Psychological Review
2016, Vol. 123, No. 4, 425–451

0033-295X/16/$12.00 http://dx.doi.org/10.1037/rev0000023

Diffusion Theory of Decision Making in Continuous Report

Philip L. Smith
The University of Melbourne

I present a diffusion model for decision making in continuous report tasks, in which a continuous, circularly distributed, stimulus attribute in working memory is matched to a representation of the attribute in the stimulus display. Memory retrieval is modeled as a 2-dimensional diffusion process with vector-valued drift on a disk, whose bounding circle represents the decision criterion. The direction and magnitude of the drift vector describe the identity of the stimulus and the quality of its representation in memory, respectively. The point at which the diffusion exits the disk determines the reported value of the attribute and the time to exit the disk determines the decision time. Expressions for the joint distribution of decision times and report outcomes are obtained by means of the Girsanov change-of-measure theorem, which allows the properties of the nonzero-drift diffusion process to be characterized as a function of a Euclidian-distance Bessel process. Predicted report precision is equal to the product of the decision criterion and the drift magnitude and follows a von Mises distribution, in agreement with the treatment of precision in the working memory literature. Trial-to-trial variability in criterion and drift rate leads, respectively, to direct and inverse relationships between report accuracy and decision times, in agreement with, and generalizing, the standard diffusion model of 2-choice decisions. The 2-dimensional model provides a process account of working memory precision and its relationship with the diffusion model, and a new way to investigate the properties of working memory, via the distributions of decision times.

Keywords: decision making, diffusion process, working memory, short-term memory (STM), continuous report

Since its introduction by Wilken and Ma (2004), the continuous report task has become an influential experimental paradigm in visual short-term memory (VSTM), or working memory, research. It has, in particular, played a central role in recent attempts to test between discrete item, or "slot," models and continuous resource models of VSTM capacity (van den Berg, Awh, & Ma, 2014). The former assume that VSTM capacity depends on the number of distinct items it can hold; the latter assume that capacity depends on an allocatable resource—much like that proposed in the capacity theory of attention of Kahneman (1973). Unlike more traditional experimental paradigms, in which the contents of memory are probed using a two (or more) alternative forced-choice task, in the continuous report task participants recall the value of a continuously distributed attribute stored in memory by matching its value to a representation of the stimulus in the display by means of an analogue device, such as a computer mouse or a trackball.

Figure 1 shows a typical example of the continuous report task, reproduced from Zhang and Luck (2008). The stimuli in their task consisted of colored patches of varying hue, defined by their position on a color wheel. On each trial, a memory set, consisting of a variable number of color patches was presented. After a 900 ms retention interval, participants reported the color of a cued item in memory by clicking a computer mouse on a representation of the color wheel in the display. Performance was quantified as the angular error between the actual and reported value of the attribute. Aggregate performance across trials was characterized using a measure of report precision, which is the reciprocal of the *SD* of the angular error of the report.

In a widely used variant of the task, the attribute to be reported is orientation rather than color. Like color, orientation is defined on a closed, circular domain, $0 < \theta \leq 360^{\circ}$. The advantage of using stimulus attributes like color or orientation that are defined on circular domains is that it avoids any problems of end effects associated with open continua, and allows performance to be characterized using a single measure of report precision. Versions of the continuous report task have been used in recent studies by Bays, Catalao, and Husain (2009); Bays, Gorgoraptis, Wee, Marshall, and Husain (2011); Fan and Turk-Brown (2013); Fougnie and Alvarez (2011); Gunseli, van Moorselaar, Meeter, and Olivers (2015); Kool, Conway, and Turk-Browne (2014); Marshall and Bays (2013); Rademaker, Tredway, and Tong (2012); and Swan and Wyble (2014), among others. The task was the focus of a recent review of progress in VSTM research by Ma, Husain, and Bays (2014).

Methodologically, the continuous report task can be viewed as a version of the method of adjustment of classical psychophysics.

This article was published Online First March 7, 2016.

