Modeling Evidence Accumulation Decision Processes Using Integral Equations: Urgency-gating and Collapsing Boundaries

Philip L. Smith,
The University of Melbourne

Roger Ratcliff
The Ohio State University

Abstract

Diffusion models of evidence accumulation have successfully accounted for the distributions of response times and choice probabilities from many experimental tasks, but recently their assumption that evidence is accumulated at a constant rate to constant decision boundaries has been challenged. One model assumes that decision makers seek to optimize their performance by using decision boundaries that collapse over time. Another model assumes that evidence does not accumulate and is represented by a stationary distribution that is gated by an urgency signal to make a response. We present explicit, integral-equation expressions for the first-passage time distributions of the urgency-gating and collapsing-bounds models models and use them to identify conditions under which the models are equivalent. We combine these expressions with a dynamic model of stimulus encoding that allows the effects of perceptual and decisional integration to be distinguished. We compare the resulting models to the standard diffusion model with variability in drift rates on data from three experimental paradigms in which stimulus information was either constant or changed over time. The standard diffusion model was the best model for tasks with constant stimulus information; the models with time-varying urgency or decision bounds performed similarly to the standard diffusion model on tasks with changing stimulus information. We found little support for the claim that evidence does not accumulate and attribute the good performance of the time-varying models on changing-stimulus tasks to their increased flexibility and not to their ability to account for systematic experimental effects.

Keywords

evidence accumulation; diffusion process; collapsing boundaries; urgency-gating

During the last few decades, evidence accumulation models like the diffusion model (Ratcliff, 1978) have successfully accounted for the speed and accuracy of two-choice decisions in a variety of experimental tasks, both in the laboratory and in applied and clinical settings (Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff, Smith, & McKoon, 2015). The diffusion model assumes that noisy evidence from the stimulus is accumulated until one of two response boundaries, or decision criteria, is reached. The first boundary reached

Address for correspondence: Philip L. Smith, Melbourne School of Psychological Sciences, The University of Melbourne, Victoria, 3010, Australia, philipls@unimelb.edu.au, Telephone: (61) (3) 8344 6343.

determines the decision outcome and the time taken to reach it determines the decision time. The evidence is assumed to be noisy either because of noise in the stimulus itself or because of moment-to-moment noise in the process of matching the perceptual representation of the stimulus to the cognitive representations of the decision alternatives. Mathematically, evidence accumulation is modeled as a Wiener or Brownian motion diffusion process, which represents a process of continuously-distributed evidence accumulating continuously in time. The model predicts the shapes of response time (RT) distributions for correct responses and errors and the associated choice probabilities (response accuracy) (Ratcliff & McKoon, 2008) as a function of the evidence in the stimulus and the decision maker's speed-accuracy tradeoff and bias settings.

Much of the success of the diffusion model is attributable to the fact that it predicts RT distributions that closely resemble those found empirically. In many perceptual and cognitive tasks the empirical distributions of RT have a characteristic unimodal, positively-skewed form that remains largely invariant, except for a change in scale, as either the difficulty of the task or the speed versus accuracy instructions are changed. If the distributions from several experimental conditions are summarized in a quantile-quantile (*Q-Q*) plot, then the resulting plot takes the form of a family or straight, or near-straight, lines (Ratcliff & McKoon, 2008; Ratcliff & Smith, 2010; Smith & Corbett, 2019). To construct such a plot, the distributions are summarized using quantiles (the values of RT that cut off specified proportions of the probability mass) and the quantiles of each member of the family are plotted against those of one member that serves as a reference distribution. Linear or near-linear *Q-Q* families are predicted by diffusion models (Ratcliff & McKoon, 2008; Figure 8; Ratcliff, 2018, Figure 14; Smith, 2016, Figure 10) and, moreover, this is all they predict. Ratcliff (2002) showed that the diffusion model was unable to predict families of plausible-looking distributions whose locations, scales, or shapes changed in ways not found in data. The success of diffusion models, then, depends on their being able to predict the forms of RT distributions that are found empirically and only to predict those forms.

There is, however, an important exception to this characterization that has provoked ongoing debate in psychology and neuroscience. This involves the RT distributions that are found under speed-stress conditions when speed is manipulated via deadlines. The issue was first raised by Ratcliff and Smith (2004), who pointed out, in relation to a deadline study reported by Van Zandt, Colonius, and Proctor (2000), that deadlines result in more symmetrical RT distributions than are found in tasks that manipulate speed-stress via instructions, and that such distributions are not well described by the diffusion model in its standard form (Murphy, Boonstra, & Nieuwenhuis, 2016; Ratcliff & Rouder, 2000). Symmetrical RT distributions have sometimes been found in experiments with awake behaving monkeys performing eye-movement decision tasks (Roitman & Shadlen, 2002), which are again not well predicted by the standard diffusion model (Ditterich, 2006a, 2006b). The animals in these experiments are highly-practiced, water-deprived, and work for juice rewards and, in the case of the Roitman and Shadlen study, had previously been trained on a deadline task. A plausible interpretation of these data is that the animals are implicitly deadlining in order to minimize the time until the next reward. This leads to the question of how they regulate their performance under these conditions.

Two proposals have emerged to explain performance in these tasks. One is that decision makers use decision boundaries that decrease or "collapse" during the course of a trial, leading them to make decisions on the basis of progressively less evidence with the passage of time. Moreover, it is optimal for them to do so (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012) because it allows them to maximize their rate of reward (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). The second is that the accumulating evidence is modified by a time-dependent "urgency signal," which makes responses increasingly more likely with the passage of time, irrespective of the accumulated evidence. These ideas, which are often treated as equivalent and interchangeable (Murphy et al., 2016; Trueblood, Heathcote, Evans, & Holmes, 2021), have become influential in the recent neuroscience literature on decision making and, increasingly, in the psychological literature as well (Evans, Hawkins, & Brown, 2020; Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015; Palestro, Weichart, Sederberg, & Turner, 2018; Voskuilen, Ratcliff, & Smith, 2016; Winkel, Keuken, van Maanen, Wagenmakers, & Forstmann, 2014).

An unresolved theoretical problem is that the equivalence of collapsing-bounds and urgency-signal models has been claimed on the basis of heuristic arguments rather than rigorous analysis. A rigorous demonstration of equivalence requires an explicit mathematical characterization of the joint first-passage time distributions of the diffusion process through the decision boundaries. These distributions provide the predicted RT distributions and choice probabilities for a model. Two models are equivalent if and only if they predict the same joint first-passage time distributions. One of the aims of this article is to provide an explicit representation of this kind, using integral equation representations of the first-passage time distributions for collapsing bound and urgency-gating models (Smith, 2000; Smith & Lilburn, 2020; Voskuilen et al., 2016). This representation allows us to provide a precise characterization of the conditions under which urgency-signal and collapsing bounds models are and are not equivalent.

Our second aim is to provide an explicit mathematical characterization and empirical evaluation of a more radical proposal that has recently come out of neuroscience, which is that evidence does not accumulate at all (Carland, Marcos, Thura, & Cisek, 2016; Carland, Thura, & Cisek, 2015; Cisek, Puskas, El-Murr, 2009; Thura, Beauregard-Racine, Fradet, & Cisek, 2012). The proposal is that, after an initial transient onset period, the evidence is represented cognitively by a statistically stationary process. This process is modulated by a time-dependent urgency signal and a response is made when the urgency modulated, or "gated," evidence crosses a decision boundary. The original proposal (Cisek et al., 2009; Thura et al., 2012) used a heuristic argument based on ordinary rather than stochastic differential equations that was shown to be mathematically incorrect (Hawkins, Wagenmakers, Ratcliff, & Brown, 2015) and led to a model whose predictions could be reliably distinguished from those of the standard diffusion model in data (Evans, Hawkins, & Brown, 2020; Hawkins et al., 2015), contrary to the earlier claims. Subsequently, however, the model was reframed in a more general form with the more nuanced claim that it and the standard diffusion model are practically indistinguishable under conditions in which stimulus information does not change during the course of a trial. Again, this argument can only be properly evaluated by formulating the model as a rigorous stochastic

model and deriving an explicit expression for its first-passage time distributions and then comparing them to those of the standard diffusion model.

In this article, we provide such a characterization and compare the standard diffusion model and a rigorously-formulated version of the urgency-gating model in five sets of empirical data. Two of them were from standard decision tasks in which the stimulus information does not change during the course of a trial. The remaining data were from the paradigm of Trueblood et al. (2021), who, like us, evaluated the urgency-gating model in its general form. One of the sets of data is from their Experiment 1, which we reanalyzed; the other two are from replications of it with modifications, as we describe below. Their paradigm is of interest because on some trials the stimulus information changes during the course of a trial. Thura (2016) claimed that the standard diffusion model and the urgency-gating model can only be distinguished in data from tasks of this kind. These tasks raise fundamental questions about processes of perceptual and decisional integration and how they should be modeled and whether they can be distinguished empirically (Smith & Lilburn, 2020), which we return to subsequently.

Urgency, Collapsing Bounds, and Optimality

The idea that decision makers under time pressure regulate their performance using time-dependent decision boundaries became popular for both empirical and theoretical reasons. Empirically, the widely-cited monkey data of Roitman and Shadlen (2002) cannot be well fit by the standard diffusion model but are well fit by a model with collapsing decision boundaries (Ditterich, 2006a, 2006b). Theoretically, the idea of collapsing bounds appeared to align well with a concept of optimality proposed by Bogacz et al. (2006), who argued that decision makers seek to maximize their rate of reward over a sequence of experimental trials. They showed that there is an optimal (fixed) boundary separation for the diffusion model that maximizes reward rate as a function of the value of correct responses, the cost of errors, the cost of sampling evidence, and the intertrial delay. The idea that there is a cost to sampling evidence was further developed by Drugowitsch et al. (2012) using the framework of stochastic dynamic programming (Ross, 1983) to derive an optimal decision "policy," which prescribes whether a decision maker should sample more evidence or make a decision immediately at each moment during a trial. In at least some situations, the optimal policy is one in which the decision boundaries decrease (converge) during the course of a trial rather than remain constant. Many researchers have found the idea of collapsing bounds attractive because it is grounded in a normative theory that prescribes how an optimal decision maker should behave.

The idea that decision makers seek to maximize their reward rate and the idea that they do so by using collapsing boundaries were investigated by subsequent researchers with mixed results. Some decision makers approach reward-rate optimality under some circumstances, but uninstructed decision makers often tend to be more conservative in their criterion settings than the theory predicts, and emphasize accuracy over reward rate (Evans, Bennett, & Brown, 2019; Holmes & Cohen, 2014; Starns & Ratcliff, 2010, 2012). Formal comparisons of fixed and collapsing bounds models have found support for collapsing bounds models only under some conditions. Hawkins et al. (2015) derived predictions for

collapsing-bounds models using Monte Carlo simulation methods and compared them to fixed-bounds models on nine different data sets from human and animal participants and found only limited support for collapsing bounds models. Most of the support for collapsing bounds came from experiments using animal participants, consistent with the idea that highly-trained animals working for juice rewards regulate their performance via implicit deadlines. Voskuilen et al. (2016) used integral-equation methods, similar to those described here, to compare fixed and collapsing bounds models on six different data sets from human participants on two different experimental tasks and again found only limited support for collapsing bounds. Evans, Hawkins, and Brown (2020) found that emphasizing decision speed through deadlines, and to a lesser extent through speed-emphasis instructions, led to the use of collapsing bounds, but instructions and experimental conditions designed to encourage reward-rate optimality did not. They concluded that collapsing bounds do not provide a good general model of human behavior.

As well as having only limited empirical support, Malhotra, Leslie, Ludwig, and Bogacz (2018) showed that, even when there is a time cost to sampling evidence, collapsing bounds models are optimal only under some circumstances, but not all. They may be optimal when experimental blocks contain a mixture of easy decisions and very hard decisions on which the expected performance is near chance. Collapsing bounds can be optimal in these circumstances because they prevent the decision maker spending too much time on decisions that are unlikely to be correct and not spending enough time on others for which the probability of being correct is greater. But if the hardest decisions are made a little easier, then fixed rather than collapsing bounds may be optimal. Overall, then, both the empirical and theoretical support for collapsing bounds models is less than it originally appeared to be.

There is, however, a more fundamental objection to the collapsing bounds idea, which is that it appears to be inconsistent with the neuroscience. A link between neural firing rates and evidence accumulation in monkeys performing eye-movement decision tasks was first made by Hanes and Schall (1996), who observed that decisions were made when the firing rates in frontal eye fields reached a fixed threshold level, consistent with the idea that the associated neurons were either implementing or reading out the results of an accumulate-to-bound decision process. The link between evidence-accumulation decision processes and the underlying neural firing rates is a theoretically productive one that has been developed by a number of researchers in the intervening period (Forstmann et al., 2016; Gold & Shadlen, 2003; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Ratcliff, Cherian, & Segraves, 2003; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007; Schall, 2002, 2003; Smith & Ratcliff, 2004). However, the idea that decision boundaries change over time is inconsistent with the empirical observation that decisions are made when neural firing rates reach a fixed threshold value.

The urgency signal model represents a possible solution to the collapsing-bounds dilemma. Churchland, Kiani, and Shadlen (2008) identified a stimulus-independent, time-dependent component of the neural firing rates in lateral interparietal area which they interpreted as a time-dependent urgency signal, and which they argued makes a decision increasingly likely with the passage of time, irrespective of the quality of the evidence extracted from the stimulus. Ditterich (2006a, 2008a) proposed an urgency model in which a time-varying

urgency signal is added to the contents of a pair of racing evidence accumulators, each of which is modeled as a diffusion process (Ratcliff & Smith, 2004). Additive urgency is a natural expression of Churchland et al.'s observation that urgency is an independent component of the neural signal, but it is inconsistent with the idea that evidence is accumulated by a single, signed evidence total between decision boundaries because the effects of urgency would then vary with the identity of the stimulus. To represent urgency in such models, Cisek and colleagues (Carland et al, 2015, 2016; Cisek et al., 2009; Thura, et al. 2012) proposed an *urgency-gating* model, in which the urgency signal is combined multiplicatively rather than additively with the accumulating evidence to make a decision. Unlike additive urgency, in multiplicative urgency models the sign of the product of the evidence and the urgency signal is automatically correct.

Stochastic Differential Equations for Evidence Accumulation

Evidence accumulation models can be characterized mathematically in either of two complementary ways: via partial differential equations or stochastic differential equations. In either case, the goal is to characterize the joint first-passage time distributions of the evidence accumulation process through the decision boundaries. These distributions are the distributions of times for the process to first reach one or other decision boundary and provide the model's predicted RT distributions and choice probabilities. The relevant partial differential equation is the so-called Kolmogorov backward equation (Bhattacharya & Waymire, 1990; Cox & Miller, 1965; Karlin & Taylor, 1981; Ratcliff, 1978, 1980), which must be solved subject to initial and boundary conditions. The initial conditions prescribe the starting state of the process (i.e., whether it is fixed or follows some probability distribution) and the boundary conditions prescribe what happens when the process reaches a boundary. Absorbing boundaries are far and away the most common form of boundary in decision models: On reaching an absorbing boundary, evidence accumulation stops and a response is made. However, some authors have also considered models with reflecting boundaries, either as models of time-controlled processing in tasks like the response-signal task (Zhang & Bogacz, 2010; Zhang, Bogacz, & Holmes, 2009) or in combination with absorbing boundaries, to constrain the sign of the evidence (Diederich, 1995; Smith & Ratcliff, 2009; Usher & McClelland, 2001). When a process reaches a reflecting boundary it is reflected back into the space and evidence accumulation continues from that point.

In contrast to the partial differential equation approach, which characterizes the evidence accumulation process indirectly via its transition probability distribution, the stochastic differential equation approach characterizes the process directly, as a random process unfolding in time. Historically, the partial differential equation approach dates from the work of Einstein (1905) and Smoluchowski (1906) and predated the stochastic differential equation approach by several decades, which had to await the development of a rigorous stochastic integral by Itô (1944, 1951). In applications, the advantages of the stochastic differential equation approach are often conceptual as much as mathematical, as they provide a direct and natural way to formalize a researcher's intuitions about a process under study. The first use of stochastic differential equations in the study of decision processes we are aware of was by Pacut (1980), who used them to characterize simple (one-choice) RT

and they were subsequently considered in detail in relation to both simple and two-choice RT by Smith (1995, 2000).

We focus here on characterizing diffusion processes by stochastic differential equations and the solution of the associated first-passage time problem via integral equations. Mathematically, a diffusion process is a continuous-time, continuous-state, Markov process, where the latter are processes whose conditional expectations at a given point in time depend solely on their value at that time and not on any earlier times. Although diffusion processes are not the most general processes that can be described by stochastic differential equations (Protter, 1990), we focus on them because of their central role in recent theorizing about decision processes. Diffusion process models arise in the study of decision-making from the assumption that evidence is represented by the pooled activity of populations of independent, noisy neurons, as proposed, for example, in the Poisson shot noise model of Smith (2010) or the Ising decision maker of Verdonck and Tuerlinckx (2014). A diffusion process representation of accumulating evidence in these models is obtained from the central limit theorem when the number of neurons in a population is large. Notationally, we follow the conventions of the applied probability literature and use subscripted upper-case Roman letters to denote stochastic processes (i.e., random variables). We use Greek or Roman letters to denote constants and nonrandom functions, with the arguments of the latter written in parentheses to distinguish them from stochastic processes.[1]

A diffusion process is defined mathematically by specifying two coefficients or functions: the *drift rate* and the *diffusion rate*. These coefficients are referred to jointly as the *infinitesimal moments* of the process. The drift rate prescribes the expected rate of change in the process and the diffusion rate prescribes the rate of change in its variability (Bhattacharya & Waymire, 1990, Ch. 7; Cox & Miller, 1965, Ch. 5; Karlin & Taylor, 1981, Ch. 15). The square root of the diffusion rate is called the *infinitesimal standard deviation*. In the most general diffusion process, the drift and diffusion rates may depend on both time, t, and the position of the process in the evidence space, x. We write such a process as

$$dX_t = A(X_t, t)dt + \sqrt{B(X_t, t)}dW_t, \tag{1}$$

where $A(x, t)$ and $B(x, t)$ are, respectively, the drift and diffusion rates, and dW_t is a zero-mean Gaussian increment whose standard deviation in a small interval of duration t is of the order $\sqrt{\Delta t}$. Stochastic differential equations are usually written in the differential form of Equation 1 rather than in the more usual form involving derivatives because of the difficulty in giving meaning to terms of the form dW_t/dt. In the usual interpretation of Equation 1, W_t is a Brownian motion, or Wiener process, whose trajectories (sample paths) are almost everywhere nondifferentiable. Articles in neuroscience sometimes express evidence accumulation equations in the form dX_t/dt, as if they were ordinary differential equations, but it is important to understand that, strictly, such expressions have no meaning.

---

[1] The reason for this notation is that stochastic processes are considered to be functions on a probability space, Ω. In this setup, a single realization of the process (i.e., a sample path on an experimental trial) will depend on the value of a point, $\omega \in \Omega$, sampled from this space. For a process X_t this relationship is written explicitly as $X_t(\omega)$, but when the details of the probability space are unimportant, as here, the dependence on ω may be omitted from the notation.

Rather, the meaning of Equation 1 comes from its corresponding expression in integrated form, which presupposes the existence of a well-defined stochastic integral.

The two most important special cases of Equation 1 in models of decision processes are the Wiener diffusion process and the Ornstein-Uhlenbeck (OU) diffusion process. The former satisfies the stochastic differential equation

$$dX_t = \mu dt + \sigma dW_t \tag{2}$$

and the latter satisfies the equation

$$dX_t = (\mu - \gamma X_t)dt + \sigma dW_t, \tag{3}$$

In both equations, the infinitesimal standard deviation, σ, and the stimulus-dependent portion of the drift rate, μ, are constant. The equations differ by the presence or absence of the term $-\gamma x$, which represents a tendency for the accumulated evidence to decay at a rate proportional to its current value, x. Because of this property, the Wiener and OU processes are often characterized as being "perfect" and "leaky" integrators, respectively. The Wiener process was proposed by Ratcliff (1978) in his original formulation of the model and is the diffusion process most widely used to model data. It is also, when augmented with various sources of across-trial variability discussed below, the process that has been implemented in third-party software packages for fitting data (Vandekerckhove & Tuerlinckx, 2008; Voss & Voss, 2007; Wiecki et al., 2013).

The OU process has been considered by several authors in varying settings. Busemeyer and Townsend (1991, 1993) used it to model evidence accumulation governed by approach-avoidance dynamics in their decision field theory. Smith (1995) used it in his sustained-and-transient channel model of simple RT and Diederich (1995) used it in her model of intersensory facilitation. The component channels in Usher and McClelland's (2001) leaky competing accumulator model are modeled as OU processes with mutual inhibition and Smith and Ratcliff (2009) assumed racing OU processes between absorbing and reflecting boundaries in their implementation of Ratcliff and Smith's (2004) dual diffusion model (see also Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, (2007)).

When the coefficients μ, σ, and γ are independent of time, as written in Equations 2 and 3, the corresponding processes are said to be *time homogeneous*. The Wiener process is also spatially homogeneous, which means that it can be translated in evidence space, simply by relabeling the boundaries and starting point, without changing any of its properties. (The OU process is not spatially homogeneous because the decay term $-\gamma x$ represents a true zero towards which evidence decays.) Time inhomogeneous versions of these processes can be obtained by making any or all of μ, σ, and γ functions of time. Such models have been used by Smith and colleagues to characterize time-dependent changes in the evidence entering the decision process (Sewell & Smith, 2012; Smith, 1995, 2000; Smith, Ellis, Sewell, & Wolfgang, 2010; Smith & Ratcliff, 2009; Smith, Ratcliff, & Sewell, 2012; Smith & Lilburn, 2020). The processes in Equations 2 and 3 can equivalently be characterized in terms of partial differential equations that, when explicitly soluble, lead to infinite-series

representations of the first-passage time distributions (Bhattacharya & Waymire, 1990; Cox & Miller, 1965; Karlin & Taylor, 1981).

Along with the accumulation process, a complete specification of a model requires decision boundaries and a starting point for evidence accumulation. Ratcliff (1978) followed Feller's (1968) "gambler's ruin" formulation of the Wiener diffusion process and denoted the boundaries as 0 and a and the starting point as z. This is the way the model is usually parameterized when it is fitted to data, but when boundaries can vary with time it is convenient to use a different parameterization. Here we denote the upper and lower boundaries as $a_1(t)$ and $a_2(t)$ with $a_2(t) < z < a_1(t)$. Two main forms of time-varying boundary functions have been considered in the literature. Voskuilen et al. (2016) followed Churchland et al. (2008) and Hanks et al. (2011) and assumed a two-parameter boundary function of the form,

$$
\begin{aligned}
a_1(t) &= \frac{a}{2}\left[1 - \kappa\frac{t}{t + t_{0.5}}\right] \\
a_2(t) &= -\frac{a}{2}\left[1 - \kappa\frac{t}{t + t_{0.5}}\right].
\end{aligned}
$$

(4)

In these equations, a denotes the boundary separation at time $t = 0$, κ denotes the rate at which the boundaries collapse (converge) and $t_{0.5}$ is a semisaturation constant that identifies the time at which the boundaries have collapsed to 50% of their starting values. Hawkins et al. (2015) assumed a more general, three-parameter function, which allowed them to compare "early collapse" and "late collapse" forms of the model. For those data sets supporting a collapsing-bounds account, their analysis favored a late collapse model.

