed a nonstationary model for these effects, in which masks interrupt stimulus processing before the information needed for a response is accumulated. This model assumes that the rate of

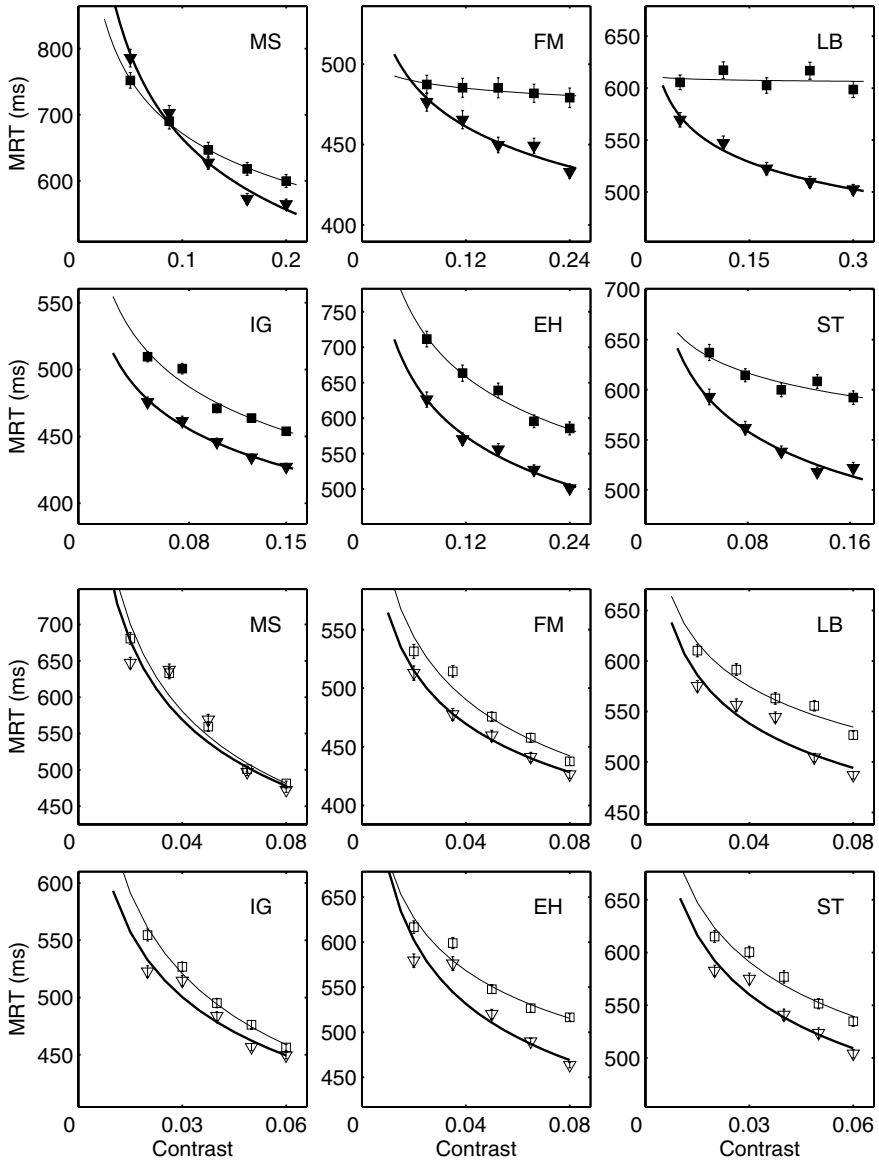

Fig. 6. Effects of cues on MRT for individual observers. Filled symbols are masked stimuli; open symbols are unmasked stimuli. Triangles are cued stimuli; squares are miscued stimuli. The continuous curves are Piéron's law fits (see Appendix A for details).

information gain from attended locations is greater than that from unattended locations so that masks depress sensitivity for unattended stimuli disproportionately. In the following section we propose a stationary model, based on the properties of the RT distributions for this task.

As well as finding mask-dependent cuing effects for d' we also obtained a spatial cuing effect in MRT. Such effects have been widely reported in the attention literature since they were first described by Posner and coworkers (Posner, 1978, 1980; Posner et al., 1980), but they have typically been obtained using suprathreshold stimuli in tasks in which the error rates are fairly low. Our results show that similar effects can be also obtained in psychophysical judgments about near-threshold stimuli. Importantly, however, the effects of cues and backward masks had different effects for d' and MRT. Whereas cues enhanced detection sensitivity only for masked stimuli, they had an unconditional effect on MRT: Responses were faster to cued than to miscued stimuli, regardless of whether they were masked. The different patterns of effects for cues and masks for d' and MRT means they cannot be explained simply by assuming that more difficult judgments are made more slowly. Rather, they must be explained by some other mechanism.

The data in Figs. 4–6 show that d' and MRT both varied systematically with stimulus contrast. Although contrast dependencies in MRT have been studied extensively (Piéron, 1920; Teichner & Krebs, 1972,

1974), this has typically again been done in tasks using suprathreshold stimuli, in which error rates have been low (an exception is Pins & Bonnet, 2000). Our data show a similar level of contrast dependency in judgments about near-threshold stimuli. This finding is important theoretically because decisions about such stimuli are frequently modelled using signal detection theory or related methods. However, signal detection theory has no mechanism to account for RT or for the joint dependence of accuracy and RT on stimulus contrast. We therefore attempt to provide an account of these data within an alternative, sequential sampling framework.

4. A diffusion process model of mask-dependent cuing effects

Along with MRTs that decreased systematically with stimulus contrast, the task yielded unimodal, positively-skewed distributions of RTs, whose variance increased with increasing MRT. Distributions of this kind are typically obtained from tasks involving the detection or discrimination of suprathreshold contrast stimuli and have been successfully modelled using sequential sampling models (e.g., Ratcliff & Rouder, 1998, 2000; Smith, 1995; see also Luce, 1986, for a review). The similarity in the distributions obtained from such tasks to those obtained here suggests that such models could also explain our data.

The organization of the remainder of this article is as follows. First we show that the families of RT distributions obtained from each of the four cue and mask conditions are well described by a stationary sequential sampling model. We then show that the changes in RT distributions and accuracy as a function of cue and mask conditions can be explained by an attention-orienting version of this model, in which the entry of stimuli into VSTM is delayed by inattention. The form of this model was suggested by the estimated parameters for fits of the diffusion model to the individual conditions. Finally, we show that the model's ability to account for the combined set of experimental data is not improved by the introduction of nonstationarity into the accumulation process.

4.1. The diffusion model of two-choice RT

The model we used to fit the empirical RT distributions and accuracy data was the Wiener diffusion model of Ratcliff (1978, 1981) and Ratcliff, Van Zandt, and McKoon (1999). In a recent evaluation of sequential sampling models, Ratcliff and Smith (in press) showed this model did as well or better than its competitors at predicting RT distributions and accuracy in a number of experimental tasks, and typically did so with greater economy of parameters. Mathematically, the diffusion process is a continuous-time version of a random walk and the model itself is a successor to early random walk models of two-choice decisions proposed by Stone (1960), Edwards (1965), Laming (1968), Link and Heath (1975) and Link (1978). The random walks are one of two main classes of sequential sampling model that have been studied in the mathematical psychology literature, the other being counter, or accumulator models (Pike, 1966; Smith & Vickers, 1988; Townsend & Ashby, 1983; Usher & McClelland, 2001; Vickers, 1970).

In random walk models, the observer's decision is based on a sequence of noisy observations of a stimulus process, each of which can take on positive or negative values. To make a decision, successive observations of the stimulus process are sampled and summed until a criterion positive or negative value is obtained. If the first criterion obtained is positive the observer makes one response; if it is negative the observer makes the other response. To model performance in an orientation discrimination task with a random walk model, the sequence of observations is identified with the moment-by-moment differences in the outputs of noise-perturbed horizontal and vertical filters.