The research in this article was supported by Australian Research Council Discovery Grant DP140102970. Background work for the research was carried out during a sabbatical visit to the Department of Psychology at Vanderbilt University in 2014. I thank the members of the Department for their hospitality during this time and, in particular, Gordon Logan, for essential research support. I thank Gordon Logan, Jeffrey Rouder, David Sewell, James Townsend, and an anonymous reviewer for helpful comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Philip L. Smith, Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC 3010, Australia. E-mail: philipls@unimelb.edu.au

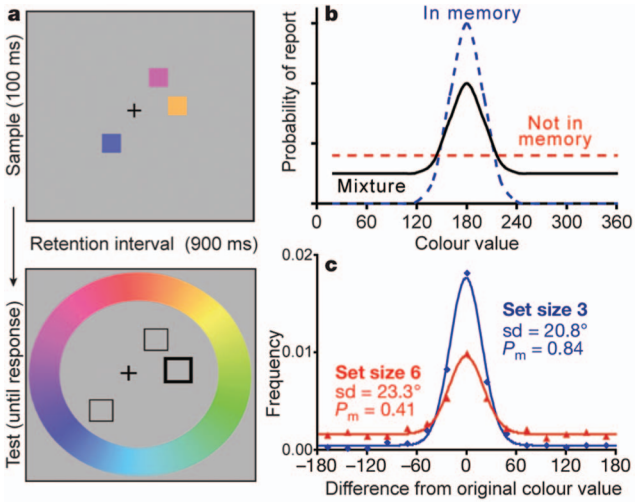

Figure 1. (a). Continuous report task. A memory set comprising a variable number of colored squares is presented and stored in memory. At the end of the retention interval, one of the display locations is probed. Participants report their memory of the probed item by clicking a mouse on a color wheel in the probe display. (b) and (c). Model and data. Distributions of report errors have a peaked, high-tailed form, symmetrical around the true stimulus value, which can be modeled either as a mixture of a von Mises distribution and a uniform distribution or as a mixture of von Mises distributions with different concentration parameters. From "Discrete Fixed-Resolution Representations in Visual Working Memory," by W. Zhang & S. J. Luck, 2008, *Nature, 453,* p. 233. Copyright 2008 by Macmillan Publishers Ltd. Reprinted with permission.

Along with the method of limits and the method of constant stimuli, the method of adjustment was one of three methods used historically to investigate sensory thresholds (Woodworth & Schlosberg, 1954). In a typical application of the method of adjustment, a point of subjective equality was elicited by asking an observer to adjust the perceived magnitude of a variable stimulus until it matched a standard. In its use of a response continuum, the method of adjustment differed from the binary responses required by the other methods. Of the other two methods, the method of constant stimuli came to be preferred over the method of limits, because it used randomized rather than predictable sequences of stimuli intensities, and so avoided the problems of bias and anticipation associated with the latter. The method of constant stimuli is the predecessor of the modern two-alternative forced-choice (2AFC) task, which is now ubiquitous in cognitive psychology.

Historically, the method of adjustment fell into disuse because it was seen as confounding sensory variability and motor variability. Green and Swets (1966, p. 122) noted that: "The method of adjustment is economical, but contaminated by 'uncertainty of the hand'." Engen (1971, p. 23) similarly noted: "He (sic) may overshoot what seems to him at the moment the equality point, and so motor skills may play an important role in the judgments, as might the amount of time the observer devotes to each judgment." Recent VSTM research has typically ignored this problem, but van den Berg et al. (2014) sought to address it in their modeling study by adding a component of variability to represent imprecision in motor output.

In this article, I present a diffusion process model of performance in the continuous report task. My aim in developing the model was to link continuous report theoretically and empirically to the successful diffusion model of 2AFC decision making (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Smith & Ratcliff, 2015a), in which evidence accumulation is modeled as a Brownian motion, or Wiener, diffusion process. My theoretical orientation in developing the model is to regard continuous report as a form of decision task, similar in kind to the 2AFC decision task, but in which there are an indefinite number of decision alternatives available, rather than just two. The idea that memory retrieval involves a decision process is of course not a new one: The diffusion model was first proposed by Ratcliff (1978) as a model of memory retrieval. Pearson, Raskevicius, Bays, Pertzov, and Husain (2014) recently argued that VSTM, or working memory, retrieval should be regarded as a decision process and Smith and colleagues (Smith & Ratcliff, 2009; Smith & Sewell, 2013) proposed a general theory of attention and decision making in which decisions about stimuli in VSTM are modeled as a diffusion process.