The Urgency-Gating Model

Urgency models were first fitted to data by Ditterich (2006a, 2006b), but the most elaborated urgency model to date is the one proposed by Cisek, Thura, and colleagues. Their model makes two core claims: One is that evidence does not accumulate unboundedly but grows to a stationary distribution; the other is that the stationary distribution of evidence is modulated (gated multiplicatively) by a time-dependent urgency function, $U(t)$, and a response is made when the urgency-gated stationary evidence reaches one of two decision boundaries.

The idea that evidence does not accumulate is of course not a new one, but is one with a long history in psychology. Dating from the work of Cartwright and Festinger (1943), a number of authors have proposed Thurstonian or signal detection models that assume a distribution of evidence that contains an interval of uncertainty bounded by two decision criteria (Swets & Green, 1961; Atkinson & Joula, 1974; Murdock, 1983). The decision maker is assumed to repeatedly sample from the distribution at a constant rate until a piece of evidence is obtained that falls outside the interval of uncertainty and to respond according to whether it falls above the upper or below the lower criterion. Hockley and Murdock (1987) proposed that the criteria converge with time, anticipating the current collapsing boundaries models in neuroscience. The addition of converging criteria improved the properties of the model, but several problems with its RT distribution predictions were identified by Gronlund and

Ratcliff (1991). Pike, McFarland, and Dalgleish (1974) proposed a counter model driven by normal distributions of evidence strength with an interval of uncertainty to model two-choice RT, in which counters were incremented only by observations falling outside the interval of uncertainty. In a similar vein, Smith and Vickers (1989) proposed a version of the Vickers accumulator model (Smith & Vickers, 1988; Vickers, 1970) in which the accumulators were driven by a signal detection process with an interval of uncertainty. Their model predicts a continuum of performance as a function of the width of the interval of uncertainty. At one end, all of the evidence is accumulated; at the other end, no evidence is accumulated and the decision is based on a single, highly-diagnostic sample. The model provided a good account of a wide range of RT distributions from a fast-paced expanded judgment task, in which the RT distributions ranged from highly skewed to highly symmetrical and were explained by treating the width of the interval of uncertainty as an individual differences parameter.

The urgency-gating model of Cisek and colleagues was formulated in continuous time in terms of ordinary rather than stochastic differential equations and its assumptions about noise were expressed in an informal way. It has also changed significantly over successive articles (cf. Cisek et al. (2009), Thura et al. (2012), and Carland et al. (2016)), which has led to confusion in the literature about its core properties. Winkel et al. (2014) reported evidence against the urgency-gating model from a task in which the stimulus information changed during a trial (see below), which was disputed by Carland et al. (2015) because Winkel's implementation omitted leakage, which the authors regarded as essential. Here we follow the presentation of the model in the later articles (Carland et al. 2015, 2016), but we express it explicitly using stochastic differential equations. We then obtain an explicit solution of the first-passage time problem for the associated process through absorbing boundaries using integral equation methods and compare the resulting model to data. Our results are important because they show that very general models with urgency, collapsing bounds, and time-varying stimulus information can be represented within a common mathematical framework that provides explicit expressions for response accuracy and RT distributions, comparable to those that exist for the standard diffusion model. The existence of such explicit expressions has been critical to the standard model's historical success.

A central feature of the urgency-gating model that purportedly distinguishes it from other decision models is that the encoded stimulus information is low-pass filtered before being accumulated. Mathematically, a low-pass filter removes high-frequency components of its input and transmits only low-frequency components. The shape of the filter determines the frequency spectrum of the output. In the urgency-gating model, the low-pass filter is implemented by an equation, which, when rigorously expressed, is identical to Equation 3. That is, the model assumes that evidence is accumulated by an OU diffusion process. While it is true that the OU process can be viewed as performing a low-pass filtering operation on its input, characterizing the process as a low-pass filter rather than an OU process has the undesirable consequence of obscuring its relationship with a model that has been well studied in the psychological literature.

The properties of the evidence accumulation equation in the urgency-gating model can best be understood by considering the process X_t that solves Equation 3, which is one of the

few stochastic differential equations that can be solved by direct methods (Karlin & Taylor, 1981, pp. 345–346). The solution may be expressed (Smith, 2000, Equation 23), as

$$X_t = \int_0^t e^{-\gamma(t-\tau)} \mu \, d\tau + \sigma \int_0^t e^{-\gamma(t-\tau)} dW_\tau, \tag{5}$$

where we assume an initial condition of $X_0 = 0$. In this form, X_t can be seen to be the sum of a continuous function and a stochastic process that are obtained, respectively, by putting a constant function, μ, and a white noise process, dW_t, through an exponential linear system with rate constant γ. Such a system is a low-pass filter. The white-noise process is the "formal derivative" of the Wiener process in the sense that it integrates to W_t. It can be thought of as a Gaussian process whose covariance at all pairs of nonidentical time points, τ and t, $\tau \neq t$ is zero.

The key properties of the process X_t can be inferred from Equation 5. First, it may be shown (Karlin & Taylor, 1981, pp. 345–346; Smith, 2000) that the second integral on the right hand side is a *martingale*, that is, a bounded stochastic process whose expected value is constant, which will equal the value of the integral at its lower bound, which is zero.[2]

The expected value of X_t will therefore simply be the value of the first integral term on the right,

$$E[X_t] = \frac{\mu}{\gamma}\left(1 - e^{-\gamma t}\right). \tag{6}$$

That is, the mean of the process grows exponentially with rate γ to an asymptote of μ/γ. It can further be shown by means of the so-called *Itô isometry* (Chung & Williams, 1983, p. 27), which also depends on the martingale proper of the stochastic integral, that the variance of X_t is

$$\mathrm{Var}[X_t] = \frac{\sigma^2}{2\gamma}\left(1 - e^{-2\gamma t}\right), \tag{7}$$

which again grows exponentially to an asymptote of $\sigma^2/(2\gamma)$. Because the process X_t is a linear combination of independent Gaussian increments, it is also Gaussian. Together, Equations 6 and 7 show that, asymptotically, the process X_t has a stationary Gaussian distribution with mean μ/γ and variance $\sigma^2/(2\gamma)$. This contrasts with the Wiener process of Equation 2, which does not possess a stationary distribution: Its mean and variance are $E[X_t] = \mu t$ and $\mathrm{Var}[X_t] = \sigma^2 t$, respectively.

Cisek and colleagues further stipulate that the approach to stationarity is rapid. Carland et al. (2015) suggested that the time constant of the filter is around 250 ms; Carland et al.

---

[2]There are two different kinds of stochastic integral in the literature: the Itô integral and the Stratonovich integral (Karlin & Taylor, 1981). The Stratonovich integral is obtained as the limit of sums taken with respect to a band-limited Gaussian process as the bandwidth is increased, so it is often seen as more directly expressing the properties of systems studied by physicists and engineers, which are necessarily band-limited. Unlike the Itô integral, however, it is not a martingale, which is why the latter is preferred for theoretical work. For simple processes like the OU process the two forms of the stochastic integral coincide.

(2016) suggested it is in the range 100–200 ms. When time is measured in seconds, a time constant of 250 ms corresponds to an OU decay of $\gamma = 4.0$, and a time constant of 125 ms corresponds to a decay of $\gamma = 8.0$.

An OU decision model with time constants in this range was investigated by Ratcliff and Smith (2004) on three different decision tasks: a perceptual task, a lexical decision task, and a recognition memory task. They found that the predictions for an OU model with $\gamma = 4.0$ could not be distinguished from the Wiener diffusion model in data, but an OU model with $\gamma = 8.0$ predicted RT distributions that were more skewed than are found in data and could be rejected. The increase in skewness was because, in the $\gamma = 8.0$ case, most of the probability mass in the stationary distribution falls inside the decision boundaries, making boundary crossings relatively infrequent and slow. Such a process behaves, asymptotically, like the discrete-time models discussed earlier in which evidence does not accumulate. In contrast, in the $\gamma = 4.0$ case, much more of the stationary distribution falls outside of the boundaries, making boundary crossings relatively frequent and yielding similar predictions to the Wiener model. When decay was allowed to vary freely in fits of the model to data, the estimated value approached $\gamma = 0$, which is the Wiener process.

In the urgency-gating model, the OU process is multiplied by an urgency function, $U(t)$, leading to an evidence accumulation function, Y_t, of the form

$$Y_t = U(t)X_t. \tag{8}$$

Some authors have proposed that urgency grows nonlinearly with time (Ditterich, 2006a, 2006b), but in the model of Cisek and colleagues the growth is linear. In early presentations of the model, an urgency function of the form $U(t) = mt$ was assumed, and this is the form of the model that has most often been evaluated in the literature (Hawkins et al., 2015; Winkel et al., 2014; Evans, Trueblood, & Holmes, 2020), but in later presentations a more general function of the form $U(t) = b + mt$ is assumed. As pointed out by Trueblood et al. (2021), this latter formulation endows the model with greater flexibility. The single-parameter function can be viewed as a "pure urgency" model, in the sense that when $m = 0$ there is no evidence accumulation. In contrast, the two-parameter function allows increasing amounts of urgency to be added to a basic OU diffusion model. When $b = 1$ and $m = 0$ the model reduces to a pure OU model, with increasing amounts of urgency as m increases.

To analyze the process of Equation 8, we ask what stochastic differential equation is satisfied by the product of functions on the right-hand side. In general, the differential of a product of stochastic processes does not follow the normal rules of calculus, because it contains an additional term called the "quadratic covariation" of the two processes (Protter, 1990), but when one of the components is a deterministic function, as here, then the normal product rule applies. We consider a general time-inhomogeneous form of the model in which the OU drift and diffusion rates both depend on time

$$dX_t = [\mu(t) - \gamma X_t]dt + \sigma(t)dW_t. \tag{9}$$

We can then write

$$
\begin{aligned}
dY_t &= U'(t)X_t dt + U(t)dX_t \\
&= U'(t)X_t dt + U(t)\{[\mu(t) - \gamma X_t]dt + \sigma(t)dW_t\} \\
&= \left\{ U(t)\mu(t) + \left[\frac{U'(t)}{U(t)} - \gamma\right]Y_t \right\}dt + U(t)\sigma(t)dW_t,
\end{aligned}
\tag{10}
$$

after making the substitution $X_t = Y_t/U(t)$. That is, the urgency-gating model satisfies a stochastic differential equation with drift rate

$$
A(y,t) = \left\{ U(t)\mu(t) + \left[\frac{U'(t)}{U(t)} - \gamma\right]y \right\}
\tag{11}
$$

and diffusion rate

$$
B(t) = U^2(t)\sigma^2(t).
\tag{12}
$$

Evans, Trueblood, and Holmes (2020) and Trueblood et al. (2021) derived similar expressions for the infinitesimal moments of the pure urgency model and the model with linear urgency signal, respectively. The derivation in Equations 9 through 12 is for a time-inhomogeneous, urgency-gated, OU process with an arbitrary urgency function, which subsumes the models analyzed in those articles as special cases.

Integral Equation Predictions for Time-Varying Diffusion Processes

The integral equation method provides an effective way to obtain predicted RT distributions and choice probabilities for a wide variety of diffusion models with time-varying drift and diffusion rates and/or with time-varying boundaries. The method was first proposed by Durbin (1971) and later developed to study the properties of integrate-and-fire neurons by Ricciardi and colleagues (Buonocore, Nobile, & Ricciardi, 1987; Buonocore, Giorno, Nobile, & Ricciardi, 1989; Giorno, Nobile, Ricciardi, & Sato, 1989). A pioneering article by Heath (1992) used Durbin's method to study a diffusion process version of McClelland's (1979) cascade model and, as noted above, the method has been used extensively by Smith and colleagues to study processes in which the evidence entering the decision process changes over time because of the action of perception, memory, and attentional processes. The method has also been used to derive predictions for diffusion models by Ditterich (2006a, 2006b), Evans, Hawkins, and Brown (2020), and Jones and Dzhafarov (2014). A detailed tutorial account, focusing on decision models and incorporating refinements of the method proposed by Gutiérrez Jáimez, Román Román, and Torres Ruiz (1995), may be found in Smith (2000).

The quantities of theoretical interest are the joint first-passage time densities for the process through the boundaries $a_1(t)$ and $a_2(t)$, which we allow to be time varying, although in the urgency-gating model they are assumed to be fixed. The reason for the extra generality is to allow us to identify conditions under which collapsing bounds and urgency-gating models are equivalent. We denote these densities as $g_A[a_1(t), t|z, 0]$, and $g_B[a_2(t), t|z, 0]$, where the subscripts "A" and "B" denote the responses associated with the upper and lower

boundaries, respectively. The conditional notation expresses the idea that these functions are first-passage time densities for a process X_t starting at z at time zero, which makes a first boundary crossing at either $a_1(t)$ or $a_2(t)$ at time t. The first-passage time densities have the integral equation representations

$$
\begin{aligned}
g_A[a_1(t), t \mid z, 0] = {}&{-}2\Psi[a_1(t), t \mid z, 0] \\
&+ 2\int_0^t g_A[a_1(\tau), \tau \mid z, 0]\Psi[a_1(t), t \mid a_1(\tau), \tau]d\tau \\
&+ 2\int_0^t g_B[a_2(\tau), \tau \mid z, 0]\Psi[a_1(t), t \mid a_2(\tau), \tau]d\tau
\end{aligned}
\tag{13}
$$

and

$$
\begin{aligned}
g_B[a_2(t), t \mid z, 0] = {}&2\Psi[a_2(t), t \mid z, 0] \\
&- 2\int_0^t g_A[a_1(\tau)\tau \mid z, 0]\Psi[a_2(t), t \mid a_1(\tau), \tau]d\tau \\
&- 2\int_0^t g_B[a_2(\tau), \tau \mid z, 0]\Psi[a_2(t), t \mid a_2(\tau), \tau]d\tau.
\end{aligned}
\tag{14}
$$

The first-passage time densities in Equations 13 and 14 are defined as the integrals of the products of their values at times $\tau < t$ and of a kernel function $\Psi[a_i(t), t|a_j(\tau), \tau]$, $i, j, = 1, 2$, which depends jointly on the boundaries and on the transition density of a diffusion process with drift and diffusion rates given by Equations 11 and 12.

For a large class of diffusion processes, specifically, those that can be transformed to a standard (zero mean, unit variance) Wiener process by a change of coordinates, Buonocore et al. (1987, 1990) showed that the kernel function has a particular form, which depends on the functions that transform the process from the old to the new coordinates. When this transformation exists, Ricciardi (1976), following Cherkasov (1957), showed the old space and time coordinates, x and t, are related to the new space and time coordinates, x^* and t^*, by a pair of functions of the form

$$
x^* = \overline{\Psi}(x, t)
\tag{15}
$$

$$
t^* = \Phi(t).
\tag{16}
$$

The new space coordinate is a function jointly of the old space and time coordinates whereas the new time coordinate is a function of the old time coordinate only. (Note carefully the overbar notation, $\overline{\Psi}(\,\cdot\,)$, that distinguishes the function mapping the space coordinate from the kernel function itself.) The conditions for the existence of the functions $\overline{\Psi}(\,\cdot\,)$ and $\Phi(\cdot)$ for a given diffusion process are given in Appendix A.

When the pair of functions $\overline{\Psi}(\,\cdot\,)$ and $\Phi(\cdot)$ exist, the kernel of the integral equations in Equations 13 and 14 can be written (Gutiérrez Jáimez et al., 1995; Smith, 2000, Equation 56) as

$$\Psi\big[a_i(t), t \mid a_j(\tau), \tau\big] = \frac{f\big[a_i(t), t \mid a_j(\tau), \tau\big]}{2}$$
$$\times \left\{ a_i'(t) + \frac{\overline{\Psi}_t'(a_i(t), t)}{\overline{\Psi}_x'(a_i(t), t)} - \frac{\big[\overline{\Psi}(a_i(t), t) - \overline{\Psi}(a_j(\tau), \tau)\big]}{\Phi(t) - \Phi(\tau)} \frac{\Phi'(t)}{\overline{\Psi}_x'(a_i(t), t)} \right\}. \quad (17)$$

In this equation, $\overline{\Psi}_x'(a_i(t), t)$ and $\overline{\Psi}_t'(a_i(t), t)$ are the partial derivatives of $\overline{\Psi}(\cdot)$ with respect to state and time, respectively, and $\Phi'(t)$ and $a_i'(t)$ are the derivatives of $\Phi(\cdot)$ and the boundary function $a_i(t)$ with respect to time. (We use subscripts to denote partial derivatives for functions of two variables but omit the subscript for derivatives of functions of a single variable.) The function $f[a_i(t),\ t|a_j(\tau),\ \tau]$ is the transition density of the process X_t, unconstrained by boundaries, expressed in terms of the functions that transform the process from the old to the new coordinates (Smith, 2000, Equation 51),

$$f\big[a_i(t), t \mid a_j(\tau), \tau\big] = \frac{1}{\sqrt{2\pi[\Phi(t) - \Phi(\tau)]}} \exp\left\{-\frac{\big[\overline{\Psi}(a_i(t), t) - \overline{\Psi}(a_j(\tau), \tau)\big]^2}{2[\Phi(t) - \Phi(\tau)]}\right\} \overline{\Psi}_x' \quad (18)$$
$$(a_i(t), t).$$

Equivalence of Urgency-Gating and Collapsing Boundaries Models

We are now ready for the main theoretical result of this article. In Appendix A it is shown that the functions that transform the diffusion process for the urgency-gating model, with drift and diffusion rates given by Equations 12 and 13, to a standard Wiener process have the form (expressed in terms of the variable y of Equations 11 to 13) as

$$y^* = \overline{\Psi}(y, t) = \frac{e^{\gamma t} y}{U(t)} - \int^t \mu(\tau) e^{\gamma \tau} d\tau \quad (19)$$

$$t^* = \Phi(t) = \int^t e^{2\gamma\tau} \sigma^2(\tau) d\tau. \quad (20)$$

In this notation, y^* represents the transformed state coordinate of the urgency-gated process, Y_t, under the transformation of Equations 15 and 15. The notation parallels the notation x^* for the transformed state coordinate of the process, X_t, without urgency, in Equation 15.

Equations 19 and 20 provide a rigorous and succinct characterization of the conditions under which collapsing boundaries and urgency-gating models are equivalent. The transformed urgency-gating process will cross a boundary when $y^* = a_i$ in Equation 19, and this expression may equivalently be interpreted as saying that a process without urgency crosses a boundary when it reaches the level $a_i/U(t)$. A collapsing boundaries model with decision boundaries $a_i(t) = a_i/U(t)$ will therefore be equivalent to an urgency-gating model with fixed boundaries a_i and urgency function $U(t)$. Trueblood et al. (2021) stated this result without proof for the urgency function $U(t) = b + mt$, and, while it is a natural and intuitive one and may strike some readers as self-evident, it is important to recognize that it is not possible to

reason about stochastic processes as if they were deterministic functions: The processes X_t and Y_t have quite different probabilistic characters because of their different diffusion rates. Mathematics and intuition agree in this case because $U(t)$ appears only in the expression for the state coordinate but not the time coordinate of the transformed process. If $U(t)$ also appeared in the transformed time coordinate then the models would no longer be equivalent. It is not self-evident from the expressions for the infinitesimal moments of Y_t in Equations 11 and 12, both of which depend on $U(t)$, that the process transforms under urgency in this way. We establish this relationship formally via the transformation equations in Appendix A. The equivalence of the models can also be shown by a direct probabilistic argument but, unlike the integral equation approach, the direct argument only establishes that the models are equivalent but it does not give explicit expressions for the first-passage time density functions.[3]

Mathematically, the equivalence of the models requires that they predict the same first-passage time densities, $g_A[a_1(t), t|z, 0]$, and $g_B[a_2(t), t|z, 0]$, in Equations 13 and 14. They will do if and only if they have the same kernel function, Equation 17. When the model is viewed as a collapsing boundaries model, the term $U(t)$ appears in the expression for the boundary, so the derivative $a_i'(t)$ is nonzero, but it does not appear in the function $\overline{\Psi}(\cdot)$ that maps the state coordinate of the process (Equation A19). Conversely, when the model is interpreted as a fixed-boundaries urgency model, the function $a'(t)$ is zero and $U(t)$ instead appears in $\overline{\Psi}(\cdot)$ (Equation A10). The resulting kernel function is the same under either interpretation, making the two models equivalent. The mathematical details may be found in Appendix A. The equivalence of collapsing boundaries and urgency-gating models holds for any boundary/urgency function, not just the function $U(t) = b + mt$ in the extended urgency-gating model. Although Equations 19 and 20 characterize the transformation of the state and time variables for an OU diffusion process, they include the Wiener process as a special case, obtained by setting $\gamma = 0$.

Explicit Predictions for Collapsing Boundaries and Urgency-Gating Models

As well as identifying conditions under which collapsing boundaries and urgency-gating models are mathematically equivalent to each another, the integral equation method provides explicit expressions for the first-passage time density functions for the model(s) which can be used to fit them to data. In applications, the solutions in Equations 13 and 14 are evaluated numerically by defining the process on a discrete time mesh, $t_i = i$, $i = 1, 2, \ldots$, and approximating the integrals with discrete sums. The discretized form of the equations can be found in several places, including Smith (2000), Smith and Lilburn (2020), and Voskuilen et al. (2016), and are reproduced in Appendix B here. Figure 1 shows some example first-passage time density functions for the urgency-gating model, for different

---

[3]A direct probabilistic approach to establishing the equivalence of the models would use the relationship $Y_t = U(t)X_t$ to show equality of the finite-dimensional distributions of the processes on a discrete set of time points, $\{t_i\}$. Because the Wiener process is of unbounded variation on any finite interval (Protter, 1990, p. 19), equality of finite-dimensional distributions does not suffice to show the models predict the same first-passage time distributions, which requires equality for all $t \in \mathbb{R}^+$, the set of positive real numbers. The kinds of sequential-limiting arguments needed to establish the equivalence of two continuous-time stochastic processes from the equality of their finite-dimensional distributions are often arduous (e.g., Smith, 2010), but, in the case of the urgency-gating model, the argument is relatively straightforward because the process Y_t is a continuous, one-to-one transformation (i.e., a homeomorphism) of a continuous process, X_t. We omit the details.

values of the urgency parameters, *b* and *m*, and compares them to the results of Monte-Carlo simulations of the model. The simulations were carried out using the Euler method (Brown, Ratcliff, & Smith, 2006), which approximates the diffusion process with a discrete-time, Gaussian random walk, using a time-step of 1 ms with a correction for the excess over the boundary on the terminating step (Smith, 1990)[4]. Each of the simulations in the figure was based on 100,000 trials. The figure shows the integral equation method is an effective way to evaluate models of this kind.

The predicted RT distributions in Figure 1 highlight a feature of the urgency-gating model that was implicit in the preceding discussion, namely, that its theoretical content involves a trade-off between two processes: an OU evidence accumulation process, which tends to lengthen RTs and increase the skewness of RT distributions as decay increases, and an urgency function, which tends to shorten RTs and decrease skewness as urgency increases. The model has two parameters, γ and *m*, which allows these two processes to be traded off against one another in a flexible way. Thura (2015) responded to criticism by Hawkins et al. (2015) of an earlier form of the model by Thura et al. (2009) by arguing that the revised urgency-gating model and the standard diffusion model were unlikely to be distinguishable in tasks in which stimulus information does not change during a trial but can be distinguished in tasks in which the information changes. Trueblood et al. (2021) reported parameter recovery simulations that appear to provide some support for Thura's claim, although the recovered parameters were highly variable for both constant and changing-stimulus conditions. We also report data from such a task, using the paradigm of Trueblood et al., although our treatment differs from theirs in two ways. First, unlike them, we compared the urgency-gating model to the standard diffusion model with across-trial variability. Variability in drift rate in the latter model is important because it leads to an ordering of correct and error RTs that is like the one predicted by models with urgency or collapsing bounds. Consequently, any unbiased comparison of models needs to include variability in drift rate in the fixed-boundary models. Second, we model the encoding of stimuli in time-varying tasks using a version of the perceptual encoding model of Smith and Lilburn (2019). This allowed us to distinguish the effects of perceptual and decisional integration and to characterize how each of these processes is affected by changes in stimulus information.

