Journal of Mathematical Psychology 59 (2014) 95-113



Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition



Journal of Mathematical



Philip L. Smith^{a,*}, Roger Ratcliff^b, David K. Sewell^a

^a Melbourne School of Psychological Sciences, The University of Melbourne, VIC 3010, Australia
^b The Ohio State University, Columbus, OH, United States

HIGHLIGHTS

- Dynamic noise impairs performance and shifts RT distributions on the time axis.
- We describe two diffusion process models for discrimination in dynamic noise.
- The integrated system model is based on a time-changed diffusion process.
- The release from inhibition model is based on known physiological processes.
- Both models gave good accounts of the RT distributions and accuracy from the task.

ARTICLE INFO

Article history: Available online 29 July 2013

Keywords: Decision-making Sequential-sampling Response time Reaction time Discrimination Signal detection

ABSTRACT

The speed and accuracy of discrimination of featurally-defined stimuli such as letters, oriented bars, and Gabor patches are reduced when they are embedded in dynamic visual noise, but, unlike other discriminability manipulations, dynamic noise produces significant shifts of RT distributions on the time axis. These shifts appear to be associated with a delay in the onset of evidence accumulation by a decision process until a stable perceptual representation of the stimulus has formed. We consider two models for this task, which assume that evidence accumulation and perceptual processes are dynamically coupled. One is a time-changed diffusion model in which the drift and diffusion coefficient grow in proportion to one another. The other is a release from inhibition model, in which the emerging perceptual representation modulates an Ornstein–Uhlenbeck decay coefficient. Both models successfully reproduce the families of RT distributions found in the dynamic noise task, including the shifts in the leading edge of the distribution and the pattern of fast errors. We conclude that both models are plausible psychological models for this task.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

In contributing an article to honor William Estes as one of the creators of mathematical psychology, we begin by reflecting on what it means to have done as Estes did, and created a discipline where none was before. Estes made numerous deep and influential contributions during his long and distinguished career, but, arguably, none had greater or more enduring significance for the future of the discipline than his original seminal work in animal learning, stimulus sampling theory (Estes, 1950, 1955a,b; Estes & Burke, 1953). In creating stimulus sampling theory, Estes not only constructed an elegant and powerful theory of learning, but also showed by example just what it means to develop and test a process model of a psychological phenomenon. Stimulus sampling

* Corresponding author. *E-mail address*: philipls@unimelb.edu.au (P.L. Smith). theory first confronted the issue that has confronted every process model since then, namely, the inherent variability of behavior: the fact that organisms, whether human or nonhuman, do not exhibit the same behavior from trial to trial or from one presentation of a stimulus to the next. Consequently, a process model for learning must be expressed at the level of operators that show how choice probabilities evolve from trial to trial. Such probabilistic variation is not just a layering of a measurement error model on top of a deterministic process, but is integral to the theory itself.

Those of us who work with process models for psychological phenomena belong to a tradition begun by Estes and are profoundly indebted to him. From his example we understand that the development of a process model is the discipline of expressing a psychological explanation in quantitative terms and, in so doing, of determining precisely what its empirical consequences might be. It is also the discipline of testing a quantitatively expressed explanation against empirical data. Like all applied mathematics, it is the art of making complex problems tractable. In this, it is

^{0022-2496/\$ –} see front matter © 2013 Elsevier Inc. All rights reserved. http://dx.doi.org/10.1016/j.jmp.2013.05.007



Fig. 1. Letter discrimination in dynamic noise. (a) Example stimulus. The upper panel shows a single frame in which 0.35 of the pixels have been inverted. The lower panel shows an average over 10 frames. Because stimuli would have been integrated by early visual filters, the lower panel provides a better indication of the perceptual experience of the task. (b) Quantile probability plots for five levels of discriminability presented under speed or accuracy instructions. The lines on the graph, bottom to top, are the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles of the RT distributions. The response probabilities on the *x*-axis are the probabilities of correct responses, *p*, and error responses, 1 - p. The five distributions on the right in each panel are the distributions of correct responses; the five distributions on the left are the distributions of errors. The data are quantile averages over participants. Note the different y-axis scaling for the two conditions.

the art of distinguishing the essential from the superfluous and the simple from the simplistic. Anyone who does work of this kind knows what the benefits of this undertaking can be. The attempt to express a psychological principle in quantitative terms is usually, in the first instance, a process of discovering that the things you thought were precise are in fact not so. It is also a way of flushing out unexamined assumptions and of exposing them to critical scrutiny.

Estes began his long career during the ascendancy of behaviorism and finished it long after the cognitive revolution had become the cognitive orthodoxy. The evolution of his research interests over time reflected the change in the conceptual landscape, from learning, which was the driving force for behaviorism, to perception, memory, categorization, and decision-making. These are topics that remain of central concern to mathematical psychologists today. A number of his later papers focused on the problem of determining whether variables that affect performance in visual recognition tasks do so by affecting perceptual or decision processes (Bjork & Estes, 1973; Estes, 1972, 1975, 1982). Estes was profoundly aware of the contribution made by decision processes, which match incoming sensory information against task representations in immediate memory, to performance in simple cognitive tasks. He was also aware of the hazards of theorizing about perceptual and decision processes in isolation, arguing that a proper understanding could only be gained by considering how they act in concert. That question, although framed in somewhat different terms, is the focus of this article.

1.1. Two-choice perceptual discrimination in dynamic noise

In a sequence of 12 experiments, Ratcliff and Smith (2010) investigated performance in a novel two-choice discrimination task in which letter stimuli were degraded by embedding them in dynamic visual noise. In their task, a randomly-chosen proportion of the pixels in the letter and the background were inverted in each consecutive frame of the display. Like other manipulations of discriminability, dynamic noise increased response time (RT) and reduced accuracy, but unlike other manipulations, it also produced significant shifts of the RT distribution on the time axis. These were manifested as changes in the distribution's leading edge, as indexed by its 0.10 quantile. Changes in the 0.10 quantile depend only on the fastest 10% of responses in the distribution and are relatively independent of changes in its variance or higher moments. Ratcliff and Smith found that dynamic noise shifted the leading edge of the distribution by more than 100 ms in the most difficult as compared to the easiest condition.

Fig. 1 shows examples of the stimuli used by Ratcliff and Smith (2010) in their Experiment 1, together with a quantile-probability plot (Ratcliff & Tuerlinckx, 2002) of group data from an unpublished experiment that used the same task. The details of the method can be found in Appendix A. Participants performed the task under speed and accuracy instructions in alternating blocks at five levels of stimulus discriminability, formed by inverting 0.35, 0.40, 0.425, 0.45, 0.475 of the pixels in the display. (When 0.5 of the pixels are inverted, the display becomes a homogeneous, random field of black and white pixels that carries no stimulus information.) In quantile probability plots, selected quantiles of the RT distributions for correct responses and errors are plotted against the choice probabilities, p_i and $1 - p_i$, for each condition, *i*. Such plots show how distribution shape, response accuracy, and the relationship between mean RTs for correct responses and errors all change as stimulus discriminability is varied. The distributions in Fig. 1 have been summarized using their 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles.

The unusual result in Fig. 1 is the systematic change in the leading edge of the distribution as a function of noise, which appears as a bowing of the curve representing the 0.1 quantile (the bottom curve in the plot) in both the speed and accuracy conditions. This is unlike the results found in the vast majority of speeded two-choice decision tasks. In most tasks, most of the changes in the distributions are in the upper quantiles; the leading edge is relatively unaffected and the curve representing the 0.1 quantile is almost flat (Ratcliff & Smith, 2004). Following Ratcliff and Smith (2010), we refer to the bowing of the 0.1 quantile function in Fig. 1 as the *leading edge effect.*

The leading edge effect in Ratcliff and Smith's (2010) study was found only with letter discrimination in dynamic noise. There was no leading edge effect in a brightness discrimination task with dynamic noise, in which participants were required to judge whether the average proportion of light pixels in the display was greater or less than 50%. There was no leading edge effect in a letter discrimination task, in which letters were degraded by a simultaneous structure mask composed of random letter fragments in the same stroke font as the stimuli. There was a smaller leading edge effect in the letter discrimination task when the noise was static rather than dynamic.

Ratcliff and Smith (2010) attributed the leading edge effect to a delay in the onset of information accumulation by a decision process until a stable perceptual representation of the stimulus had formed. The phenomenological basis for this interpretation is compelling: When letters are viewed in dynamic noise, they appear to emerge slowly out of the noise. The perceptual experience is quite unlike that in the masking-by-structure discrimination task, in which the stimuli seem to appear instantaneously.

The data in Fig. 1 can be well fitted by a version of Ratcliff's (1978) diffusion model in which the non-decision time, or time for other processes, T_{er} , varies systematically with the level of noise in the display. Ratcliff's model assumes that RT can be additively decomposed into a decision time, T_D , and a time for other processes,

$$RT = T_D + T_{er}.$$
 (1)

The decision time is the first passage time for a Wiener or Brownian motion diffusion process through one of two absorbing boundaries that represent decision criteria. The absorbing boundaries represent upper and lower limits of diffusion: once a boundary is reached, diffusion ceases. Formally, if X(t) is a diffusion process starting at zero, X(0) = 0, and a_1 and a_2 are absorbing boundaries, with $a_2 < 0 < a_1$, then we define the first passage times, $T(a_1)$ and $T(a_2)$, as

 $T(a_1) = \min\{t : X(t) \ge a_1\}$ $T(a_2) = \min\{t : X(t) < a_2\}.$

The decision time, T_D , in Eq. (1) is the first of these events to occur:

$$T_D = \min\{T(a_1), T(a_2)\}.$$

Either $T(a_1)$ or $T(a_2)$ may be infinite, but T_D is finite with probability one (Cox & Miller, 1965). That is, the process is guaranteed to terminate in finite time. The boundaries are defined as absorbing by the relations

$$P[X(t) = a_1 | T(a_1) \le t] = 1$$

$$P[X(t) = a_2 | T(a_2) \le t] = 1.$$

These equations state that once the process has reached an absorbing boundary its value does not change any further. Absorbing boundaries are one of several kinds of possible boundary for diffusion processes, which may be either *accessible* or *inaccessible* and *absorbing*, *reflecting*, or *sticky*. A discussion of the varieties of boundary behavior may be found in Karlin and Taylor (1981, pp. 226–242). A combination of absorbing and reflecting boundaries has been used in decision models with racing, parallel diffusion processes (Ratcliff & Smith, 2004; Usher & McClelland, 2001) and in models of simple RT (Diederich, 1995).

The information accumulation process in Ratcliff's model, again denoted as X(t), can be described by a stochastic differential equation,

$$dX(t) = \xi \, dt + s \, dW(t). \tag{2}$$

In this equation, ξ is a (random) drift coefficient, *s* is the square root of the diffusion coefficient, and W(t) is a standard Brownian motion process. The square root of the diffusion coefficient is also termed the infinitesimal standard deviation. A standard Brownian motion process has zero drift, unit variance, independent increments, covariance function $cov[W(\tau), W(t)] = min(\tau, t)$, and possesses a version which is almost surely continuous and is almost everywhere non-differentiable (Karlin & Taylor, 1981). In the psychological model of Eq. (2), it describes a process in which evidence is accumulated continuously in time and perturbed by broad spectrum Gaussian noise, idealized as white noise. The drift is assumed to be normally distributed, $\xi \sim N(\nu, \eta)$, with mean ν and standard deviation η . The standard deviation η describes the between-trial variability in stimulus quality, like the noise in signal detection theory. In most applications of diffusion process models, the model can be fitted with a single value of T_{er} , but Ratcliff and Smith (2010) found that a separate value of $T_{\rm er}$ was needed for each condition to account for the data from the dynamic noise task. We have used ξ and s to denote the drift and infinitesimal standard deviation in Eq. (1) to emphasize the link with Ratcliff's work, but elsewhere in the article we denote them by v and σ , respectively.

The additive decomposition in Eq. (1) is consistent with the kind of discrete stages model proposed by Sternberg (1969), in

which the process of stimulus identification does not begin until after the process of stimulus encoding is complete. If we make such an interpretation, and if we identify the process of stimulus identification with the accumulation of evidence by a diffusion process or some other sequential-sampling process, then Ratcliff and Smith's (2010) results imply that the effect of dynamic noise is to delay the onset of evidence accumulation: the noisier the stimulus, the more evidence accumulation is delayed. This is consistent with the phenomenology, but it begs the deeper question of how the decision process knows when to "turn on." To express this in less homuncular terms, what is the trigger signal or other mechanism that initiates the process of evidence accumulation, and how is it linked to the process of perceptual encoding?