The diffusion model represents decision making as a stochastic evidence accumulation process, in which the decision-maker accumulates noisy evidence until a criterion amount of evidence has been obtained. The particular criterion that is reached determines the response and the time to reach criterion determines the decision time. Performance depends jointly on the quality of the information in the stimulus, described mathematically as the *drift rate* of the diffusion process, and the choice of the criteria, which are assumed to be under the decision-maker's control and to reflect the speed–accuracy trade-off settings for the task (Luce, 1986, chap. 6). Because it takes longer to accumulate a criterion amount of evidence when criteria are higher, and because the process is less likely to terminate at the incorrect criterion, higher criteria lead to slower and more accurate responding.

The principal theoretical advantage of the diffusion model, as has been described in many places (e.g., Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004, 2015; Smith & Ratcliff, 2004, 2015a), is that it successfully predicts all of the main features of performance from speeded two-alternative decision tasks. It predicts distributions of response times (RT) for correct responses and errors, the relationship between RT and accuracy, and the relative speeds of mean RTs for correct responses and errors, as functions of the experimental conditions. The model has been successfully applied to a variety of tasks investigating basic cognitive processes in perception, attention, memory, and language, and to a large number of tasks investigating cognitive deficits in special participant populations, such as the elderly, the sleep-deprived, and dyslexics (Ratcliff & Smith, 2015; Ratcliff, Smith, & McKoon, 2015).

In this article, I model the continuous report task as a two-dimensional (2D) diffusion process with drift on a circular domain. Despite its comparative complexity, the model is surprisingly analytically tractable. I obtain explicit, analytic expressions for choice probabilities—viewed as elements of a continuous outcome set—mean decision times, and the associated decision-time distributions. My goal in this theoretical development is to show how 2AFC decision making and continuous report can be viewed as two different manifestations of a psychological process of evidence accumulation to a criterion and to characterize the relationship between the two tasks in a precise way. The use of the method of adjustment and the method of constant stimuli in classical

psychophysics as equivalent ways of eliciting sensory thresholds presupposed that two-alternative decisions and continuous report were alternative ways of measuring and characterizing the same underlying psychological processes. The 2D diffusion model shows precisely in what way this is true.

Although RTs have typically not been measured in recent continuous report studies, the quote from Engen (1971) above emphasizes that continuous report has an intrinsic time component, which can be expected to covary with the accuracy of the report. Rather than being a nuisance variable, or a source of experimental confounds, the distribution of decision times in the 2D model is potentially a rich source of information about the underlying decision process, just as it is in 2AFC decisions. Because the model predicts entire distributions of decision times, not just means, and these have a simple and tractable form, the 2D model offers researchers a new set of tools for characterizing performance in VSTM tasks. I address the practical issue of how one might obtain distributions of RT in continuous report tasks in the Discussion. The following six points summarize the main theoretical results in the article:

1. **Von Mises distribution of report outcomes.** In the 2D diffusion model, the predicted choice probabilities form a continuous distribution of report accuracy on a circular domain $(0,2\pi)$ and follow a von Mises distribution, which is a circular counterpart of the normal distribution. Report accuracy in the VSTM literature is typically modeled using discrete or continuous mixtures of von Mises distributions (e.g., Bays et al., 2009; Fougnie & Alvarez, 2011; van den Berg, Shin, Chou, George, & Ma, 2012; van den Berg et al., 2014; Zhang & Luck, 2008). The 2D diffusion model provides a theoretical rationale for why von Mises distributions should be observed empirically. Rather than just being a convenient quantitative summary of the distribution of report, in the 2D model the von Mises distribution is a signature of the underlying retrieval process.

2. **Theoretical decomposition of precision.** Predicted report accuracy in the von Mises distribution is controlled by a precision or concentration parameter. The 2D diffusion model provides a theoretical decomposition of precision into a product of two terms, one representing the quality of the information in the stimulus and the other representing the amount of evidence needed for a response. Precision in the 2D model is thus an expression of the same psychological processes that determine performance in models of 2AFC decision making. The decomposition of the von Mises concentration parameter into evidence quality and evidence quantity parallels a similar decomposition of detection sensitivity for diffusion and random walk models, originally derived by Link (1975).

3. **Independence of decision times and decision outcomes.** All the main features of the standard diffusion model, in which decision making is modeled as a Wiener process diffusion on a line (i.e., 1D diffusion), carry over to the 2D model. If the process is unbiased and if the only source of variability is moment-to-moment diffusive variability in the decision process, both models predict that decision times will be independent of decision outcomes. The 1D model predicts identical distributions of decision times for correct responses and errors; the 2D model predicts identical distributions of decision times for all report outcomes.