In the diffusion process model, information accumulates in continuous time from a stimulus representation that is continuously perturbed by white noise (i.e., broad-spectrum Gaussian noise). The main features of the model are shown in Fig. 7. The accumulated information on any trial is represented by a continuous, but highly irregular sample path, whose mean and variance are both linear functions of time. This path is described by two parameters, the *drift*, ξ and the *diffusion coefficient*, s^2. The drift determines the average rate of increase (or decrease) of the sample path; the diffusion coefficient determines the magnitude of the random fluctuations of individual sample paths around the average path. In the model, the drift describes the quality of the encoded stimulus information and the diffusion coefficient describes noise within the observer. For given values of the drift and the diffusion coefficient, the population of sample paths as a function of time, t, has mean ξt and variance $s^2 t$, when the process is not constrained by absorbing boundaries.

On any trial the process begins at time $t = 0$ at a value of z, which is located between two *absorbing boundaries* located at 0 and a. The boundaries represent criterion amounts of information needed for positive and negative responses. [3] On presentation of (say) a vertical stimulus the process begins to drift in an upward direction at a rate that depends on the stimulus contrast.

[3] The $0 < z < a$ parameterization of starting point and boundaries is arbitrary. Because the Wiener process is translation invariant, it may alternatively be parameterized to start at zero, with $-z$ and $a - z$ as absorbing boundaries.

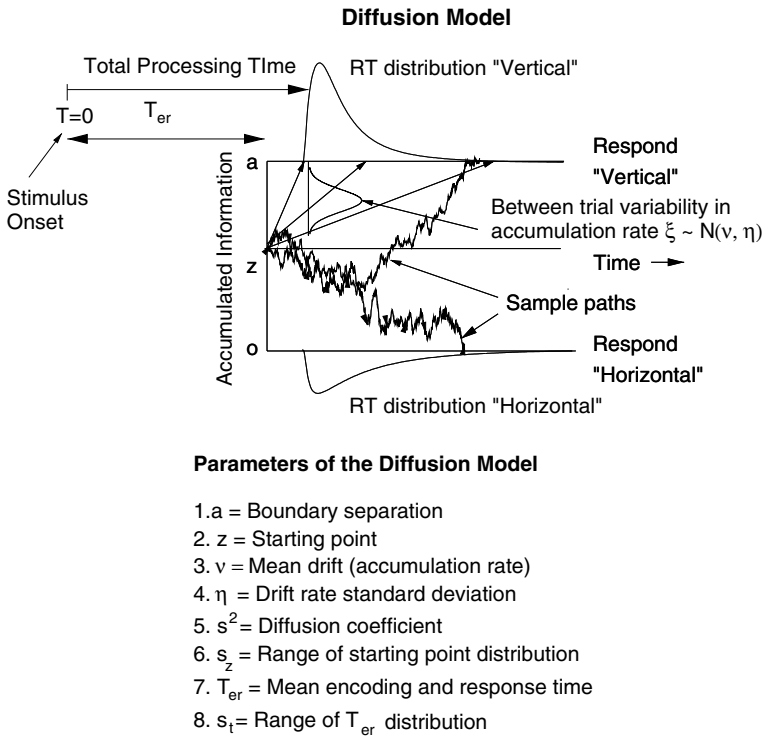

Fig. 7. Parameters of the diffusion model. Information accumulation begins at time $t = 0$ and starting point z. The process drifts towards one of two absorbing boundaries at 0 and a at a rate ξ that depends on the encoded stimulus information. For a given value of stimulus contrast, ξ is normally distributed with mean v and standard deviation η. The variability of individual sample paths around the average path, ξt, depends on the diffusion coefficient, s^2. The RT on any trial depends on the time at which the sample path first crosses one of the two absorbing boundaries. The nondecision component of RT is T_{er}.

This continues until the accumulated information reaches a, at which point the observer responds "vertical." Because of random variations in the accumulation process, however, on some proportion of trials the process will cross the boundary at 0 before it is reaches the boundary at a. On these trials the observer incorrectly responds "horizontal." The situation is the same when a horizontal stimulus is presented, except that the drift of the process is now in a downward direction. In general, errors are more likely to occur when the drift rate is small, or when the separation between the boundaries is small—as occurs when the observer is trying to respond rapidly—or when the starting point, z, is biased towards the incorrect response boundary. The latter may occur when the prior probabilities of the two kinds of stimulus are unequal and the observer elects to make the two responses based on different amounts of evidence. All of these conditions increase the probability that the first boundary crossed by the process is the wrong one and the decision is incorrect.

The decision time component of RT is identified in the model with the *first passage time* of the process, that is, the time at which the process first reaches one or other absorbing barrier. Mathematically, the predicted distributions of decision times are given by the distributions of first passage times of the diffusion process through the upper and lower barriers. For stationary models like the one considered here, these distributions may be obtained using either partial differential equation methods or numerical integral equation methods (see Ratcliff, 1978; Ratcliff & Smith, in press; Smith, 2000b, for detailed accounts).

An important feature of Ratcliff's diffusion model is that the stochastic process representing the encoded stimulus is perturbed by two sources of noise, one varying moment-to-moment and the other varying trial-to-trial. In applications of the model to psychophysical data, these two sources of noise correspond, respectively, to the decision noise and encoding noise in Fig. 1. The decision noise, as just discussed, may be viewed as arising from noise within the observer, and has variance s^2. In addition, the drift, ξ, which represents the encoded stimulus information, is assumed to vary randomly from trial to trial for any fixed level of stimulus contrast, because of random variations in the efficiency of visual transduction and encoding. In the model, drift is assumed to be normally distributed with mean v and variance η^2, that is, $\xi \sim N(v, \eta)$. In many applications of the model, including the one described here, it suffices to assume that the two kinds of stimuli have equal and

opposite effects on the distributions of drift. That is, if ξ_V, the drift on trials on which a vertical stimulus is presented is distributed as $N(v,\eta)$, then $\xi_H \sim N(-v,\eta)$.

The inclusion of drift variance in the model serves two important functions. First, it allows the model to predict that correct responses will be made, on average, more rapidly than errors. This relationship is typically found in tasks in which stimulus discriminability is low and accuracy of responding is stressed (see Swensson, 1972, Vickers, 1980, and Wilding, 1971, for reviews; also Luce, 1986, Chap. 6). In addition, it ensures that observers cannot attain arbitrarily high levels of accuracy by using widely-spaced absorbing boundaries and trading accuracy for speed. Because drift in the model is normally distributed, on some proportion of trials the encoded stimulus information will be such that the sign of the drift is wrong (i.e., will be negative when a vertical stimulus is presented, or vice versa). On such trials the process is more likely to terminate at the error boundary than the correct response boundary, regardless of boundary separation. This imposes a limit on the observer's ability to trade speed for accuracy, and means that some proportion of responses will inevitably be made incorrectly when the task is difficult, regardless of the observer's decision strategy.

Fig. 7 shows the parameters that must be estimated to fit the diffusion model to data. As well as those already described, the nondecisional components of RT are amalgamated into a single quantity, denoted T_{er} (time for encoding and response). In early applications of the model this time was treated as a constant, but recently it has been found that the model yields better fits to the leading edge of RT distributions if the nondecisional component of RT is treated as a random variable (see Ratcliff & Tuerlinckx, 2002, for a discussion). Because the variance of this component is assumed to be small relative to the variance of the decision times, however, its precise distributional form has a negligible effect on the shape of the predicted distributions of RTs. For computational reasons, it is convenient to treat it as uniformly distributed with mean T_{er} and range s_t.

Finally, the starting point of the process, z, may be assumed to vary randomly with range s_z. The introduction of starting point variability into the model allows it to predict fast errors, that is, error responses that are faster than correct responses. Fast errors are often found in tasks in which accuracy is high and speed of responding is stressed (Swensson, 1972; Wilding, 1971). Laming (1968) proposed that such variability might arise because of the observer's temporal uncertainty about stimulus onset. As a result, when speed of responding is stressed, perceptual sampling may begin prematurely and the decision process will accumulate noise from the pre-exposure field. Laming showed that a random walk model with starting point variability could predict the fast errors obtained experimentally from speeded responses to suprathreshold stimuli. In the present application, however, in which errors were slower than correct responses, the fit of the model was not improved by introducing variability of this kind. Consequently, the parameter s_z plays no further role in subsequent discussion.