1.2. Two models for discrimination in dynamic noise

Ratcliff and Smith (2010) proposed two general mechanisms that could adaptively couple the onset of evidence accumulation to the time course of stimulus encoding. One mechanism was based on the integrated system model of Smith and Ratcliff (2009), which is a form of stochastic continuous-flow system, like the cascade model of McClelland (1979). In the integrated system model, the onset of evidence accumulation is gradual rather than abrupt. The decision process becomes active as soon as stimulus information becomes available, but the rate of accumulation increases as the stimulus representation develops. The rate of evidence accumulation is controlled by a time-dependent diffusion coefficient that sets the clock of the process: the larger the diffusion coefficient, the more rapidly the process diffuses towards the absorbing boundaries. The coupling of the encoding and decision processes provided by the time-dependent diffusion coefficient avoids the need for a mechanism that initiates evidence accumulation based on an assessment of stimulus quality. The accumulation process in the integrated system is described by the stochastic differential equation

$$dX(t) = v(t)dt + \sigma(t)dW(t).$$
(3)

Here v(t) is a time-dependent drift and $\sigma(t)$ is a time-dependent infinitesimal standard deviation. The time dependency in the coefficients reflects the time course of perceptual encoding. In tasks with brief stimulus exposures the encoded stimulus information is identified with the contents of visual short-term memory (VSTM). In fitting the model to data, we again assume the additive decomposition of Eq. (1), but make a slightly different interpretation of $T_{\rm er}$. In Ratcliff's model, $T_{\rm er}$ is an aggregate of the times for perceptual encoding, response selection, and response execution. In the model of Eq. (3), stimulus information becomes available part way through the encoding process and begins to drive the decision process. The component of encoding that begins when the decision process becomes active and ends when the drift and diffusion coefficients have reached their maximum value is excluded from $T_{\rm er}$.

To obtain a well-behaved model, the drift and diffusion coefficients are assumed to be proportional to one another: $\nu(t) \propto \sigma^2(t)$. Here "well-behaved" means a model that predicts distributions of RT and orderings of correct responses and errors that resemble those found in empirical data. The drift is assumed to depend on the stimulus condition whereas the diffusion coefficient is the same for all conditions. In a neurally-inspired version of the model, the drift depends on the difference between an excitatory and an inhibitory process and the diffusion coefficient depends on their sum (Smith, 2010; Smith & McKenzie, 2011). The proportionality between the drift and diffusion coefficients is then only approximate rather than exact, but suffices to yield a well-behaved model. In either case, the rate of evidence accumulation

depends on the diffusion coefficient. When no stimulus information is present, v(t) and $\sigma^2(t)$ are both zero and no accumulation takes places.

The second mechanism proposed by Ratcliff and Smith (2010) was release from inhibition, which they conceptualized as a stimulus-dependent modulation of decay in an Ornstein–Uhlenbeck (OU) diffusion process (Busemeyer & Townsend, 1992, 1993; Smith, 1995, 2000; Usher & McClelland, 2001). The information accumulation process in this model can be described by the stochastic differential equation

$$dX(t) = [\nu(t) - \lambda(t)X(t)] dt + \sigma dW(t).$$
(4)

In this equation, v(t) and $\lambda(t)$ are, respectively, time-dependent stimulus information and decay coefficients. Unlike Eq. (3), the diffusion coefficient, σ^2 , is constant. In Eq. (4), evidence accumulation is controlled by $\lambda(t)$ rather than by the diffusion coefficient.

The release from inhibition mechanism relies on the properties of the stationary distribution of the OU process. Unlike the Wiener process in Eqs. (2) and (3), the OU process possesses a stationary distribution. For an OU process with constant stimulus information, $v(t) \equiv v$, and constant decay, $\lambda(t) \equiv \lambda$, the mean and variance of the process are

$$E[X(t)] = \frac{\nu}{\lambda} \left[1 - e^{-\lambda t} \right]$$
(5)

and

$$\operatorname{var}[X(t)] = \frac{\sigma^2}{2\lambda} \left[1 - e^{-2\lambda t} \right], \tag{6}$$

respectively (Karlin & Taylor, 1981; Smith, 2000). Because the process X(t) is Gaussian, its finite dimensional distributions are completely characterized by its first two moments, together with its covariance function. At large values of t, X(t) has a stationary Gaussian distribution, $N(\nu/\lambda, \sigma/\sqrt{2\lambda})$. If λ is large, most the probability mass will be concentrated in the vicinity of the starting point, X(0) = 0. There is a non-negligible probability that the process can reach an absorbing boundary and trigger a response, but when λ is large this probability will be small.

In Ratcliff and Smith's (2010) proposed release from inhibition model, at the beginning of the trial $\lambda(t)$ is large. Because the diffusion coefficient is constant, the process accumulates information, but on the majority of trials it stays near its starting point. At a point at which the quality of the information provided by the perceptual encoding process is sufficient, the inhibition is released, and the decay coefficient changes from a large to a small value. An OU process with small decay approximates a Wiener process (Ratcliff & Smith, 2004). Consequently, once inhibition is released, the accumulation process will satisfy a stochastic differential equation like Eq. (2). With $\xi = \nu$ and $s = \sigma$, the mean and variance of this process are

$$E[X(t)] = \nu t \tag{7}$$

and

$$\operatorname{var}[X(t)] = \sigma^2 t. \tag{8}$$

That is, once inhibition is released, the mean and variance of the process increase linearly with time as occurs in Ratcliff's (1978) model. The resulting model would be expected to behave similarly to Ratcliff's model with a random starting point and a value of $T_{\rm er}$ that depends on the time at which release from inhibition occurs.

Ratcliff and Smith (2010) discussed these mechanisms only in general qualitative terms. Our aim in this article is to describe formal implementations of them and to report fits to experimental data. To foreshadow our results, both models provide a good quantitative account of performance in the dynamic noise task. They accurately capture the leading edge effect and also the pattern of fast errors in Fig. 1. Notably, they do so without the assumption of between-trial variance in starting point. In the diffusion model the starting point of the accumulation process, *z*, is assumed to be uniformly distributed with range s_z (Ratcliff, Van Zandt, & McKoon, 1999). Starting point variability allows the model to capture the pattern of fast errors that is often found when discriminability is high and speed is stressed (Luce, 1986). Our models are able to capture this pattern without assuming trial-to-trial variation in starting point. Before we describe fits of the models we first characterize their qualitative properties to give the reader insight into the way in which they are able to predict the patterns of performance found in the data.

2. The integrated system model as a time-changed diffusion process

The accumulation process in the integrated system model can be viewed as a time-changed diffusion process, in which the instantaneous rate of evidence accumulation depends on the time-dependent diffusion coefficient, $\sigma^2(t)$. Useful insights into the properties of such processes can be obtained by considering the transformation that changes them into a standard Brownian motion process. This transformation is the basis for numerical integration equation methods for solving first-passage time problems for time-inhomogeneous diffusion processes developed by Ricciardi and co-workers (Buonocore, Giorno, Nobile, & Ricciardi, 1990; Buonocore, Nobile, & Ricciardi, 1987) and described by Smith (1995, 2000). Consideration of this transformation provides insights into how a model differs qualitatively from the fixed drift, constant absorbing barrier, model of Ratcliff (1978).

Under fairly general conditions on the drift and diffusion coefficients, there exists a coordinate transformation of the diffusion process X(t) of the form $X(t) \rightarrow X^*(t^*)$ such that the process $X^*(t^*) \equiv B(t^*)$ is a standard Brownian motion process. If it exists, the coordinate transformation is of the form

$$x^* = \bar{\Psi}(x, t) \tag{9}$$

$$t^* = \Phi(t). \tag{10}$$

Here stars denote transformed coordinates; the original coordinates are unstarred. The function $\bar{\Psi}$ that maps the old state coordinate to the new state coordinate, $x \rightarrow x^*$, is a function jointly of the old state coordinate, x, and the old time coordinate, t. The function Φ that maps the old time coordinate to the new time coordinate, $t \rightarrow t^*$, is a function of the old time coordinate only. A constructive proof for the existence of this transformation was given by Cherkasov (1957). It was put into a convenient form for applications by Ricciardi (1976) and stated more succinctly by Ricciardi and Sato (1983). The details may be found in Appendix B.

For a time-inhomogeneous Wiener process with drift v(t) and constant diffusion coefficient, σ^2 , which satisfies the stochastic differential equation

$$dX(t) = v(t) dt + \sigma dW(t), \tag{11}$$

the transformation is of the form (Smith, 2000, p. 446):

$$x^* = \frac{1}{\sigma} \left[x - \int^t v(s) \, ds \right] \tag{12}$$

$$t^* = t. \tag{13}$$

In this example, the transformation of the time coordinate is the identity; only the state coordinate changes. These equations state that the first passage time of the process X(t) through the constant absorbing boundaries a_1 and a_2 is the same as that of a standard Brownian motion process through the time-varying absorbing boundaries $a_1^*(t^*) = \bar{\Psi}(a_1, t)$ and $a_2^*(t^*) = \bar{\Psi}(a_2, t)$. In



Fig. 2. Coordinate transformation for a time-changed diffusion process. The panel on the left shows the transformation of the time coordinate; the panel on the right shows the transformation of the absorbing boundaries for a process starting at X(0) = 0. The process terminates when it first crosses one of the transformed boundaries, $a_1^* = \bar{\Psi}(a_1, t)$ and $a_2^* = \bar{\Psi}(a_2, t)$. The first passage times of the transformed process through the time-dependent boundaries are the same as those of the original process through constant boundaries, a_1 and a_2 . The dashed lines are for a high discriminability stimulus; the continuous lines are for a low discriminability stimulus. The diffusion coefficient, σ^2 , was set to unity.



Fig. 3. Coordinate transformation for the release from inhibition model. The dashed lines shown the transformation for a fast-release process; the continuous lines show the transformation for a slow-release process. The transformation of the state coordinate has been plotted against the transformed time coordinate, t*, to facilitate comparison.

the special case where $v(t) \equiv v$ (constant), the first passage time distributions are the same as those of a standard Brownian motion through the linear boundaries $(a_1 - vt)/\sigma$ and $(a_2 - vt)/\sigma$. We have followed Ricciardi and Sato (1983) and written Eq. (12) and similar equations without a lower bound of integration because the lower bound contributes only an inessential constant to the transformation (Eq. (B.3)). An implied lower bound of zero was used to generate Figs. 2 and 3.

For the accumulation process in the integrated system model in Eq. (3), both the drift and diffusion coefficient depend on time. When the drift is proportional to the diffusion coefficient, $\sigma^2(t) \equiv \sigma^2 v(t)$, as assumed in the model, it is shown in Appendix B that the transformation that maps the process to a standard Brownian motion process is

$$x^* = \frac{1}{\sigma} \left[x - \int^t v(s) \, ds \right] \tag{14}$$

$$t^* = \int_0^t v(s) \, ds. \tag{15}$$

Here *s* is the variable of integration; it is not the infinitesimal standard deviation of Eq. (2).

The transformation of the state coordinate is the same as that for the time-inhomogeneous Wiener process with constant diffusion coefficient in Eq. (11), but the time coordinate now changes. Specifically, the new time coordinate is the integral of the drift with respect to the old time coordinate. The larger the drift, the larger the value of the new time coordinate and the more rapidly the clock of the process will run. The proportionality

between the drift and diffusion coefficient in Eq. (3) means that the clock of the process, which determines how rapidly it diffuses towards the absorbing boundaries, will depend on the stimulus. This introduces an additional source of stimulus dependency into the temporal properties of the model, independent of that arising from the stimulus dependency in drift.

Fig. 2 shows the effects of the transformation in Eqs. (14) and (15) when the drift grows smoothly to an asymptote. The parameters of the function v(t) used to generate this figure were similar to those used to fit the data in Fig. 1, as described subsequently. Fig. 2 shows the resulting transformation of the absorbing boundaries and time coordinate under conditions in which stimulus information becomes available rapidly or slowly. The asymptotic stimulus discriminability, $v(\infty)$, was the same in both conditions. The process was assumed to start at zero, X(0) = 0, and to have positive drift, so responses made at the upper boundary were correct; responses made at the lower boundary were errors. The specific mechanism that controlled the rate at which stimulus information became available was based on the integrated system model and is described subsequently.

Fig. 2 shows that the absorbing boundaries of the transformed process are asymptotically linear with slopes that depend on processing rate: the more rapidly stimulus information becomes available the steeper the slope. Functionally, this means that the process will terminate more rapidly when stimulus information becomes available more rapidly. The transformed time coordinate is also asymptotically linear with a slope that depends on discriminability. The effect of the transformation is to map the non-zero-drift process, X(t), to a zero-drift process, $B(t^*)$, with $t^* < t$.