4. **Slow and fast errors.** In the 1D diffusion model, variability in drift rates results in slow errors; variability in decision criteria results in fast errors. In the 2D model, errors are not defined in a categorical way because the model has a continuous outcome set. Instead, the relationship can be characterized by looking at the relationship between report accuracy and decision time in a continuous way. The 2D model then exhibits the same properties as the 1D model: Trial-to-trial variability in drift rate leads to fast accurate responses and slow inaccurate responses; trial-to-trial variability in decision criterion leads to slow accurate responses and fast inaccurate responses.

5. **Distributions of decision times.** The 2D model, like the 1D model, predicts unimodal, positively skewed distributions of decision times. Also like the 1D model, distributions of decision times from slower experimental conditions are related to distributions from faster conditions, to a close approximation, by dilations of the time scale. When the only source of variability in the model is diffusive variability, predicted families of decision time distributions, obtained by varying a stimulus or a response parameter and represented in the form of a quantile-quantile (Q-Q) plot, are strongly linear.

6. **Von Mises mixture distributions.** Trial-to-trial variability in either drift rate or decision criterion leads to distributions of report outcomes that are mixtures of von Mises distributions. Either discrete and continuous mixture distributions are possible, depending on the assumptions of the model. A zero-drift process, in which there is no effective stimulus information and the retrieval process is driven solely by decision noise, leads to a uniform distribution of report outcomes. This provides a single-process account of a "guessing" process, as used to fit data by Bays et al. (2009, 2011); Zhang and Luck (2008), and others. Mixtures of drift rates and mixtures of decision criteria lead to similar distributions of report outcomes, but qualitatively different families of decision time distributions that can be used to distinguish between them in RT data.

A Note on Terminology

I use the term "standard diffusion model" to refer to the model introduced by Ratcliff (1978) and developed subsequently by him and his colleagues (Ratcliff & McKoon, 2008). This model has been implemented in several freely available software packages and used by many applied researchers to distinguish among the effects of stimulus quality, decision criterion, and nondecision times. Ratcliff and Smith (2015) and Ratcliff et al. (2015) recently

reviewed some of these applications. The model assumes within-trial diffusive variability in evidence accumulation, augmented with several sources of across-trial variability. Within-trial variability is represented as a 1D Wiener process with constant drift rate and constant diffusion coefficient. Drift rates, decision criteria—represented by the starting point of the process—and nondecision times vary across trials. I refer to this model and related models as "1D diffusion" models, although they are typically depicted in a 2D plot, with the horizontal axis representing time and the vertical axis representing accumulated evidence. The terminology refers to the evidence states in the model, which are points on a line.

Continuous Report as a 2D Diffusion Process

Figure 2 shows the main elements of the 2D diffusion model. The decision process is represented as a 2D diffusion process on a disk, which is a subset of the Cartesian plane, $\mathbb{R}^2$. The process, viewed as a function of time, is denoted $\mathbf{X}_t$ and its components, in Cartesian coordinates, are denoted (X_t^1, X_t^2). For simplicity, the components are assumed to be independent of each other with common infinitesimal *SD* σ, or otherwise expressed, with diffusion coefficient σ^2. The average rate of evidence growth depends on a drift vector, $\boldsymbol{\mu}$, with components (μ_1, μ_2). Because drift is a vector quantity, it has both a direction, or phase angle, and a magnitude, or norm. Together, these two quantities define the value of the drift rate in polar coordinates. The phase angle is the inverse tangent of the ratio of the vertical and horizontal components of the drift rate, $\theta_\mu = \arctan(\mu_2/\mu_1)$, and its magnitude is the Euclidean norm, $\|\boldsymbol{\mu}\| = \sqrt{\mu_1^2+\mu_2^2}$. In applications, the phase angle of the drift rate depends on the identity of the stimulus (i.e., the color or orientation of the probed element of the memory set) and the magnitude depends on how well represented the item is in memory. The quality of the representation will depend on memory set size and on other experimental variables that affect VSTM trace strength, such exposure duration, retention interval, and use of backward masks.