Like other sequential sampling models, the parameters of the diffusion process model are identified only to the level of a ratio; that is, they may be multiplied by a constant without affecting any of the model's predictions. It is therefore customary in such models to treat the diffusion coefficient as a fundamental scaling parameter of the model and express other parameters in units of s. Some investigators, in recognition of the diffusion coefficient's role as a scaling parameter, choose to set it to unity. Here, however, we follow the convention used in previous articles by Ratcliff and set $s = 0.1$ so that the estimated parameters are comparable with those reported elsewhere.

4.2. Unconstrained fits of the model

As a first step, we sought to ascertain whether the diffusion model could fit the sets of RT distributions from each of the four Cue × Mask conditions separately. We refer to these fits as "unconstrained" because they made no attempt to explain why performance changes across the four conditions as a function of attentional cues or backward masks. Rather, our goal was simply to ascertain whether a stationary decision model could provide a reasonable account of the empirical RT distributions. Subsequently, we use the estimated parameters from these fits to develop a theoretical model of how cues and masks influence performance.

The procedure we used to fit the diffusion model was essentially that described by Ratcliff and Smith (in press) and Ratcliff and Tuerlinckx (2002). Initially, we fitted the model to group data obtained by averaging the quantiles of RT distributions for each condition across observers. Parameter estimates obtained by fitting the model to quantile-averaged data show reasonable agreement with the average of parameters obtained by fitting the model to individual observers (e.g., Ratcliff, Thapar, & McKoon, 2003). Denoting the RT in any condition by T, the ith distributional quantile, $Q(i)$, is defined by

$$P[T \leqslant Q(i)] = q_i.$$

That is, $Q(i)$ is the inverse probability transformation of the conditional RT distribution function. It is that value of time such that the proportion of the distribution's mass falling below $Q(i)$ is q_i. In direct rather than inverse terms, q_i is the fraction of responses in the distribution that are faster than $Q(i)$.

We used the quantile set $q_i \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ in our evaluation. Previous fits of the model have shown

that five quantiles provide a reasonable tradeoff between resolution and stability: It suffices to resolve the main features of a distribution's shape but is relatively insensitive to artifacts in the leading and trailing edges of the distribution in the data for individual observers (Ratcliff & Smith, in press; Ratcliff & Tuerlinckx, 2002). For each distribution of correct responses and errors at each of the five contrast levels, the corresponding quantiles of the empirical RT distributions were averaged across observers and across horizontal and vertical stimuli. Group accuracy measures were obtained by averaging the proportions of correct responses at each contrast level across stimulus orientations and observers. This resulted in 10 group RT distributions (i.e., a correct and an error distribution for each contrast level) and their associated response probabilities for each cell of the Cue × Mask design.

To fit the model, the G^2 statistic (i.e., the likelihood ratio chi-square) was minimized iteratively using the Nelder–Mead Simplex algorithm (Nelder & Mead, 1965), using the empirical RT quantiles as bounds to group the data into bins (see Appendix A for details). The sample size used in calculating G^2 was 2000 (i.e., 400 observations per contrast condition), the number of trials obtained from each observer in each cell of the Cue × Mask design. This is consistent with our interpretation of G^2 as assessing the fit of the model to an "average observer." Each such G^2 is based on 55 degrees of freedom and so, formally at least, may be tested for significance against a chi-square distribution with $55 - m$ degrees of freedom, where m is the number of free parameters in the model. As these statistics were based on quantile averages rather than independent observations, however, we do not attach strong probability interpretations to them, and we report the associated p values for reference purposes only. Instead, we focus on the model's ability to capture the shapes of the empirical RT distributions and response probabilities as a function of contrast. Subsequently, to allow fits based on different numbers of parameters to be compared, we also report a penalized likelihood statistic, the Bayesian Information Criterion, or BIC (Schwarz, 1978), which, for binned data, may be defined as

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

where m again denotes the number of free parameters in the model and N is the total number of observations on which G^2 was based. The BIC penalizes a model for its number of free parameters in a sample-size-dependent way. For a set of competing models, the model that is chosen as the best is the one with the smallest BIC.

Fig. 8 shows fits of the diffusion model to each of the four conditions of the experiment; Table 2 gives the associated fit statistics and estimated parameters. The fits are shown in the form of quantile-probability

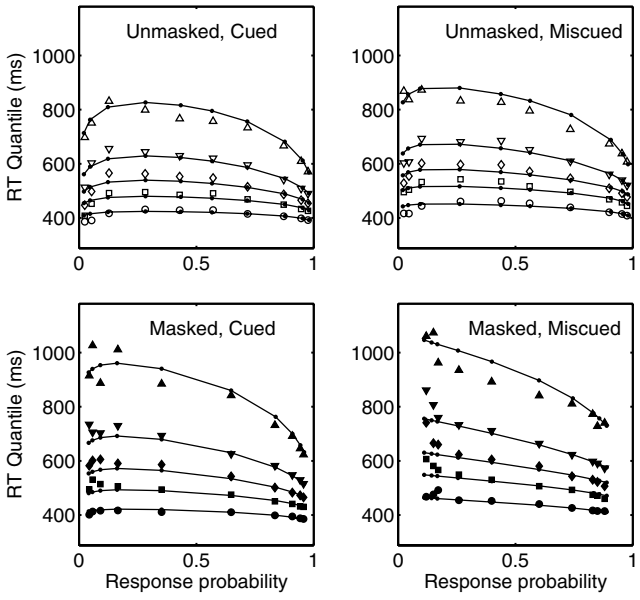

Fig. 8. Fits of the unconstrained model. Quantiles of the RT distributions for correct and error responses are plotted against their associated response probabilities for the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles for each contrast condition. Changes in the shape of the RT distribution with stimulus contrast are shown by changes in the vertical separation between quantiles. Only a single parameter, the mean drift, v, varies within each panel.

Table 2
Unconstrained diffusion model parameters and fit statistics

Drift estimates	v_1	v_2	v_3	v_4	v_5
Unmasked, cued	0.0412	0.1410	0.2879	0.4223	0.5225
Unmasked, miscued	0.0439	0.1781	0.3725	0.4875	0.5864
Masked, cued	0.1026	0.2668	0.3631	0.4319	0.4759
Masked, miscued	0.0919	0.2376	0.3488	0.3863	0.4414
Other parameters	a	η	T_{er}	s_t	$G^2(46)$
Unmasked, cued	0.0997	0.1728	0.3262	0.0983	34.04
Unmasked, miscued	0.1041	0.2228	0.3351	0.1406	47.95
Masked, cued	0.1189	0.2196	0.2996	0.1284	27.50
Masked, miscued	0.1263	0.3294	0.3095	0.1850	46.31

Note. The starting point for each condition, z, was set equal to $a/2$; s_z was set to zero.

plots, in which the quantiles of the predicted and observed RT distributions are plotted against their associated response probabilities on the abscissa. That is, if $p_1 < p_2 < \cdots < p_5$ are the observed probabilities of a correct response at each of five, increasing levels of contrast, then the quantiles of the correct response and error distributions for the lowest level of contrast are plotted against p_1 and $1 - p_1$, respectively. Those for the next lowest contrast are plotted against p_2 and $1 - p_2$, and those for the highest contrast are plotted against p_5 and $1 - p_5$. The innermost pair of distributions in the plot are therefore from the lowest level of contrast; the outermost pair are from the highest level. The shapes of the RT distributions are expressed in the plot by the vertical separations between adjacent pairs of quantiles.

The quantile-probability plot provides a way to represent the fit of the model to the RT distributions and to the associated response probabilities in a single, parametric plot. It also allows salient features of the data to be clearly depicted, such as how the shapes of the RT distributions change with stimulus contrast and how correct and error RTs are ordered relative to each other (Ratcliff & Smith, in press; Ratcliff & Tuerlinckx, 2002).