This means that a process with time-varying drift is equivalent—in the sense of having the same first passage statistics—to a process with a constant diffusion coefficient evaluated at some earlier time. In other words, the process with time-varying diffusion coefficient is slowed relative to a process with constant diffusion coefficient, with the extent of the slowing depending on the rate at which stimulus information becomes available. The left hand panel in Fig. 2 shows that the function $\Phi(t)$ that characterizes the clock of the time-inhomogeneous process approximates a shifted straight line, with larger shifts for lower discriminability stimuli. This property is the basis of the model's ability to predict the leadingedge effect in Fig. 1.

3. The release from inhibition model

Fig. 3 shows the transformation that maps the release from inhibition model to a standard Wiener process. In Appendix B it is shown that this transformation is

$$x^{*} = \frac{1}{\sigma} \left\{ x \exp\left[\int^{t} \lambda(s) ds\right] - \int^{t} \nu(s) \exp\left[\int^{s} \lambda(z) dz\right] ds \right\}$$
(16)
$$t^{*} = \int_{0}^{t} \exp\left[2\int^{s} \lambda(z) dz\right] ds.$$
(17)

This transformation generalizes the well-known result (Cox & Miller, 1965, p. 229), that a standard OU process, with drift $-\lambda x$ and unit variance, can be realized from the Wiener process by an exponential expansion of the time variable and an exponential contraction of the state variable. For an OU process with constant decay λ , the transformation of the time axis is of the form, $\Phi(t) = [\exp(2\lambda t) - 1]/(2\lambda)$ where the process is assumed to start at time t = 0 (Ricciardi, 1976, p. 195; Smith, 2000, p. 447). The release from inhibition model has an initial pre-release segment with large λ followed by a post-release segment with zero or near-zero λ . The transformation in Eq. (17) is then approximately of the form $t^* = kt$, where the constant k is an increasing function of the time at which inhibition is released.

The coordinate transformations shown in Fig. 3 are for slow (continuous lines) and fast (dashed lines) release from inhibition processes, respectively. The implementation of the release from inhibition process is described subsequently (Eqs. (20)–(22)). The parameters of the fast process were chosen so that the release was virtually instantaneous. Under these conditions, the resulting process approximates a constant drift Wiener process. The transformation of the time axis approximates the identity, $t^* \approx t$ (i.e., $k \approx 1$), and the transformed boundaries are $a_1^* \approx (a_1 - \nu t)/\sigma$ and $a_2^* \approx (a_2 - \nu t)/\sigma$.

The continuous lines in the figure are for a slow process, in which the release from inhibition took around 100 ms to complete. The transformation of the time axis is again approximately linear, with k > 1.0. The transformed boundaries on the right of the figure have an initial exponential segment that represents the pre-release phase followed by a linear segment that represents the post-release phase. The exponential expansion of the initial segment of the boundaries is a reflection of the fact that the OU process is obtained from the Wiener process by an exponential contraction of the state variable. Consequently, the probability that an OU process will pass through a given level, x, is equal to the probability that the Wiener process will pass through some other level, x^* , which is exponentially further away.

Because of the comparatively large difference in the transformation of the time coordinate for slow and fast processes, for ease of interpretation the transformed boundaries in Fig. 3 have been plotted against the transformed time coordinate, t^* , rather than the original time coordinate, t. The shallower slopes of the linear segments of the transformed boundaries for the slow process is a



Fig. 4. Simulated sample paths for the release from inhibition model and the Wiener diffusion model. The function $\lambda(t)$ in the upper panel shows the time course of inhibition for the progressive release model of Eq. (22). (The amplitude of this function is not shown to scale.) The dashed horizontal lines a_1 and a_2 represent decision criteria. The simulated sample paths are for a process with drift parameter $\mu = 0.5$ and infinitesimal standard deviation (*s* in Eq. (2) and σ in Eq. (4)) of 0.1. The parameters of $\lambda(t)$ were $\lambda(0) = 50.0$, $\alpha = 5.0$, and $\beta = 0.016$ (Eqs. (20) and (21), and Table 2).

reflection of the slower rate at which it terminates. During the initial OU segment of the slow process, in which the boundaries are curvilinear, the correct response boundary is comparatively further away from the starting point than is the corresponding boundary for the fast process. This means the process is more likely to terminate at the incorrect response boundary and produce an error than is the fast process. As a result, the process can generate fast errors like those in Fig. 1.

To provide further insight into the properties of the model, the upper panel of Fig. 4 shows 25 simulated sample paths for the release from inhibition model of Eq. (4). For comparison purposes, the lower panel shows 25 simulated paths for the Wiener diffusion model of Eq. (2). The function $\lambda(t)$ in the upper panel shows the time course of inhibition (Eq. (22)). The time course is typical of what we have obtained from fits to data, but the amplitude is not shown to scale. For the parameters of $\lambda(t)$ in Fig. 4, the release from inhibition begins around 50 ms after stimulus onset and is largely completed 100 ms later. In fitting the data in Fig. 1 we used a starting value of inhibition of $\lambda(0) = 15.0$, but for the simulations in Fig. 4 we used a larger value of $\lambda(0) = 50.0$ in

order to emphasize the difference between the pre-release and post-release phases of the inhibition process. It is evident from a comparison of the two panels that the effect of inhibition is to slow responding. None of the processes in the upper panel terminate until after inhibition is fully released at around t = 150 ms, whereas a significant proportion of the processes in the lower panel terminate in the first 100 ms. At a distribution level, this slowing manifests itself as a shift in the leading edge.

4. Fitting the integrated system model

4.1. Assumptions of the model

The integrated system model (Sewell & Smith, 2012; Smith, Ellis, Sewell, & Wolfgang, 2010; Smith & Ratcliff, 2009) combines a time-inhomogeneous diffusion decision process with a process model of drift (Appendix C). The drift model seeks to characterize the combined effects of perception, memory, and attention on performance in speeded two-choice tasks. The model was developed to account for the effects of spatial attention in nearthreshold visual tasks with briefly presented stimuli and, to that end, it assumes that stimuli are encoded perceptually by spatiotemporal visual filters (Watson, 1986). Stimuli that are relevant to the current task are transferred to visual short-term memory (VSTM) under the control of spatial attention. The VSTM representation of the stimulus forms the basis for the perceptual decision. In tasks that use noisy, response-terminated stimuli, like the dynamic noise task, we assume that the decision process is driven by a stable representation of the stimulus that is extracted from the dynamically changing perceptual input. We identify this stable representation with the VSTM trace in the model.

The process of VSTM trace formation in the integrated system model is described computationally by a shunting differential equation (Eq. (C.4)). Equations of this kind have been proposed as neurocomputational models of short-term memory by Grossberg (1980, 1988) and as cognitive models by Loftus and colleagues (Busey & Loftus, 1994; Loftus & Ruthruff, 1994). In the integrated system model, the time-varying VSTM trace, v(t), determines both the drift and the diffusion coefficient of the accumulation process in Eq. (3). Some of the features of the integrated system model are arguably not required for the dynamic noise task, in which stimuli were always presented at a known location in the center of the display. However, we chose to fit the model in much the same form as described by Smith and Ratcliff (2009), to facilitate comparison with their work.

Apart from the dynamic noise task, we know of at least one other situation in which a leading edge has been found, and that is when very low contrast stimuli are presented directly on a uniform field. In Smith and Ratcliff's (2009) article, they compared performance in two superficially similar attentional cuing paradigms, in which the task was to discriminate the orientations of briefly presented, low contrast, Gabor patches presented at either cued or uncued locations. In one study, reported by Smith, Ratcliff, and Wolfgang (2004), the stimuli were presented on top of suprathreshold (15%) contrast luminance disks, or pedestals, whose purpose was to localize stimuli perceptually in the display. In the other study, reported by Gould, Wolfgang, and Smith (2007), stimuli were presented directly against a uniform field and were localized by surrounding them with four arms of a high contrast fiducial cross. Despite the similarities in these two tasks, the fiducial cross task produced a marked leading edge effect, but the pedestal task did not. The 0.1 quantile functions in the pedestal task were relatively flat, like those found in the majority of other tasks we have studied. One of our aims in this article was to investigate whether Smith and Ratcliff's model for the leading edge effect in the fiducial cross task could also account for the effect in the dynamic noise task.

To fit the integrated system model to data, parameters of four different subprocesses must be specified: the sensory response function; the attention and VSTM processes; the decision process, and the remaining, nondecision processes. A total of 14 parameters are needed to fit the data in Fig. 1, three of which can be specified a priori, leaving a total of 11 parameters that must be estimated from the data. These subprocesses and their associated parameters are described below.

4.2. Sensory response function

The model assumes that stimuli are encoded perceptually by visual filters, possibly in cortical area V1. The time course of perceptual encoding is characterized by a sensory response function, $\mu(t)$ (Eq. (C.1)). As is common in visual psychophysics, the sensory response function is modeled as a series of cascaded exponential stages (Busey & Loftus, 1994; Smith, 1995; Sperling & Weichselgartner, 1995; Watson, 1986). Following Smith and Ratcliff (2009), we set the number of cascaded stages to n = 3, and estimated the stage rate, β_{on} , from the data. The fit of the model is relatively insensitive to the value of n because it mainly affects the time at which $\mu(t)$ begins to provide sensory information and this time can be absorbed into the value of $T_{\rm er}$. The subscript on the rate refers to the time course of sensory onset. When brief stimulus exposures are used, a second, offset rate must also be estimated (Eq. (C.1)), whose value varies depending on whether or not stimuli are backwardly masked. When stimuli are response-terminated, as here, an offset rate is not required.

The amplitude of the sensory response is assumed to be a nonlinear function of the stimulus contrast or intensity. Unlike most applications of diffusion models, in which separate drift parameters are estimated for each stimulus condition, the integrated system model assumes the drifts in all conditions can be described by a single Naka–Rushton function (Eq. (C.3)). The Naka–Rushton function is a sigmoid function, which has an expansive nonlinearity at low contrasts and a compressive nonlinearity at high contrasts. Such functions have been shown to provide good models of both the visual contrast response of cortical neurons (Boynton, 2005; Heeger, 1991; Kaplan, Lee, & Shapley, 1990) and of psychophysical contrast sensitivity (Foley, 1994; Foley & Schwarz, 1998). Functions of this form arise theoretically from so-called divisive inhibition models of contrast sensitivity, in which visual mechanisms tuned to a particular spatial frequency and orientation are inhibited by mechanisms with different tunings at the same or surrounding locations (Foley, 1994; Reynolds & Heeger, 2009). They are obtained naturally from shunting equations like the one in the integrated system model (Smith et al., 2010).

To fit the data in Fig. 1, we assumed the amplitude of the sensory response, *I*, was a Naka–Rushton function of stimulus contrast (Eq. (C.3)),

$$I = \frac{\Delta_I^2}{\Delta_I^2 + I_{\rm in}},\tag{18}$$

where $\Delta_I = 1 - 2\pi$, π is the proportion of pixels inverted in the letter and the background, and I_{in} is a divisive inhibition term that determines the horizontal position of the function on the logcontrast axis. The quantity $1-2\pi$ is the Michelson contrast, Δ_I / Σ_I , where Δ_I is the luminance difference between the letter and the background and Σ_I is the luminance sum (De Valois & De Valois, 1990, p. 6). In the dynamic noise task, the average luminance of the letter is proportional to $1 - \pi$; the average luminance of the background is proportional to π , so the resulting contrast is

$$\frac{\Delta_l}{\Sigma_l} = \frac{(1-\pi) - \pi}{(1-\pi) + \pi} = 1 - 2\pi.$$

In tasks in which low-contrast stimuli are presented under conditions of spatial uncertainty, the exponent in Eq. (18) often takes values other than 2. We found we did not need this extra generality to account for the data in Fig. 1 and have constrained the exponent accordingly. To fit the model to data, we therefore needed to estimate two parameters of the sensory response function: the stage rate, β_{on} , and the divisive inhibition term, I_{in} .

4.3. Attention and the VSTM trace

In the second stage of the model, the stimulus information in the sensory response function is encoded in a durable form in VSTM. The strength of the VSTM trace determines the drift of the diffusion process. Three additional parameters are needed to specify the dynamics of VSTM trace formation: an asymptotic VSTM amplitude, θ ; a rate parameter, γ , and a stimulus saliency parameter, I_0 . The VSTM amplitude parameter determines the scaling between trace strength and drift; the rate parameter determines the overall rate at which the VSTM trace is formed, and the saliency parameter characterizes the relationship between stimulus contrast and the rate of trace formation.