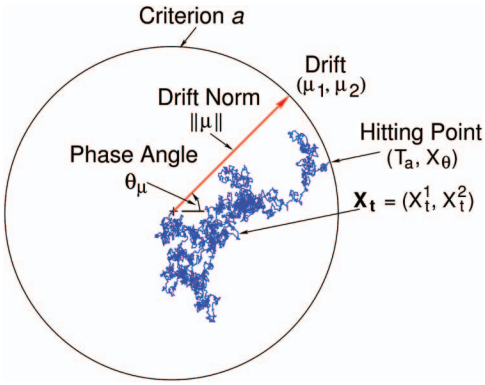

Figure 2. Diffusion model of continuous report. Memory retrieval is modeled as a two-dimensional (2D) diffusion process on a disk, of radius a, whose bounding circle represents the decision criterion. The process $\mathbf{X}_t = (X_t^1, X_t^2)$ consists of two independent components, both with infinitesimal *SD* σ. The drift rate, $\boldsymbol{\mu} = (\mu_1, \mu_2)$, is vector-valued, with length, or norm, $\|\boldsymbol{\mu}\|$ and phase angle θ_μ. The phase angle specifies the direction of the drift vector in polar coordinates. A decision is made when the process $\mathbf{X}_t$ reaches the criterion. The decision time, T_a, is the time to reach the boundary and the response, X_θ, is the hitting point on the boundary.

The process $\mathbf{X}_t$ satisfies the stochastic differential equation

$$\mathbf{dX}_t = \boldsymbol{\mu} dt + \sigma \mathbf{dW}_t, \tag{1}$$

or in components,

$$\begin{bmatrix} dX_t^1 \\ dX_t^2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} dt + \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dW_t^1 \\ dW_t^2 \end{bmatrix}, \tag{2}$$

where $\mathbf{dW}_t = [dW_t^1, dW_t^2]^T$ is the differential of a 2D Brownian motion, or Wiener process, and where T denotes the matrix transpose. The quantity $\mathbf{dW}_t$ describes the horizontal and vertical components of the random change in the process $\mathbf{X}_t$ during a small interval of length dt. A discussion of the use of stochastic differential equations in diffusion models of decision making may be found in Smith (2000).

The process $\mathbf{X}_t$ that solves Equations 1 and 2 is a 2D Brownian motion, or Wiener process, with drift. A simulated realization, or sample-path, of the process on a single experimental trial is shown in Figure 2. The sample paths of the process are continuous but highly irregular. Mathematically, the sample paths are almost nowhere differentiable (i.e., the set on which they are differentiable has probability measure zero). They also possess a self-similar, or fractal, property, that is, they exhibit the same irregularity on any time scale, regardless of how highly the process is magnified. Like the 1D Wiener process in Ratcliff's (1978) model, the process $\mathbf{X}_t$ is a Gaussian process whose mean and variance both grow linearly in time. For the 2D process, the mean, or expected value, vector and the covariance matrix are

$$E(\mathbf{X}_t) = \boldsymbol{\mu} t \tag{3}$$

$$\mathrm{cov}(\mathbf{X}_t) = \sigma^2 \mathbf{I} t, \tag{4}$$

where $\mathbf{I}$ is the identity matrix. In the psychological model, the diffusive variability in $\mathbf{X}_t$ represents noise in the VSTM retrieval and decision process. A standard reference for multidimensional diffusion processes is Stroock and Varadhan (1979).

On any trial of an experiment, probing a location in the poststimulus array initiates a noisy retrieval process. The accumulating evidence is represented by a diffusion process starting at the origin, (0, 0). Evidence accumulates until the process hits a circular criterion boundary, with radius a. The value of a is under the decision-maker's control and reflects the speed-accuracy settings for the task. Like the 1D model, larger values of a mean that more evidence is required before a decision is made.

Evidence is accumulated until the process $\mathbf{X}_t$ hits the boundary, a. The time at which this occurs is a random variable, T_a. (I use lower case letters to denote fixed values and upper case letters to denote random variables, that is, values that depend on the particular realization of the process. Lower case t denotes a fixed value of time; upper case T denotes a random time.) The point at which the process hits the boundary is also random. The value of the process at the point at which it hits the boundary is denoted $\mathbf{X}_T$. The angle at which it exits the bounded region is likewise a random variable, θ_T, equal to $\theta_T = \arctan(X_T^2/X_T^1)$. Performance on the task on any trial is, therefore, a random pair, (T_a, θ_T). The time at which the process hits the boundary determines the decision time and the point where it hits the boundary determines the

reported value of the stimulus attribute. If the phase angle of the drift is a veridical representation of the probed stimulus, then the error of the report is equal to $\theta_T - \theta_\mu$. As in the standard model, I assume that RT is the sum of the decision time, T_a, and a time for other processes, including sensory transduction, response selection, and motor output time. In Ratcliff's (1978) model, the time for other processes is denoted T_{er}, for "encoding and response" processes.