As may be seen from Fig. 8 and from the fit statistics in Table 2, the model provides a good description of the families of RT distributions obtained in this task. It predicts the unimodal, positively-skewed shapes of the distributions; it predicts the way these shapes change as a function of contrast, and it predicts that error responses are slower than correct responses at each level of contrast. (This last feature is reflected in the overall tendency of the plot to slope upwards, right to left.) Moreover, it does so with only a single parameter, the mean drift, v, varying with contrast. There are some minor discrepancies in fit apparent in the 0.9 quantile for some conditions, but this quantile is estimated empirically with comparatively low reliability. (In general, errors of estimate within the plot increase from right to left and from bottom to top, so that the largest errors of estimate are associated with the 0.9 quantile for error responses to high contrast stimuli. See Ratcliff and Tuerlinckx (2002), for graphical illustration of this point.)

The important result from these fits is that they show that the distributions of RT, and the associated accuracy values, are well described by a stationary decision model, even when the stimulus icon is erased after 60 ms by a backward mask. Although responding was slower with masked presentation and the RT distributions were more variable and more skewed, the distributions for both masked and unmasked stimuli could all be explained by a model in which there was no decay in effective stimulus information during the time taken to make a decision. Estimates of T_{er}, the mean nondecisional component of RT, were of the order of 300 ms in each of the four conditions. This implies that observers required anywhere between 100 and 700 ms to accumulate sufficient information to make a decision, depending on stimulus contrast and the quality of the encoded stimulus information available on a given experimental trial. The fits of the diffusion model imply that the observer's stimulus representations remained stable during this time. We interpret this result as implying that the observers' decisions are based on stable stimulus representations in VSTM.

The estimated model parameters in Table 2 show the following features: The separation of the absorbing boundaries (a) are somewhat higher under masked than unmasked conditions (i.e., observers use higher decision criteria when identifying masked stimuli). Mean drift rates (v) increase systematically with contrast and span a larger range of values for unmasked than for masked conditions. The relationship with cuing is less consistent. There is some tendency for the drift rates to be larger for cued than for miscued stimuli in the masked condition, but the pattern is reversed under unmasked conditions. (The latter is likely to reflect estimation artifact of some kind, rather than to be a real property of the underlying process.) Drift variance (η^2) increases with miscuing and with masking. The nondecisional component of RT, T_{er}, also increases with miscuing, as does its variability, s_t.

4.3. An attention-orienting model of diffusion process drift

A simple model, consistent with these parameters and with the evidence for the stationarity of the decision process, is shown in Fig. 9a. The model is a formalization of the attention-window model described in Section 1. It assumes that drift in the diffusion process depends on the quality of the stimulus information represented in VSTM. Access to VSTM is controlled by an attention window, $\gamma(t)$, that admits, or "latches" a visually-encoded stimulus into VSTM. Mathematically, this operation is represented by taking an inner product of the stimulus waveform and the attention window,

$$v = \int_{t_g}^{t_g+w} \mu(t)\gamma(t)\,\mathrm{d}t,$$

where t_g is the time at which the attention window opens and w is its width. This operation results in the formation of a stable stimulus representation in VSTM whose value is equal to the portion of the stimulus waveform falling within the window. We identify the mean drift of the diffusion process with this value.

To explain the mask-dependent cuing effects, we make two assumptions. First, the time at which the attention window opens, t_g, is delayed for unattended stimuli, relative to attended stimuli. This delay reflects the time taken to orient to the stimulus. Second, the physical stimulus trace, or icon, persists for some time after stimulus offset when stimuli are unmasked, but is

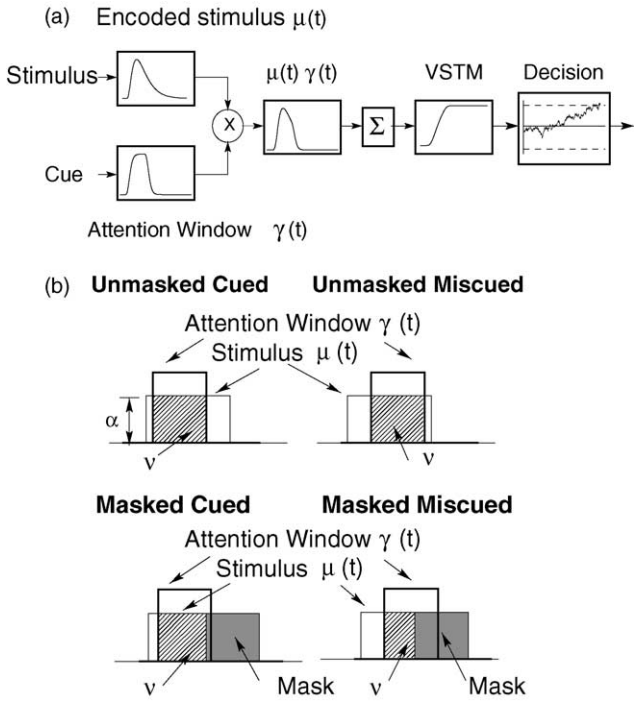

Fig. 9. (a) Attention-window model. The strength of the VSTM trace is equal to the proportion of the encoded stimulus, $\mu(t)$, that falls within the attention window, $\gamma(t)$. The drift of the diffusion process, v, is proportional to the strength of the VSTM trace. The integral of $\mu(t)\gamma(t)$ over the window is denoted by Σ. (b) Assumptions used to model mask-dependent cuing. Miscuing delays the entry of the stimulus into VSTM. When the stimulus is unmasked, the decay of the perceptual trace is negligible relative to the miscuing delay (estimated to be 16 ms). The values of v for cued and miscued stimuli (the shaded regions) are therefore the same. When the stimulus is masked, the suppression of the perceptual trace by the mask is virtually immediate. Values of v are therefore reduced in proportion to the miscuing delay.

rapidly erased when they are masked. This assumption is also made in the nonstationary model of Smith and Wolfgang (in press), but here we use it in a different way to develop a stationary model.

The implications of these assumptions are depicted, in a highly simplified way, in Fig. 9b. In this figure, the stimulus waveform, $\mu(t)$, is assumed to be constant for the duration of the stimulus. When stimuli are unmasked, the stimulus information persists, with negligible decay, for some period after stimulus offset. When stimuli are masked, the stimulus is suppressed by the mask almost immediately. The effect of a delay in the opening of the attentional window will have little effect with unmasked stimuli, because the window computes a value of drift from the information in the icon, which decays relatively slowly. When stimuli are masked, however, the effect of a delay will be appreciable, because the rapid suppression of the stimulus by the mask reduces the amount of the physical stimulus trace that falls within the window. We also considered the possibility, suggested by the estimated parameters in Table 2, that masking increases drift variance because noise from the mask is integrated into the attention window. These assumptions lead to the prediction of reduced drift for masked, unattended stimuli, increased T_{er} for unattended stimuli and possibly also increased drift variance for both masked stimuli and unattended stimuli.