The rate of VSTM trace formation is assumed to be jointly a function of attention gain and the perceptual saliency of the stimulus. Attention gain describes the proportion of the participant's resources that is allocated to a particular location in the visual field at a given time during an experimental trial. In spatial cuing tasks, gain varies depending on whether the stimulus is presented at a cued or an uncued location. In tasks in which attention is reallocated during a trial, gain will also vary with time (Sewell & Smith, 2012). In the dynamic noise task, gain can be treated as a constant.

The effects of saliency are needed to account for the differences in the leading edges of the RT distributions of Smith et al. (2004), in which low contrast stimuli were presented on top of suprathreshold contrast luminance pedestals, and those of Gould et al. (2007), in which stimuli were presented against a uniform field. Smith and Ratcliff (2009) attributed the differences in performance in these tasks to differences in the perceptual saliency of the stimuli. They used an energy-based summation function to describe saliency that reflected the physical properties of their stimuli. Its theoretical substance was that the combined effects of stimulus saliency and attention can be characterized by a function that has both contrast-dependent and contrastindependent components and which increases roughly linearly with contrast. For the dynamic noise task, we found we could characterize this relationship using a function of the form, $\gamma =$ $\gamma_g(I + I_0)$, where I is the transduced stimulus contrast, γ_g is the attention gain, and I_0 is a constant. With these assumptions, three attention/VSTM parameters are required to fit the model to data: the attention gain, the saliency parameter, I_0 , and the VSTM amplitude, θ . The combination of gain, saliency, and the Naka–Rushton inhibition parameter, *I*_{in}, predict the magnitude of the leading edge effect.

4.4. The decision process

To fit the dynamic noise data we assumed a symmetric decision process, in which drift rates for the two stimuli and the criteria for the two responses were equal in magnitude and opposite in sign. We assumed one criterion for the speed condition, a_s , and another for the accuracy condition, a_a . We also assumed, like other applications of diffusion models (Ratcliff & Smith, 2004), that drift rates were normally distributed with standard deviation, η . In the integrated system model, η describes the distribution of the asymptotic VSTM trace strengths, $v(\infty)$.

As a time-changed Wiener process, the decision process in Eq. (3) is unable to predict the fast errors in Fig. 1. Like the

Wiener process in Ratcliff's (1978) diffusion model, the model predicts equal mean RTs for correct responses and errors in the absence of other sources of variability. When there is variability in drift rates it can predict slow errors, but it cannot predict fast errors. Following the work of Laming (1968), fast errors are often attributed to variability in the starting point of the diffusion process. Laming argued that participants' uncertainty about the precise time of stimulus onset leads them to sample noise from the prestimulus field. When the stimulus is presented, the decision process begins to accumulate evidence from a random starting point that depends on the sample of prestimulus noise. Trials on which the prestimulus noise drives the process towards the incorrect response boundary are more likely to terminate rapidly and with an error. Ratcliff et al. (1999) showed that the addition of starting point variability to Ratcliff's (1978) diffusion model allowed it to predict fast errors, as Laming proposed. When the model has both drift variability and starting point variability, it can predict fast errors and slow errors, and the crossover pattern found in some data sets, in which errors are faster than correct responses in some conditions and slower in others (Ratcliff & Smith, 2004).

Rather than attributing the fast errors in Fig. 1 to starting point variability, we used the model of Smith and Ratcliff (2009), which was also motivated by Laming's (1968) theoretical ideas. In the evidence accumulation model of Eq. (3), the diffusion coefficient, $\sigma^2(t)$, depends on the encoded stimulus information and increases progressively over time. To model performance in the Gould et al. (2007) study, Smith and Ratcliff assumed an additional source of constant diffusive noise, independent of the noise in the stimulus representation. This noise is like the noise from the prestimulus field proposed by Laming, but is assumed to persist throughout the trial. The evidence accumulation function for this extended model can be written

$$dX(t) = v(t) dt + \sigma_1 v(t) dW_1(t) + \sigma_2 dW_2(t).$$
(19)

Because of the additive properties of Brownian motion processes, the model can be viewed as a process with a single coactive source of noise, W(t), with infinitesimal standard deviation $\sqrt{\sigma_1^2 v(t) + \sigma_2^2}$. Following the conventions used in Ratcliff's work,

we set $\sigma_1 = 0.1$ and estimate σ_2 from the data.

4.5. Nondecisional processes

To complete the model, we make the same assumptions as are made in Ratcliff's work (e.g., Ratcliff & Smith, 2004), that RT is the sum of decision and nondecision times, as stated in Eq. (1). The nondecision time is assumed to be uniformly distributed with mean $T_{\rm er}$ and range s_t . Because nondecisional processes contribute only a small part of the overall variance of RT, the particular form assumed for the distribution of $T_{\rm er}$ has a minimal effect on the shapes of the predicted RT distributions.

4.6. Fitting and estimation

We obtained first passage time and first passage probability statistics for the model using the integral equation methods proposed by Buonocore et al. (1990), Gutiérrez Jáimez, Román Román, and Torres Ruiz (1995), and described by Smith (2000). The code for the model was implemented in C, called from Matlab, as described by Smith and Ratcliff (2009). To fit the model, we minimized the likelihood ratio statistic, G^2 ,

$$G^2 = 2 \sum_{i=1}^{10} n_i \sum_{j=1}^{12} p_{ij} \log \frac{p_{ij}}{\pi_{ij}},$$

on the quantile-averaged group data in Fig. 1, using the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles to group the data into bins. In this equation,



Fig. 5. Fit of the integrated system model. The data and fitted values are shown as a quantile probability plot. The symbols are the experimental data; the continuous curves are the fitted values.

the outer summation over *i* runs over the five discriminability levels in the speed and accuracy conditions in Fig. 1. The inner summation over *j* extends over the 12 bins formed by each pair of joint distributions of correct responses and errors. (There were five quantiles per distribution, resulting in six bins per distribution, or 12 bins in total, with 11 degrees of freedom, for each distribution pair.) In this equation, p_{ij} is the proportion of probability mass falling in the *j*th bin in the *i*th condition, and π_{ii} is the proportion of mass predicted by the model. (It is not, as we have used it elsewhere, the proportion of inverted pixels in the display.) The number n_i is the number of observations in each condition of the experiment. As multinomial sampling assumptions are not satisfied for G^2 computed on quantile-averaged data, we follow Ratcliff and Smith (2004) and set n_i (arbitrarily) to 100 to obtain a convenient scale for fitting. We do not interpret the resulting fit statistic as a chi-square, but simply treat it as a comparative measure of fit. To fit the model we minimized G^2 iteratively using the Matlab implementation of the Nelder and Mead (1965) simplex algorithm. We chose to fit the model to binned data rather than by maximum likelihood both for reasons of computational efficiency and because Ratcliff and Tuerlinckx (2002) showed that fits to

Table 1

Parameters	of the	integrated	system	model
i ai ai ii c c c i s	or the	megrateu	System	mouci.

Parameter	Symbol	Value
Sensory response function		
Onset rate	β_{on}	26.68
Number of stages	n	3 ^a
Naka-Rushton inhibition	I _{in}	0.064
Naka-Rushton exponent	ρ	2 ^a
Attention/VSTM		
Attention gain	γ_g	19.93
VSTM amplitude scaling	θ	1.226
Saliency constant	I ₀	0.509
Decision process		
Boundary separation (speed)	$2a_s$	0.083
Boundary separation (accuracy)	$2a_a$	0.136
Drift variability	η	0.171
Diffusion coefficient (square root)	σ_1	0.100 ^a
Stimulus independent diffusion	σ_2	0.013
Nondecision processes		
Mean nondecision time	Ter	0.301
Nondecision time range	s _t	0.134
lote Time-based parameters are in units of s	econds: state-based n	arameters are in

Note. Time-based parameters are in units of seconds; state-based parameters are in units of infinitesimal standard deviations per second.

^a denotes a fixed parameter.

binned data are less susceptible to the effects of contaminants and outliers in the data than is maximum likelihood.

The fit of the integrated system model is shown in Fig. 5. The best-fitting model, with 11 free parameters, yielded $G^2(99) =$ 15.59. The estimated model parameters are shown in Table 1. When we used the more complex saliency function assumed by Smith and Ratcliff (2009), the fit was only slightly better, $G^2(99) =$ 14.57. As shown in Fig. 5, the model does a good job of accounting for the main features of the empirical data. The only systematic discrepancies in fit are that it slightly underpredicts the 0.9 quantiles at the lowest levels of stimulus discriminability. Critically, the model predicts a leading-edge effect that closely matches that found in the data in both the speed and accuracy conditions. The predicted 0.1 quantile functions exhibit the same bowed shape as those in the data. The assumption that asymptotic drift rate is a Naka-Rushton function of stimulus contrast predicts response accuracy in both the speed and the accuracy condition. This represents an appreciable gain in model parsimony; instead of assuming separate drift rates for each of the five discriminability conditions the model captures the differences in accuracy between conditions with only two parameters: the Naka-Rushton inhibition parameter, I_{in} , and the VSTM amplitude scaling parameter, θ .

Fig. 5 also shows that the addition of a second source of diffusive variability allows the model to accurately predict the fast errors in the speed condition in Fig. 5, represented by the bowing downward to the left of the 0.1 and higher quantiles. The amount of additional variability needed to predict fast errors is small: Table 1 shows that the estimated value of σ_2 is only one tenth the magnitude of σ_1 . Nevertheless, it is crucial to the performance of the model. When σ_2 is set to zero, the fit appreciably worsens, $G^2(100) = 33.4$. This is worse than the fit in Fig. 5 by a factor of two.

5. Fitting the release from inhibition model

5.1. The time course of perceptual encoding

The two key assumptions of the release from inhibition model are: (a) dynamic noise delays the process of forming a representation of the information in the stimulus, and (b) the process of evidence accumulation is controlled by a time-dependent OU decay coefficient that is time-locked to the developing representation. Instead of assuming the VSTM-based drift model of the integrated system model, we sought to implement these assumptions in the simplest possible way, to allow us to evaluate the performance of the release from inhibition mechanism unencumbered by assumptions about perceptual encoding. To this end, we assumed the time course of sensory encoding was described by an incomplete gamma function (Abramowitz & Stegun, 1964, p. 260),

$$\mu(t;\alpha) = \frac{1}{\Gamma(\alpha)} \int_0^t s^{\alpha-1} e^{-s} \, ds. \tag{20}$$

Our choice of notation here serves to emphasize the link with the sensory encoding function in the integrated system model (Eq. (C.1)). We assumed the stimulus-dependent component of the drift in Eq. (4) was of the form

$$\nu(t) = \theta I \mu(t/\beta; \alpha), \tag{21}$$

where *I* is the transduced contrast of Eq. (18), β is a dispersion parameter, and θ is a parameter that maps stimulus information to drift amplitude (cf. Eq. (C.4)). For integral α , the normalizing constant, $\Gamma(\alpha)$, in Eq. (20) reduces to the factorial function, $(\alpha - 1)!$, and the function then has the same form as the $(\alpha - 1)$ -stage, cumulative gamma probability distribution. When Eq. (20) is viewed as a deterministic function rather than a probability distribution, it describes the output of a linear filter composed of $\alpha - 1$ cascaded exponential stages (cf. Eq. (C.2)). The incomplete gamma function generalizes the linear filter model to allow the number of "stages" to be non-integral. We chose this more general representation to allow us to smoothly parametrize the model for ease of estimation. In so doing, we are using Eq. (20) simply to describe the time at which stimulus information becomes available to the decision process; we are not assuming that the number of processing stages changes as a function of external noise.

Fig. 6 shows examples of the function in Eq. (21) for different values of α , which represent different levels of noise in the stimulus, obtained by fitting the release from inhibition model to our data. The differences in the amplitudes of these functions reflects differences in the transduced contrast, *I*; the differences in the time at which stimulus information becomes available is reflected by differences in α . As the release from inhibition model was not based on an a priori theory of sensory encoding, we treated the time of release as a free parameter of the model, giving five free α parameters. The other free parameters of the encoding process were the Naka–Rushton inhibition, *I*_{in}, the drift amplitude scale, θ , and the incomplete gamma dispersion, β , giving a total of eight free parameters to specify the sensory response. As previously, we fixed the Naka–Rushton exponent to $\rho = 2$.

5.2. Time course of the release from inhibition

We considered two versions of the release from inhibition process: a progressive release model, in which inhibition decreases progressively as stimulus information becomes available, and an all-on-none model, in which inhibition is released when the stimulus encoding function reaches some threshold. The progressive release model is in the spirit of continuous flow models like the cascade model of McClelland (1979); the all-or-none model is in the spirit of discrete-stages models like the additive factors model of Sternberg (1969).