Representational Assumptions of the 2D Model

The model of Figure 2 can be viewed as a generalization of the 1D Wiener process assumed in the standard diffusion model and related models. In the standard model, the accumulating evidence is modeled as a unidimensional process, X_t, that represents the relative strength of evidence for the two decision alternatives. In memory tasks, X_t is interpreted as a measure of familiarity, which expresses the strength of the global match between a probe item and the contents of memory. In perceptual tasks, X_t is a measure of the relative match strengths of a test stimulus to the memory representations of the decision alternatives. The natural domain of application of the 2D model is to stimuli like color, orientation, and motion, which can be represented statistically in a 2D vector space, in which both dimensions of the space are relevant to the judgment. In the 2D model, the drift rate is decomposed into separate phase angle and drift norm components that characterize the identity of the stimulus and how well it is represented in memory. The model presupposes that, in the intended domain of application, the identity of the stimulus and the quality of its representation in memory are independent attributes that can be experimentally manipulated. For typical applications of the continuous report paradigm using color and orientation, such a decomposition is a natural one and an obvious dynamic generalization of multidimensional models of stimulus representation, like multidimensional signal detection theory (Macmillan & Creelman, 2005) and general recognition theory (Ashby & Townsend, 1986), but it will not be so in all applications. Some tasks will not be amenable to being represented as evidence accumulation in a closed circular domain. I discuss these limitations and ways to extend the theory to handle them in the section Extensions and Generalizations of the Model.

For tasks in which the 2D representational assumptions are satisfied, the accumulating evidence process $\boldsymbol{X}_t$ can be given a likelihood ratio interpretation, similar to the one in the 1D diffusion model (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Laming, 1968). In the 1D model with normally distributed stimulus increments, the accumulating evidence is a linear function of the log-likelihood ratio that the observed sequence of evidence was produced by a stimulus with one mean, μ_1, versus a stimulus with another, μ_2. As discussed in the following section, in the 2D model, the accumulating evidence is a linear function of the log-likelihood ratio that the observed sequence of evidence was produced by a stimulus with mean $\boldsymbol{\mu}$ versus a stimulus with mean $\mathbf{0}$ — that is, by an uninformative stimulus. The likelihood ratio (and its logarithm) is maximized when θ_T and θ_μ are equal, which occurs when the accumulating evidence hits the criterion circle a at the point that corresponds to the phase angle of the drift. This point is the most probable point for the accumulating evidence to exit the bounding circle and an observer who reports the exit point as the value of the stimulus attribute is a maximum likelihood observer.

Obtaining Predictions for the 2D Model

The assumption that the evidence accumulation process starts from the origin means that the process must travel the same distance in any direction to reach the boundary circle, a. The assumption that the diffusion coefficient is the same in the horizontal and vertical directions means that the stimulus-independent spread of evidence toward the boundary occurs at the same rate in all directions. Because the drift rate is nonzero, however, the process is more likely to hit the boundary at some points than others. The zero-drift process is radially symmetrical, but the symmetry is broken when the drift rate is nonzero, which makes the process harder to analyze than it would be otherwise. Smith (2000) discussed possible psychological applications of 2D diffusion on a circular domain, but did not obtain predictions for the model.

In this article, I provide an explicit solution to this problem. To do so, I use a theorem from the stochastic processes literature called the Girsanov theorem (or Cameron-Martin-Girsanov theorem; Karatzas & Shreve, 1991; Revus & Yor, 1990; Rogers & Williams, 1987/2000). The Girsanov theorem is a change-of-measure theorem that makes it possible to characterize the relationship between the probability measure (i.e., the distribution) of the nonzero-drift process and that of the zero-drift process. Among its applications, the Girsanov theorem is used in the mathematical finance literature to model derivative pricing (Baxter & Rennie, 1996).