Table 3 lists the parameters that were estimated to fit the attention-window model to all four conditions of the Cue × Mask design simultaneously (i.e., 40 different RT distributions). In the model, the amplitudes of the stimulus waveforms, $\mu(t)$, depend on contrast. To reduce the number of free parameters, the amplitudes, α_i, were assumed to be related to contrast, c, by a Naka-Rushton function

$$\alpha_i = A\left(\frac{c_i^{\rho}}{c_i^{\rho} + c_{0.5}^{\rho}}\right), \quad i = 1, 2, \ldots, 10.$$

This allowed the effects of 10 free drift parameters (5 each for masked and unmasked stimuli) to be represented by three Naka-Rushton parameters: a saturation and a semisaturation constant (A and $c_{0.5}$, respectively), and a shape parameter, ρ. A relationship of this form is often assumed in models of contrast transduction and gain control (e.g., Foley, 1994). [4]

Clearly, more elaborate assumptions could be made about the shape of the stimulus waveforms and attention window in Fig. 9 by identifying them with the outputs of early visual filters and ascribing the appropriate smooth dynamics to them (cf. Sperling & Weichselgartner, 1995). Indeed, the nonstationary models of Smith (1995) and Smith and Wolfgang (in press) were developed in this way, in part because the integral equation-techniques used to derive predictions require smoothness assumptions of this kind. Here, however, such elaboration is unnecessary, because the product $\mu(t)\gamma(t)$ is integrated twice to obtain predictions: once to obtain the drift and once to derive the first passage time statistics. Consequently, the effect of the precise shapes assumed for $\gamma(t)$ and $\mu(t)$ on the predictions will be minimal. Because drift depends on the integral of the stimulus waveform within the attention window, the parameters A and w may trade off to produce the same value of drift, making them difficult to identify individually. Preliminary fits confirmed this, at least for a range of theoretically-plausible values of w, of around 60–100 ms, as suggested by typical estimates of the perceptual integration time in Bloch's law (Smith, 1998a; Watson, 1986). We therefore set w equal to the stimulus duration of 60 ms.

[4] We used the Naka-Rushton function here as a convenient and flexible way to represent the relationship between contrast and mean drift. Potentially, other functional forms may characterize this relationship equally well. We have not explored this possibility as it is incidental to the main purpose of the article.

Table 3
Attention-window model parameters

	Parameter	Group	Individual (M)	Individual (SD)
Boundary separation (a)				
$a(u)$	Unmasked	0.099	0.099	0.012
$a(m)$	Masked	0.131	0.130	0.046
Drift standard deviation (η)				
$\eta(u)$	Unmasked	0.157	0.126	0.048
$\eta(m)$	Masked	0.376	0.299	0.090
Nondecisional time (T_{er})				
$T_{er}(u,c)$	Unmasked, cued	0.320	0.322	0.025
$T_{er}(u,m)$	Unmasked, miscued	0.336	0.335	0.029
$T_{er}(m,c)$	Masked, cued	0.296	0.296	0.028
$T_{er}(m,m)$	Masked, miscued	0.312[a]	0.310[a]	0.031
Nondecisional time variability (range) (s_t)				
$s_t(u,c)$	Unmasked, cued	0.104	0.100	0.016
$s_t(u,c)$	Unmasked, miscued	0.129	0.122	0.010
$s_t(u,c)$	Masked, cued	0.163	0.150	0.023
$s_t(u,c)$	Masked, miscued	0.188	0.172	0.039
Naka-Rushton drift parameters				
A	Saturation constant	0.709	0.642	0.174
$c_{0.5}$	Semisaturation constant	0.071	0.067	0.010
ρ	Shape	2.333	2.527	0.392

[a] The value of $T_{er}(m,m)$ was set equal to $T_{er}(m,c) + [T_{er}(u,m) - T_{er}(u,c)]$ to constrain the miscuing delay to be equal for masked and unmasked stimuli. This delay determined the magnitude of the drift reduction for masked, miscued stimuli.

Qualitatively, the attention-window model provides almost as good a description of the data as does the unconstrained model (Fig. 10), and does so with many fewer parameters (14 vs. 36). The goodness-of-fit statistics for the two models were $G^2(184) = 155.81$, $p = 0.93$, for the unconstrained model and $G^2(206) = 202.85$, $p = 0.55$, for the attention-window model. (The value for the unconstrained model was obtained by summing G^2 values across conditions.) Although the fit statistic for the attention-window model is around 30% larger than that for the unconstrained model, the ordering of the BIC values, which impose a penalty for model freedom, was reversed (BIC = 429.34 for the unconstrained model and BIC = 328.70 for the attention-window model). According to this criterion, of the two models, the attention-window model provides the more parsimonious account of the data.

As may be seen from Table 3, the attention-window model provided a good description of the data under the assumption that boundary separations and drift variance are both increased by masking but are unaffected by attention. Allowing the boundary separation to vary with attention did not improve the fit. Contrary to our expectations and to the results of the fits in Table 2, the fit of the attention-window model was not improved by assuming that drift variance increased with both inattention and masking. Rather the mask-dependent cuing effect was explained by a selective reduction in mean drift for masked, miscued stimuli. The model attributes this reduction to the extra time needed to orient attention, resulting in a delay in the formation of the VSTM

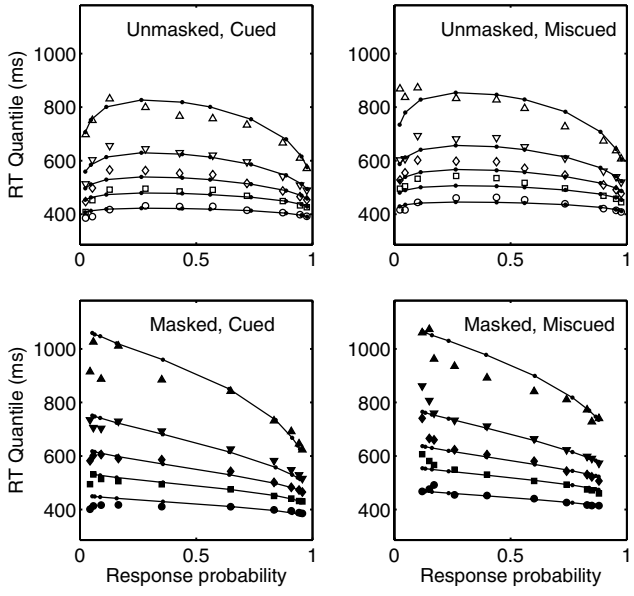

Fig. 10. Fits of the attention-window model. The mean drift of the diffusion process is determined by a single Naka-Rushton function for all four conditions. Differences in drift between cued and miscued conditions are determined solely by differences in, t_g, the time at which the attention window opens. Differences in T_{er} and s_t affect the locations and leading edges of the distributions but have a negligible effect on their shape.

Table 4
Attention-window model fit statistics for individual observers

Observer	Unconstrained model		Attention-window model	
	$G^2(184)$	BIC	$G^2(206)$	BIC
MS	531.16	834.71	616.02	752.72
FM	407.76	710.77	641.33	765.38
LB	359.91	663.93	495.50	617.46
IG	622.32	923.77	681.08	805.26
EH	409.54	712.95	482.93	608.64
ST	317.37	621.40	433.00	558.08

Note. The upper 5% points for the G^2 statistic with 184 and 206 degrees of freedom are 216.7 and 240.5, respectively.

trace. For unmasked stimuli, this delay, estimated at 16 ms, manifests itself as an increase in the mean and variance of the nondecision component of RT, with no effect on accuracy. For masked stimuli, the delay resulted in a reduction in effective stimulus information to around 72% of its value for cued stimuli.

4.4. Individual observers

To check whether these conclusions were consistent with effects at the individual observer level, we refitted the unconstrained model and the attention-window model to the data for each observer (Tables 2 and 4 and Fig. 11). The individual quantile-probability plots are noisier than are the group plots, particularly the error distributions for high-contrast stimuli, which reflect the small samples on which they were based. (For example, if accuracy at a particular contrast level is 95%, the correct response distribution is based on 380 observations whereas the error distribution is based on only 20 observations.) As expected, the individual G^2 statistics were systematically larger than those for the group fits. This occurs because quantile-averaging yields smoothed estimates of the underlying distribution. The individual G^2 statistics indicate some degree of failure to fit although, qualitatively, the model captures the main features of the data for most observers.

In general, large goodness-of-fit statistics may be obtained even for a true model because of interblock or intersession variability that is not represented in the model. When sample sizes are large, as here, fit statistics may be substantially inflated by variability of this kind (see Smith, 1998b, for further discussion). The largest discrepancy between the predictions and data were for the 10% quantile for the masked, miscued condition for observer EH. For EH, miscuing increased the fastest masked RTs more than it increased the fastest unmasked RTs. We interpret this as reflecting differences in orienting times across phases of the experiment for this observer.