To implement the progressive release model, we assumed that the decay function in Eq. (4) was modulated by the encoded stimulus information,

$$\lambda(t) = \lambda_0 [1 - \mu(t/\beta; \alpha)], \tag{22}$$

where λ_0 is the initial inhibition, which we set to an arbitrary, large value, $\lambda_0 = 15.0$. The function $\lambda(t)$ decreases smoothly from λ_0 to zero at a rate that is time-locked to the sensory encoding process. The effect of increasing α is to delay the time at which stimulus information becomes available, as shown in Fig. 6. Eq. (22) states



Fig. 6. Time course of perceptual encoding estimated from the release from inhibition model. (a) Drift functions, v(t), as a function of time. (b) Drift amplitude, *I*, as a function of contrast. (c) Release from inhibition time, α , as a function of contrast.

(a) that the release from inhibition is delayed by a corresponding amount, and (b) that the rate at which inhibition is released depends on the rate at which encoded stimulus information becomes available, which is controlled by the dispersion parameter, β .

We implemented the all-or-none model in a very simple way, by assuming that inhibition was released once the encoded stimulus information reached 50% of its maximum,

$$\lambda(t) = \begin{cases} \lambda_0 & \mu(t/\beta; \alpha) < 0.5\\ 0 & \mu(t/\beta; \alpha) \ge 0.5. \end{cases}$$
(23)

The release from inhibition mechanism defined in this way uses a relative threshold; the large variation in the amplitudes of the estimated encoding functions in Fig. 6 suggests it would have been difficult to specify a well-behaved model that used an absolute threshold. The choice of a 50% threshold was arbitrary; very similar results would have been obtained with other choices. We also considered a stochastic version of an absolute threshold model, in which the time of release from inhibition was controlled by a single-barrier diffusion process, which follows a Wald distribution (Luce, 1986, p. 509). This model performed comparatively poorly, so we have not reported it. With these assumptions both versions of the release from inhibition model were fully specified by the parameters of the sensory response function, together with the fixed λ_0 parameter.

5.3. Decision and nondecision processes

The assumptions we made about the decision and nondecision processes were identical to those we made about the integrated system model. Specifically, we assumed that decisions were made by a two-barrier diffusion process, described by Eq. (4). We assumed one criterion for speed and one criterion for accuracy and a single value of the drift standard deviation, η . We also assumed a single value for the mean nondecision time, $T_{\rm er}$, and for the nondecision time range, s_t .

5.4. Fitting and estimation

We obtained first passage time statistics for this process by approximating it as finite-state Markov chain, as described by Diederich and Busemeyer (2003) and as implemented by Smith and Ratcliff (2009). In this approach, a continuous-time, continuous-state process is approximated by a discrete-time, discrete-state birth-death process. We approximated the process on a time scale of h = 0.0025 s using a state vector of $\{\pm i\Delta\}$, i =0, 1, 2, ..., $\Delta = \sigma \sqrt{h}$, which is the natural scaling for a diffusion process. We then defined a time-dependent transition matrix that specified the probability of an increment or a decrement of size Δ at each step, conditional on the current state of the process and the inhibition function $\lambda(t)$, using the well-known relationship between the coefficients of a continuous time diffusion process and the transition probabilities of the approximating Markov chain (Bhattacharya & Waymire, 1990, Ch. V.5), together with appropriate boundary conditions. To approximate a diffusion process without starting point variability, we defined the initial state of the process as a Kronecker delta function, which defines the distribution of a process all of whose mass is concentrated at the state X(0) at time t = 0. We then obtained the first passage time distributions for the approximating process by matrix multiplication. Potentially, we could have used integral equation methods for the progressive release from inhibition model, but not for the all-or-none model, as these methods require smoothness of the underlying functions (specifically, two continuous derivatives, Smith, 2000). We therefore chose to use a Markov-chain approximation for both models.

Fig. 7 shows a quantile–probability plot of the fitted progressive release model; Table 2 shows the estimated model parameters. Fig. 6b shows the estimated drift amplitudes, *I*, as a function of contrast and Fig. 6c shows the estimated α parameters, which



Fig. 7. Fit of the release from inhibition model.

determine the time of release from inhibition. The drift amplitudes are fairly linear in contrast; the release times are a decreasing, negatively accelerating function of contrast. Of the two versions of the model, the best-fitting was the progressive release model. For this model, $G^2(97) = 17.18$; for the all-or-none model, $G^{2}(97) = 28.05$, which is more than 60% worse. The quality of the fit of the progressive release model was similar to the integrated system model, although it required three more free parameters. Potentially, the parsimony of the model could be improved by seeking a simpler functional characterization of the release time parameters in Fig. 6c, but we have not attempted to do this. Overall, then, the progressive release from inhibition model provides a comparably good account of the leading edge effect and fast error data to that of integrated system model, but does so using different assumptions about the processes that couple sensory encoding and decision-making.

6. Generalizing to other dynamic noise tasks

Ratcliff and Smith (2010) attributed the leading edge effect to the time needed to form a perceptual representation of the features of a noisy stimulus. Although they showed the effect occurs in

Ta	ble	2
	DIC.	~

Parameter	Symbol	Value
Drift function		
Naka-Rushton inhibition	I _{in}	0.049
Naka-Rushton exponent	ρ	2 ^a
Drift amplitude scaling	θ	0.829
Incomplete gamma dispersion	β	0.016
Incomplete gamma location	α_1	7.71
Incomplete gamma location	α_2	6.39
Incomplete gamma location	α ₃	4.82
Incomplete gamma location	α_4	3.46
Incomplete gamma location	α_5	2.41
Decision process		
Boundary separation (speed)	$2a_s$	0.073
Boundary separation (accuracy)	$2a_a$	0.114
Drift variability	η	0.128
Diffusion coefficient (square root)	σ	0.100 ^a
Prerelease inhibition	λ_0	15.0 ^a
Nondecision processes		
Mean nondecision time	T _{er}	0.353
Nondecision time range	<i>s</i> _t	0.139

Note. Time-based parameters are in units of seconds; state-based parameters are in units of infinitesimal standard deviations per second.

^a denotes a fixed parameter.

letter discrimination, they did not investigate whether it occurs in other tasks in which stimulus features are presented in dynamic noise. Letter discrimination is a relatively complex task because the stimuli are comprised of multiple features. To perform it, people must form representations of the stimulus configuration as a whole, not just the individual features. To test the generality of Ratcliff and Smith's conclusions, we investigated discrimination in dynamic noise in two simpler feature processing tasks. The stimuli for these tasks are shown in Fig. 8; details of the methods can be found in Appendix A.

In the first task, participants discriminated the orientation of three white parallel bars, presented vertically or horizontally on a black background. Discriminability was varied by randomly inverting some proportion of the pixels in the bars and the background in each 16.7 ms frame of the display. On half of the trials, the proportion of inverted pixels was less than 50%, creating the appearance of noisy light bars on a dark background, as shown in the example in Fig. 8a. On these trials, 0.35, 0.43, 0.46, or 0.475 of the pixels were inverted in each frame. (Like the letter discrimination task, when 50% the pixels are inverted the display becomes a homogeneous field of noise that carries no discriminative information.) On the remaining half of the trials, the proportion of inverted pixels was more than 50%. On these trials, 0.65, 0.57, 0.54, or 0.525 of the pixels were inverted, creating the appearance of dark bars on a light background. We included both kinds of trials in order to increase stimulus uncertainty, which Ratcliff and Smith (2010) found increased the magnitude of the leading edge effect. For the purposes of analysis, we averaged the data from trials with dark and light and vertical and horizontal bars within each contrast level after verifying there were no systematic differences in performance on the various trial types. The data, shown in the top half of Fig. 8, are quantile-averaged over 20 participants.

In the second task, participants discriminated the orientation of vertical or horizontal Gabor patches (Gaussian vignetted sinusoidal gratings). Gabor patches differ from bars in that they are bandlimited in their spatial frequency composition, whereas bars are broadband (see Graham, 1989, for a discussion). Band-limited stimuli are useful if one wishes to target visual mechanisms tuned to a single spatial scale; broadband stimuli target multiple mechanisms tuned to different scales. We ran experiments using both band-limited and broadband stimuli in order to establish whether the leading edge effect would occur with feature stimuli



Fig. 8. Discrimination of (a) oriented bars and (b) Gabor patches in dynamic noise. The figure shows example stimuli and fits of the integrated system model to the quantile averaged group data. For each task, the stimuli are (top to bottom): a noise-free vertical stimulus, a noisy single frame, and a 10-frame average.

of different kinds. As shown in Fig. 8a, the stimuli were constructed by sinsoidally modulating the luminance of the display around its mean value to produce a Gabor patch with a peak contrast of 20%. (Peak contrast is defined as the maximum depth of the luminance modulation divided by the luminance of the background.) Stimulus discriminability was varied by replacing 0.35, 0.50, 0.65, or 0.80 pixels in the display with uniformly-distributed grayscale noise pixels. The noise had a mean equal to the mean luminance of the display and a range of \pm 50%. For the purposes of analysis, we averaged the data from trials with vertical and horizontal stimuli within each discriminability level after verifying there were no systematic differences in performance on the two trial types. The data, quantile-averaged across 20 participants, are shown in Fig. 8b.

As Fig. 8 shows, there were leading edge effects in both tasks and fast errors for high discriminability stimuli, replicating the findings from the letter discrimination task. For the bars task, there was a 62 ms difference in the 0.1 quantiles for the highest and lowest noise conditions for correct responses and a 105 ms difference for errors. For the Gabor patch task, there was a 68 ms difference in the 0.1 quantiles for correct responses and a 79 ms difference for errors. The finding of a leading edge effect in these tasks is consistent with Ratcliff and Smith's (2010) claim that the onset of the decision process is delayed until a stable perceptual representation of the stimulus is formed, which is slowed by dynamic noise. The occurrence of a leading edge effect in these simpler feature processing tasks shows that it does not depend upon higher-order configural properties of the stimuli,

Table 3

Model parameters for discriminating bars and Gabor patches in noise.

Parameter	Symbol	Model			
		Integrated system model		Release from inhibition model	
		Bars	Gabors	Bars	Gabors
Stimulus encoding					
Onset rate	$\beta_{\rm on}$	18.69	33.03	-	-
Number of stages	n	3 ^a	3 ^a	-	-
Naka-Rushton inhibition	I _{in}	0.019	<0.001	0.041	< 0.001
Naka-Rushton exponent	ρ	2 ^a	3.14	2 ^a	2.89
Incomplete gamma dispersion	β	-	-	0.158	0.164
Incomplete gamma location	α_1	-	-	11.73	8.40
Incomplete gamma location	α_2	-	-	11.53	8.83
Incomplete gamma location	α_3	-	-	9.02	7.58
Incomplete gamma location	α_4	-	-	6.81	5.06
Attention/VSTM					
Attention gain	γg	20.89	12.77	_	-
Drift amplitude scaling	θ	0.625	0.634	0.335	0.426
Saliency constant	I ₀	0.449	0.627	-	-
Decision process					
Boundary separation	2a	0.143	0.155	0.106	0.117
Drift variability	η	0.157	0.140	0.011	0.076
Diffusion coefficient (square root)	σ_1	0.100 ^a	0.100 ^a	0.100 ^a	0.100 ^a
Stimulus independent diffusion	σ_2	0.019	0.010	-	-
Prerelease inhibition ^a	λ_0	-	-	15.0	15.0
Nondecision processes					
Mean nondecision time	T _{er}	0.208	0.267	0.227	0.271
Nondecision time range	s _t	0.184	0.116	0.158	0.147
3 1 4 6 1 4					

^a denotes a fixed parameter.

such as feature conjunctions. The finding of an effect with both bars and Gabor patches shows it is not a function of the broadband nature of the stimuli and, in particular, that it does not depend on information carried by high spatial frequencies, which encode information about perceptual edges. This is consistent with the occurrence of a leading edge effect in letter discrimination, which is mediated by visual filters with an intermediate tuning of around three cycles per letter and is relatively unaffected by the presence of high spatial frequency information (Solomon & Pelli, 1994).

To assess the generality of our modeling results we fitted the integrated system model and the release from inhibition model to the group RT distributions and choice probabilities for both tasks. The fits of the integrated system model are shown in Fig. 8. Estimated parameters for both models are shown in Table 3. As in the previous fits, we assumed that drift was a Naka–Rushton function of stimulus intensity. For the Gabor patch stimuli, we assumed that the effective stimulus intensity was proportional to the fraction of unoccluded pixels in the display and set the transduced stimulus intensity, *I*, in Eq. (C.3) to

$$I = \frac{[(1-\pi)c_P]^{\rho}}{[(1-\pi)c_P]^{\rho} + I_{\rm in}},$$
(24)

where $c_P = 0.20$ is the peak contrast of the Gabor patch and π is the fraction of noise pixels. We considered other, more complex, representations of stimulus transduction, based on the stimulus signal-to-noise ratio, which depends on the space-averaged RMS (root mean square) contrast of the stimulus, but found they performed more poorly than the simpler model of Eq. (24). We found we needed larger values of the Naka-Rushton exponent, ρ , to characterize drift than the value of $\rho = 2$ assumed for the letter discrimination task (Eq. (18)). This is consistent with some previous fits of the integrated system model to data obtained with Gabor patch stimuli (e.g., Sewell & Smith, 2012; Smith & Ratcliff, 2009). For the bars experiment, stimulus transduction was well described by Eq. (18), with an exponent of 2.