Method

Experimental Studies

We compared the standard diffusion model and the urgency-gating model on the data from three different experimental paradigms. Two of them were standard RT tasks, in which the stimulus information did not change during the course of a trial and third of them

---

[4]A diffusion process terminates as soon as a boundary is reached but a random walk simulated with the Euler method terminates only once a boundary has been exceeded. On average, the walk travels further before terminating than does the corresponding diffusion process. The random-walk approximation to a diffusion process is improved if the boundaries of the diffusion are adjusted for the excess of the walk over the boundary on its terminating step. For the simulations in Figure 1, the boundary separation was increased by $\sqrt{h}\sigma$, where $h = 0.001$ s was the time step used in the simulation and $\sigma = 0.1$ was the infinitesimal standard deviation of the diffusion process. Further discussion of corrections for two-boundary diffusion processes may be found in Smith (1990). Discussion of an analogous correction for the circular diffusion model may be found in Footnote 4 of Smith (2016).

was a task in which stimulus information could change. The two standard tasks were a numerosity discrimination study by Ratcliff (2008), in which participants made judgments about the number of stimulus elements in a display, and an attentional cuing study by Smith, Ratcliff, and Wolfgang (2004), in which participants discriminated the orientations of grating patches presented at cued or uncued locations. The third was the paradigm of Trueblood et al. (2021) in which participants made decisions about the dominant hues of flashing grids of blue and orange squares. We report data and model fits from three experiments using this paradigm. In their version of the task, trials timed out after 2 s if participants did not respond and the experimental program went on to the next trial. This procedure effectively imposes a deadline on responding — albeit a comparatively long one. Our first set of fits is from a reanalysis of their Experiment 1. The second is from a replication of their Experiment 1 that was run without the deadline. The third is from a version of the task that used luminance rather than color stimuli, again run without a deadline. One of our aims was to compare models in which perceptual and decisional integration could vary with time, in which perceptual integration was represented mathematically by a time-varying stimulus information function, $\mu(t)$, in Equation 11. The comparison between luminance and color is of theoretical interest when evaluating these models because color is processed more slowly than luminance (Ingling & Martinez, 1983), so we might expect to see the difference reflected in the estimated model parameters. All of the experiments collected large samples of data from individual participants at different levels of stimulus discriminability, permitting a detailed analysis of the RT distributions.

The participants in Ratcliff's (2008) study were asked to decide whether the number of randomly-placed dots in a 10×10 grid was greater or less than 50. There were eight nominal discriminability conditions, in which the numbers of dots were: 31–35, 36–40, 41–45, 46–40, 41–45, 46–50, 61–65, and 66–70, crossed with speed versus accuracy instructions. On half the trials participants were instructed to respond rapidly and on the other half they were told to respond accurately and they were given feedback to encourage them to perform as instructed. Data were collected from 19 college-aged participants and 19 older participants who performed both a standard RT task, in which they responded as soon as they had sufficient evidence, and a response-signal task, in which they responded to a random external deadline. We restrict our analysis to the younger participants and the standard RT task. After eliminating fast and slow outliers, there were about 850 valid trials in each cell of the design, yielding around 13,600 valid trials per participant.

Participants in the study of Smith et al. (2004) performed an attentional cuing task in which low contrast Gabor patches were presented for 60 ms at either a cued location or at one of two uncued locations. On each trial, participants decided whether the orientation of the patch was vertical or horizontal. The cue consisted of four corners of a square marking a stimulus location that were flashed for 60 ms, 140 prior to stimulus onset. In one condition of the experiment, the stimuli were backwardly masked with high-contrast checkerboards and in the other condition they were briefly flashed and then extinguished. Data were collected from six participants, who performed the task at five different levels of contrast, which were chosen for each participant individually during practice to span a range of performance from just above chance ($\approx$ 55% correct) to near-perfect ($\approx$ 95% correct). They were encouraged to be as accurate as possible but not to deliberate for too long and were given auditory accuracy

feedback on each trial. There were 400 valid trials for each participant in each cell of the Cue × Mask × Discriminability design, yielding 8000 trials per participant.

The third set of experiments used the flashing grid task of Trueblood et al. (2021). The task resembles the dynamic noise tasks of Ratcliff and Smith (2010), but uses different stimulus elements, as discussed subsequently. We report a reanalysis of Trueblood et al.'s Experiment 1 and data from two new experiments using versions of their task. In Trueblood's task, participants made judgments about the dominant hues of flashing 20×20 grids of random blue and orange squares that changed every 50 ms (20 Hz). We replicated this experiment (see below) and also carried out a brightness discrimination of version of it that used black and white instead of colored squares, in which participants made judgments about whether black or white squares predominated. In both the color and the luminance versions of the task, on half the trials, the stimulus information, represented by the proportion of squares of the dominant attribute in each frame, stayed constant during the trial and on the other half it changed. In addition, the discriminability of the stimuli was manipulated by varying the proportion of squares of the dominant attribute in each frame. Following the procedure of their Experiment 1, on half of the constant information trials, 0.53 of the squares were of the same color or lightness and on the other half 0.57 were of the same color or lightness. On the changing information trials, on half the trials the stimulus information changed from low to high discriminability (0.47 to 0.57) after 350 ms and on the other half it changed to high to low (0.43 to 0.53) after 350 ms. On changing information trials, responses consistent with the dominant attribute after the switch were deemed to be correct and participants were given average RT and accuracy feedback at the end of each block of 72 trials.

In Trueblood et al.'s experiments, trials timed out after 2000 ms and a nonresponse was recorded. We followed their procedure in our replication except we removed the requirement that responses had to be made within 2000 ms, for both psychological and statistical reasons. Psychologically, limiting the time for which the display can be viewed may incline participants to use a collapsing boundaries or urgency-based decision strategy to a greater extent than they might otherwise. Statistically, terminating trials at 2000 ms is a data-censoring process that can bias estimation if its effects are not modeled explicitly. To investigate the effects of censoring, we carried out a parametric bootstrap cross-validation study (Wagenmakers, Ratcliff, Gomez, & Iverson, 2004; Voskuilen et al., 2016), in which we cross-fit the standard diffusion model and the urgency-gating model to simulated data from the other model. Using parameters derived from fits of the two models to Trueblood et al.'s published data, we found that there was a bias towards the urgency-gating model that was increased by censoring the data at 2000 ms, so we ran our replication without a deadline to minimize the bias.

Initially, we reanalyzed the data from Trueblood et al.'s (2021) Experiment 1, which are publicly available from the Open Science Foundation, but features of the data made it difficult to undertake the detailed analysis of RT distributions we wished. One feature was an unusually high proportion of fast guesses; the other was that many participants showed poor across-trial stability, manifested as extended runs of fast or slow responses. Of the 34 participants in Trueblood et al.'s Experiment 1, the average fast-guess rate for 22 of them was 25.5%, which we defined as responses with RTs of less than 350 ms

and chance-level (0.491) accuracy. Figures 10 and 11 in Appendix C show trial-by-trial plots of the RTs on the experimental trials for the 34 participants. The horizontal dashed line at 350 ms shows the fast-guess threshold and RTs on timed-out trials are plotted as 2000 ms. This representation provides a graphical way to identify those participants whose performance was dominated by fast guesses and those whose performance was unstable over time. Trueblood et al. reported data from four experiments using this task, all of which showed similar features to those reproduced here. The overall fast-guess rates for their four experiments were 16.8%, 5.6%, 12.3%, and 19.7%, respectively. In our reanalysis of their data we restricted the analysis to the 12 participants from their Experiment 1 who did not show high fast-guessing rates and who showed stable performance across trials.

In our replication of the task we included a time-out penalty for very fast responding, which was effective in controlling fast guesses. We collected data from 19 participants, each of whom provided around 350 trials in each of the four experimental conditions after fast guesses were excluded. Of those 19 participants, three of them showed similar across-trial instability to that observed in the data of Trueblood et al., so we restricted our analysis to the 16 remaining participants. In our luminance version of the task we collected data from 16 participants, each of whom provided a similar number of trials after excluded fast guesses.

Modeling Drift Rates in the Flashing Grid Task

The flashing-grid task raises theoretical questions not raised by either the numerosity or spatial cuing experiments about how to model the evidence entering the decision process, represented in the model by the drift and diffusion rates. In the standard diffusion model, the drift and diffusion rates are modeled as random step functions: After a random time, which occupies some part of the nondecision time, T_{er}, the drift and diffusion rates change from zero to nonzero values and remain constant for the duration of the trial. This abrupt-onset assumption suffices to model the data from many decision tasks, including our numerosity and spatial cuing tasks (Ratcliff, 2008; Smith et al., 2004), especially if the nondecision time is allowed to vary randomly across trials (Ratcliff, 2002) because it can accommodate non-abruptness in the stimulus onset if the time scale is not too long. The same abrupt-onset assumption has been made in theoretical and empirical treatments of changing-information paradigms (Carland et al., 2015; Thura, 2016; Zhang and Bogacz, 2010) and was made by Trueblood et al. (2021) in modeling the flashing-grid task, although they did not include nondecision variability. However, it is not self-evident that the abrupt-onset assumption is an appropriate one for tasks of this kind.

Stimulus information in the flashing-grid task is carried by the proportion of blue and orange squares in grids presented at a rate of 20 Hz. The decision tasks this task most resembles are dynamic noise tasks, in which stimulus information is perturbed by dynamic, external noise (Ratcliff & Smith, 2010; Smith, Ratcliff, & Sewell, 2012). Performance in these tasks is often not well described by the diffusion model in its standard form (Ratcliff & Smith, 2010). Instead, it is better characterized by a model in which the drift and diffusion rates grow smoothly to an asymptote over several hundred milliseconds (Smith et al., 2012). The modeling results agree with the perceptual experience of doing the task, in which the stimuli

(letters, bars, gratings, etc.) appear to emerge progressively from the noise over a period of around half a second or so.

The most widely studied dynamic noise task is the random dot motion (RDM) task, in which participants identify the direction of coherent motion in clouds of randomly-moving dots. Psychophysical studies using classical temporal-integration paradigms suggest that these kinds of tasks may have much longer perceptual integration times than simple stimuli like spots of light or gratings. Watamaniuk and Sekular (1992) obtained threshold-versus-duration functions for the RDM task, in which they measured the level of motion coherence needed to achieve a criterion level of accuracy for different exposure durations, and obtained an integration time of 400–450 ms. The integration time was the same for high and low coherence stimuli, suggesting it reflected perceptual rather than decisional integration. Smith and Lilburn (2020) fit the choice probabilities and RT distributions from an RDM task reported by Dutilh et al. (2019) with a time-varying diffusion model, in which drift and diffusion rates grow smoothly to an asymptote, which they estimated to be at around 400 ms after stimulus onset, in agreement with the temporal integration times estimated by Watamaniuk and Sekuler. They found that a gradual-onset model provided a better overall fit than an abrupt-onset one and showed fewer violations of assumptions about how model parameters should vary with experimental conditions.

It is not the case, however, that all dynamic noise tasks have long perceptual integration times. Ratcliff and Smith (2010) reported data from a brightness discrimination task in which participants made judgments about the proportions of black and white pixels in 60 Hz dynamic noise arrays. The data from this task were well-described by the standard diffusion model with abrupt-onset drift and diffusion rates. Ratcliff, Voskuilen, and Teodorescu (2018) studied a version of the task that used larger stimulus elements, which required comparison of the brightness of pairs of 15×15 grids of four-pixel black and white squares presented at 60 Hz, which was also well-described by the standard diffusion model. Although the brightness discrimination tasks resemble the flashing-grid task in that the discriminative information in both tasks is carried by global rather than local features of the display, the spatial, temporal, and chromatic properties of the stimuli in the tasks make them very different perceptually. Specifically, they differ in the size of the individual stimulus elements, the presentation rate (60 Hz vs. 20 Hz), and in whether the stimulus information is encoded perceptually by luminance or color channels. The latter distinction is relevant because the color system is slower than the luminance system. (The stimuli in the flashing grid task are not constrained to be isoluminant and show pronounced 20 Hz luminance flicker as a result, but it is unlikely that the presence of correlated color and luminance changes contributes to the way in which the task is performed because color and luminance are processed at different rates.) It is not clear from viewing the stimuli whether the information in them is integrated perceptually over successive frames or not, so it is hard to determine a priori what the time scale of drift rate computations should be. If there is no perceptual, as distinct from decisional, integration across successive frames, then the abrupt-onset assumption should suffice, but if there is perceptual integration across frames, then a gradual-onset model, like the one of Smith et al. (2012) and Smith and Lilburn (2020) may be more appropriate.

One of our aims was to compare the extended urgency-gating model of Trueblood et al. (2021) to a version of the diffusion model with time-varying drift and diffusion rates. To this end, we used a variant of the model of Smith and Lilburn (2020), which represents the evidence entering the decision process as the output of a linear filter, composed of a cascade of exponential stages (Watson, 1986). The model is loosely based on the sustained-plus-transient channel diffusion model of Smith (1995) and the integrated system model of Smith and Ratcliff (2009), which were both motivated by the classical literature on visual temporal sensitivity (de Lange, 1952, 1954, 1958; Kelly, 1961, 1969; Roufs, 1972, 1974; Sperling & Sonhdi, 1968; Watson & Nachmias, 1977). Exponential filter models of perceptual processing are commonly used in the temporal sensitivity literature to model temporal integration, pulse-pair summation, flicker-fusion perception, and related perceptual phenomena. Like the filter in the model of Cisek et al., the exponential filter cascade in Smith and Lilburn's model endows the system with low-pass filter characteristics whose effect is to remove sharp transients from the input. Unlike Cisek's model, however, the low-pass filter is associated with perceptual rather than decision processes: it affects the way in which stimuli are perceived rather than the way in which perceptual information is integrated. The main effect of identifying the low-pass filtering operation with perceptual rather than decisional processes is that, unlike Cisek's model, the evidence in the decision process does not grow to a stationary distribution. In his model, the low-pass filtering operation is identified with the decay term in an OU process; in the model we consider here it is instead identified with time-varying drift and diffusion rates in a Wiener process.

Mathematically, we assumed that the drift and diffusion rates of a stimulus whose amplitude changes at time t_0 from of v_1 to v_2 (where the signs of the amplitudes encode the stimulus identity before and after the change) can be represented by a function of the form

$$\Theta(t) = v_1\theta(t) + (v_2 - v_1)\theta(t - t_0), \tag{21}$$

where

$$\theta(t) = \frac{1}{\Gamma(n)}\int_0^{\beta t} e^{-s}s^{n-1}ds; \ t \geq 0 \tag{22}$$

is the incomplete gamma function (Abramowitz & Stegun, 1965; p, 260, Equation 6.5.1).

Equation 21 describes the output of a linear system with impulse response function $d\theta/dt$ to a step-change stimulus whose identity changes from v_1 to v_2 at time t_0. In these equations β specifies the encoding rate and has units of encoding strength per unit time. An example of the encoding function is shown in Figure 2. Larger values of β lead to sharper encoding functions that more closely resemble the step-change profile of the stimulus.

When implementing a time-inhomogeneous diffusion model based on Equation 21, there are two possible scaling relationships for the infinitesimal moments of the process. Assuming a process in which $A(x, t) = \mu(t)$ and $B(x, t) = \sigma^2(t)$ in Equation 1 (i.e., a time-inhomogeneous Wiener process), both scalings assume that drift rate grows in proportion to the output of the perceptual encoding process, $\mu(t) \propto \Theta(t)$, but they differ in their assumptions about the diffusion rate. One scaling assumes that the diffusion rate grows in proportion to the

underlying temporal encoding function $\sigma^2(t) \propto \theta(t)$; the other assumes that the infinitesimal standard deviation grows in proportion to it, $\sigma(t) \propto \theta(t)$. In either case we assume that, asymptotically, $\sigma(t) \to 0.1$, in order to obtain an identifiable model whose scaling is compatible with the standard diffusion model.

Smith and Lilburn (2020) followed Smith and Ratcliff (2009) and Smith et al. (2012) and assumed the first form, which they augmented with an additional source of constant diffusion noise to capture the tendency to make fast errors, while Ditterich (2006a, 2006b) assumed the second form. In the second form, the variance of the evidence entering the decision process grows more rapidly than does its mean (as, indeed, is also the case for the OU process), which again leads to a tendency to make fast errors. This kind of scaling relationship is plausible if the diffusion process reflects the mass-action properties of underlying neural processes. An example is the Poisson shot-noise model of Smith (2010), in which the drift rate depends on the difference between pairs of excitatory and inhibitory shot noise processes and the diffusion rate depends on their sum. The diffusion rate grows more rapidly than the drift rate as a result. We assumed the second scaling relation on pragmatic grounds, to capture any tendency to make fast errors without the need for an additional constant-diffusion parameter like the one in Smith and Lilburn's model. In our implementation we assumed that

$$\mu(t) = \Theta(t) \tag{23}$$

$$\sigma(t) = \sigma_{\mathrm{asy}}\, \theta(t), \tag{24}$$

with $\sigma_{\mathrm{asy}} = 0.1$, where the parameters v_1 and v_2 of $\Theta(t)$ vary with the stimulus condition. Equations 23 and 24 state that the diffusion rate grows smoothly to a constant asymptote while the drift rate grows to an asymptote that changes as the magnitude and identity of the stimulus changes.

Results

In the standard diffusion model, within-trial noise in evidence accumulation is augmented with three sources of across-trial noise or variability: in drift rate, in starting point, and in nondecision time (Ratcliff & McKoon, 2008). Drift rate is normally distributed with mean v and standard deviation η; starting point is uniformly distributed with range s_z, and the nondecision time, T_{er}, is uniformly distributed with range s_t. Variability in drift rate and starting point allow the model to predict the ordering of RTs for correct responses and errors, while variability in nondecision time allows it to better capture the shapes of RT distributions when accuracy is high and responses are fast. In general, RTs for errors tend to be longer than RTs for correct responses when stimulus discriminability is low and accuracy is stressed and shorter when discriminability is high and speed is stressed (Luce, 1986). Variability in drift rate and starting point allow the model to predict slow errors and fast errors, respectively.

The inclusion of across-trial variability in the diffusion model has sometimes been criticized, especially in neuroscience, where it has been argued that mechanisms like collapsing

decision boundaries provide an alternative way to predict slow errors, although none of the proposed alternatives has been shown to predict the full pattern of RT orderings that are found experimentally (Ratcliff & Smith, 2004; Ratcliff & McKoon, 2008). In our model evaluation, we compared three versions of the urgency-gating model, the pure urgency model with $U(t) = mt$ (Thura et al., 2012), the extended urgency-gating model with $U(t) = b + mt$ (Carland et al., 2015, 2016), and an extended model in which the urgency rate, m, was allowed to vary with speed versus accuracy instructions. This last model tested a natural prediction from the theory that urgency should increase under speed instructions. For the flashing-grid task, we also considered a time-varying diffusion model, with drift and diffusion rates as described in the preceding section. The parameters of the models and their interpretation are listed in Table 1. We included drift-rate variability in the diffusion model for all three tasks and starting-point variability for the numerosity task, in which there was a speed-accuracy manipulation. We included variability in nondecision time in both diffusion and urgency-gating models. Nondecision time variability is not a feature of published urgency-gating models but we included it in our model comparisons in order to make them as fair as possible. Trueblood et al. (2021), did not include across-trial variability in their models.

Fitting Methods

There are differences in opinion among modelers about how best to fit models to RT data, particularly in relation to classical versus Bayesian estimation and hierarchical versus nonhierarchical fitting methods. (See the diversity of methods used in the blinded validity study of Dutilh et al. (2019) for example.) We chose to use methods that were similar to those used in the original studies of Ratcliff (2008) and Smith et al. (2004). We minimized the likelihood-ratio chi-square statistic (G^2) for the response proportions in the bins formed by the .1, .3, .5, .7, and .9 RT quantiles for the distributions of correct responses and errors. When bins are formed in this way, there are a total of 12 bins (11 degrees of freedom) in each pair of joint distributions of correct responses and errors. The resulting G^2 statistic can be written as

$$G^2 = 2 \sum_{i=1}^{M} n_i \sum_{j=1}^{12} p_{ij} \log\left(\frac{p_{ij}}{\pi_{ij}}\right).$$

In this equation, p_{ij} and π_{ij} are, respectively, the observed and predicted proportions in the bins bounded by the quantiles and "log" is the natural logarithm. The inner summation over j extends over the 12 bins formed by each pair of joint distributions of correct responses and errors. The outer summation over i extends over the M experimental conditions. For the numerosity study, $M = 16$ (2 instruction conditions $\times$ 8 dot proportions). For the cuing study, $M = 20$ (masked/unmasked $\times$ cued/uncued $\times$ 5 contrasts). The quantity n_i is the number of experimental trials in each condition. For the numerosity study, $n_i \approx 850$ and for the cuing study, $n_i = 400$. We fit the models to the individual participants' data by minimizing G^2 using the Nelder-Mead simplex algorithm (Nelder & Mead, 1965) as implemented in Matlab (fminsearch). The fit statistics we report are the minimum G^2 values obtained from six runs of simplex using randomly-perturbed estimates from the preceding run as the starting

point for the next run. Ratcliff and Childers (2015) showed that minimum chi-square fits to individual participant data yielded good parameter recovery in large samples like those we fit here.

To compare models with different numbers of parameters, we used standard model selection methods based on the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC, Schwarz, 1978). The first of these statistics is derived from classical principles whereas the second is Bayesian, but we use them in the spirit in which they are typically used in the modeling literature, as penalized likelihood statistics that impose more or less severe penalties on the number of free parameters in a model. As is well known, the AIC tends to gravitate towards more complex models with increasing sample sizes more quickly than does the BIC (Kass & Raftery, 1995), although for the large samples we used here they were in close agreement. For binned data, the AIC and BIC may be written as

$$
\begin{aligned}
AIC &= G^2 + 2p \\
BIC &= G^2 + p\log N,
\end{aligned}
$$

where p is the number of free parameters in the model and $N = \sum_i n_i$ is the total number of observations on which the fit statistic was based.

Numerosity Study (Ratcliff, 2008)

Table 2 lists the four models we compared using the data from the Ratcliff (2008) numerosity study, together with their identifying parameters. These are the parameters that distinguished a model from the other models under comparison. All of the models had a common set of mean drift rates, decision boundaries, starting points, and nondecision times, as shown at the bottom of the table. We treated the diffusion model, with across-trial in variability in drift rate, η, starting point, s_z, and nondecision time, s_t, as the reference model, and compared three versions of the urgency-gating model to it. All of the urgency-gating models had OU decay, γ, and, to make them comparable to the diffusion model in their assumptions about nondecision processes, we included nondecision time variability, s_t, in all models. (Although all models had nondecision time variability, we include it in the notation to better reflect the model semantics.) In the model in which the urgency rate was permitted to vary with instructions, the rate parameters in the speed and accuracy conditions are denoted m_s and m_a.