Because of its radial symmetry, the zero-drift process can be fully characterized in terms of the radial, or Euclidean distance process, R_t,

$$R_t = \sqrt{(X_t^1)^2 + (X_t^2)^2}, \tag{5}$$

that describes the distance of the process $\mathbf{X}_t$ from its starting point. The radial process is known as the Bessel process, which is defined on the positive half-line, $\mathbb{R}^+$, and whose properties are well understood (Karlin & Taylor, 1981). The advantage of characterizing the process in terms of the Euclidean distance process is that it allows a 2D problem to be a reduced to a 1D problem, which appreciably simplifies it. The first passage time distribution for the Bessel process gives the probability that the process will have crossed a boundary a by time t. This distribution has an infinite-series representation (Borodin & Salminen, 1996; Hamana & Matsumoto, 2013; Kent, 1978, 1980) that can be used to obtain predicted decision-time distributions. The task of analyzing the 2D model therefore has two parts: first, to obtain first passage time distributions for the zero-drift process, and second, to obtain first passage time distributions for the nonzero-drift process as functions of those for the zero-drift process via the Girsanov theorem. I consider the second of these problems first.

Relating Zero-Drift and Nonzero-Drift Processes

The Girsanov theorem, which characterizes of the relationship between the probability distributions for the zero-drift and nonzero-drift processes, is an abstract change-of-measure theorem that is expressed in the language of martingales. Martingales are

pervasive in modern path-space treatments of stochastic processes, but they will not be familiar to all readers. In general terms, a martingale is a stochastic process whose conditional expectation as a function of time remains constant and which, in addition, satisfies certain technical integrability conditions. As an example of the theoretical importance of martingales, the Wald identity, which was used to analyze discrete random walk models by Laming (1968) and Link and Heath (1975), is a special case of the martingale optional sampling theorem, as was noted by Smith (1990). An accessible treatment of martingale theory can be found in Williams (1991). Because of the abstract nature of the Girsanov theorem, it is not immediately obvious on first encounter how it can be used to obtain first-passage statistics for the nonzero-drift process. To help the reader gain an understanding of the theorem and how it is applied, I give a compact statement of the main result and then show how it applies to the simpler and more familiar 1D model. I then show how the theorem can be used to obtain the first passage statistics of the 2D model. A formal statement of the theorem can be found in Appendix A.

Let $P_t(A)$ be the probability that a d-dimensional, zero-drift, Brownian motion process with unit infinitesimal SD (i.e., $\sigma_i = 1$ for all components i) will be found in a designated set $A \subset \mathbb{R}^d$ at time t. Let $\tilde{P}_t(A)$ be the corresponding probability for a nonzero-drift Brownian motion process, with drift rate γ_t. In both the standard diffusion model and in the theory presented here the drift rate is constant during a trial, but the Girsanov theorem allows it to be a stochastic process, which the time dependency in the notation serves to emphasize. The theorem states that the probability measures for the zero-drift and nonzero-drift processes are related via an exponential martingale, $\boldsymbol{Z}_t(\boldsymbol{X})$,

$$d\tilde{P}_t(A) = \boldsymbol{Z}_t(\boldsymbol{X}) dP_t(A), \tag{6}$$

where

$$\boldsymbol{Z}_t(\boldsymbol{X}) = \exp\left[\sum_{i=1}^{d} \int_0^t \gamma_s^i dX_s^i - \frac{1}{2}\int_0^t \|\gamma_s\|^2 ds\right], \tag{7}$$

and where $d\tilde{P}_t(A)$ and $dP_t(A)$ are the probability density functions corresponding to the probability measures $\tilde{P}_t(A)$ and $P_t(A)$, expressed in differential notation. (The superscript i's in Equation 7 index the dimensions of the process; they are not powers.) An important feature of the change-of-measure in Equation 6 and the key to its power in applications is that it holds at random times, not just at fixed times. It holds, in particular, at times at which a process first exits or enters a designated set, such as when it first hits a boundary or criterion, denoted here as T_a.