We found that the integrated system model and the release from inhibition model both gave good accounts of the data from the bars and the Gabor patch tasks and that they nicely captured the leading edge effects and the fast errors in both data sets. Like the letter discrimination task, the models were able to capture the fast errors without the assumption of trial-to-trial variation in starting point. For the bars task, the integrated system model yielded $G^2(34) = 6.60$ and the release from inhibition model yielded $G^2(33) = 6.75$. The fits of the release from inhibition model have fewer degrees of freedom because we treated the release times as free parameters (four in total), as we did for the letter discrimination task. For the Gabor patch task, the integrated system model yielded $G^2(32) = 5.06$. These latter fits have fewer degrees of freedom because the exponent in Eq. (24) was treated as a free parameter.

A comparison of the estimated parameters in Tables 1–3 shows there are differences in the magnitudes of the estimated parameters between experiments, particularly for the integrated system model rates, β_{on} , γ_g , and the asymptotic VSTM trace strength, θ , which characterizes the maximum amplitude of drift. The cascaded structure of the integrated system model means that in some data sets these parameters can be traded off to yield similar fits. This is particularly so in experimental designs with only a single treatment factor, like the bars and Gabor patches experiments. Estimation will typically be better in designs like the one in the letter discrimination task, in which a stimulus manipulation is crossed with another experimental factor, like speed versus accuracy instructions.

7. The overconstrained estimation view

Compared to the standard diffusion model of Eq. (2), the integrated system model and the release from inhibition model make relatively complex assumptions about the time course of processing within a trial. We were motivated to consider such complex models because the standard diffusion model is unable to account for data like those in Fig. 1. An alternative view was proposed by Donkin, Brown, and Heathcote (2009), who argued that poor fits like those reported by Ratcliff and Smith (2010) may be a byproduct of overconstraining the parameters of



Fig. 9. Fit of the unconstrained time-homogeneous diffusion model. Drift and diffusion coefficients were permitted to vary freely across stimulus conditions.

the model rather than any fundamental limitation of the model itself. In particular, Donkin et al. questioned the usual practice of constraining diffusion coefficients to be equal across stimulus conditions.

Because parameters of models like the diffusion model are only identified to the level of a ratio, at least one parameter needs to be fixed arbitrarily to provide a scale in which other parameters can be estimated. Typically, following the example of signal detection theory, the value of the diffusion coefficient (s^2 or σ^2) is fixed. Most applications of diffusion models make the further assumption that the diffusion coefficient is the same in every stimulus condition, but, as Donkin et al. (2009) argued, this is more than is needed to make the model identifiable. They advocated fixing the diffusion coefficient in one condition only and estimating the values in the other conditions from the data. Comparison of more and less restricted versions of the model can then provide a test of the empirical hypothesis that the diffusion coefficient is invariant across conditions.

As an example of this approach, Donkin et al. (2009) refitted one of the conditions from the Gould et al. (2007) study, in which a large leading edge effect was found, using a time-homogeneous diffusion model (Eqs. (1) and (2)) with between-trial variability

Table 4

Parameters of the unconstrained time-homogeneous diffusion model.

Parameter	Symbol	Value	
Stimulus encoding			
Mean drift rate	ν_1	0.027	
Mean drift rate	ν_2	0.101	
Mean drift rate	ν_3	0.231	
Mean drift rate	ν_4	0.385	
Mean drift rate	ν_5	0.635	
Diffusion coefficient (square root)	<i>s</i> ₁	0.100 ^a	
Diffusion coefficient (square root)	<i>s</i> ₂	0.105	
Diffusion coefficient (square root)	\$ ₃	0.111	
Diffusion coefficient (square root)	\$4	0.123	
Diffusion coefficient (square root)	\$ ₅	0.147	
Drift variability	η	0.077	
Decision process			
Boundary separation (speed)	$2a_s$	0.080	
Boundary separation (accuracy)	$2a_a$	0.130	
Starting point variability	Sz	0.008	
Nondecision processes			
Mean nondecision time	Ter	0.472	
Nondecision time range	s _t	0.008	
late Time based parameters are in units of seconds; state based parameters are in			

Note. Time-based parameters are in units of seconds; state-based parameters are in units of infinitesimal standard deviations per second.

^a denotes a fixed parameter.

in drift and starting point, and allowed the diffusion coefficient to vary freely in four of the five stimulus conditions. They found that by relaxing the constraint on the diffusion coefficient they could substantially improve the fit — although there were still some systematic discrepancies in the predicted 0.1 quantiles. They also considered an unconstrained version of the linear ballistic accumulator (Brown & Heathcote, 2008) and showed that the fit of this model was also significantly improved. Indeed, the unconstrained form of the ballistic accumulator model provided a better fit than did the unconstrained diffusion model.

In the light of Donkin et al.'s (2009) results, we asked whether the dynamic noise data could be explained by systematic variation in the diffusion coefficient across conditions. This hypothesis is plausible because, as noted previously, the diffusion coefficient sets the clock of the process. Processes with smaller diffusion coefficients accumulate evidence more slowly, dilating their time scale. This increases the spread of the RT distribution and shifts its leading edge. Unlike the integrated system model or the release from inhibition model, the variable diffusion coefficient model is not a process model; it is simply an alternative, less restrictive, set of constraints on the estimation procedure. One of the benefits of relaxing the assumed constraints on the parameters is that systematic patterns of variation in empirical parameter estimates may stimulate new theoretical developments. Accordingly, we refitted the dynamic noise data with the time-homogeneous diffusion model of Eqs. (1) and (2), with variation in drift and starting point, and allowed the diffusion coefficient to vary freely in four of the five discriminability conditions. We used the computational routines described by Tuerlinckx (2004) to obtain first passage time distributions and associated response probabilities for the model. The resulting model fit is shown in Fig. 9; the estimated parameters are shown in Table 4.

In agreement with the results of Donkin et al. (2009), allowing the diffusion coefficient to vary across conditions substantially improved model fit. The fit statistic for the unconstrained model was $G^2(96) = 30.37$, compared to $G^2(100) = 63.16$ for the constrained model, with all diffusion coefficients equal. The quantile probability plot of the data shows that the unconstrained model successfully captures important features of the data, particularly the tail quantiles for low discriminability stimuli in the accuracy condition. It slightly underpredicts the magnitude of the leading effect in the accuracy condition, but the approximation is reasonable, and is markedly better than the constrained model. However,

it fails to capture the leading edge effect in the speed condition, exhibiting systematic misses in the 0.1 quantiles for error distributions. It also underpredicts accuracy at the highest levels of stimulus discriminability in the speed condition. Quantitatively, the fit of unconstrained model is almost 2–2.5 times worse than either of our process models.

We also considered the fits of the constrained and unconstrained models to the bars and Gabor patch data in Fig. 8. For the unconstrained model, we allowed the diffusion coefficient to vary freely in the three highest discriminability condition. The picture that emerged was much the same as the one emerging from the letter discrimination data. For both tasks the unconstrained model performed appreciably better than the constrained model, but still failed to capture some aspects of the data. The fit statistics for the constrained model were $G^{2}(35) = 21.93$ and $G^{2}(35) = 15.01$ for the bars task and the Gabor patches task, respectively. The corresponding fit statistics for the unconstrained model were $G^{2}(32) =$ 9.29 and $G^2(35) = 5.95$. The differences among the fit statistics for the unconstrained model, the integrated system model, and the release from inhibition model are not as pronounced as they were for the letter discrimination task. This is unsurprising, as the design of the bars and Gabor patches tasks did not cross discriminability and instructions, and so provided fewer constraints on model fit. The fit to the bars data was around 40% worse than the best of the two process models; the fit to the Gabor patch data was comparably good. Like the fit of the model to the letter discrimination task, freeing the diffusion coefficient allowed the model to capture the upper quantiles of the RT distributions well, but it underpredicted the leading edge effect. This was true for both the bars task, in which the fit is poorer than either of the process models, and the Gabor patches task, in which the fit is comparably good.

Overall, we conclude that a time-homogeneous model with stimulus intensity dependent diffusion coefficients does not provide an adequate model for discrimination in dynamic noise. Not only do our process models provide a better quantitative characterization of the data, they also provide a theoretically principled account of the psychological processes that underlie the leading-edge effect. The unconstrained diffusion model, in contrast, produces improvements in fit simply by an increase in the number of free parameters. Potentially, intensity-dependent variation in the diffusion coefficient could be motivated theoretically by arguing that the underlying neural processes are Poisson or Poisson-like in nature, as suggested by Smith (2010) and Smith and McKenzie (2011). However, the estimated drift and diffusion coefficients in Table 4 do not follow a simple Poisson law with equal means and variances. The estimated drift coefficients in Table 4 change by a factor of more than 20 across conditions while the estimated diffusion coefficients (the squares of the s_i parameters) only change by factor of two. To be plausible, a neural interpretation of the stimulusdependency in the diffusion coefficient would need to explain why the estimated parameters in Table 4 do not scale in the expected way. It might be possible to do this by assuming that the underlying processes are mixed Poisson processes, composed by randomly sampling Poisson processes with different intensity parameters on different trials (Grandell, 1997), or by aggregating across neurons that are differentially responsive to the signal. Such processes are more variable than the simple Poisson process because of the additional variability contributed by the mixing process. Although this idea has some a priori plausibility, we can foresee difficulties in making it rigorous and convincing, so we have not attempted to develop it in any formal way.

8. Discussion

In this article, we considered two process models for decisionmaking in the dynamic noise task. Our larger theoretical concern in investigating this task was to try to understand the relationship between perceptual and decision processes in two-choice discrimination. Diffusion models have been extremely successful in accounting for performance in such tasks, but they do so by subsuming all of the processes prior to the decision process into a single value of drift, which is most often treated as a free parameter. Given the success of such models, our theoretical focus in this article was to try to characterize the processes involved in the computation of drift, and to characterize how drift processes and decision processes are coupled. The phenomenology of the dynamic noise task provides compelling grounds for assuming that noise affects the time course of perceptual processing. This makes the task an ideal test-bed for investigating the coupling of perceptual and decision processes.

Both of our models for the dynamic noise task are generalizations of the time-homogeneous Wiener diffusion model of Ratcliff (1978), which has been successfully applied to many experimental tasks (e.g. Ratcliff & Smith, 2004). Our focus has been on the shifts in the RT distributions produced by noise, and the associated pattern of fast errors in some conditions. Both of our models assume that noise delays evidence accumulation by a decision process until after a stable stimulus representation has formed. In characterizing performance in this way-as comprising distinct stimulus encoding and decision-making stages-we are implicitly rejecting single-stage models, which assume that decisions are made simply by summing successive, noisy stimulus states. Single-stage models have provided successful accounts of performance in expanded judgment tasks, in which people are required to make decisions about the statistical properties of a stream of discrete stimulus elements (Smith & Vickers, 1989), but they do not predict shifts in RT distributions with changes in stimulus discriminability, and so do not appear to be appropriate models for the dynamic noise task. The focus of our theoretical work has therefore been to try to characterize how the onset of evidence accumulation could be adaptively coupled to the stimulus encoding process.

The first of our models was based on the integrated system model of Smith and Ratcliff (2009). The key assumption of this model is that the rate of evidence accumulation depends on a time-dependent diffusion coefficient, whose temporal properties depend on attention gain and on the perceptual saliency of the stimulus. Reducing saliency, assumed to occur under conditions of high noise, slows the growth of the stimulus representation and is accompanied by a corresponding slowing of the growth of the diffusion coefficient. The second model was a release from inhibition model. Release from inhibition is an important mechanism for action selection by the basal ganglia, which have a central role in decision-making (Berns & Sejnowski, 1996), so the proposed mechanism is a physiologically plausible one. The key assumption of the release from inhibition model was that inhibition modulates the decay coefficient in an OU diffusion model. We considered two versions of this model: a progressive release model and an all-or-none model, in which inhibition is released when the emerging stimulus representation reaches a threshold value.