With both forms of urgency function, the number of identifiable parameters in the model is one fewer than the number of parameters in the function $U(t)$. It is common practice to treat the infinitesimal standard deviation as a scaling parameter in diffusion models and to fix it to an arbitrary value because the parameters of the model are identified only to the level of a ratio. We set $\sigma = 0.1$, which is the most common scaling convention in the literature. For the urgency models the infinitesimal standard deviation is either $\sqrt{B(t)} = mt\sigma$ or $\sqrt{B(t)} = (b + mt)\sigma$ (Equation 12) and in either case one of the parameters may be eliminated by redefining (rescaling) σ (cf. Evans, Trueblood, & Holmes, 2020). For the pure urgency model we set $m = 1$ and for the extended urgency model we set $b = 1$.

Otherwise, with one exception, we parameterized the models in the same way as in Ratcliff's (2008) study. There were separate boundary separations and starting point parameters for speed and accuracy conditions, a_s, a_a, z_s, z_a. There was a single parameter for each of drift rate variability, η, starting point variability, s_z, nondecision time, T_{er}, and nondecision time variability, s_t. Ratcliff allowed mean drift rate to vary freely for the eight dot-numerosity conditions and fit the data from the standard RT task and the response-signal task simultaneously using the same mean drift rates. We found that the standard RT task could be well fit with a symmetrical drift rate model, $v_i = -v_{9-i}$, in which the drift rates for conditions with less than 50 dots were mirror images of those for conditions with more than 50 dots. This reduced the number of mean drift rates from eight to four. The estimates in Ratcliff's Table 4 when mean drift rates were free to vary across conditions show a similar symmetry.

Table 3 shows the fit statistics and Table 4 shows the parameter estimates for the four models, averaged across the fits to the individual participants. The columns #AIC and #BIC in Table 3 are the numbers of participants for whom the given row model was preferred to the diffusion model according to either the AIC or BIC. The fits to the diffusion model are similar to those reported by Ratcliff (2008). (Ratcliff used Pearson χ^2, which, like G^2, is distributed asymptotically as a chi-square random variable under independent, multinomial sampling assumptions. The averaged individual participant fits in his Table 2 are combined χ^2 fits for the RT task and the response-signal task.) A quantile probability plot of the fit of the diffusion model is shown in Figure 3. These plots show the quantiles of the RT distributions for correct and errors, plotted against the choice probabilities, for a range of stimulus discriminabilities. Readers who are unfamiliar with this way of representing model fits are referred to Ratcliff and Smith (2004) or Ratcliff and McKoon (2008), among other sources. The data in Figure 3 are quantile-averaged group data and the fitted values are quantile-averaged individual fits. Although there were some minor procedural differences between our treatment of the data and Ratcliff's, the two sets of fits are in close agreement.

The second model in Table 3 is the pure urgency-gating model, UG(γ, m, s_t) with $m = 1$. The bottom panels of Figure 3 show a quantile probability plot of the fit of this model. Qualitatively and quantitatively, the pure urgency-gating model fares badly. The average G^2 is more than double that for the diffusion model and the fitted model fails to capture the shape of the quantile-probability functions. In general, a quantile probability function that is canted upwards towards the left indicates a slow-error pattern, while one that is canted downwards towards the left indicates a fast-error pattern. Sometimes both can be present in the same data set (see Ratcliff and Smith (2004), Figure 7, for an example). The quantile probability plot for a data set in which correct and error RTs are the same will be symmetrical across its vertical midline. The plot for the urgency-gating model shows that it predicts a strong slow-error pattern, but the quantile probability functions it predicts are insufficiently bowed to match the data, and, unlike the diffusion model, it has no mechanism to predict the fast errors in the speed-instructions condition (see below). The failure of the model confirms the conclusions of Winkel et al. (2014), who used a task in which stimulus information changed over time. As noted above, Winkel's analysis was criticized by Carland et al. (2015) for not including low-pass filtering (OU decay), but we came to the same conclusions as Winkel et al. with OU decay in the model. Indeed, the estimate in Table 3, γ

= 0.012, shows that when decay was allowed to vary freely, on average it approached zero, which is a pure Wiener process, in agreement with the earlier comparison of the OU and Wiener models by Ratcliff and Smith (2004).

In comparison, the extended urgency-gating model UG(γ, b, m, s_t) with $b = 1$ fared much better. Although it was preferred to the diffusion model for only 5 of the 19 participants by either the AIC or the BIC, quantitatively and qualitatively its fit was fairly similar to the diffusion model. Again, however, the estimate of $\gamma = 0.780$ in Table 4 suggests that OU decay made a negligible contribution to the fit. We have not shown a full quantile probability plot of the fitted model because the differences between it and the diffusion model are not easy to discern in the plot. Instead, in Figure 4a we have shown the empirical and predicted 0.1 RT quantiles for the diffusion model and the urgency-gating models. The 0.1 RT quantile characterizes the fastest responses (the leading edge) in the distribution, and is of theoretical interest when speed versus accuracy is manipulated because the fast-error pattern under speed stress is often evident in the leading edge. The figure shows that the 0.1 quantile function under accuracy instructions is bowed, but relatively symmetrical, whereas the 0.1 quantile function under speed instructions is bowed downward to the left, which is the fast-error pattern.

The figure makes clear that the three models make qualitatively different predictions for the 0.1 quantile function. The diffusion model, which has a mechanism for predicting both fast and slow errors fares the best: It predicts both the symmetrical, bowed function under accuracy instructions and the downward bow on the left under speed instructions. The pure urgency-gating model fares the worst: It predicts a monotonic 0.1 quantile function that does not correspond to the pattern in the data. The performance of the extended urgency-gating model was better. Under accuracy conditions, it predicts a symmetrical function that is almost indistinguishable from the one predicted by the diffusion model, but under speed instructions it also predicts a symmetrical function that again does not correspond to the pattern in the data. This error in prediction is unsurprising because, unlike the diffusion model, neither urgency-gating model has a mechanism for predicting fast errors.

The final model in the table is an extended urgency-gating model, UG(γ, b, m_s, m_a, s_t), in which the urgency rate varies with experimental instructions. The average G^2, AIC, and BIC for this model are the smallest of the four models, although it was preferred to the diffusion model for only six of the 19 participants. (The data for two of the participants were poorly fit by both models and for them the G^2 for the diffusion model was around 65% worse than for the urgency-gating model, inflating the model average.) Like the other versions of the urgency-gating model, the averaged OU decay, $\gamma = 0.861$, suggests that decay contributed to the fits in only a minor way. Although both versions of the extended urgency-gating model performed well numerically, the estimated parameters of the model suggest that much of its success derives from the trading-off of parameters discussed earlier. Evans, Trueblood, and Holmes (2020) investigated the identifiability of parameters in the urgency-gating model in a simulation study and found its parameters could be recovered accurately, but they considered only the simpler pure urgency model. The extended form of the model is much more flexible and, we suspect, for the reasons given below, its parameters are less well identified.

Figure 4b shows histograms of the estimates of γ and m for the individual participants for the extended urgency-gating model, UG(γ, b, m, s_t), Although the most common estimate of γ was close to zero, individual estimates varied on a range from zero to almost 5.0, which, according to the model, implies that the time constant in the low-pass filter for individual participants varied from 200 ms to infinity (no decay). (All of the time parameters in our models are expressed in seconds, so $\gamma = 5.0$ implies that the average evidence strength is equal to 63% of the mean of the stationary distribution by 0.2 s.) It is hard to know how to interpret this variability if low-pass filtering is viewed as a hard-wired property of the cognitive system. In addition, the distribution of m was strongly bimodal across participants. We imposed a penalized upper bound of 2.0 on m to improve convergence and, as can be seen in the figure, the estimates tended to cluster towards either the upper or lower bounds. The scatterplot shows that virtually all participants had either near-zero urgency and nonzero decay or vice versa. These estimates suggest that urgency and decay are being traded off significantly in the fits, as would be expected if the parameters are not well-identified and the model were overfitting the data. Overfitting is a likely consequence of the greater flexibility of the urgency-gating model. The diffusion model predicts RT distributions that resemble those found in the majority of experimental tasks and predicts only those distributions, but the urgency-gating model predicts distributions that are both more skewed and less skewed than these, as Figure 1 shows.

A second reason for thinking that the model is overfitting comes from the estimates of the rate parameters in the model UG(γ, b, m_s, m_a, s_t) under speed and accuracy instructions, which were $m_s = 0.182$ and $m_a = 0.923$, respectively. At a group level, the difference in these estimates is significant by a (classical) t-test: $t(18) = -3.31$, $p = 0.004$, but the ordering is the opposite of what would be expected from the model semantics, which would predict increased urgency under speed stress conditions. These differences reinforce the impression that some part of the fit of the urgency-gating models is due to parameter tradeoffs rather than to them capturing unique structure in the data.

Spatial Cuing Study (Smith et al., 2004)

Although the Smith et al. (2004) task did not include a speed versus accuracy manipulation, the effects of decay and urgency should nevertheless be identifiable in the data from this task if the urgency-gating model is true. A core claim of the urgency model is that the evidence entering the decision process quickly becomes statistically stationary, implying it is subject to large OU decay. Decay should therefore be a hard-wired property of the cognitive system and be independent of speed versus accuracy instructions. Urgency should similarly be present under both forms of instruction, because large decay without urgency leads to RT distributions that do not resemble those found in data, as Ratcliff and Smith (2004) showed and Figure 1 makes clear.

Because there was no speed versus accuracy manipulation in the Smith et al. (2004) study, the models are simpler than those for the Ratcliff et al. (2008) study because the starting point variability parameter could be omitted without worsening the fit. Also, the speed and accuracy of decisions to vertical and horizontal grating stimuli were sufficiently similar that they could be pooled to obtain a single distribution of correct responses and a single

distribution of errors for each stimulus condition. This symmetry implies a symmetry constraint on the starting point, $z = a/2$. Smith et al. compared an unconstrained model, with a separate submodel for each cell of the Cue × Mask design, to an attention orienting model, which prescribed a relationship between the nondecision times, T_{er}, and the mean drift rates, ν_i, across conditions. The theoretical substance of the orienting model was that cuing shortens nondecision times because it allows attention to be focused on the target location prior to its presentation. For masked stimuli, cuing also leads to higher drift rates because it allows more stimulus information to be extracted from the display before the perceptual representation is suppressed by the mask. Cuing has no effect on the drift rates for unmasked stimuli because of their greater visual persistence. As a result, the quality of information that can be extracted from the display is unaffected by whether or not attention is focused on the target location prior to stimulus onset. Because we are interested in the task as a decision task rather than an attention task, we refit the unconstrained model, with a separate submodel for each cell of the Cue × Mask design for each of the six participants. We fit the same models as for the Ratcliff (2008) study other than the model with instruction-dependent urgency, as summarized in Table 5. The fit statistics we report are the sums of the G^2, AIC, and BIC values and the degrees of freedom for the submodels for the four cells of the design.

One feature of the results of the Smith et al. (2004) study should be highlighted here. They found, as Ratcliff and Rouder (2000) found in an earlier study, that the RT distributions and choice probabilities were well fit by a diffusion model in which the drift and diffusion rates remained constant throughout a trial, even though the stimuli were physically present for only 60 ms. Like Ratcliff and Rouder, nothing in their data suggested that the evidence entering the decision process decays after stimulus offset, as a simple OU model would predict. Smith et al. interpreted their findings as showing that the decision process is driven by stable stimulus representations in visual short-term memory (VSTM) that preserve stimulus information without degradation for the second or so needed to make a decision about it. Smith and Ratcliff (2009) subsequently incorporated a linear-system model of VSTM encoding into their integrated system model of decision making to account for the time course of VSTM formation. The model assumes that people form a time-dependent perceptual representation of the stimulus, as described by Equation 21, which drives formation of the VSTM trace while the stimulus is physically present and which stops changing once the stimulus is extinguished. As well as casting doubt on the need for decay, their findings highlight the importance of theorizing about perceptual and VSTM processes if we wish to understand decision-making in tasks of this kind.

Table 6 shows the average fit statistics for the individual participants for the models and Table 7 shows the estimated model parameters averaged across the four conditions in the Cue × Mask design and across participants. As in Table 3, the columns #AIC and #BIC are the numbers of participants for whom one of the urgency-gating models was preferred to the diffusion model according to each of the two criteria. The upper panels of Figure 5 show a quantile probability plot for performance in the task and the predictions of the diffusion model. The lower panels show the predictions for the best-fitting urgency model, UG(γ, b, m, s_t).

The fit statistics in Table 6 and the plot in Figure 5 show that the diffusion model, DIFF(η, s_t), with across-trial variability in drift rate and nondecision time, provides a good account of the choice probabilities and RT distributions in the task. Differences in the shapes of the quantile probability functions and the range of accuracy values in the four conditions are accounted for by differences in the mean drift rates, v_{ji}, $j = 1, \ldots, 5$, nondecision times, $T_{\mathrm{er},i}$, and drift rate standard deviations, η_i. The parameter estimates in Table 7 are averages across the four cells of the design. Readers who are interested in how the parameters vary across cue and mask conditions are referred to Table 2 of the original article.

As in the numerosity study, the pure urgency-gating model, UG(γ, m, s_t), fared badly, for similar reasons as there. Like the fit to those data, the model was unable to capture the shapes of the RT distributions across the four conditions, and was not preferred to the diffusion model for any of the participants by either the AIC or BIC. As in Table 4, the estimate of decay, $\gamma = 0.027$, suggests that OU decay contributed almost nothing to these fits. These results confirm the findings of Winkel et el. (2014) and the majority of the fits reported by Hawkins et al. (2015) in showing that the pure urgency-gating is in general not an appropriate model for these kinds of data.

The extended urgency-gating model, UG(γ, b, m, s_t), performed better, although again, the estimated decay of $\gamma = 2.307$ implies that it contributed to the fit in only a minor way. A decay of 2.5 corresponds to a low-pass filter time constant of 400 ms, which is materially longer than the a priori range of values stipulated by Carland et al. (2015, 2016). The average G^2 was around 34% worse than that for the diffusion model and the model was preferred to the diffusion model for only two of the six participants according to the AIC and the BIC. Although urgency-gating provides the model with a mechanism for predicting slow errors, the quantile probability plot suggests that it does not predict the slow error pattern in the data as well as does the diffusion model with across-trial variability in drift rate. As in the numerosity study, there was evidence of substantial variation and tradeoffs in the estimates of γ and m. Figure 6 shows histograms and a scatterplot of γ and m for the 24 estimates (four cells of the design for the six participants). As in Figure 4, most of the estimates of γ clustered around the lower bound of zero (i.e., the Wiener process). Although there was not the same bimodality in the distribution of m as in Figure 4b, many of the estimates clustered around the upper bound of 2.0. Across the 24 estimates, the correlation in γ and m was $r = -0.435$, $t(22) = -2.266$, $p = 0.033$. As in the numerosity task, this significant correlation appears to be a reflection of the greater flexibility of the urgency-gating model, which allows it to predict similar patterns of performance using different combinations of urgency and decay parameters.

Flashing-Grid Task

Table 8 summarizes the models for the flashing grid task. For this task, we compared two versions of the diffusion model and two versions of the urgency-gating model. For both models, we compared more and less restricted versions of them. For the standard diffusion model, the models were DIFF(η, s_t) and DIFF(β, η, s_t). The first model assumes abrupt-onset drift and diffusion rates; the second assumes rates that increase according to Equations 23 and 24. For the urgency-gating model the two models were UG(γ, b, m, s_t)

and UG(b, m, s_t). These models compare the effects of urgency, with and without decay. To ensure the models were comparable, we implemented the abrupt-onset models using the same code as used for the time-varying diffusion model with a large, fixed rate constant, $\beta = 100$. This value of β represents a perceptual integration time that is within the 100 ms Bloch's law critical duration that characterizes the majority of perceptual tasks (Bloch, 1885; Gorea, 2015; Smith, 1998; Smith & Lilburn, 2020; Watson, 1986).

Like Trueblood et al. (2021), we allowed the drift rates to differ for constant and changing stimulus conditions and, for the changing conditions, to differ for the early (t .35 s) and later ($t > .35$ s) portions of the stimulus presentation interval (i.e., four drift rates in all). To characterize the differences in the speed and accuracy of the two responses shown by some participants, we allowed the starting point for evidence accumulation, z, to vary and the drift rates for the two stimuli (orange vs. blue or black vs. white) to differ in magnitude. Instead of assuming that the drift rates were equal in magnitude and opposite in sign, we allowed them to depend on a drift criterion, c, such that $v_{A,i} = -v_{B,i} - c$, $i = 1, \ldots 4$ (Ratcliff, 1985; Ratcliff & Smith, 2004).

Table 9 summarizes the fit statistics for the four models. Instead of comparing all of the other models to the standard diffusion model, as we did for the previous experiments, we compared the more and the less restrictive versions of the diffusion and urgency-gating models to each other (DIFF(η, s_t) vs. DIFF(β, η, s_t) and UG(γ, b, m, s_t) vs. UG(b, m, s_t)) and compared the most general diffusion and urgency-gating models to each other (DIFF(β, η, s_t) vs. UG(γ, b, m, s_t)). The number of participants favored by these model comparisons are tabulated as #AIC and #BIC. As well as comparing models via the AIC and the BIC, we compared the most general models using a parameteric bootstrap cross-validation procedure (Wagenmakers et al., 2004; Voskuilen et al., 2016) in which we cross-fit the models DIFF(β, η, s_t), UG(γ, b, m, s_t) to 50 sets of simulated data generated by each of the two models, using the average of the estimated parameters from the fits to the individual participants to generate the simulations. Figure 7 shows kernel density estimates of the distribution of G^2[DIFF(β, η, s_t)] $- G^2$[UG(γ, b, m, s_t)], the difference in the G^2 (or AIC or BIC) statistics for the two models when the model generating the data was the diffusion model or the urgency-gating model. Using the point at which the two density functions cross each other to classify the models maximizes classification accuracy. For all three data sets the crossover point was positive, indicating that the urgency gating model is somewhat more flexible than the diffusion model (i.e., produces smaller G^2 values). For the three data sets, the crossover points were: Trueblood et. al, 5.35; Color, 5.14; Luminance, 4.03. The column labeled #BOOT shows the number of participants for whom the model DIFF(β, η, s_t) or UG(γ, b, m, s_t) was preferred, using the crossover point as the classification criterion. According to the BIC, 24 of the 44 participants for the three experiments favored the urgency-gating model, but according to the parametric bootstrap, 26 of the 44 favored the diffusion model.

Figure 8 shows histograms and a scatterplot of the γ and m parameters for the model UG(γ, b, m, s_t) for the three experiments. Both γ and m have bimodal frequency distributions with estimates that cluster at the extreme ends of their ranges. The joint distribution of (m, γ) in the scatterplot is fairly uniform across the range and lacks an identifiable mode of the kind that would be expected if the urgency-gating model were true and the individual

parameter estimates were sampled from a joint distribution with nonzero marginal means. The distributions in Figure 8 reinforce the impression from the earlier experiments that the good fit of the urgency gating model is more a reflection of its flexibility than of essential structure in the data that the model is capturing. Contrary to the claims of Thura (2016), changing-stimulus conditions do not seem to provide a reliable way to distinguish between the urgency-gating model and the diffusion model. Indeed, our fits suggest the converse: The models were better distinguished in tasks with fixed stimuli.

Like Trueblood et al. (2021), we found the drift rate parameters, ν_i, varied as a function of whether the stimulus information was constant or changed during a trial. As shown in Table 10, in most cases the magnitudes of the estimated drift rates were reduced if they were preceded by 350 ms of the other stimulus, i.e., $|\nu_3| < |\nu_2|$ and $|\nu_4| < |\nu_1|$, consistent with integration or averaging across stimulus-change boundaries. This was so for both the diffusion model with time-varying perceptual integration, DIFF(β, η, s_t), and the remaining three models that assumed rapid changes in drift rate after a stimulus change.

Figure 9 shows quantile probability plots for the fits of the models DIFF(β, η, s_t) and UG(γ, b, m, s_t) for the three experiments. The plots show that the diffusion model and the urgency-gating model both capture the fairly challenging pattern of RT distributions in these experiments, which show substantial changes in the leading edges (the 0.1 quantiles) as a function of the stimulus condition, as well as in the distribution tails (the 0.7 and 0.9 quantiles). There is also a consistent slow error pattern, although its effects are somewhat masked by the large differences between stimulus conditions. Trueblood et al. (2021) evaluated the qualitative performance of the models at the level of mean RT and conditional and unconditional response accuracy only, but the recent psychological literature has emphasized that models should be able to account for entire distributions of correct responses and errors. Our model fits show that the diffusion model and urgency-gating model provide satisfactory accounts of the RT distributions in data that has been screened for fast guesses. Although we fit the data for only 12 of the 34 participants in Trueblood et al.'s Experiment 1, the main patterns in their data were replicated in our color task, which had a manipulation to discourage fast guessing.

The other fits in Table 9 compare diffusion models with and without time-varying encoding and urgency models with and without OU decay. There was only equivocal support for time-varying stimulus encoding in the diffusion model, unlike the findings for the random dot motion task reported by Smith and Lilburn (2020). There were a similar numbers of participants for whom the time-varying and abrupt-onset diffusion models were preferred for all three experiments, although the numbers depend on whether the AIC or BIC is used because the models have different numbers of parameters. The estimated β parameters in Table 10 correspond to perceptual integration times between 100 ms and 200 ms, which is shorter than the 400 ms estimate for the random dot motion task reported by Smith and Lilburn (2020) and, at its lower bound, is consistent with a perceptual integration process governed by Bloch's law. We conjectured that the perceptual integration times in the color task might be longer than in the luminance task because of the slower perceptual response of the color system and, while the ordering of the β parameters is consistent with this idea, the

difference in the estimates from the two experiments using the color task is greater than the difference between the color and luminance tasks.

The comparison of the two versions of the urgency-gating model in Table 9 is much clearer: For all three data sets, the model without decay was preferred. This is especially so if the BIC is used. Using the BIC, the number of participants for whom $UG(b, m, s_t)$ was preferred to $UG(\gamma, b, m, s_t)$ were: Trueblood et al., 8/12; Color, 14/16; Luminance 13/16. Like the numerosity and spatial cuing experiments, then, there is no strong evidence of decay during the evidence accumulation period. Rather, the estimated average decay parameters of $\gamma = 3.94$, 2.56, and 4.71 suggest that any decay that is present is at most moderate in magnitude, consistent with the results of the comparison of the OU and Wiener diffusion models reported by Ratcliff and Smith (2004).

Discussion

Our use of the integral equation method allowed us to derive explicit expressions for the first-passage time densities for urgency-gating and collapsing boundary models and to show precisely how and when the two kinds of models are equivalent. We have shown, in an explicit, formal way, that the conditions under which the models appear intuitively to be equivalent hold rigorously. Specifically, an urgency-gating model with urgency function $U(t)$ is equivalent to a model without urgency in which the decision boundaries converge in inverse proportion to $U(t)$. Although our results show that the mathematics and intuition agree with one another, the formal justification is essential and cannot be omitted. The recent literature on decision processes has shown the hazards of trying to reason in an intuitive way about the first-passage time distributions of stochastic processes as if they were deterministic functions. The results we have presented apply both to the Wiener process, which is used to model accumulation in the diffusion model, and the OU process, which can be interpreted either as a leaky accumulator or as a low-pass filtered evidence process. As well providing a characterization of the equivalence of the two kinds of models, the integral equation method provides a method for fitting the models to data that avoids the need for recourse to Monte Carlo simulation.