Table 4 shows that for the majority of observers, the attention-window model provided a better description of the data, in a BIC sense, than did the unconstrained model. The exception was observer FM, although for this observer also, the attention-window model captures the main feature of the data in Fig. 11. Table 2 shows that the estimated model parameters, averaged across observers, are in good agreement with parameters estimated from group data, as has been reported for the diffusion model previously. The mean drift rates, v, in the masked, miscued condition, expressed as proportions of those in the cued condition, were estimated for the individual observers to be: MS, 0.79; FM, 0.37; LB, 45; IG, 0.90; EH, 0.87, and ST, 0.65. Qualitatively, these estimates appear fairly consistent with the cuing effects in Fig. 5.

4.5. Evidence against nonstationarity

The model fits reported in the previous sections showed that a stationary diffusion model provides a good description of the RT distributions from our experimental task. However, they do not exclude the possibility that the fits could be improved by introducing nonstationarity of some kind. In general, nonstationarity in a sequential sampling model like that of Fig. 1 can arise in either of two ways. It may arise because the decision process is driven directly by the output of a visual filter, whose mean value decays progressively after stimulus offset, or it may arise because the decision mechanism is an imperfect integrator, from which information decays as it is being accumulated. In general, the effects of these two kinds of nonstationarity on observed performance will be similar, because both result in a slower-than-linear rate of information accumulation on any trial (see Smith, 1995, for further discussion).

Continuous-time sequential sampling models with decay at the decision stage have been modelled as an Ornstein–Uhlenbeck (OU) diffusion model (Busemeyer & Townsend, 1993; Diederich, 1995; Heath, 1992; Smith, 1995, 1998a, 2000b). In the OU diffusion process, the accumulated information decays in proportion to the distance of the process from its starting point (i.e., to the amount accumulated). The drift in the model is equal to $\xi - \beta(x - z)$, where ξ is the information from the stimulus, x is the total accumulated information, z is the starting point, and β is a decay parameter. This

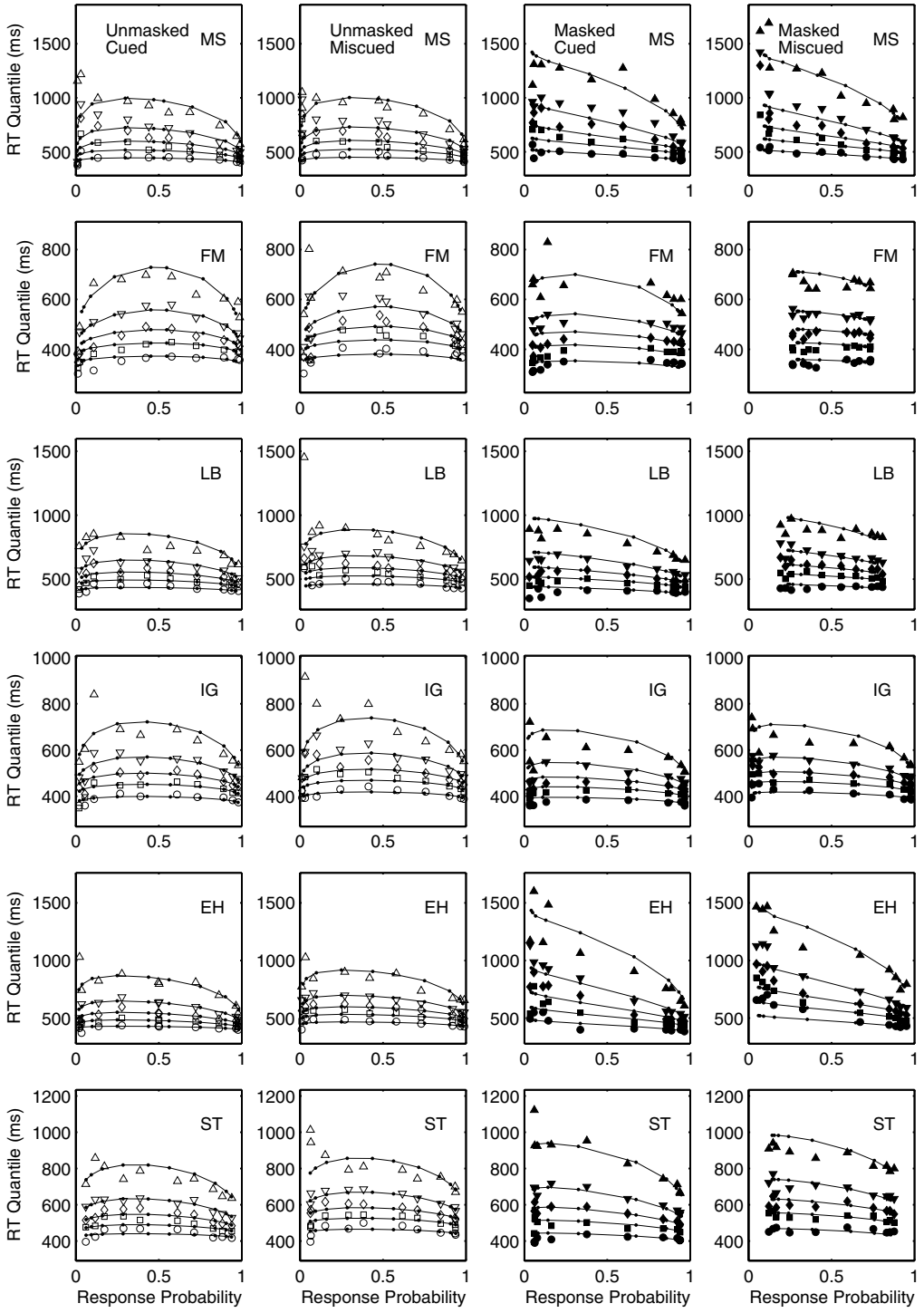

Fig. 11. Fits of the attention-window model to individual observers. Details are as for Fig. 10.

contrasts with the Wiener diffusion model, in which the drift is simply, ξ. In comparison to the OU process, the Wiener process may be characterized as a "perfect integrator."

To determine whether our data could be better described by assuming slower-than-linear integration of this kind, we fitted the OU model to the group quantile-probability functions, in the same way as we fitted the Wiener model. The methods used to obtain first passage time distributions for this model are described in Smith (2000b). We considered two variants of this model, one in which the OU decay parameter, β, differed for masked and unmasked stimuli, and another in which the β values were equal. When these parameters were allowed to vary freely in the fitting routine, in either case they converged to $\beta \approx 0$ and $G^2 \approx 202$, that is, to the Wiener model.

The results of these fits agree with those of Ratcliff and Smith (in press) who showed that an OU process with small-to-moderate decay could not be distinguished empirically from perfect integration. They also found that, for each of their three data sets, when the decay parameter was allowed to vary freely it approached a value of $\beta \approx 0$. These results support the proposition that psychophysical decisions are based on stationary stimulus information—or, at least, on information that remains approximately stationary while a decision is being made. As the physical stimulus trace is highly nonstationary during this time, especially under masked conditions, these findings imply that stimulus information is maintained in some kind of durable form until the decision is completed. We identify such durable traces with stable representations in VSTM.

5. General discussion

We obtained the same mask-dependent cuing effects with horizontal–vertical discrimination judgments as were found with yes–no detection by Smith (2000a), Smith and Wolfgang (in press), and Smith et al. (in press). Further, responses were faster to cued stimuli, irrespective of whether they were masked. The RT distributions and accuracy data from this task were well described by a diffusion process model in which information entering the decision process does not degrade appreciably during the time taken to make a decision. The cuing effects were well described by a model in which the drift of the diffusion process depends on the strength of the VSTM trace and miscuing delays the entry of stimuli into VSTM. When stimuli are masked, this delay results in degraded stimulus information. In its identification of miscuing with a delay in central stimulus processing this model is consistent with orienting theories of attention. In its presumption of a relationship between visual encoding, the effects of attention, and the contents of VSTM, the model is consistent with a large body of research initiated by Sperling (1960) on the relationship between attention and VSTM. Although these ideas have long had currency in the attention literature, we have shown they suffice to predict entire families of RT distributions and associated response probabilities for a range of mask, contrast, and cue conditions, simultaneously.