Both the integrated system model and the progressive release from inhibition model provided satisfactory accounts of the RT distributions and choice probabilities in the dynamic noise task. In particular, the models accurately characterized the leading edge effect in the RT distributions and the fast errors in the speed condition. The integrated system model requires the assumption of an additional source of stimulus-independent diffusive noise to predict fast errors; the release from inhibition model naturally predicts fasts errors because of the dynamics of the decision process. The integrated system model is somewhat more parsimonious in the number of free parameters it requires, but the parsimony of the release from inhibition model could be improved by augmenting it with a process model of the release times in Fig. 6c. We conclude that both models provide satisfactory accounts of the dynamic noise data. Although we have not been able to distinguish between these models empirically, we see the value of our work is that it illuminates the relationship between the evidence accumulation process of decision-making and the early sensory and memory processes involved in the computation of the evidence. We have shown, in particular, that the integrated system model, which was developed to account for performance in a quite different domain, also successfully predicts performance in the dynamic noise task. In addition, we have shown that release from inhibition, which is a known physiological process, leads to a wellbehaved decision model that successfully predicts performance at the detailed level of choice probabilities and RT distributions.

One question on which our work is silent is why a leading edge effect is found only in some dynamic noise tasks. Specifically, the effect is found only in tasks in which features such as letters, bars, or Gabor patches are presented in dynamic noise. It is increased by changing the task to a letter versus digit discrimination that maximizes featural uncertainty (Ratcliff & Smith, 2010), but it is not found in a brightness discrimination task in which there are no stimulus features to be encoded. Our hypothesis is that the stimulus attributes used in these different judgments may be carried by different cortical systems, with different temporal integration properties. Specifically, the feature information used in the letter discrimination task may be carried by a perceptually sustained system, with a long temporal integration time, whereas the information used in the brightness discrimination task may be carried by a perceptually transient system, with a short temporal integration time (Smith, 1995, 1998). According to this hypothesis, dynamic noise produces no shift in the leading edge of the RT distribution in brightness discrimination because the perceptual system that encodes brightness is only able to integrate information over a short time scale, so no improvement in the fidelity of the stimulus representation is obtained by delaying the decision process for longer.

Our general conclusion is that models like the ones we have introduced here are likely to be required for any task in which the process of formation a perceptual representation of the stimulus is extended in time and varies as a function of stimulus discriminability. The standard model of Eqs. (1) and (2) assumes that the stimulus information encoded in the drift becomes available all at once after a variable time delay. This is clearly an idealization, but it is one that comes reasonably close to reality whenever perception (as distinct from stimulus identification) is rapid. Most of the tasks to which diffusion models have been successfully applied have used suprathreshold stimuli, which are perceived rapidly, but which take varying amounts of time to identify or to discriminate. For such tasks, the additive decomposition of decision and nondecision processes, combined with a constant drift diffusion model, provides a good approximation to the dynamics of performance. Two situations that we are aware of in which this model fails are the external noise task, investigated here, and the task investigated by Gould et al. (2007), in which very low contrast grating patches were presented directly against a uniform field. Conceivably, under conditions of very low contrast, noise in the visual system may act in a similar way to external noise in the dynamic noise task, to delay the formation of a perceptual representation of the stimulus and to produce a leading-edge effect.

9. Coda

As mathematical psychologists working today we often tend to take for granted the idea that psychological processes can be characterized mathematically. We are comfortable that our program of developing psychological explanations for behavior, expressing them in mathematical form, and then testing the resulting model against empirical data—is a meaningful one. It is therefore easy to forget that when William Estes and others created mathematical psychology in the 1950s it was by no means obvious that any of this was the case. The theory and practice of experimental psychology in the 1950s had a distinctly anti-quantitative cast, and Estes' program was very far from behaviorist orthodoxy. The success of his program was, first and foremost, a profound demonstration of the value of a mathematical theory of behavior. It was a demonstration that, as in other sciences, psychological phenomena are amenable to, and can greatly benefit from, mathematical analysis. His legacy to subsequent generations is no less than this and, accordingly, it is difficult to overstate its importance.

Acknowledgments

The research in this article was supported by Australian Research Council Discovery Grant DP 110103406 to Philip Smith and Air Force Office of Scientific Research Grant FA9550-11-1-0130 to Roger Ratcliff.

Appendix A. Experimental methods

A.1. Letter discrimination in dynamic noise

A.1.1. Stimuli

Participants were presented with one of two noise-degraded capital letters and made a choice between the two. The letters were 0.85° high and 0.6°-1.1° wide, presented in white on a black 64×64 pixel background subtending $3.0^{\circ} \times 3.0^{\circ}$ of visual angle. The letters were degraded by randomly inverting the contrast polarity of a proportion of pixels in the letter and the background in each 16.7 ms frame of the display. Letters were presented at the center of the random array and remained present until the response. Participants responded with the/key for the right-hand letter choice and with the Z key for the left-hand letter choice. The same two letters were used for a block of trials; these were then replaced with another letter pair for the next block. Each pair of letters was reused in a later block, but with the mapping of the letters to the left and right hands reversed. The letter pairs were FQ, PL, WK, TX, GR, and BN. To indicate the mapping between letters and response keys, two letters corresponding to the two choices were displayed on the left and right sides of the screen (left letter for the Z and right letter for the/key) and remained on the screen throughout the block of trials.

The experiments were run by a real-time Linux system on Pentium 4 class computers. Stimuli were presented on Dell Ultrascan 0780 CRT monitors with 17 in. (43.18 cm) viewing areas. Participants viewed the stimuli at a distance of 57 cm; at this distance, 1 cm on the screen subtended 1° of visual angle.

A.1.2. Participants

Twenty undergraduate students from the Ohio State University each participated in two experimental sessions. In each session, they performed 12 blocks, each of 100 experimental trials, with speed and accuracy instructions alternating in consecutive blocks.

A.1.3. Procedure

Each trial began with a fixation point in the center of the screen, displayed for 500 ms, then the target letter was displayed in dynamic noise. The onset of the noise and the onset of the stimulus coincided. Participants were instructed to press the / key on the keyboard if the right-hand letter was presented and the Z key if the left-hand letter was presented. There were five levels of discriminability produced by inverting 0.35, 0.40, 0.425, 0.45, 0.475 of the pixels in each consecutive frame of the display. In

speed blocks, participants were instructed to respond as quickly as possible; in accuracy blocks they were instructed to make as few errors as possible. In speed blocks, responses longer than 800 ms were followed by a "TOO SLOW" message; in accuracy blocks, an "ERROR" message was given for error responses. After slow responses in the speed condition and error responses in the accuracy condition, a 300 ms time penalty was given before proceeding to the next trial.

A.1.4. Data screening

Responses faster than 300 ms and slower than 3000 ms were excluded from the analysis. There were 2026 such responses, or 4% of trials. Correct responses to left-hand letter stimuli were pooled with correct responses to right-hand stimuli for each level of discriminability. Error responses were combined in a similar way. Quantile RTs were computed from the pooled RTs for correct responses and errors and averaged over participants to obtain the data in Fig. 1.

A.2. Bars in dynamic noise

Participants were presented with three, parallel, vertical or horizontal white bars on a black background and discriminated their orientations via a keypress response. The bars were 40 pixels long and 4 pixels wide and had an inter-center spacing of 14 pixels. Viewed from 57 cm, the length, width, and spacing of the bars subtended 2.0°, 0.2° wide and 0.7° of visual angle, respectively. The stimuli were degraded by randomly inverting the contrast polarity of some proportion of the pixels in the bars and background in each 16.7 ms frame of the display. On half of the trials, the proportion of inverted pixels was less than 50%, producing the appearance of light bars on a dark background. On the other half of the trials the proportion of inverted pixels was greater than 50%, producing the appearance of dark bars on a light background. On light bar trials, the proportion of inverted pixels was 0.35, 0.43, 0.46, or 0.475; on dark bar trials, it was 0.65, 0.57, 0.54, or 0.525. Because stimuli in which the proportions of inverted pixels are π and $1-\pi$ have the same contrast, this resulted in four discriminability conditions in which stimuli had the same contrast magnitude but opposite contrast polarity. We pooled across contrast polarity and bar orientation in the data analysis. All other details of apparatus, participants, procedure, and data screening were the same as in the letter discrimination task, except that participants completed only one experimental session and trials were not blocked according to speed and accuracy instructions.

A.3. Gabor patches in dynamic noise

Participants were presented with circularly-symmetrical, sinephase, Gabor patches with horizontal or vertical sinusoidal carriers and discriminated the carrier orientation via a keypress response. The mathematical form of the stimuli was as given by (Graham, 1989, p. 53). The sinusoidal carrier had a period of 8 pixels; the Gaussian envelope had a space constant (full width at half height) of 20 pixels. Viewed from 57 cm, the spatial frequency of the sinusoid was 2.5 cycles/deg and the width of the Gaussian envelope was 1.0°. The peak contrast of the stimuli was 0.20. The stimuli were degraded by replacing 0.35, 0.50, 0.65, or 0.80 of the pixels in the display with uniformly-distributed grayscale noise pixels. The noise had a mean equal to the mean luminance of the display and a range of \pm 50%. The location and the contrast of the noise pixels was chosen randomly in each consecutive 16.7 ms frame of the display. We pooled across orientations within discriminability levels in the data analysis. All other details of apparatus, participants, procedure, and data screening were the same as for the bars task.

Appendix B. Transformation of a diffusion to a standard Brownian motion process

Theorem (*Ricciardi & Sato*, 1983; see also Smith, 2000, p. 442). Let X(t) be a diffusion process with drift $\mu(x, t)$ and diffusion coefficient $\sigma^2(x, t)$ satisfying the stochastic differential equation

$$dX(t) = \mu(x, t) dt + \sigma(x, t) dW(t).$$
(B.1)

Let $\sigma_x^{2'}(x,t) = (\partial/\partial x)\sigma^2(x,t)$ and $\sigma_t^{2'}(x,t) = (\partial/\partial t)\sigma^2(x,t)$ be the first partial derivatives of the diffusion coefficient with respect to its state and time coordinates, respectively. If there exists a pair of functions, $c_1(t)$ and $c_2(t)$, such that

$$\mu(x,t) = \frac{\sigma_x^{2'}(x,t)}{4} + \frac{\sigma(x,t)}{2} \left\{ c_1(t) + \int^x \frac{c_2(t)\sigma^2(y,t) + \sigma_t^{2'}(y,t)}{\sigma^3(y,t)} dy \right\}, \quad (B.2)$$

then there exists a coordinate transformation, $X(t) \rightarrow X^*(t^*)$ of the form

$$x^* = \bar{\Psi}(x, t) \tag{B.3}$$

$$t^* = \Phi(t), \tag{B.4}$$

such that $X^*(t^*) = B(t^*)$ is a standard Brownian motion. If it exists, this transformation is of the form

$$x^* = \bar{\Psi}(x,t) = \exp\left[-\frac{1}{2}\int^t c_2(s)ds\right]\int^x \frac{dy}{\sigma(y,t)} -\frac{1}{2}\int^t c_1(s)\exp\left[-\frac{1}{2}\int^s c_2(z)dz\right]ds$$
(B.5)

$$t^* = \Phi(t) = \int_{t_0}^t \exp\left[-\int^s c_2(z)dz\right]ds,$$
 (B.6)

where t_0 is the starting time of the process.

For the special case in which the drift may depend on time and state, and the diffusion coefficient may depend on time, but is independent of state, the condition in Eq. (B.2) takes the simpler form

$$\mu(x,t) = \frac{\sigma(t)}{2}c_1(t) + \frac{x}{2}\left[c_2(t) + \frac{\sigma^{2'}(t)}{\sigma^2(t)}\right].$$
(B.7)

B.1. The integrated system model

For the integrated system model of Eq. (3), the drift is v(t) and the infinitesimal standard deviation is $\sigma \sqrt{v(t)}$. The unknown functions $c_1(t)$ and $c_2(t)$ in the coordinate transformation therefore satisfy the relationship

$$\nu(t) = \frac{\sigma\sqrt{\nu(t)}}{2}c_1(t) + \frac{x}{2}\left[c_2(t) + \frac{\nu'(t)}{\nu(t)}\right].$$

Equating coefficients on the left and right yields

$$c_1(t) = \frac{2\sqrt{\nu(t)}}{\sigma}$$
$$c_2(t) = -\frac{\nu'(t)}{\nu(t)}.$$

Evaluating the exponentiated integral terms in Eqs. (B.5) and (B.6) yields

$$\exp\left[-\frac{1}{2}\int^t c_2(s)\ ds\right] = \sqrt{\nu(t)}$$

and

$$\exp\left[-\int^t c_2(s) \, ds\right] = v(t).$$

We substitute these values in Eqs. (B.5) and (B.6) to obtain

$$x^* = \frac{x}{\sigma} - \frac{1}{\sigma} \int^t \nu(s) \, ds \tag{B.8}$$

$$t^* = \int_{t_0}^t v(s) \, ds. \tag{B.9}$$

This is the transformation in Eqs. (14) and (15) and shown in Fig. 2, for initial condition $t_0 = 0$.