Fixed-Stimulus Experiments

Our model fits using fixed stimuli clarify and extend the results of previous studies. We found the pure urgency-gating model, with $U(t) = mt$, performed worse than the diffusion model. This confirms the results of studies using both constant (Voskuilen et al., 2016) and changing stimulus information (Winkel et al., 2014). Winkel's results were criticized by Carland et al. (2015) because his models did not include OU decay (low-pass filtering), but our estimates of γ show that the contributions of decay to the fit of the model were negligible. Other researchers, such as Hawkins et al. (2015) and Evans, Hawkins, and Brown (2020) have investigated similar models and found support for them only under speed-stress conditions, in which it is likely that participants were deadlining.

In comparison, the extended urgency-gating model, with $U(t) = b + mt$, performed better. This model can be interpreted as adding variable amounts of urgency to an OU evidence process, which allows it to predict a larger range of distribution shapes than does the model

in its pure form, which more closely correspond to what is found in data. Although the model in its extended form captures many of the features of the RT distributions and choice probabilities, for the constant-stimulus experiments it did not do so as well as the diffusion model with across-trial variability. In the numerosity data of Ratcliff (2008) the model lacks a mechanism for predicting fast errors and in the spatial cuing study of Smith et al. (2004) it did not capture the empirical pattern of slow errors as well as did the diffusion model with drift-rate variability. In addition, the variability in the estimates of urgency and decay and the correlations between them suggests that these parameters are not well identified in data. Changes in these parameters change the shapes of the predicted RT distributions and, as shown in Figure 1, these changes can be traded off against other. We would therefore expect identification of the associated parameters to be challenging. Moreover, when we allowed the level of urgency to vary with experimental instructions in our fits to speed versus accuracy data, the fit of the model was improved but the ordering of the estimated urgency parameters was the opposite to the predicted one. This reinforces the impression that the parameters are not well identified and that much of the model's ability to fit data relies on parameter trade-offs.

One of the core claims made by the urgency-gating model is that evidence does not accumulate. Rather, it grows rapidly to a stationary distribution that is gated multiplicatively by an urgency signal to make a response. We emphasized that the low-pass filtering assumptions of the urgency-gating model are tantamount to representing evidence accumulation by an OU diffusion process — an identification that has not always been made explicit in previous discussions of the models. When we fit an urgency-gated OU process, most of the estimates of OU decay were small, consistent with the findings of Ratcliff and Smith (2004). We did not find any consistent evidence for the proposition that evidence does not accumulate.

Changing-Stimulus Experiments

The model fits from the experiments in which stimulus information changed during a trial presented a somewhat more complex picture, with greater evidence for urgency. Unlike previous studies of changing-stimulus tasks, we considered a diffusion model in which the drift and diffusion rates depended on the outputs of a time-varying perceptual encoding process. Previous theoretical and empirical treatments of these kinds of tasks have assumed that drift rates change abruptly when the stimulus changes. In contrast, we used an explicit model of perceptual encoding that allowed us to distinguish temporal integration in perceptual encoding from evidence accumulation in the decision process. Under these circumstances, we found that the diffusion and urgency-gating models performed similarly and that the preferred model depended on the model-selection criterion. When we used the AIC or BIC, the urgency-gating model was the preferred model for the majority of participants by a small margin, but when we used the parametric bootstrap, which takes account of model flexibility, the ordering was reversed and the diffusion model was preferred. The nonzero value of the optimal classification point in Figure 7 is a reflection of the comparatively greater flexibility of the urgency-gating model when decay and urgency are both allowed to vary, which was highlighted in Figure 2.

Although the fits to the changing-stimulus experiments did not clearly distinguish between the urgency-gating and diffusion models, they agreed with the results from the fixed-stimulus experiments in finding no strong support for decay. The core claim of the urgency-gating model is that evidence grows rapidly to a stationary distribution, determined by the OU decay parameter, and decisions are made, not by evidence accumulation, but by an urgency function acting on the stationary distribution. The rationale for using changing-stimulus tasks is to identify decay in the accumulation process, which should be expressed as a recency weighting of the evidence. Our model comparisons showed that the best urgency-gating model for the majority of participants was the model UG(b, m, s_t), in which variable amounts of urgency are added to a Wiener diffusion process, without decay. As we have shown in an explicit, formal way, this model can be interpreted as a collapsing-bounds model. The comparatively better performance of the urgency model on the changing-stimulus tasks may be because participants implicitly deadline on these tasks, because they use extended sequences of stimulus information rather than single stimuli. We discuss differences among tasks in the following section.

Our conclusions in relation to decay diverge from those of Trueblood et al. (2021) who argued for a model with both urgency and decay. Their estimate of the OU time constant from their Experiment 1 was around 250 ms which agrees with the mean estimate of $\gamma \approx 4$ in Table 10, but they did not compare models with and without decay. Rather, their conclusions were based on estimates of the mode of the posterior density from a Bayesian hierarchical model fitted to the entire set of data (including the 22 participants who showed high fast-guessing rates). These estimates are population estimates and do not distinguish variations in performance at the individual participant level. At the individual level, the estimates of γ were bimodal, with some participants showing little or no decay and some participants showing large decay, and there was evidence of trade-offs in urgency and decay in all three experiments. (The correlations between γ and m for the three experiments were: Trueblood et al., $r = -.366$; color, $r = -.347$; luminance $r = -.291$.) The presence of these kinds of trade-offs, in which there are large estimates of decay for some participants and small estimates for others, likely underlies the nonzero posterior modes in Trueblood et al.'s hierarchical fits. One of the stated advantages of hierarchical Bayesian models, which prescribe population distributions for the parameters a priori, is that they can help stabilize estimation under conditions in which the individual participant data is sparse or noisy. However, they cannot overcome of the problem of underidentified parameters at the individual participant level. When parameters are underidentified, and the likelihood surface of the model is locally flat or near-flat for the parameters in question, Bayesian models will prefer those values from a set of equally-likely parameter values that are assigned the highest prior probabilities. This can lead to estimates that are better behaved statistically but does not aid in establishing the scientific ground truth unless the ground truth and the researcher's prior beliefs happen to coincide.

The other consistent finding from the changing-stimulus experiments is that we found no strong evidence for extended perceptual integration, as expressed by time-varying changes in drift rate. The estimates of perceptual integration from the β parameter in the model DIFF(β, η, s_t) were in the range 100 ms - 200 ms, the lower bound of which falls in the Bloch's law regime that characterizes the majority of perceptual tasks. These estimates suggest

that the flashing-grid task is similar to the dynamic brightness tasks of Ratcliff and Smith (2010) and Ratcliff, Voskuilen, and Teodorescu (2018), which also require decisions about dynamic random arrays and which can be modeled by an abrupt-onset evidence process. These kinds of tasks are markedly different from the dynamic form-discrimination tasks studied by Ratcliff and Smith and Smith et al. (2012), which are not well-described by abrupt-onset models and are better characterized by models in which the evidence entering the decision process increases progressively over time. Smith and Lilburn (2020) found that the widely-studied RDM task was also better characterized by a model of this kind. A reasonable conjecture, based on the available evidence, is that whether the decision process is better characterized as an abrupt-onset or a gradual-onset process will depend on the way in which evidence is encoded perceptually. When the evidence is carried by the statistics of the noise itself, the data appear to favor an abrupt-onset process, but when it is carried by a signal — whether static form or coherent motion — embedded in the noise, then the data appear to favor a gradual-onset process. Unlike Smith and Lilburn (2020, Figure 7), we found appreciable individual differences in the estimates of the β parameter in the flashing-grid experiments, suggesting it may not be well-identified under conditions in which encoding is fast, but the fact that we have obtained fairly consistent evidence for a near-abrupt-onset model implies this is a meaningful distinction to make theoretically and is one that can be tested in data using the methods we have presented here.

Varieties of Decision Tasks

The three tasks we considered here, as well as differing in whether stimulus information remained fixed or changed over time, differed in the way in which noise enters the decision process. Like many of the perceptual, language, and memory tasks to which diffusion models have been applied, the noise in the numerosity and attentional cuing tasks is unobserved and internal. Theoretically, it arises from moment-to-moment variability in the process of matching the encoded stimulus to the cognitive representations of the decision alternatives. In contrast, the noise in the flashing-grid task is external: information is carried by the statistics of a noisy sequence of stimulus elements. There is a long tradition in psychophysics of treating external and internal sources of noise as if they are the same (Ratcliff, Voskuilen, & McKoon, 2018) and, in the study of evidence accumulation models, the same mathematical models have been used to characterize decisions about single stimuli and sequences of stimulus elements (Edwards, 1965; Stone, 1960). Tasks that require decisions about sequences of stimulus elements are known as expanded-judgment tasks and the literature on them dates back several decades (Cisek et al., 2009; de Gardelle & Summerfield, 2011; Edwards, 1965; Piestch & Vickers, 1997; Summerfield & Tsetsos, 2015; Vickers, Caudrey & Willson, 1971).

Despite the formal resemblance between expanded-judgment and other kinds of decision tasks, in a recent review of diffusion models Ratcliff, Smith, Brown, and McKoon (2016) cautioned against assuming they are psychologically equivalent. They pointed out that there are phenomena found in expanded-judgment tasks that do not appear to have any direct counterparts in decision tasks using single stimuli. These include differential weighting of stimulus elements near to and far from a category boundary (Summerfield & Tsetsos, 2015), increased engagement of visual working memory (Piestch & Vickers, 1997), and a

greater variety in the shapes of RT distributions for individual participants than is found in single-stimulus tasks (Smith & Vickers, 1989). If decision tasks are viewed as lying on a continuum with single-stimulus tasks at one end and expanded-judgment tasks at the other, then the flashing-grid task can be viewed as an "edge of expanded judgment" task: The presentation rate is sufficiently high that the stimuli form a continuously changing grid of contrasting elements rather than a discrete sequence and the temporal changes are experienced perceptually as random spatial displacements.

It is an open question whether these kinds of tasks are best viewed as limiting cases of expanded judgment tasks, in which the noise driving the decision process is external, or whether they should be viewed as versions of single-stimulus decision tasks, in which a drift rate is computed from the aggregated perceptual properties of the stimulus and the noise arises internally from a cognitive matching process, as in the dynamic brightness discrimination task. The substantial individual differences in our estimates of urgency for the flashing-grid task, which reflect variations in the shapes of the RT distributions, are reminiscent of the large range of individual variations in the shapes of the RT distributions reported by Smith and Vickers (1989) for an expanded-judgment task using sequences of normally-distributed line segments. In either case, the variability in the RT distributions suggests that tasks involving sequences of stimulus elements may engage strategic processes that control the way evidence is sampled to a greater extent than do tasks involving single stimuli. If so, then we should be cautious about generalizing from these kinds of decision tasks to others.

Conclusion

In this article, we have shown that the integral equation method provides a natural theoretical framework for representing dynamic decision models in which decision boundaries change over time or in which the accumulating evidence is gated by an urgency signal. Unlike previous treatments of these models in the literature, we obtained explicit mathematical representations of the RT distributions and choice probabilities for models of both kinds and provided a precise characterization of the conditions under which they are equivalent. We compared the diffusion model and versions of the urgency-gating model on five sets of data from three decision tasks that provided large samples of data from individual participants. One of the tasks used response-terminated stimuli; the second used stimuli that were briefly flashed and then masked or extinguished, and the third used stimuli whose identity remained fixed or changed after 350 ms. For the two single-stimulus tasks, the simplest, pure urgency model performed poorly, as found by previous investigators, but an extended urgency model, in which varying amounts of urgency were added to an underlying OU process, performed better. It did not, however, perform as well as the diffusion model. Unlike the diffusion model, the urgency-gating model lacks a mechanism to predict fast errors and, while it can predict slow errors, the account of the slow-error pattern in the data was not as good as that provided by the diffusion model. For the changing-stimulus task, both the diffusion model and the urgency-gating model provided comparably good accounts of the RT distributions and choice probabilities. Critically, we found little evidence for the core claim of the urgency-gating model that evidence does not accumulate. Under these circumstances, the urgency-gating model can alternatively be

viewed as a Wiener process with collapsing boundaries. Our theoretical results provide the first mathematically explicit characterization of the relationship between these two kinds of models.

Appendix A

The Kernel of the Integral Equation for Time-Inhomogeneous Diffusion Processes

Urgency-Gating Model

This appendix provides a derivation of the kernel of the integral equation in Equation 17 and shows the equivalence of an urgency-gating model with fixed boundaries, a_i, $i = 1,2$, and urgency function $U(t)$ and a model with time-varying boundaries $a_i/U(t)$. The expression for the kernel relies on the existence of a pair of functions, $\overline{\Psi}(x, t)$ and $\Phi(t)$, that transform an arbitrary diffusion process, X_t, with drift rate $A(x, t)$ and diffusion rate $B(x, t)$ to a standard Wiener process. When this transformation exists, the kernel, $\Psi[a_i(t), t|a_i(\tau), \tau]$, goes to zero as $\tau \to t$, which guarantees that the integral equations in Equations 13 and 14 will be numerically stable. Ricciardi (1976), following Cherkasov (1957), showed that this transformation exists if and only if there exists a pair of functions, $c_1(t)$ and $c_2(t)$, of time only, which relate the drift and diffusion rates in a prescribed way. The relationship is most simply expressed in the form in which it was given by Ricciardi and Sato (1983).

An arbitrary diffusion process, with drift and diffusion rates both depending on state, x, and time, t, can be transformed to a standard Wiener process if functions $c_1(t)$ and $c_2(t)$ can be found that satisfy the following relationship

$$A(x, t) = \frac{B'_x(x, t)}{4} + \frac{\sqrt{[B(x, t)]}}{2}\left\{c_1(t) + \int^x \frac{c_2(t)B(y, t) + B'_t(y, t)}{\sqrt{B^3(y, t)}}dy\right\}, \tag{A1}$$

where $B'_x(x, t)$ and $B'_t(y, t)$ are, respectively, the partial derivatives of the diffusion rate with respect to its state and time coordinates. When the drift rate may depend on both time and state, but the diffusion rate depends only on time, as in Equations 11 and 12, this relationship has the simpler form (Smith & Lilburn, 2020; Voskuilen et al., 2016),

$$A(x, t) = \frac{\sqrt{[B(t)]}}{2}c_1(t) + \frac{x}{2}\left[c_2(t) + \frac{B'(t)}{B(t)}\right]. \tag{A2}$$

If functions $c_1(t)$ and $c_2(t)$ can be found that satisfy this equation, then the functions transforming the process X_t into a zero-drift, unit variance, Wiener process have the form

$$\begin{aligned} x^* = \overline{\Psi}(x, t) = \exp\left[-\frac{1}{2}\int^t c_2(s)ds\right]\left[\int^x \frac{dy}{\sqrt{B(t)}} \right. \\ \left. - \frac{1}{2}\int^t c_1(s)\exp\left[-\frac{1}{2}\int^s c_2(z)dz\right]ds\right. \end{aligned} \tag{A3}$$

$$t* = \Phi(t) = \int^t \exp\left[-\int^s c_2(z)dz\right]ds, \tag{A4}$$

where x^* and t^* are the new state and time coordinates, respectively.

The drift and diffusion rates of the urgency-gating model of Equation 8 are given by Equations 11 and 12,

$$A(y,t) = \left\{U(t)\mu(t) + \left[\frac{U'(t)}{U(t)} - \gamma\right]y\right\} \tag{A5}$$

$$B(t) = U^2(t)\sigma^2(t), \tag{A6}$$

where we are allowing the drift rate, $\mu(t)$, and the diffusion rate, $\sigma^2(t)$, of the underlying Wiener process to be time-varying for the sake of maximum generality. Substituting the drift rate and the infinitesimal standard deviation of Equations A5 and A6 into Equation A2 shows that the functions $c_1(t)$ and $c_2(t)$ must satisfy the following equation

$$U(t)\mu(t) + \left[\frac{U'(t)}{U(t)} - \gamma\right]y = \frac{U(t)\sigma(t)}{2}c_1(t) + \frac{y}{2}\left\{c_2(t) + \frac{\left[U^2(t)\sigma^2(t)\right]'}{U^2(t)\sigma^2(t)}\right\}, \tag{A7}$$

where primes denote derivatives with respect to time. Equating coefficients on the left and right hand sides of this equation yields the functions

$$c_1(t) = \frac{2\mu(t)}{\sigma(t)} \tag{A8}$$

$$c_2(t) = 2\left[\frac{U'(t)}{U(t)} - \gamma\right] - \frac{\left[U^2(t)\sigma^2(t)\right]'}{U^2(t)\sigma^2(t)}. \tag{A9}$$

Substituting these functions into Equations A3 and A4 and evaluating them yields, after some algebra,

$$\overline{\Psi}(y,t) = \frac{e^{\gamma t}y}{U(t)} - \int_0^t \mu(s)ds \tag{A10}$$

and

$$\Phi(t) = \int_0^t e^{2\gamma s}\sigma^2(s)ds. \tag{A11}$$

The kernel of the integral equation, $\Psi[a_i(t), t|a_j(\tau), \tau]$, in Equation 17 and the transition density of the unconstrained process, $f[a_i(t), t|a_j(\tau), \tau)]$, in Equation 18, depend on the

partial derivatives of $\overline{\Psi}(\,\cdot\,)$, with respect to its state and time coordinates, evaluated at the boundaries, and the derivative of $\Phi(\cdot)$, with respect to time. These functions are

$$\overline{\Psi}'_t = e^{\gamma t}\left\{\frac{[\gamma U(t) - U'(t)]}{U^2(t)}y - \mu(t)\right\}; \quad \overline{\Psi}'_x = \frac{e^{\gamma t}}{U(t)}; \quad \Phi'(t) = e^{2\gamma t}\sigma^2(t).$$

Equation 17 states that the kernel function is

$$\Psi[a_i(t), t \mid a_j(\tau), \tau] = \frac{f[a_i(t), t \mid a_j(\tau), \tau]}{2}$$
$$\times \left\{ a_i'(t) + \frac{\overline{\Psi}'_t(a_i(t), t)}{\overline{\Psi}'_x(a_i(t), t)} - \frac{[\overline{\Psi}(a_i(t), t) - \overline{\Psi}(a_j(\tau), \tau)]}{\Phi(t) - \Phi(\tau)} \frac{\Phi'(t)}{\overline{\Psi}'_x(a_i(t), t)} \right. \tag{A12}$$
$$\left. \vphantom{\frac{1}{1}}\right\},$$

with transition density

$$f[a_i(t), t \mid a_j(\tau), \tau] = \frac{1}{\sqrt{2\pi[\Phi(t) - \Phi(\tau)]}}\exp\left\{-\frac{[\overline{\Psi}(a_i(t), t) - \overline{\Psi}(a_j(\tau), \tau)]^2}{2[\Phi(t) - \Phi(\tau)]}\right\}\overline{\Psi}'_x \tag{A13}$$
$$(a_i(t), t).$$

For the urgency-gating model, the boundaries are constant, so the $a_i'(t)$ term in Equation A12 is zero and the expression in braces on the right hand side evaluates to

$$\left\{ a_i\left[\gamma - \frac{U'(t)}{U(t)}\right] - U(t)\mu(t) - \frac{[e^{\gamma t}(a_i - a_j) - U(t)\int_\tau^t \mu(s)e^{\gamma s}ds]}{\int_\tau^t e^{2\gamma s}\sigma^2(s)ds}e^{\gamma t}\sigma^2(t)\right\}. \tag{A14}$$

The rest of the kernel function is obtained by substituting terms into the expression for $\bar{f}[a_i(t), t \mid a_j(\tau), \tau]$ as indicated. In the time-homogeneous case, $\mu(t) \equiv \mu$ and $\sigma(t) \equiv \sigma$, and the integral terms in Equation A14 reduce to the expressions for the mean and variance of the time-homogeneous OU process in Equations 6 and 7.

Collapsing Boundaries Model

Evidence accumulation in the collapsing boundaries model is described by an OU process through time-varying boundaries, $a_i(t)$, $i = 1, 2$. In the time-inhomogeneous case, the drift and diffusion rates for this model are

$$A(x, t) = \mu(t) - \gamma x \tag{A15}$$

$$B(t) = \sigma^2(t), \tag{A16}$$

from which we obtain, via Equation A2,

$$c_1(t) = \frac{2\mu(t)}{\sigma(t)} \tag{A17}$$

$$c_2(t) = -2\gamma - \frac{\sigma^{2'}(t)}{\sigma^2(t)}, \tag{A18}$$

and after substituting in Equations A3 and A4,

$$\overline{\Psi}(x,t) = e^{\gamma t}x - \int_0^t e^{\gamma s}\mu(s)ds \tag{A19}$$

and

$$\Phi(t) = \int_0^t e^{2\gamma s}\sigma^2(s)ds. \tag{A20}$$

When $\overline{\Psi}(x,t)$ is evaluated at a boundary, $x = a_i(t)$. If $a_i(t) = a_i/U(t)$, then Equations A10 and A19 are identical, as are Equations A11 and A20. In other words, when the boundaries are inversely proportional to the urgency function, the transformations that map the urgency gated OU process with constant boundaries and the ungated process with time-varying boundaries to a standard Wiener process are the same. To show that first-passage time densities for the two models are also the same, we need to show that when the transformations mapping the process to a Wiener process are the same, then the kernels of the integral equations are also the same. This is not completely self-evident because the expression for the kernel contains a term $a_i'(t)$, for the derivative of the boundary, which will be zero for the urgency-gating model but not for the collapsing boundaries model.

For the collapsing boundaries model $a_i'(t) = -a_iU'(t)/U^2(t)$ and the derivatives of $\overline{\Psi}(\cdot)$ and $\Phi(\cdot)$ are

$$\overline{\Psi}_t' = \gamma e^{\gamma t}\frac{a_i}{U(t)} - \mu(t)e^{\gamma t}; \ \overline{\Psi}_x' = e^{\gamma t}; \ \Phi'(t) = e^{2\gamma t}\sigma^2(t),$$

where we have evaluated $\overline{\Psi}_t'(x,t)$ at $x = a_i(t)$. Substituting these expressions into the expression for the kernel, Equation A12, yields for the term in braces

$$\left\{a_i\left[\frac{\gamma}{U(t)} - \frac{U'(t)}{U^2(t)}\right] - \mu(t) - \frac{\left[e^{\gamma t}(a_i - a_j)/U(t) - \int_\tau^t \mu(s)e^{\gamma s}ds\right]}{\int_\tau^t e^{2\gamma s}\sigma^2(s)ds}e^{\gamma t}\sigma^2(t)\right\}. \tag{A21}$$

This is equal to the corresponding expression for the urgency-gating model, Equation A14, up to a scale factor, $U(t)$. The kernel in Equation A12 is obtained by multiplying the term in braces by the transition density, $f[a_i(t), t|a_j(\tau), \tau)]$, in Equation A13. The last term in the transition density is the partial derivative, $\overline{\Psi}_x'(t)$, which is $e^{\gamma t}$ for the OU process and $e^{\gamma t}/U(t)$

for the urgency-gating model. In the kernel, the term in braces for the urgency-gating model will therefore be divided by $U(t)$, making the terms for the two models identical. The other terms in the kernel depend on the values of $\overline{\Psi}(x, t)$, evaluated at a_i, or $a_j(t)$, respectively, and $\Phi(t)$, which are the same for the two models. This result shows that the first passage time densities for an urgency-gating model with boundaries a_i and urgency function $U(t)$ and a collapsing boundary model with boundaries $a_j(t) = a_i/U(t)$ are the same. We have shown this equivalence for an OU diffusion process, which was interpreted by Carland et al. (2015, 2016) as the output of a low-pass filter, but it also holds for the Wiener process, which is obtained from the OU process with $\gamma = 0$ as a special case.