5.1. Orienting, gain, and rate models of attention

The orienting model of attentional cuing developed here contrasts with the class of attention gain models, described earlier. In gain models, increased attentional gain may be the result of signal enhancement, noise reduction, or some combination of the two, the net result being an increase in the signal-to-noise ratio for attended stimuli. Simple gain accounts are not consistent with mask-dependent cuing because they predict unconditional sensitivity advantages for cued stimuli, regardless of whether they are masked. Gain models that embody multiple attentional mechanisms, like the perceptual template model of Lu and Dosher (1998) and Dosher and Lu (1998), may be able to predict mask dependencies through processes such as external noise exclusion.

Dosher and Lu (2000) and Lu, Lesmes, and Dosher (2002) reported that one way cues can improve sensitivity is by excluding noise at the target location. This manifests itself as an increased cuing effect in high external noise displays. To the extent that pattern masks can be viewed as sources of external noise, the mask dependencies found here and elsewhere may be simply another manifestation of external noise exclusion. Against this interpretation is the finding of Turvey (1973), corroborated by Smith and Wolfgang (in press), that backward pattern masks reduce stimulus identifiability by interrupting processing rather than by adding noise to visual channels. Whether gain models like the perceptual template model can be elaborated to account for interruption masking effects by augmenting them with more complex temporal dynamics is an open question. Also, the perceptual template model in its current form accounts only for accuracy but not RT.

Closely related to the idea that cuing affects attentional gain is the idea that it affects rate of processing. In diffusion process models, an increase in attentional gain is modelled as an increase in ξ/s^2, the ratio of the drift and the diffusion coefficient of the accumulation process. Such an increase produces increased sensitivity (d') and a reduction in RT. In contrast, a proportional increase in ξ and s^2 that leaves their ratio unaltered produces an increased rate of processing. This results in a reduction in RT, with no change in sensitivity. [5] As noted previously, Carrasco and McElree (2001) reported that both rate of processing and sensitivity were affected by cues in an orientation discrimination task in which they used a response-signal procedure to measure processing time.

[5] This feature of diffusion models contrasts with signal detection theory. In signal detection theory, the ratio of the mean and the standard deviation of the decision variable is an invariant of the model. Changes to the scaling of the decision variable that preserve this ratio leave sensitivity unaltered. The corresponding invariant of diffusion models is the ratio of ξ/s^2, the ratio of the drift and the diffusion coefficient, that is, the ratio of the infinitesimal mean and variance of the distribution of increments to the process. (Expressed in terms of average drift for a condition instead of drift on an individual trial, the ratio is v/s^2.) This difference between diffusion models and signal detection theory is a reflection of the fact that, in the former, response probabilities depend jointly on the distribution of increments and the position of the decision criteria rather than on the distribution of increments alone.

The model of mask-dependent cuing described by Smith and Wolfgang (in press) is a rate-based model of this kind. The model predicts a cuing advantage for masked stimuli because the rapid rate of accumulation from cued locations means more information can be obtained from such locations before the stimulus is suppressed by the mask. No cuing advantage is predicted for unmasked stimuli because of their visual persistence, which allows the accumulation process to run to completion before the icon decays, for both cued and uncued stimuli. The model provides a quantitative account of mask-dependent cuing as a function of stimulus contrast and exposure duration, but relies in an essential way on the assumed nonstationarity of the decision mechanism. Our results here yielded no evidence for nonstationarity, but we have not investigated all possible ways in which nonstationary may enter the system. Further research is needed to test between rate-based and orienting accounts in a systematic way.

5.2. Detection and discrimination

The mask dependencies found in yes–no detection by Smith (2000a) and in orthogonal discrimination here contrast with the findings of Carrasco et al. (2000) who reported that cues increased sensitivity in these tasks in the absence of masks. Carrasco et al. (2000) pointed out that whereas their study used pure peripheral cues, Smith (2000a) used mixed central-peripheral cues. Consequently, differences in cue effectiveness may have been responsible for the difference in results. However, Smith et al. (in press) subsequently obtained a mask-dependent cuing effect in yes–no detection using the pedestal task and the same peripheral cues as used by Carrasco et al. (2000).

Smith et al. (in press) attributed the difference between the two studies to the use of a pedestal. They argued that when near-threshold stimuli are presented directly against a uniform field, as was done in the Carrasco et al. (2000) study, uncertainly effects may have a significant influence on performance because they covary with cue condition. Consistent with this, Smith et al. (in press) found that when no pedestal was used, some observers showed a significant cuing effect with unmasked stimuli. Averaged across observers, this yielded a significant cuing effect for the experiment as a whole.

In a later study, Cameron et al. (2002) found that cues also enhanced sensitivity in an unmasked orientation discrimination task in which observers discriminated between Gabor patches oriented at either ±15° or ±4° to the vertical. In the 15° task, the observers' ability to discriminate the stimuli covaried with their ability to localize them, whereas in the 4° task, stimulus localization at the higher levels of contrast was virtually perfect. Cameron et al. (2002) argued that performance is unlikely to be influenced by uncertainty reduction when localization is near ceiling and suggested their cuing effects could be attributed to signal enhancement.

We have not investigated orientation discrimination using the pedestal task, but the results of Cameron et al. (2002) combined with our own are evidence for a Cue × Mask × Task interaction: In detection, cues produce signal enhancement only for backwardly-masked stimuli, whereas in discrimination—that is, judgments in which performance is limited in part by the similarity of the stimuli rather than by contrast alone—cues produce signal enhancement unconditionally. (Orthogonal discrimination in these terms is functionally equivalent to detection, as argued previously.) Support for the idea that cues increase sensitivity unconditionally in discrimination was provided by Carrasco, Williams, and Yeshurun (2002) in a study using masked and unmasked Landolt squares.

The rate-based model of Smith and Wolfgang (in press) predicts this three-way interaction under the assumption that the decision mechanism accumulates information more slowly in discrimination than in detection (see also Smith et al., in press). This is consistent with the typical finding that RTs for detection are shorter than those for discrimination (Luce, 1986). Smith and Wolfgang (in press) argued that information accumulates more slowly in discrimination than detection because discrimination has a lower signal-to-noise ratio. This is because it requires comparison of the outputs of filters whose bandwidths overlap, whereas detection and orthogonal discrimination do not. They showed that under these circumstances, a rate-based model predicts similar cuing effects for unmasked discrimination and masked detection, resulting in the predicted three-way interaction.

The attention-window model described here also predicts this interaction under the assumption that larger samples of perceptual information are required for discrimination than for detection. In the attention-window model, stimulus discriminability is identified with the drift of the diffusion process, which is computed by integrating the output of a visual filter over an attention window. This initial period of perceptual integration precedes and is distinct from the subsequent accumulation of information by the diffusion process to make a decision. According to the attention-window account, the task of discriminating between stimuli that activate filters with overlapping bandwidths requires a longer period of perceptual integration than does detection or orthogonal discrimination because of their relatively poor signal-to-noise ratio. With unmasked stimuli, a delay in opening the attentional window will have a greater effect in discrimination than detection because the longer period of perceptual integration needed to compute the drift of the diffusion process means it will be disproportionately affected by iconic decay.

Acknowledgements

Preparation of this article was supported by Australian Research Council Discovery Grant DP0209249 and NIMH grants R37-MH44640 and K05-MH01891. We thank Marisa Carrasco and John Palmer for helpful comments on an earlier version of the manuscript and Zhong-Lin Lu and Joshua Solomon for useful discussions.