B.2. The release from inhibition model

The drift and infinitesimal standard deviation of the release from inhibition model are $\mu(x, t) \equiv \nu(t) - \lambda(t)x$ and $\sigma(x, t) \equiv \sigma$, respectively. The unknown functions in the transformation to the Wiener process therefore satisfy Eq. (B.7) in the form

$$\nu(t) - \lambda(t)x = \frac{\sigma}{2}c_1(t) + \frac{c_2(t)x}{2},$$

because the term $\sigma^{2'}(t)$ vanishes. Equating coefficients yields

$$c_1(t) = \frac{2\nu(t)}{\sigma}$$
$$c_2(t) = -2\lambda(t).$$

Evaluating the exponentiated integral terms in Eqs. (B.5) and (B.6) yields

$$\exp\left[-\frac{1}{2}\int^{t}c_{2}(s)\ ds\right] = \exp\left[\int^{t}\lambda(s)\ ds\right]$$

and

$$\exp\left[-\int^t c_2(s) \ ds\right] = \exp\left[2\int^t \lambda(s) \ ds\right].$$

We substitute these values into Eqs. (B.5) and (B.6) to obtain the transformation

$$x^{*} = \frac{1}{\sigma} \left\{ x \exp\left[\int^{t} \lambda(s) \, ds\right] - \int^{t} \nu(s) \exp\left[\int^{s} \lambda(z) \, dz\right] \right\} (B.10)$$
$$t^{*} = \int_{t_{0}}^{t} \exp\left[2\int^{s} \lambda(z) \, dz\right] \, ds. \tag{B.11}$$

This is the transformation in Eqs. (16) and (17) and shown in Fig. 3.

Appendix C. The integrated system model

The integrated system model is defined by a set of equations that characterize the time course of each of its subprocesses. These equations describe the amplitude and duration of the sensory response function, $\mu(t)$, the growth of the VSTM trace, $\nu(t)$, the attention gain function, $\gamma(t)$, and the accumulation of evidence by the decision process, X(t).

The sensory response to a stimulus of duration *d* is of the form

$$\mu(t) = \Gamma(t; \beta_{\text{on}}, n) [1 - \Gamma(t - d; \beta_{\text{off}}, n)], \qquad (C.1)$$

where $\Gamma(t; \beta, n)$ is the output of a linear filter composed of *n* identical exponential stages,

$$\Gamma(t;\beta,n) = 1 - e^{-\beta t} \sum_{j=0}^{n-1} \frac{(\beta t)^j}{j!}.$$
(C.2)

The quantities β_{on} and β_{off} are time constants that determine the onset (rise) and offset (decay) time of the filter response. For response-terminated stimuli like those used by Ratcliff and Smith (2010), there is no stimulus offset term and $\mu(t)$ reduces to Eq. (C.2) with $\beta = \beta_{on}$.

The amplitude of the sensory response is a function of the contrast of the stimulus, which we write for convenience as a Michelson contrast, Δ_I / Σ_I , where Δ_I is the luminance difference between the stimulus and the background and Σ_I is the luminance sum. In tasks with near-threshold stimuli, the amplitude of the contrast response, denoted *I*, is a Naka–Rushton function of the form

$$I = \frac{(\Delta_{I}/\Sigma_{I})^{\rho}}{(\Delta_{I}/\Sigma_{I})^{\rho} + I_{0.5}^{\rho}} = \frac{\Delta_{I}^{\rho}}{\Delta_{I}^{\rho} + I_{\rm in}}.$$
 (C.3)

In this equation, the exponent ρ characterizes the nonlinearity of contrast transduction in the early visual system and $I_{0.5}$ is a socalled semi-saturation constant that specifies the value of contrast at which the function attains 50% of its maximum value of 1.0. In the expression on the right, the sum Σ_l has been absorbed into a general divisive inhibition term, I_{in} , by writing $I_{in} = (I_{0.5} \Sigma_l)^{\rho}$. As described in the text, we characterize the stimulus information using a value of contrast equal to $1 - 2\pi$, where π is the proportion of inverted pixels in the letter and the background. We found that our data were well fitted by a model with $\rho = 2$ (contrast energy scaling). This is consistent with the results of physiological studies of visual contrast sensitivity (Boynton, 2005).

VSTM trace formation is described by an excitatory-inhibitory shunting equation

$$\frac{d\nu}{dt} = \gamma \{ I\mu(t)[\theta - \nu(t)] - (1 - I)\mu(t)\nu(t) \}.$$
 (C.4)

The excitatory and inhibitory coefficients in this equation, I and 1 - I respectively, represent a form of balanced center–surround interaction.

The constant θ determines the scaling of VSTM trace strength; the function γ , is a rate function that determines the rate at which stimulus information is transferred to VSTM. As described in the text, the rate is assumed to depend jointly on attention gain and stimulus saliency. The solution to Eq. (C.4) is

$$\nu(t) = \theta I \left\{ 1 - \exp\left[-\gamma \int_0^t \mu(s) \, ds\right] \right\}.$$
(C.5)

This equation states that the VSTM trace grows to an asymptote that is proportional to the (transduced) stimulus intensity, *I*, where the constant of proportionality is the VSTM trace scaling parameter, θ . The drift of the diffusion process, *X*(*t*), in Eq. (3) is assumed to be equal to the VSTM trace strength, v(t).

References

- Abramowitz, M., & Stegun, I. (1964). *Handbook of mathematical functions*. New York: Dover.
 Berns, G. S., & Sejnowski, T. J. (1996). How the basal ganglia make decisions. In A.
- Berns, G. S., & Sejnowski, T. J. (1996). How the basal ganglia make decisions. In A. R. Damasio, H. Damasio, & Y. Christen (Eds.), *Neurobiology of decision-making* (pp. 101–113). New York: Springer.
- Bhattacharya, R. B., & Waymire, E. C. (1990). Stochastic processes with applications. New York, NY: Wiley.
- Bjork, E. L., & Estes, W. K. (1973). Letter identification in relation to linguistic context and masking conditions. *Memory & Cognition*, 1, 217–223.
- Boynton, G. (2005). Attention and visual perception. Current Opinion in Neurobiology, 15, 465–469.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178.
- Buonocore, A., Giorno, V., Nobile, A. G., & Ricciardi, L. (1990). On the twoboundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability*, 27, 102–114.
- Buonocore, A., Nobile, A. G., & Ricciardi, L. (1987). A new integral equation for the evaluation of first-passage-time probabilities densities. *Advances in Applied Probability*, 19(6), 784–800.

- Busemeyer, J., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. Mathematical Social Sciences, 23, 255-282.
- Busemeyer, J., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, 100, 432-459.
- Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. Psychological Review, 101, 446-469.
- Cherkasov, I. D. (1957). On the transformation of the diffusion process to a Wiener process. Theory of Probability and its Applications, 2, 373-377.
- Cox, D. R., & Miller, H. D. (1965). The theory of stochastic processes. London, UK: Chapman & Hall.
- De Valois, R. L., & De Valois, K. K. (1990). Spatial vision. New York: Oxford.
- Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. Journal of Mathematical Psychology, 39, 197-215.
- Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, 47, 304–322.
- Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: rethinking the scaling problem. Psychonomic Bulletin & Review, 16 1129-1135
- Estes, W. K. (1950). Toward a statistical theory of learning. Psychological Review, 57, 94-107.
- Estes, W. K. (1955a). Statistical theory of spontaneous recovery and regression. Psychological Review, 62, 145-154.
- Estes, W. K. (1955b). Statistical theory of distributional phenomena in learning, Psychological Review, 62, 369-377
- Estes, W. K. (1972). Interactions of signal and background variables in visual processing. Perception & Psychophysics, 12, 278-286.
- Estes, W. K. (1975). The locus of inferential and perceptual processes in letter identification. Journal of Experimental Psychology: General, 2, 122-145.
- Estes, W. K. (1982). Similarity-related channel interactions in visual processing. Journal of Experimental Psychology: Human Perception and Performance, 8, , 353–382.
- Estes, W. K., & Burke, C. J. (1953). A theory of stimulus variability in learning. Psychological Review, 60, 276-286.
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. Journal of the Optical Society of America, 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
- Gould, I. C., Wolfgang, B. J., & Smith, P. L. (2007). Spatial uncertainty explains exogenous and endogenous attentional cuing effects in visual signal detection. Journal of Vision, 7(13), 1–17. 4.
- Graham, N. V. S. (1989). Visual pattern analyzers. New York, NY: Wiley
- Grandell, J. (1997). Mixed Poisson processes. London, UK: Chapman and Hall.
- Grossberg, S. (1980). How does a brain build a cognitive code? Psychological Review, 87. 1-51.
- Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. Neural Networks, 1, 17-61.
- Gutiérrez Jáimez, R., Román Román, P., & Torres Ruiz, F. (1995). A note on the Volterra integral equation for the first-passage-time probability. Journal of Applied Probability, 32, 635-648.
- Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In M. S. Landy, & J. A. Movshon (Eds.), Computational models of visual processing (pp. 119-134). Cambridge, MA: MIT Press.
- Kaplan, E., Lee, B. B., & Shapley, R. M. (1990). New views of primate retinal function. In N. N. Osborne, & G. J. Chader (Eds.), Progress in retinal research, vol. 9 (pp. 273–336). Oxford, England: Pergamon Press.
- Karlin, S., & Taylor, H. M. (1981). A second course in stochastic processes. Orlando, FL: Academic Press. Laming, D. R. J. (1968). Information theory of choice reaction time. New York: Wiley.

Loftus, G. R., & Ruthruff, E. (1994). A theory of visual information acquisition and

visual memory with special application to intensity-duration trade-offs. Journal of Experimental Psychology: Human Perception and Performance, 20, 33-49.

- Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization. New York: Oxford University Press.
- McClelland, J. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. Psychological Review, 86, 287-330.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. Computer Journal, 7, 308-313.
- Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85, 59-108.
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: the temporal relationship between perceptual encoding and decision making. Journal of Experimental Psychology: General, 139, 70-94.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential-sampling models for two choice reaction time. Psychological Review, 111, 333-367.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating 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.
- Reynolds, J. H., & Heeger, J. H. (2009). The normalization model of attention. Neuron, 61, 168-185.
- Ricciardi, L. (1976). On the transformation of diffusion processes into the Wiener process. Journal of Mathematical Analysis and Applications, 54, 185-199.
- Ricciardi, L., & Sato, S. (1983). A note on the evaluation of first-passage-time probability densities. Journal of Applied Probability, 20, 197–201.
- Sewell, D. K., & Smith, P. L. (2012). Attentional control in visual signal detection: effects of abrupt-onset and no-onset stimuli. Journal of Experimental Psychology: Human Perception and Performance, 38, 1043–1068.
- Smith, P. L. (2010). From Poisson shot noise to the integrated Ornstein–Uhlenbeck process: neurally-principled models of diffusive evidence accumulation in decision-making and response time. Journal of Mathematical Psychology, 54, 266-283
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. Psychological Review, 102, 567-591.
- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: a foundational primer. Journal of Mathematical Psychology, 44, 408-463.
- Smith, P. L. (1998). Bloch's law predictions from diffusion process models of detection. Australian Journal of Psychology, 50, 139-147. Special issue on Mathematical Psychology in Australia.
- Smith, P. L., Ellis, R., Sewell, D. K., & Wolfgang, B. J. (2010). Cued detection with compound integration-interruption masks reveals multiple attentional mechanisms. Journal of Vision, 10, 1-28.
- Smith, P. L., & McKenzie, C. R. L. (2011). Diffusive information accumulation by minimal recurrent neural models of decision-making. Neural Computation, 23, 2000-2031.
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. Psychological Review, 116, 283-317.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. Vision Research, 44, 1297–1320.
- Smith, P. L., & Vickers, D. (1989). Modeling evidence accumulation with partial loss in expanded judgment. Journal of Experimental Psychology: Human Perception and Performance, 15, 797-815.
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediated letter identification. Nature, 369, 395-397.
- Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. Psychological Review, 102, 503-532.
- Sternberg, S. (1969). The discovery of processing stages: extension of Donders' method. In W. G. Koster (Ed.), Acta psychologica: vol. 30. Attention and performance II (pp. 276-315).
- Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and density functions in the diffusion model. Behavior Research Methods, Instruments, & Computers, 36, 702-716.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychological Review, 108, 550-592.
- 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.