Appendix B

Numerical Solution of the Integral Equations

To evaluate Equations 13 and 14 numerically, we discretize them and evaluate them on the mesh k , $k = 1, 2, \ldots$. The discretized forms of the equations (Buonocore et al., 1990; Smith, 2000, pp. 440–441) are

$$
\begin{aligned}
g_A(a_1, k\Delta \mid z, 0) = & -2\Psi(a_1, k\Delta \mid z, 0) \\
& + 2\Delta \sum_{j=1}^{k-1} g_A(a_1, j\Delta \mid z, 0)\Psi(a_1, k\Delta \mid a_1, j\Delta) \\
& + 2\Delta \sum_{j=1}^{k-1} g_B(a_2, j\Delta \mid z, 0)\Psi(a_1, k\Delta \mid a_2, j\Delta),
\end{aligned}
\tag{B1}
$$

and

$$
\begin{aligned}
g_B(a_2, k\Delta \mid z, 0) = & \, 2\Psi(a_2, k\Delta \mid z, 0) \\
& - 2\Delta \sum_{j=1}^{k-1} g_A(a_1, j\Delta \mid z, 0)\Psi(a_2, k\Delta \mid a_1, j\Delta) \\
& - 2\Delta \sum_{i=1}^{k-1} g_B(a_2, j\Delta \mid z, 0)\Psi(a_2, k\Delta \mid a_2, j\Delta),
\end{aligned}
\tag{B2}
$$

for $k = 2, 3, \ldots$. For $k = 1$, the equations reduce to

$$
g_A(a_1, \Delta \mid z, 0) = -2\Psi(a_1, \Delta \mid z, 0)
\tag{B3}
$$

and

$$
g_B(a_2, \Delta \mid z, 0) = 2\Psi(a_2, \Delta \mid z, 0).
\tag{B4}
$$

Equations B1 and B2 represent the first-passage time densities at time k as functions of their values at preceding times j , $j < k$, and of the kernel function Equation A12. Buonocore et al. (1990) proved that if the kernel is chosen according to Equation A12, then the discrete approximations converge to the true first-passage densities as $\rightarrow 0$. Equations B1 to B4 provide a computationally efficient and numerically stable way to obtain predictions for models with time-varying drift and diffusion rates or time-varying

boundaries. Voskuilen et al. (2016, Appendix B) gave versions of the equations for a Wiener diffusion process with constant drift and diffusion rates through time-varying boundaries, which they used to evaluate collapsing boundary models.
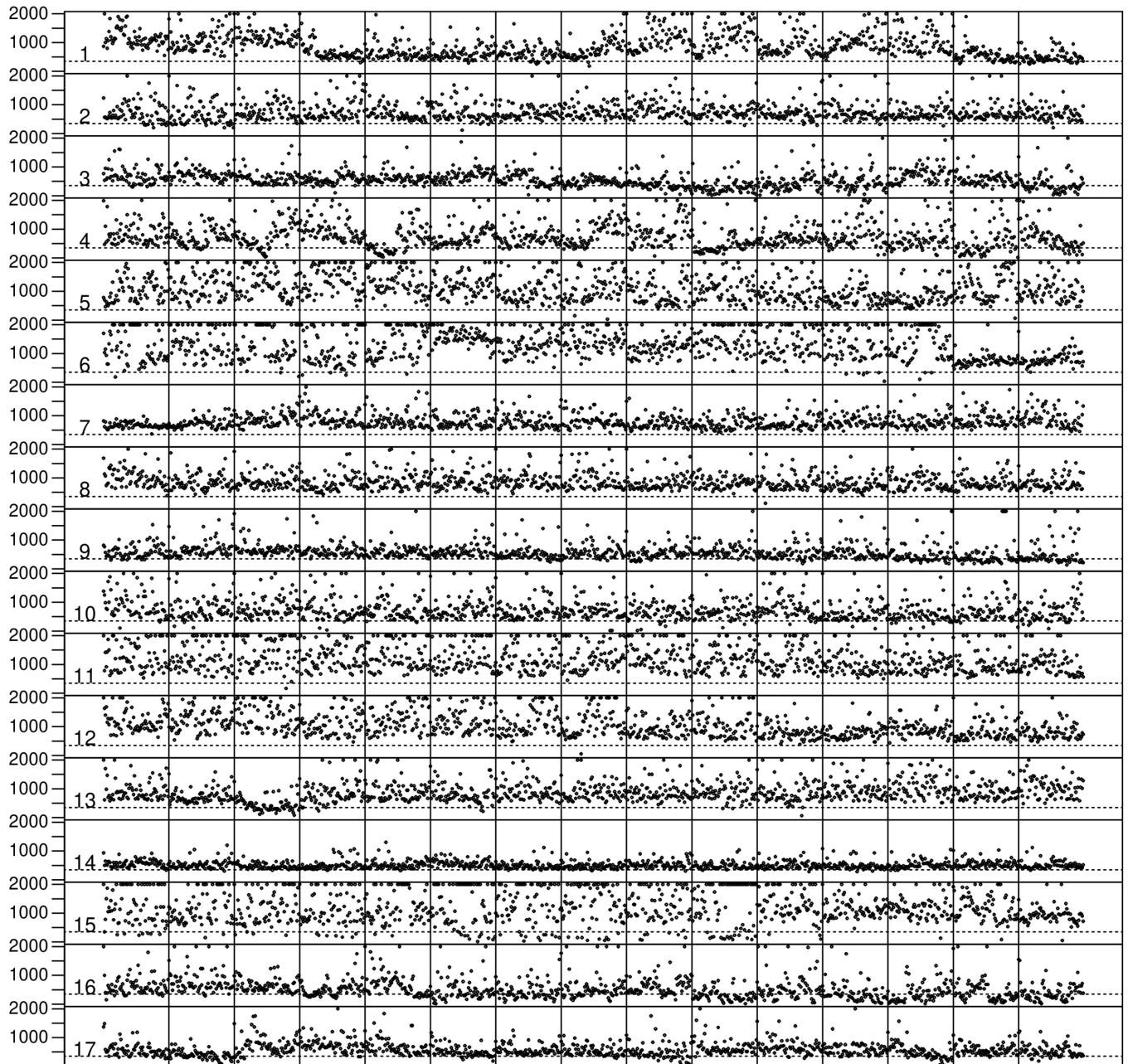
Appendix C

Data from Trueblood et al. (2021)

Figure 10:

RTs for Participants 1–17 of Experiment 1 of Trueblood et al. (2021). RTs for individual trials are plotted on the y-axis against trial number on the x-axis. Row numbers identify the

participants. RTs on timed-out trials are shown as 2000 ms. The horizontal dashed line is the 350 ms fast-guess threshold.
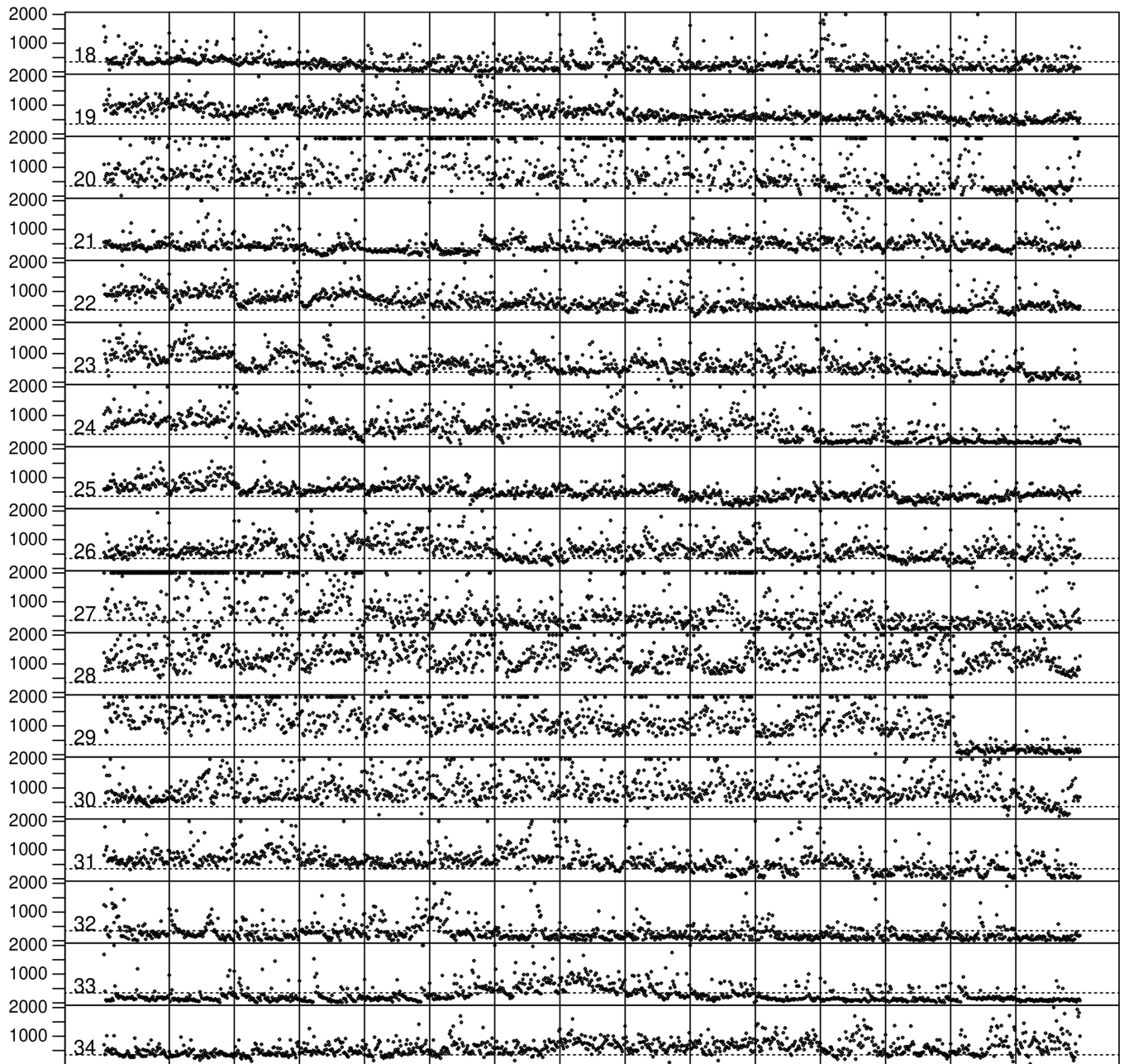

Figure 11:
Trial-to-trial RTs for Participants 18–34 of Experiment 1 of Trueblood et al. (2021). The participants used in the analyses in this article were 2, 5, 7, 8, 10, 11, 12, 13, 14, 19, 28, 30.

References

Abramowitz M, & Stegun I (1965). Handbook of mathematical functions. New York, NY: Dover.

Akaike H (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control AC 19, 716–723. 10.1109/TAC.1974.1100705

Atkinson RC, & Juola JF (1974). Search and decision processes in recognition memory. In Krantz DH, Atkinson RC, Luce RD, & Suppes P. (Eds.), Contemporary developments in mathematical psychology (Vol. 1, pp. 243–293). San Francisco, CA: Freeman.

Bloch A-M (1885). Expériences sur la vision. Computes Rendus du Séances de le Société de Biologie, 37, 493–495.

Bhattacharya RN, Waymire EC (1990). Stochastic processes with applications, New York, N.Y.: Wiley.

Breitmeyer BG (1984). Visual masking: An integrative approach. Oxford, U.K.: Clarendon Press.

Bogacz R, Brown E, Moehlis J, Holmes P, & Cohen JD (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. Psychological Review, 113, 700–765. 10.1037/0033-295X.113.4.700 [PubMed: 17014301]

Brown SD, Ratcliff R, & Smith PL (2006). Evaluating methods for approximating stochastic differential equations. Journal of Mathematical Psychology, 50, 402–410. 10.1016/j.jmp.2006.03.004 [PubMed: 18574521]

Buonocore A, Giorno V Nobile AG, & Ricciardi L (1990). On the two-boundary first-crossing-time problem for diffusion processes. Journal of Applied Probability, 27, 102–114. 10.2307/3214598

Buonocore A, Nobile AG, & Ricciardi L (1987). A new integral equation for the evaluation of first-passage-time probabilities densities. Advances in Applied Probability, 19, 784–800. 10.2307/1427102

Busemeyer J, & Townsend JT (1991). Fundamental derivations from decision field theory. Mathematical Social Sciences, 23, 255–282. 10.1016/0165-4896(92)90043-5

Busemeyer J, & Townsend JT (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, 100, 432–459. 10.1037/0033-295X.100.3.432 [PubMed: 8356185]

Carland MA, Marcos E, Thura D, & Cisek P (2016). Evidence against perfect integration of sensory information during perceptual decision making. Journal of Neurophysiology, 115, 915–930. 10.1152/jn.00264.2015 [PubMed: 26609110]

Carland MA, Thura D, & Cisek P (2015). The urgency-gating model can explain the effects of early evidence. Psychonomic Bulletin & Review, 22, 1830–1838. 10.3758/s13423-015-0851-2 [PubMed: 26452377]

Cartwright D, & Festinger L (1943). A quantitative theory of decision. Psychological Review, 50, 595–621. 10.1037/h0056982

Cisek P, Puskas GA, & El-Murr S (2009). Decisions in changing conditions: The urgency-gating model. The Journal of Neuroscience, 29, 11560–11571. 10.1523/JNEUROSCI.1844-09.2009 [PubMed: 19759303]

Cherkasov ID (1957). On the transformation of the diffusion process to a Wiener process. Theory of probability and its applications, 2, 373–377.

Chung KL & Williams RJ (1983). Introduction to stochastic integration. Boston: Birkhäuser.

Churchland AK, Kiani R, & Shadlen MN (2008). Decision-making with multiple alternatives. Nature Neuroscience, 11, 693–702. 10.1038/nn.2123 [PubMed: 18488024]

Cox DR, & Miller HD, (1965). The theory of stochastic processes. London, UK: Chapman and Hall.

Diederich A (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. Journal of Mathematical Psychology, 39, 197–215. 10.1006/jmps.1995.1020

de Gardelle V, & Summerfield C (2011). Robust averaging during perceptual judgment. Proceedings of the National Academy of Sciences of the United States of America, 108, 13341–13346. 10.1073/pnas.1104517108 [PubMed: 21788517]

de Lange H (1952). Experiments on flicker and some calculations on an electrical analogue of the foveal system. Physica, 18, 935–950. 10.1016/S0031-8914(52)80230-7

de Lange H (1954). Relationship between critical flicker frequency and a set of low frequency characteristics of the eye. Journal of the Optical Society of America, 44, 380–389.) 10.1364/JOSA.44.000380 [PubMed: 13163770]

de Lange H (1958). Research into the dynamic nature of the fovea-cortex system with intermittent and modulated light. I. Attenuation characteristics with white and colored lights. Journal of Optical Society of America, 48, 777–784. 10.1364/JOSA.48.000777

Ditterich J (2006a). Evidence for time-variant decision making. European Journal of Neuroscience, 24, 3682–3641. 10.1111/j.1460-9568.2006.05221.x

Ditterich J (2006b). Stochastic models of decisions about motion direction: Behavior and physiology. Neural Networks, 19, 981–1012. 10.1016/j.neunet.2006.05.042 [PubMed: 16952441]

Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, & Pouget A (2012). The cost of accumulating evidence in perceptual decision making. The Journal of Neuroscience, 32, 3612–3628. 10.1523/JNEUROSCI.4010-11.2012 [PubMed: 22423085]

Durbin J (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. Journal of Applied Probability, 8, 431–453. 10.2307/3212169

Dutilh G, Annis J, Brown SD, Cassey P, Evans NJ, Grasman RPPP, et al. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. Psychonomic Bulletin & Review, 26, 1051–1069. 10.3758/s13423-017-1417-2 [PubMed: 29450793]

Edwards W (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. Journal of Mathematical Psychology, 2, 312–329. 10.1016/0022-2496(65)90007-6

Einstein A (1905). Über die von der molekularinetischen Theorie der Wärne geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik, 17, 549–560.

Evans NJ, Bennett AJ, & Brown SD (2019). Optimal or not; depends on the task. Psychonomic Bulletin & Review, 26, 1027–1034. 10.3758/s13423-018-1536-4 [PubMed: 30411197]

Evans NJ, Hawkins GE, & Brown SD (2020). The role of passing time in decision-making. Journal of Experimental Psychology: Learning, Memory, and Cognition, 46, 316–326. 10.1037/xlm0000725

Evans NJ, Trueblood JS, & Holmes WR (2020). A parameter recovery assessment of time-variant models of decision-making. Behavior Research Methods, 52, 193–206. 10.3758/s13428-019-01218-0 [PubMed: 30924107]

Forstmann BU, Ratcliff R, & Wagenmakers E-J (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. Annual Review of Psychology, 67, 641–666. 10.1146/annurev-psych-122414-033645

Gold JI, & Shadlen MN (2003). The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. The Journal of Neuroscience, 23, 632–651. 10.1523/JNEUROSCI.23-02-00632.2003 [PubMed: 12533623]

Gorea A (2015). A refresher of the original Bloch's law paper (Bloch, July, 1885), i-Perception, 6(4), 1–6, 2015. 10.1177/2041669515593043 [PubMed: 26034566]

Gronlund SD, & Ratcliff R (1991). Analysis of the Hockley and Murdock decision model. Journal of Mathematical Psychology, 35, 319–344. 10.1016/0022-2496(91)90051-T

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. 10.2307/3215118

Feller W (1968). An introduction to probability theory and its applications (3rd. ed). New York: N.Y.: Wiley.

Hanes DP, & Schall JD (1996). Neural control of voluntary movement initiation. Science, 274, 427–430. 10.1126/science.274.5286.427 [PubMed: 8832893]

Hawkins GE, Forstmann BU, Wagenmakers E-J, Ratcliff R, & Brown SD (2015). Revising the evidence for collapsing boundaries and urgency signals in perceptual decision-making. The Journal of Neuroscience, 35, 2476–2484. 10.1523/JNEUROSCI.2410-14.2015 [PubMed: 25673842]

Heath RA (1992). A general nonstationary diffusion model for two-choice decision making. Mathematical Social Sciences, 23 283–309. 10.1016/0165-4896(92)90044-6

Hockley WE, & Murdock BB (1987). A decision model for accuracy and response latency in recognition memory. Psychological Review, 94, 341–358. 10.1037/0033-295X.94.3.341

Holmes P, & Cohen JD (2014). Optimality and some of its discontents: Successes and shortcomings of existing models for binary decisions. Topics in Cognitive Science, 6, 258–278. 10.1111/tops.12084 [PubMed: 24648411]

Holmes WR, Trueblood JS, & Heathcote A (2016). A new framework for modeling decisions about changing information: The piecewise linear ballistic accumulator model. Cognitive Psychology, 85, 1–29. 10.1016/j.cogpsych.2015.11.002 [PubMed: 26760448]

Ingling CR, & Martinez E (1983). The spatiochromatic signal of the r-g channel. In Mollon JD & Sharpe LT (Eds.), Color vision: Physiology and psychophysics (pp. 433–444). London: Academic Press. 10.1016/0042-6989(85)90077-X

Itô K (1944). Stochastic integral. Proceedings of the Imperial Academy Tokyo, 20, 519–524. 10.3792/pia/1195572786

Itô K (1951). On a formula concerning stochastic differentials. Nagoya Mathematical Journal, 3, 55–65. 10.1017/S0027763000012216

Jones M, & Dzhafarov EN (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. Psychological Review, 121, 1–32. 10.1037/a0034190 [PubMed: 24079307]

Karatzas I, & Shreve SE (1991). Brownian motion and stochastic calculus. New York, N. Y.: Springer

Karlin S, & Taylor HM (1975). A first course in stochastic processes. New York, N.Y.: Academic Press.

Karlin S, & Taylor HM (1981). A second course in stochastic processes. Orlando, FL.: Academic Press.

Kass RE, & Raftery AE (1995). Bayes factors. Journal of the American Statistical Association, 90, 773–795. 10.1080/01621459.1995.10476572

Kelly DH (1961). Visual response to time-dependent stimuli. II. Single-channel model of the photopic visual system. Journal of the Optical Society of America, 51, 747–754. 10.1364/JOSA.51.000747

Kelly DH (1969). Flickering patterns and lateral inhibition. Journal of the Optical Society of America, 59, 1361–1370. 10.1364/JOSA.59.001361

Kelly DH (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. Journal of the Optical Society of America, 69, 1340–1349. 10.1364/JOSA.69.001340 [PubMed: 521853]

Luce RD (1986). Response times: Their role in inferring elementary mental organization. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195070019.001.0001

Malhotra G, Leslie DS, Ludwig CJH, & Bogacz R (2018). Time-varying decision bounds: insights from optimality analysis. Psychonomic Bulletin & Review, 25, 971–996. 10.3758/s13423-017-1340-6 [PubMed: 28730465]

Mazurek ME, Roitman JD, Ditterich J, & Shadlen MN (2003). A role for neural integrators in perceptual decision making. Cerebral Cortex, 13, 1257–1269. 10.1093/cercor/bhg097 [PubMed: 14576217]

McClelland J (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86, 287–330. 10.1037/0033-295X.86.4.287

Murdock BB (1983). A distributed memory model for serial-order information. Psychological Review, 90, 316–338. 10.1037/0033-295X.90.4.316

Murphy PR, Boonstra E, Nieuwenhuis S (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. Nature Communications, 7, Art. 13526. 10.1038/ncomms13526

Nelder JA, & Mead R (1965). A simplex method for function minimization. The Computer Journal, 7, 308–313. 10.1093/comjnl/7.4.308

Oberhettinger F, & Badii L (1973). Tables of Laplace transforms. Berlin: Springer-Verlag. 10.1002/zamm.19750551022

Pacut A (1980). Mathematical modelling of reaction latency: The structure of the models and its motivation. Acta Neurobiologiae Experimentalis, 40, 199–213. [PubMed: 7424578]

Palestro JJ, Weichart E, Sederberg PB, & Turner BM (2018). Some task demands induce collapsing bounds: Evidence from a behavioral analysis. Psychonomic Bulletin & Review, 25, 1225–1248. 10.3758/s13423-018-1479-9 [PubMed: 29845433]

Pietsch A & Vickers D (1997). Memory capacity and intelligence: Pnovel techniques for evaluating rival models of a fundamental information processing mechanism. Journal of General Psychology, 124, 229–339. 10.1080/00221309709595520

Pike AR, McFarland K, & Dalgleish L (1974). Speed-accuracy tradeoff models for auditory detection with deadlines. Acta Psychologica, 38, 379–399. 10.1016/0001-6918(74)90042-0 [PubMed: 4432771]

Protter P (1990). Stochastic integration and differential equations: A new approach. Berlin: Springer-Verlag.