Appendix A. Model fitting procedures

Psychometric functions. The psychometric functions in Fig. 5 were fitted with three-parameter Weibull functions of the form

$$F(c) = \alpha\left\{1 - \exp\left[-\left(\frac{c}{\beta}\right)^{\gamma}\right]\right\},$$

where c is stimulus contrast and α, β, and γ are amplitude, dispersion, and shape parameters. The functions were fitted to the empirical estimates of d' by iteratively minimizing the chi-square statistic

$$\chi^2 = \sum \frac{[d'(c) - F(c)]^2}{\mathrm{var}[d'(c)]},$$

using the Nelder–Mead Simplex algorithm (Nelder & Mead, 1965). The term in the denominator of this equation is the asymptotic variance estimator of Gourevitch and Galanter (1967):

$$\mathrm{var}(d') = \frac{P_{\mathrm{H}}(C)[1 - P_{\mathrm{H}}(C)]}{2n_{\mathrm{H}}\phi^2\{z[P_{\mathrm{H}}(C)]\}} + \frac{P_{\mathrm{V}}(C)[1 - P_{\mathrm{V}}(C)]}{2n_{\mathrm{V}}\phi^2\{z[P_{\mathrm{V}}(C)]\}},$$

where n_{H} are n_{V} are the number of horizontal and vertical stimuli in each condition, and $\phi(\cdot)$ is the standard normal density function evaluated at the specified abscissa. The other quantities are as defined in the text. (The factor of 2 in the denominator is a reflection of the $\sqrt{2}$ term in the denominator in the definition of d'.) The fits in Fig. 5 are for a two-function model in which a pair of Weibulls with identical dispersion and shape parameters but different amplitudes were fitted to the psychometric functions for cued and miscued stimuli.

Piéron's law. The values of MRT in Fig. 6 were fitted with two-parameter power law functions

$$F(c) = \alpha c^{-\beta},$$

by minimizing a chi-square statistic, as was done for the psychometric functions, with the square of the estimated standard error of the mean used as a variance term in the denominator. Piéron's law is often written with an additional asymptote parameter, the so-called "irreducible minimum RT" (Luce, 1986, p. 58). This three-parameter form did not yield any improvement in fit and showed poorer stability.

Diffusion processes. The diffusion process models in Figs. 8, 10, and 11 were fitted by minimizing the G^2 statistic (i.e., the likelihood ratio chi-square) using the empirical RT quantiles as bounds to group the data into bins. Letting $p_{jk}(c)$ denote the proportion of probability mass in bin j of the empirical joint RT distribution for response k (correct or error) at contrast level c, and letting $\pi_{jk}(c)$ denote the corresponding mass predicted by the model, G^2 was defined as

$$G^2 = 2\sum_{c=1}^{5} n_c \sum_{k=1}^{2}\sum_{j=1}^{6} p_{jk}(c)\log\left[\frac{p_{jk}(c)}{\pi_{jk}(c)}\right]$$

(Agresti, 1990, p. 48). That is, G^2 measures the discrepancy between the observed and predicted mass in each bin, summed across contrasts, responses, and quantiles. (Note that the use of five quantiles to partition the distributions results in six bins per distribution.) The values of n_c in this equation are the numbers of experimental trials on which each pair of correct and error distributions was based. For each pair of correct and error distributions, this yields $12 - 1 = 11$ degrees of freedom per distribution pair.

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Averbach, E., & Sperling, G. (1960). Short-term storage of information in vision. In C. Cherry (Ed.), *Information theory*. London: Butterworth.

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.

Bonnel, 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.

Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon.

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.

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 Science, 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*, 467–479.

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. (2003). MDFT account of decision making under time pressure. *Psychonomic Bulletin & Review, 10*, 157–166.

Di Lollo, V., & Woods, E. (1981). Duration of visible persistence in relation to range of spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 754–769.

Dobkins, K. R., & Bosworth, R. G. (2001). Effect of set-size and selective spatial attention on motion processing. *Vision Research, 41*, 1501–1517.

Dosher, B. A., & Lu, Z.-L. (1998). External noise distinguishes attentional mechanisms. *Vision Research, 9*, 1183–1198.

Dosher, B. A., & Lu, Z.-L. (2000). 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., Thomas, J. P., Palmer, J., Shimozaki, S., & Steven, 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.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology, 2*, 312–329.

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.

Gegenfurtner, K., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 845–866.

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.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika, 32*, 25–33.

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, XI* (pp. 269–284). Hillsdale, NJ: Erlbaum.

Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. 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 process model for two choice decision making. *Mathematical Social Sciences, 23*, 283–309.

Kawahara, J., Di Lollo, V., & Enns, J. T. (2001). Attentional requirements in visual detection and identification: Evidence from the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 969–984.

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.

Link, S. W. (1978). The relative judgment theory of the psychometric function. In J. Requin (Ed.), *Attention and performance, VII* (pp. 619–630). Hillsdale, NJ: Erlbaum.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40*, 77–105.

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attentional mechanisms. *Vision Research, 9*, 1183–1198.

Lu, Z.-L., Lesmes, L. A., & Dosher, B. A. (2002). Spatial attention excludes external noise at the target location. *Journal of Vision, 2*, 312–323.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. 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 electrophysical evidence. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 887–904.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

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.

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.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 308–313.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, NJ: Erlbaum.

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.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283–290.

Piéron, H. (1920). Nouvelles recherches sur l'analyse du temps de latence sensorialle et sur la loi qui relie ce temps à l'intensité de l'excitation. *Année Psychologique, 22*, 58–142.

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.

Pins, D., & Bonnet, C. (2000). The Pieron function in the threshold region. *Perception & Psychophysics, 62*, 127–136.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsale, NJ: Erlbaum.

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). Visual attention modulates metacontrast masking. *Nature, 373*, 66–68.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology, 21*, 178–184.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88*, 552–572.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347–356.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 127–140.

Ratcliff, R., & Smith, R. (in press). A comparison of sequential-sampling models for two-choice reaction time. *Psychological Review*.

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., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93*, 180–206.

Robinson, D. A. (1965). The mechanics of human smooth pursuit eye movements. *Journal of Physiology, 180*, 569–591.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Shaw, M. I. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and the detection of luminance increments. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance, X* (pp. 109–121). Hillsdale, NJ: Erlbaum.

Shih, S.-I., & Sperling, G. (2003). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109*, 260–305.

Shiu, L.-P., & Pashler, H. (1994). Negligible effects of spatial precuing on the 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*, 567–593.

Smith, P. L. (1998a). Bloch's law predictions from diffusion process models of detection. *Australian Journal of Psychology, 50*, 139–147.

Smith, P. L. (1998b). 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., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32*, 135–168.

Smith, P. L., & Wolfgang, B. J. (in press). The attentional dynamics of masked detection. *Journal of Experimental Psychology: Human Perception and Performance*.

Smith, P. L., Wolfgang, B. J., & Sinclair, A. (in press). Mask-dependent attentional cuing in visual signal detection: The psychometric function for contrast. *Perception & Psychophysics*.

Sperling, G. (1960). The information available in brief stimulus presentations. *Psychological Monographs, 74*, 1–29.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102*, 503–532.

Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*, 251–260.

Swensson, R. G. (1972). The elusive speed–accuracy tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception & Psychophysics, 12*, 16–32.

Tata, M. S. (2002). Attend to it now or lose it forever: Selective attention, metacontrast masking, and object substitution. *Perception & Psychophysics, 64*, 1038–1082.

Teichner, W. H., & Krebs, M. J. (1972). Laws of the simple visual reaction time. *Psychological Review, 79*, 344–358.

Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review, 81*, 75–98.

Thomas, J. P., & Gille, J. (1979). Band widths of orientation channels in human vision. *Journal of the Optical Society of America, 69*, 652–660.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

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 integrator model. *Psychological Review, 34*, 273–287.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13*, 37–58.

Vickers, D. (1980). Discrimination. In A. T. Welford (Ed.), *Reaction times* (pp. 25–72). London: 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.

Wilding, J. M. (1971). The relation between latency and accuracy in the identification of visual stimuli. II. The effects of sequential dependencies. *Acta Psychologica, 35*, 399–415.