Ratcliff R (1978). A theory of memory retrieval. Psychological Review, 85, 59–108. 10.1037/0033-295X.85.2.59

Ratcliff R (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. Journal of Mathematical Psychology, 21, 178–184. 10.1016/0022-2496(80)90006-1

Ratcliff R (1985). Theoretical interpretation of the speed and accuracy of positive and negative responses. Psychological Review, 92, 212–225. 10.1037/0033-295X.92.2.212 [PubMed: 3991839]

Ratcliff R (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. Psychological Review, 95, 238–255. 10.1037/0033-295X.95.2.238 [PubMed: 3375400]

Ratcliff R (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. Psychonomic Bulletin & Review, 9, 278–291. 10.3758/BF03196283 [PubMed: 12120790]

Ratcliff R (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. Psychology and Aging, 23, 900–916. 10.1037/a0013930 [PubMed: 19140659]

Ratcliff R (2018). Decision making on spatially continuous scales. Psychological Review, 125, 888–935. 10.1037/rev0000117 [PubMed: 30431302]

Ratcliff R, Cherian A, & Segraves M (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. Journal of Neurophysiology 90, 1392–1407. 10.1152/jn.01049.2002 [PubMed: 12761282]

Ratcliff R, & Childers R (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. Decision, 2, 237–279. 10.1037/dec0000030

Ratcliff R, Hasegawa Y, Hasegawa R, Smith PL, & Segraves M, (2007). A dual diffusion model for single cell recording data from the superior colliculus in brightness discrimination task. Journal of Neurophysiology, 97, 1756–1797. 10.1152/jn.00393.2006 [PubMed: 17122324]

Ratcliff R, & McKoon G (2008). The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation, 20, 873–922. 10.1162/neco.2008.12-06-420 [PubMed: 18085991]

Ratcliff R, & Rouder JN (2000). A diffusion model account of masking in letter identification. Journal of Experimental Psychology: Human Perception and Performance, 26, 127–140. 10.1037/0096-1523.26.1.127 [PubMed: 10696609]

Ratcliff R, & Smith PL (2004). A comparison of sequential-sampling models for two choice reaction time. Psychological Review, 111, 333–367. 10.1037/0033-295X.111.2.333 [PubMed: 15065913]

Ratcliff R, & Smith PL (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. 10.1037/a0018128 [PubMed: 20121313]

Ratcliff R, Smith PL, Brown SD, & McKoon G (2016). Diffusion decision model: Current issues and history. Trends in Cognitive Sciences, 20, 260–281. 10.1016/j.tics.2016.01.007 [PubMed: 26952739]

Ratcliff R, Smith PL, & McKoon G (2015). Modeling response time and accuracy data. Current Directions in Psychological Science, 24, 458–470. 10.1177/0963721415596228 [PubMed: 26722193]

Ratcliff R, Voskuilen C, & McKoon G (2018). Internal and external sources of variability in perceptual decision-making. Psychological Review, 125, 33–46. 10.1037/rev0000080 [PubMed: 29035076]

Ratcliff R, Voskuilen C, & Teodorescu A (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. Cognitive Psychology, 103, 1–22. 10.1016/j.cogpsych.2018.02.002 [PubMed: 29501775]

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. 10.2307/3213736

Roitman JD, & Shadlen MN (2002). Response of neurons in the lateral interparietal area during a combined visual discrimination reaction time task, Journal of Neuroscience, 22, 9475–9489. 10.1523/JNEUROSCI.22-21-09475.2002 [PubMed: 12417672]

Robson JG (1966). Spatial and temporal contrast-sensitivity function. Journal of the Optical Society of America, 56, 1141–1142. 10.1364/JOSA.56.001141

Ross SM (1983). Introduction to stochastic dynamic programming. San Diego, CA: Academic Press.

Roufs JAJ (1972). Dynamic properties of vision — II. Theoretical relationship between flicker and flash thresholds. Vision Research, 12, 279–292. 10.1016/0042-6989(72)90118-6 [PubMed: 5033690]

Roufs JAJ (1974). Dynamic properties of vision — IV. Thresholds of decremental flashes, incremental flashes and doublets in relation to flicker fusion. Vision Research, 14, 831–851. 10.1016/0042-6989(74)90148-5 [PubMed: 4422125]

Schall JD (2002). The neural selection and control of saccades by the frontal eye field. Philosophical Transactions of the Royal Society of London, B, 357, 1073–1082. 10.1098/rstb.2002.1098

Schall JD (2003). Neural correlates of decision processes: neural and mental chronometry. Current Opinions in Neurobiology, 13, 182–186. 10.1016/s0959-4388(03)00039-4

Schwarz G (1978). Estimating the dimension of a model. Annals of Statistics 6, 461–464. 10.1214/aos/1176344136

Sewell DK, & Smith PL (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. 10.1037/a0026591 [PubMed: 22201464]

Smith PL (1990). A note on the distribution of response times for a random walk with Gaussian increments. Journal of Mathematical Psychology, 34, 445–459. 10.1016/0022-2496(90)90023-3

Smith PL (1995). Psychophysically principled models of visual simple reaction time. Psychological Review, 102, 567–591. 10.1037/0033-295X.102.3.567

Smith PL (1998). Bloch's law predictions from diffusion process models of detection. Australian Journal of Psychology, 50, 139–147. 10.1080/00049539808258790

Smith PL (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. Journal of Mathematical Psychology, 44, 408–463. 10.1006/jmps.1999.1260 [PubMed: 10973778]

Smith PL (2010). From Poisson shot noise the integrated Ornstein-Uhlenbeck process: Neurally principled models of information accumulation in decision-making and response time, Journal of Mathematical Psychology, 54, 266–283. 10.1016/j.jmp.2009.12.002

Smith PL (2016). Diffusion theory of decision making in continuous report. Psychological Review, 123, 425–451. 10.1037/rev0000023 [PubMed: 26949831]

Smith PL, & Corbett EA (2019). Speeded multielement decision making as diffusion in a hypersphere: Theory and application to double-target detection. Psychonomic Bulletin & Review, 26, 127–162. 10.3758/s13423-018-1491-0 [PubMed: 29968205]

Smith PL, Ellis R, Sewell DK, & Wolfgang BJ (2010). Cued detection with compound integration-interruption masks reveals multiple attentional mechanisms. Journal of Vision, 10(5), Art 3., 1–28. 10.1167/10.5.3

Smith PL & Lilburn SD (2020). Vision for the blind: visual psychophysics and blinded inference for decision models. Psychonomic Bulletin & Review, 27, 882–910. 10.3758/s13423-020-01742-7 [PubMed: 32514800]

Smith PL, & Ratcliff R (2004). Psychology and neurobiology of simple decisions. Trends in Neurosciences, 27, 161–168. 10.1016/j.tins.2004.01.006 [PubMed: 15036882]

Smith PL, & Ratcliff R (2009). An integrated theory of attention and decision making in visual signal detection. Psychological Review, 116, 283–317. 10.1037/a0015156 [PubMed: 19348543]

Smith PL, Ratcliff R, & Sewell DK (2012). Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition. Journal of Mathematical Psychology, 59, 95–113. 10.1016/j.jmp.2013.05.007

Smith PL, Ratcliff R, & Wolfgang BJ (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. Vision Research, 44, 1297–1320. 10.1016/j.visres.2004.01.002 [PubMed: 15066392]

Smith PL, & Vickers D (1988). The accumulator model of two-choice discrimination. Journal of Mathematical Psychology, 32, 135–168. 10.1016/0022-2496(88)90043-0

Smith PL, & Vickers D (1989). Modeling evidence accumulation with partial loss in expanded judgment. Journal of Experimental Psychology: Human Perception and Performance, 15, 797–815. 10.1037/0096-1523.15.4.797

Smoluchowski M. von (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. Annalen der Physik, 326, 756–780. 10.1002/andp.19063261405

Sperling G, & Sondhi MM (1968). Model for visual luminance discrimination and flicker detection. Journal of the Optical Society of America, 58, 1133–1145. 10.1364/JOSA.58.001133 [PubMed: 5668364]

Starns JJ, & Ratcliff R (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. Psychology & Aging, 25, 377–390. 10.1037/a0018022 [PubMed: 20545422]

Starns JJ, & Ratcliff R (2012). Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. Psychonomic Bulletin & Review, 19, 139–145. 10.3758/s13423-011-0189-3 [PubMed: 22144142]

Stone M (1960). Models for choice reaction time. Psychometrika, 25, 251–260. 10.1007/BF02289729

Summerfield C, & Tsetsos K (2015). Do humans make good decisions? Trends in Cognitive Sciences, 19, 27–34. 10.1016/j.tics.2014.11.005 [PubMed: 25488076]

Swets JA, & Green DM (1961). Sequential observations by human observers of signals in noise. In Cherry C (Ed.), Information theory: Proceedings of the Fourth London Symposium (pp. 177–195). London: Butterworths.

Thura D (2016). How to discriminate conclusively among different models of decision making? Journal of Neurophysiology, 115, 2251–2254. 10.1152/jn.00911.2015 [PubMed: 26538611]

Thura D, Beauregard-Racine J, Fradet C-W, & Cisek P (2012). Decision making by urgency-gating: theory and experiment. Journal of Neurophysiology, 108, 2912–2930. 10.1152/jn.01071.2011 [PubMed: 22993260]

Tolhurst DJ (1975a). Reaction times in the detection of gratings by human observers: A probabilistic mechanism. Vision Research, 15, 1143–1149. 10.1016/0042-6989(75)90013-9 [PubMed: 1166615]

Tolhurst DJ (1975b). Sustained and transient channels in human vision. Vision Research, 15, 1151–1155. 10.1016/0042-6989(75)90014-0 [PubMed: 1166616]

Trueblood JS, Heathcote A, Evans NJ, Holmes WR (2021). Urgency, leakage, and the relative nature of information processing in decision-making. Psychological Review, 128, 160–186. 10.1037/rev0000255 [PubMed: 32852976]

Usher M, & McClelland JL (2001). The time course of perceptual choice: The leaky, competing accumulator model. Psychological Review, 108, 550–592. 10.1037/0033-295X.108.3.550 [PubMed: 11488378]

Vandekerckhove J, & Tuerlinckx F (2008). Diffusion model analysis with MATLAB: A DMAT primer. Behavior Research Methods, 40, 61–72. 10.3758/BRM.40.1.61 [PubMed: 18411528]

Van Zandt T, Colonius H, & Proctor RW (2000). A comparison of two response time models applied to perceptual matching. Psychonomic Bulletin & Review, 7, 208–256. 10.3758/BF03212980 [PubMed: 10909132]

Verdonck S, & Tuerlinckx F (2014). The Ising decision maker: A binary stochastic network model for choice response time. Psychological Review, 121, 422–462. 10.1037/a0037012 [PubMed: 25090426]

Vickers D (1970). Evidence for an accumulator model of psychophysical discrimination. Ergonomics, 13, 37–58. 10.1080/00140137008931117 [PubMed: 5416868]

Vickers D, Caudrey D, & Willson RJ (1971). Discriminating between the frequency of occurrence of two alternative events. Acta Psychology, 35, 151–172. 10.1016/0001-6918(71)90018-7

Voskuilen C, Ratcliff R, & Smith PL (2016). Comparing fixed and collapsing boundary versions of the diffusion model. Journal of Mathematical Psychology, 73, 59–79. 10.1016/j.jmp.2016.04.008 [PubMed: 28579640]

Voss A, & Voss J (2007). Fast-dm: A free program for efficient diffusion model analysis. Behavior Research Methods, 39, 767–775. 10.3758/BF03192967 [PubMed: 18183889]

Voss A, & Voss J (2008). A fast numerical algorithm for the estimation of diffusion model parameters. Journal of Mathematical Psychology, 52, 1–9. 10.1016/j.jmp.2007.09.005

Wagenmakers EJ, Ratcliff R, Gomez P, & Iverson GJ (2004). Assessing model mimicry using the parametric bootstrap. Journal of Mathematical Psychology, 48, 28–50. 10.1016/j.jmp.2003.11.004

Watamaniuk SNJ, & Sekuler R (1992). Temporal and spatial integration in dynamic random-dot stimuli. Vision Research, 32, 2341–2347. 10.1016/0042-6989(92)90097-3 [PubMed: 1288010]

Watson AB (1986). Temporal sensitivity. In Boff KR, Kaufman L, & Thomas JP (Eds.), Handbook of perception and performance (Vol. 1, pp. 6.1–6.85). New York: Wiley.

Watson AB, & Nachmias J (1977). Patterns of temporal interaction in the detection of gratings. Vision Research, 17, 893–902. 10.1016/0042-6989(77)90063-3 [PubMed: 595394]

Wiecki TV, Sofer I, & Frank MJ (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. Frontiers in Neuroinformatics, 7, 14. 10.3389/fninf.2013.00014 [PubMed: 23935581]

Winkel J, Keuken MC, van Maanen L, Wagenmakers E-J, & Forstmann BU (2014). Early evidence affects later decisions: Why evidence accumulation is required to explain response time data. Psychonomic Bulletin & Review, 21, 777–784. 10.3758/s13423-013-0551-8 [PubMed: 24395093]

Zhang J, Bogacz R, & Holmes P (2009). A comparison of bounded diffusion models for choice in time controlled tasks. Journal of Mathematical Psychology, 53, 231–241. 10.1016/j.jmp.2009.03.001 [PubMed: 19812713]

Zhang J, & Bogacz R (2010). Bounded Ornstein-Uhlenbeck models for two-choice time controlled tasks. Journal of Mathematical Psychology, 54, 322–333. 10.1016/j.jmp.2010.03.001
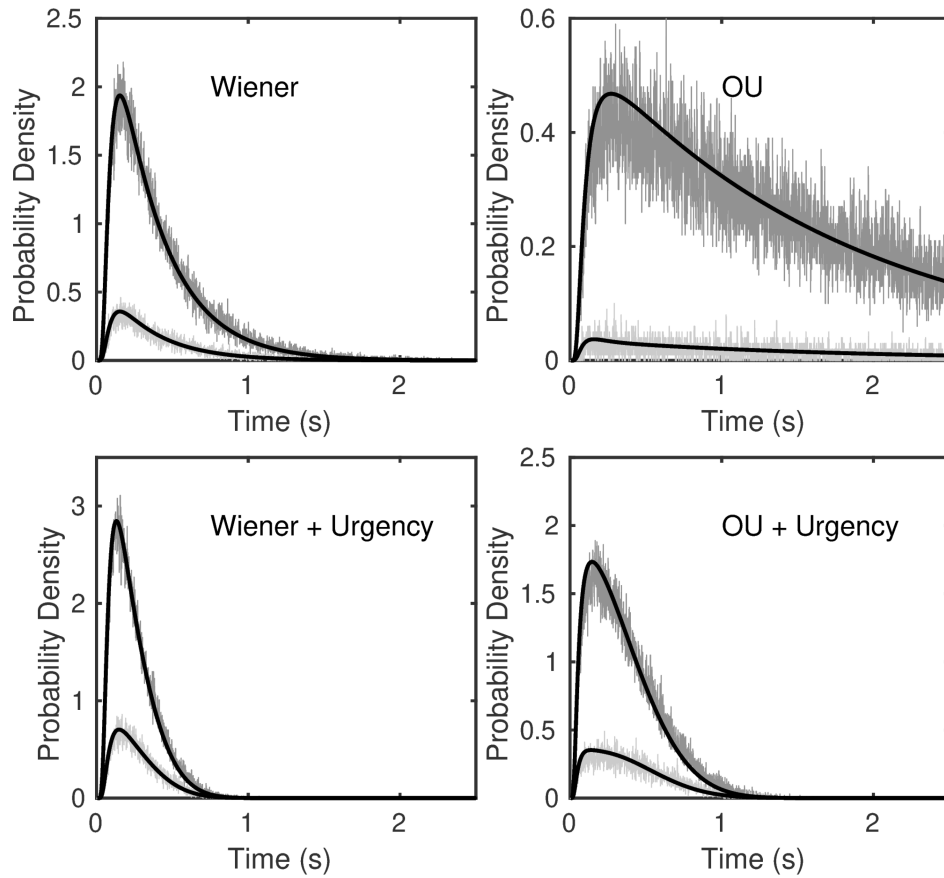
Figure 1:
Simulated and predicted joint distributions of correct responses and errors. Left to right and top to bottom the distributions are for: (a) a pure Wiener process; (b) a pure OU process; (c) a pure Wiener process with urgency-gating, and (d) an OU process with urgency-gating. For all models, $\mu = 0.12$, $a_1 = 0.07$, $a_2 = -0.07$, $z = 0$, and $\sigma = 0.1$. For the Wiener process models $\gamma = 0$ and for the OU models $\gamma = 8$. For the Wiener plus urgency model $b = 1.0$ and $= 1.5$ and for the OU plus urgency model $b = 1.3$ and $m = 0.35$.
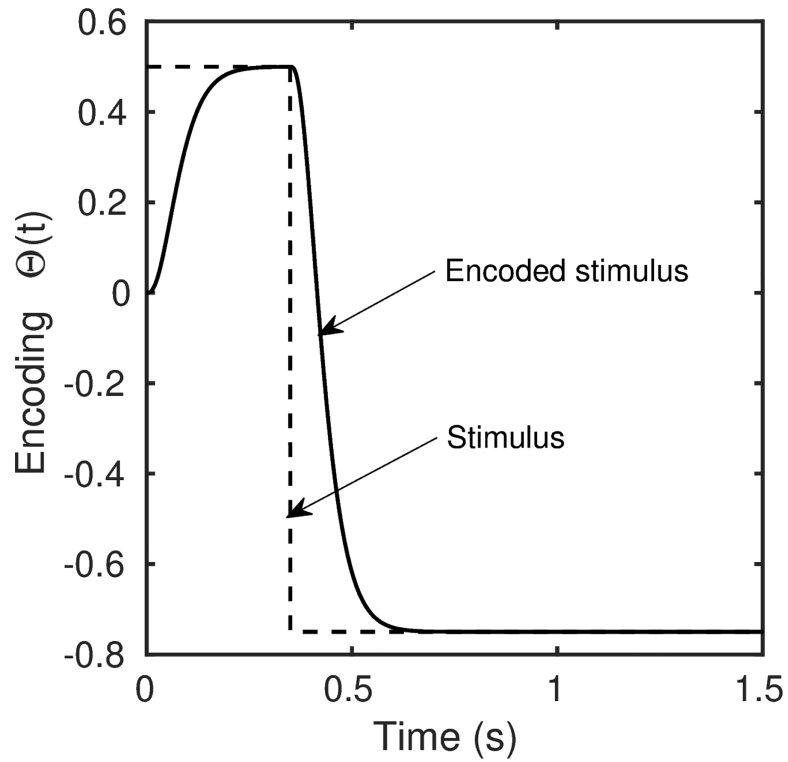
Figure 2:
Perceptual encoding function, $\Theta(t)$. The drift rate is proportional to the output of a linear filter composed of a cascade of exponential stages. The dashed line shows a stimulus waveform that changes from 0.5 to −0.75 at time $t = 0.35$ s and the continuous line shows the output of a three-stage exponential filter cascade with stage rate constant $\beta = 35$ u · s^{-1}, where "u" is the unit of encoding strength that maps stimulus discriminability (proportion of blue vs. orange squares) to drift rate.
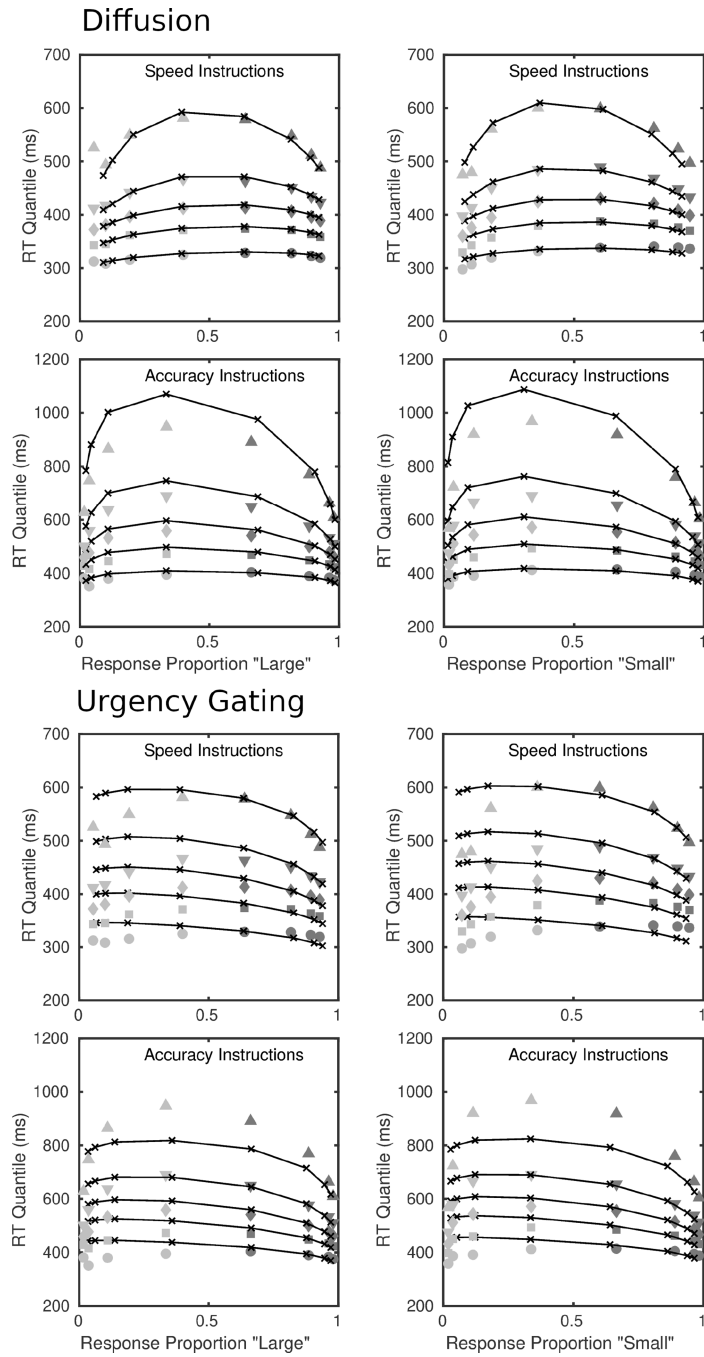
Figure 3:

Quantile probability functions for "large" and "small" responses for speed and accuracy conditions for the diffusion model DIFF(η, s_z, s_t) and the urgency-gating model UG(γ, m, s_t). The quantile RTs in order from the bottom to top are the .1, .3, .5, .7, and .9 quantiles (circles, squares, diamonds, inverted triangles, upright triangles, respectively). The dark gray symbols are the quantiles for correct responses and the light gray symbols are the quantiles for errors. The continuous curves and x's are the predictions from the model. For the data

and models the quantile RTs are plotted on the *y*-axis against the observed and predicted response proportions on the *x*-axis.
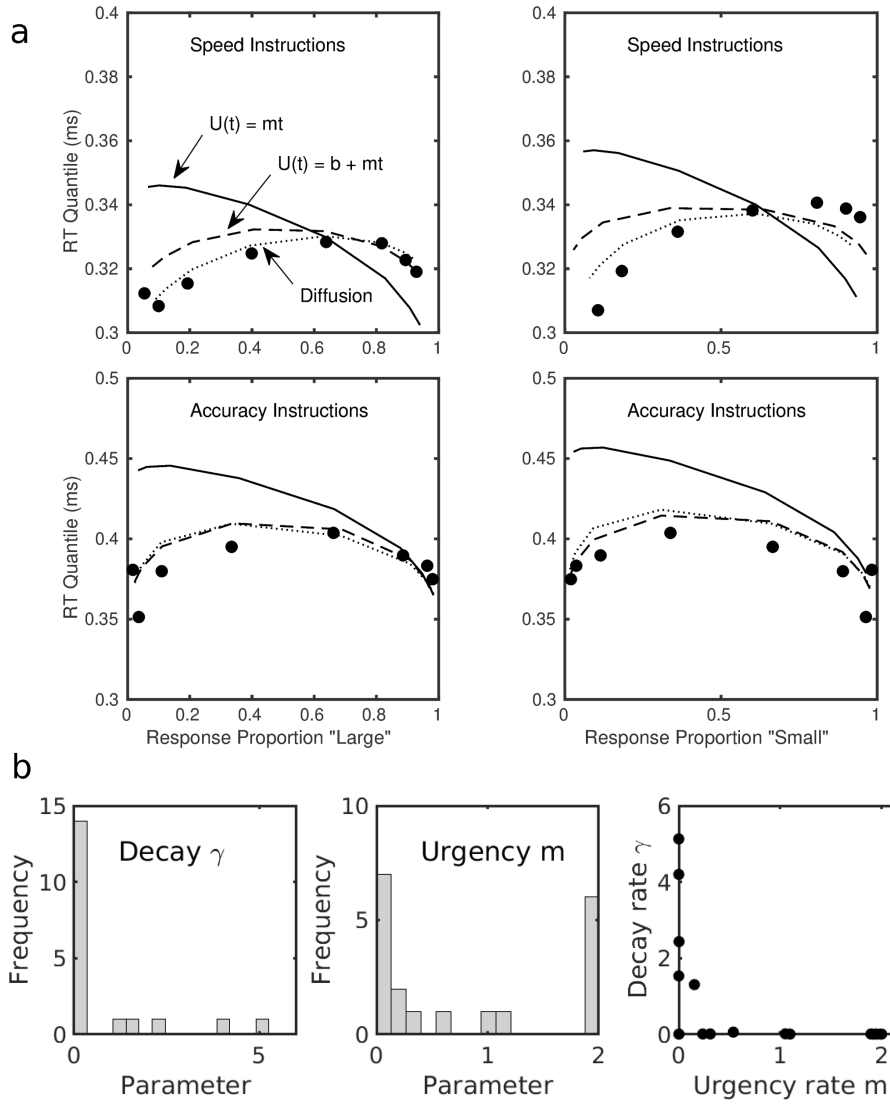
Figure 4:
(a) 0.1 RT quantiles for data (filled circles), diffusion model DIFF(η, s_z, s_t) (dots), urgency-gating model UG(γ, m, s_t) (continuous line), and urgency-gating model UG(γ, b, m, s_t) (dashed line) for the numerosity study. (b) Frequency distributions and scatterplot of parameter estimates for the urgency-gating model UG(γ, b, m, s_t).
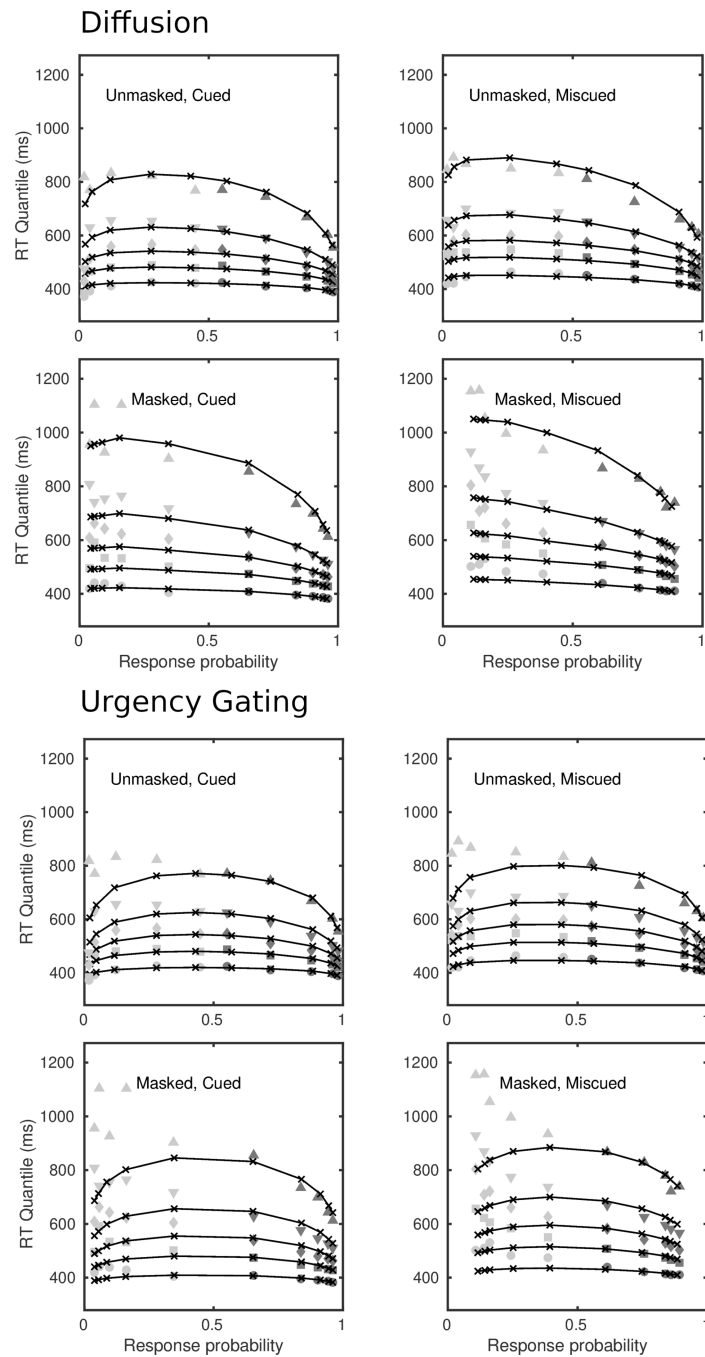
Figure 5:
Quantile probability functions for correct responses and errors in the Cue × Mask design for the diffusion model DIFF(η, s_t) and the urgency-gating model UG(γ, b, m, s_t) for the spatial cuing study. The meaning of the symbols in the plot is the same as in Figure 3.

Figure 6:
Frequency distributions and scatterplot of parameter estimates for the urgency-gating model UG(γ, b, m, s_t) in the spatial cuing study.
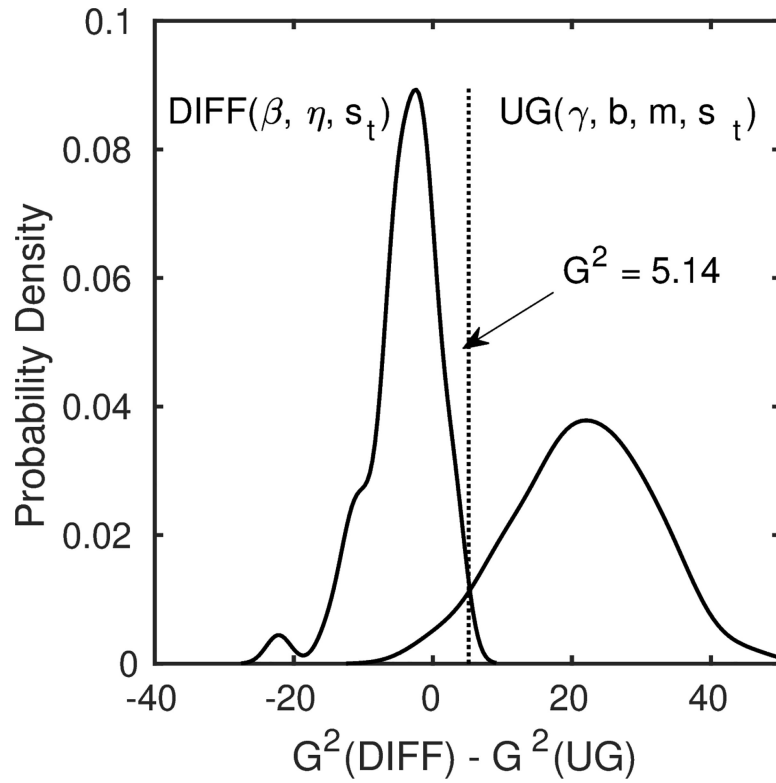
Figure 7:
Distributions of G^2 differences, $G^2[\text{DIFF}(\beta, \eta, s_t)] - G^2[\text{UG}(\gamma, b, m, s_t)]$, for models DIFF($\beta, \eta, s_t$) and UG($\gamma, b, m, s_t$) cross-fit to simulated data for the color experiment, generated using the means of the estimated parameters from fits to the empirical data in Table 10. Classification accuracy is maximized by setting the classification criterion equal to the G^2 difference at the intersection of the two distributions, at $G^2 = 5.14$.
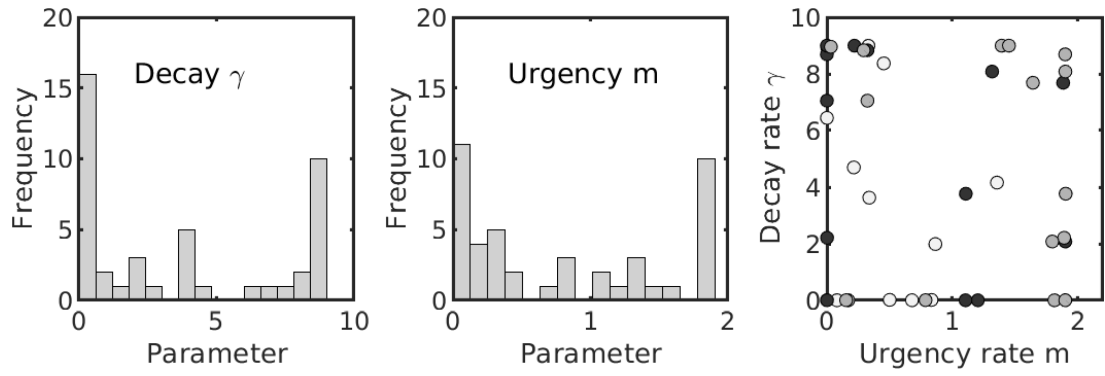
Figure 8:
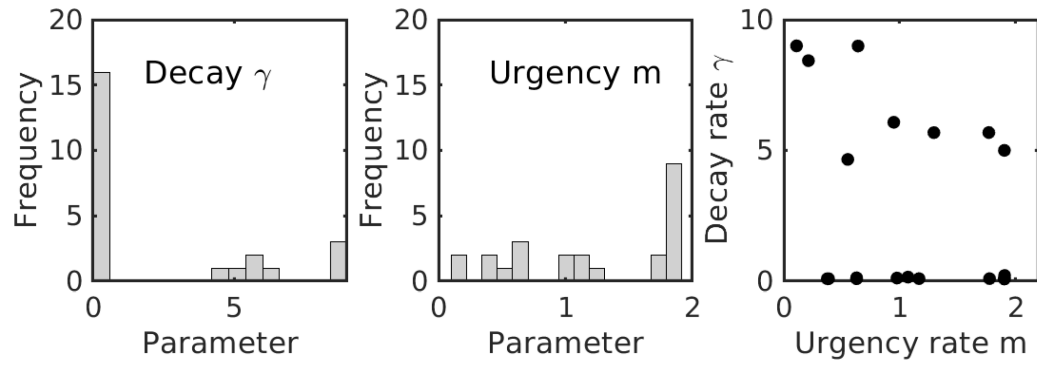Frequency distributions and scatterplot of parameter estimates for the urgency-gating model UG(γ, b, m, s_t) in the flashing-grid experiments. White = Trueblood et al; Gray = Color; Black = Luminance.
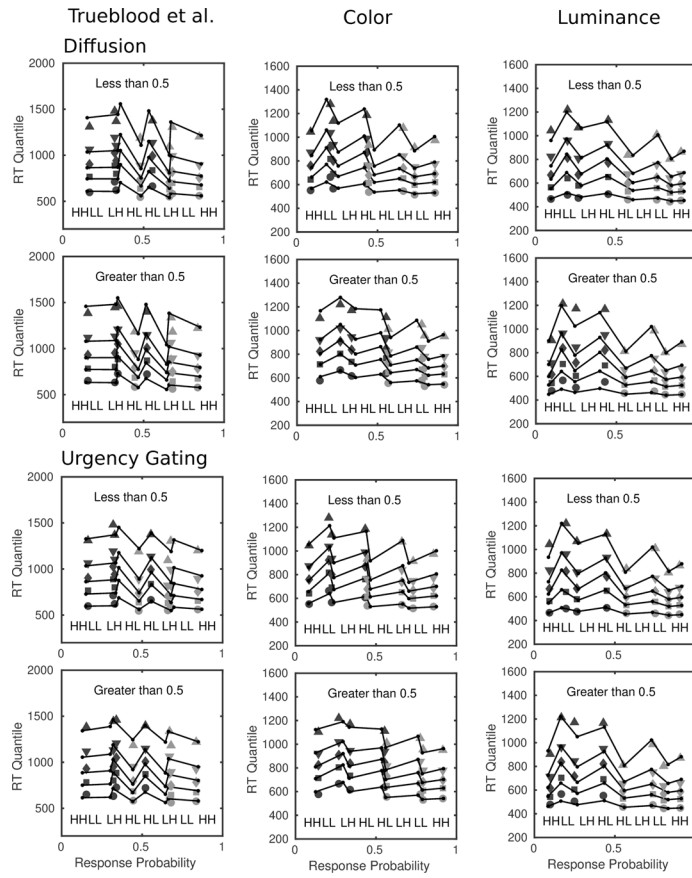
Figure 9:

Quantile probability plots of the diffusion model DIFF(β, η, s_t) and urgency-gating model UG(γ, b, m, s_t) to Experiment 1 of Trueblood et al. (2021) and the color and luminance replications. The data are conditioned on the response ("less than 0.5" and "greater than 0.5"). The pure low-discriminability and high-discriminability conditions are denoted LL and HH; the high-to-low and low-to-high stimulus-change conditions are denoted HL and LH, respectively.

Table 1

Parameters of the Decision Models

Parameter	Symbol
Boundary separation speed	a_s
Boundary separation accuracy	a_a
Starting point speed	z_s
Starting point accuracy	z_a
Mean drift rate	ν_i
Nondecision time	T_{er}
Drift rate variability	η
Starting point variability	s_z
Nondecision time variability	s_t
Infinitesimal standard deviation	σ
OU decay rate	γ
Urgency offset	b
Urgency growth rate	m
Drift rate growth	β^a

Note. a = Flashing-grid task only

Table 2

Models for the Ratcliff (2008) Numerosity Study

Model	p	Properties
1	12	Standard diffusion, DIFF(η, s_z, s_t)
2	11	Pure urgency-gating, UG(γ, m, s_t); $m = 1$
3	12	Extended urgency-gating, UG(γ, b, m, s_t); $b = 1$
4	13	Instruction-dependent urgency, UG(γ, b, m_s, m_a, s_t); $b = 1$

Common parameters

$a_s, a_a, z_s, z_a, v_1, v_2, v_3, v_4, T_{er}; \sigma = 0.1$

Note. p = number of free parameters

Table 3

Fit Statistics for the Ratcliff (2008) Numerosity Study

Model	Properties	G^2	df	AIC	BIC	#AIC	#BIC
1	DIFF(η,s_z,s_t)	820.5	164	844.5	934.7	—	—
2	UG(γ,m,s_t)	1,972.0	165	1,994.0	2,076.7	1	1
3	UG(γ,b,m,s_t)	828.6	164	852.6	942.6	5	5
4	UG(γ,b,m_s,m_a,s_t)	766.8	163	792.8	890.5	6	6

#AIC and #BIC are numbers favoring the indicated model to Model 1 out of 19

Table 4

Parameters for the Ratcliff (2008) Numerosity Study

Model	Properties	a_s	a_a	ν_1	ν_2	ν_3	ν_4	z_s	z_a
1	DIFF(η,s_z,s_t)	0.076	0.132	0.516	0.423	0.287	0.097	0.040	0.068
2	UG(γ,m,s_t)	0.026	0.049	0.267	0.224	0.155	0.053	0.014	0.025
3	UG(γ,b,m,s_t)	0.076	0.134	0.370	0.303	0.202	0.066	0.040	0.068
4	UG(γ,b,m_s,m_a,s_t)	0.073	0.135	0.372	0.304	0.203	0.067	0.038	0.069

		γ	m_s	m_a	η	s_z	T_{er}	s_t
1	DIFF(η,s_z,s_t)	—	—	—	0.137	0.046	0.340	0.133
2	UG(γ,m,s_t)	0.012	—	—	—	—	0.188	0.096
3	UG(γ,b,m,s_t)	0.780	0.794	—	—	—	0.318	0.124
4	UG(γ,b,m_s,m_a,s_t)	0.861	0.182	0.923	—	—	0.319	0.122

Note. $m_S = m$ for urgency models with a single rate parameter.

Table 5

Models for the Smith et al. (2004) Spatial Cuing Study

Model	p	Properties
1	36	Standard diffusion, $\text{DIFF}(\eta, s_t)$
2	36	Pure urgency-gating, $\text{UG}(\gamma, m, s_t)$; $m = 1$
3	40	Extended urgency-gating, $\text{UG}(\gamma, b, m, s_t)$; $b = 1$
Common parameters		
		$a_i, v_{1i}, v_{2i}, v_{3i}, v_{4i}, v_{5i}, T_{er,i}, z_i = a_i/2, i = 1,...,4, \sigma = 0.1$

Note. p = number of free parameters summed across four conditions

Table 6

Fit Statistics for the Smith et al. (2004) Spatial Cuing Study

Model	Properties	G^2	df	AIC	BIC	#AIC	#BIC
1	DIFF(η,s_t)	449.0	184	561.0	704.6	—	—
2	UG(γ,m,s_t)	790.1	184	862.1	1,005.8	0	0
3	UG(γ,b,m,s_t)	604.2	180	684.2	843.8	2	2

#AIC and #BIC are numbers favoring the indicated model to Model 1 out of 6

Table 7

Parameters for the Smith et al. (2004) Spatial Cuing Study

Model	Properties	a_i	ν_{1i}	ν_{2i}	ν_{3i}	ν_{4i}	ν_{5i}
1	DIFF(η,s_t)	0.113	0.067	0.206	0.337	0.429	0.504
2	UG(γ,m,s_t)	0.046	0.034	0.102	0.172	0.217	0.251
3	UG(γ,b,m,s_t)	0.123	0.042	0.124	0.216	0.282	0.336

		γ	m	η	T_{er}	s_{ti}
1	DIFF(η,s_t)	—	—	0.184	0.380	0.136
2	UG(γ,m,s_t)	0.027	—	—	0.195	0.062
3	UG(γ,b,m,s_t)	2.307	1.236	—	0.344	0.083

Table 8

Models for the Flashing Grid Task

Model	p	Properties
1	11	Standard diffusion, DIFF(η, s_t)
2	12	Inhomogeneous diffusion, DIFF(β, η, s_t)
3	12	Urgency-gating, decay, UG(γ, b, m, s_t); $b = 1$
4	11	Urgency-gating, no decay, UG(b, m, s_t); $b = 1$

Common parameters

a, z, $\nu_1, \dots \nu_4$, c, T_{er}, $\sigma = 0.1$

ν_1=high; ν_2=low; (ν_1, ν_3)=high→low; (ν_2, ν_4)=low→high

Note. p = number of free parameters

Table 9

Fit Statistics for the Flashing Grid Task

Model	Properties	G^2	df	AIC	BIC	#AIC	#BIC	#BOOT
Trueblood et al., Experiment 1								
1	DIFF(η,s_t)	135.9	77	155.9	205.1	—	—	
2	DIFF(β,η,s_t)	127.7	76	149.6	203.7	6	5	
3	UG(γ,b,m,s_t)	119.8	76	141.8	195.9	7	7	5
4	UG(b,m,s_t)	124.6	77	144.6	193.8	6	8	
Color								
1	DIFF(η,s_t)	169.9	77	189.9	241.8	—	—	
2	DIFF(β,η,s_t)	143.5	76	165.5	222.6	11	10	
3	UG(γ,b,m,s_t)	113.5	76	135.5	192.6	10	10	9
4	UG(b,m,s_t)	135.4	77	155.4	207.3	11	14	
Luminance								
1	DIFF(η,s_t)	152.9	77	172.9	224.6	—	—	
2	DIFF(β,η,s_t)	138.9	76	160.9	217.8	9	5	
3	UG(γ,b,m,s_t)	137.5	76	159.5	216.4	7	7	4
4	UG(b,m,s_t)	142.1	77	162.1	213.8	8	13	

Note. Comparisons are 2 vs 1; 3 vs. 2; 4 vs. 3; table entries are numbers out of 12, 16, and 16, respectively, for the three experiments favoring the second model of a pair.

Table 10

Parameters for Flashing Grid Task

Model	Properties	a	z	ν_1	ν_2	ν_3	ν_4	c
Trueblood et al., Experiment 1								
1	DIFF(η,s_t)	0.158	0.084	0.176	0.0751	−0.043	−0.129	0.012
2	DIFF(β,η,s_t)	0.150	0.080	0.166	0.0727	−0.044	−0.125	0.011
3	UG(γ,b,m,s_t)	0.153	0.080	0.144	0.0622	−0.038	−0.089	0.004
4	UG(b,m,s_t)	0.187	0.098	0.134	0.0577	−0.043	−0.091	0.007
Color								
1	DIFF(η,s_t)	0.127	0.072	0.251	0.102	−0.061	−0.277	0.054
2	DIFF(β,η,s_t)	0.112	0.063	0.210	0.089	−0.126	−0.260	0.047
3	UG(γ,b,m,s_t)	0.167	0.093	0.191	0.082	−0.050	−0.133	0.044
4	UG(b,m,s_t)	0.187	0.103	0.182	0.078	−0.059	−0.146	0.037

		β	γ	m	η	T_{er}	s_t
Trueblood et al., Experiment 1							
1	DIFF(η,s_t)	—	—	—	0.095	0.438	0.271
2	DIFF(β,η,s_t)	67.97	—	—	0.082	0.359	0.244
3	UG(γ,b,m,s_t)	—	3.94	0.471	—	0.399	0.167
4	UG(b,m,s_t)	—	—	0.482	—	0.384	0.163
Color							
1	DIFF(η,s_t)	—	—	—	0.130	0.477	0.294
2	DIFF(β,η,s_t)	38.92	—	—	0.070	0.354	0.270
3	UG(γ,b,m,s_t)	—	2.56	1.209	—	0.394	0.169
4	UG(b,m,s_t)	—	—	1.188	—	0.398	0.203
Luminance							
1	DIFF(η,s_t)	0.129	0.065	0.261	0.118	−0.018	−0.210	−0.034
2	DIFF(β,η,s_t)	0.122	0.062	0.241	0.112	−0.040	−0.212	−0.027
3	UG(γ,b,m,s_t)	0.133	0.067	0.225	0.103	−0.030	−0.165	−0.022
4	UG(b,m,s_t)	0.164	0.083	0.210	0.097	−0.032	−0.171	−0.025

		β	γ	m	η	T_{er}	s_t
Luminance							
1	DIFF(η,s_t)	—	—	—	0.090	0.369	0.205
2	DIFF(β,η,s_t)	76.72	—	—	0.073	0.318	0.205
3	UG(γ,b,m,s_t)	—	4.71	0.685	—	0.340	0.155
4	UG(b,m,s_t)	—	—	0.759	—	0.327	0.163