The exponential martingale $\boldsymbol{Z}_t(\boldsymbol{X})$ of Equation 7 can be interpreted as the likelihood ratio associated with the evidence accumulation process $\boldsymbol{X}_t$ up to and including time t. The ratio of probability densities

$$\frac{d\tilde{P}_t}{dP_t} = Z_t(\boldsymbol{X}) \tag{8}$$

is known as the Radon-Nikodym derivative of the probability measure $\tilde{P}_t$ with respect to the measure P_t. The Radon-Nikodym derivative is similar to the familiar likelihood ratio of statistics, but instead of describing a finite or countable sequence of noisy observations, it describes the entire history of the continuous process $\boldsymbol{X}_t$. Equation 8 states that the martingale $\boldsymbol{Z}_t(\boldsymbol{X})$ can be interpreted as a likelihood ratio; specifically, it is the likelihood of obtaining the observed evidence process $\boldsymbol{X}_s$, $s \in [0, t]$, given that the drift was γ_s, versus the likelihood of obtaining the same process, given the drift was $\mathbf{0}$, that is, an uninformative process.

Applications to the 1D model. An appreciation of the power of the Girsanov theorem can most easily be gained by considering its application to a two-boundary, one-dimensional diffusion process on the line. Smith (1990) provided expressions for the first passage time density functions for the one-dimensional diffusion process (see also Ratcliff, 1978, Appendix). For a process X_t, starting at $X_0 = 0$, constrained by (possibly asymmetric) boundaries at $-b \le X_t \le a$, with drift μ and unit infinitesimal SD, the first passage time densities at the upper and lower boundaries are

$$d\tilde{P}_t(a) = \exp\left(\mu a - \frac{\mu^2 t}{2}\right)\frac{\pi}{(a+b)^2} \times \sum_{k=1}^{\infty} k \exp\left[-\frac{k^2\pi^2 t}{2(a+b)^2}\right] \sin\left(\frac{k\pi a}{a+b}\right) \tag{9}$$

and

$$d\tilde{P}_t(-b) = \exp\left(-\mu b - \frac{\mu^2 t}{2}\right)\frac{\pi}{(a+b)^2} \times \sum_{k=1}^{\infty} k \exp\left[-\frac{k^2\pi^2 t}{2(a+b)^2}\right] \sin\left(\frac{k\pi b}{a+b}\right). \tag{10}$$

These functions are joint densities: $d\tilde{P}_t(a) = P[X_T = a; T \in dt]$ and $d\tilde{P}_t(-b) = P[X_T = -b; T \in dt]$, where $T \in dt$ means that the first boundary crossing occurs in a small interval $(t, t + dt)$. Summing the joint densities gives the marginal density; integrating them over time gives the probabilities of responding at the upper and lower boundaries, respectively. As in Equation 8, the tilde in the notation indicates that the process has nonzero drift.

Equations 9 and 10 are sums of products of terms. The dependence on drift rate in these equations is confined to the first exponential term; the remaining exponential and sine terms depend on time and boundary settings, but are independent of drift rate. When the drift rate is zero, the initial exponential terms are equal to unity and the expressions become

$$dP_t(a) = \frac{\pi}{(a+b)^2}\sum_{k=1}^{\infty} k \exp\left[-\frac{k^2\pi^2 t}{2(a+b)^2}\right] \sin\left(\frac{k\pi a}{a+b}\right) \tag{11}$$

and

$$dP_t(-b) = \frac{\pi}{(a+b)^2}\sum_{k=1}^{\infty} k \exp\left[-\frac{k^2\pi^2 t}{2(a+b)^2}\right] \sin\left(\frac{k\pi b}{a+b}\right). \tag{12}$$

When the boundaries are equidistant from the starting point, $a = b$, these functions are identical.

The relationship between the first passage densities for the nonzero-drift process in Equations 9 and 10 and those for the zero-drift process in Equations 11 and 12 is given by the Girsanov theorem. For the 1D process, with $\gamma_t = \mu$, the exponential martingale in Equation 7 is simply

$$Z_t = \exp\left(\int_0^t \mu\, dX_s - \frac{1}{2}\mu^2 t\right) = \exp\left(\mu X_t - \frac{1}{2}\mu^2 t\right). \tag{13}$$

Equation 6 states that the first-passage time probability density for the nonzero-drift process is equal to the density for the zero-drift process multiplied by Z_T, the value of Z_t evaluated at X_T, the value of the process at the time at which it first hits a boundary. The value of X_T will be either a or $-b$, depending on whether it first hits the upper or lower boundary. Substituting these values into Equation 6, yields

$$d\tilde{P}_t(a) = \exp\left(\mu a - \frac{1}{2}\mu^2 t\right) dP_t(a) \tag{14}$$

and

$$d\tilde{P}_t(-b) = \exp\left(-\mu b - \frac{1}{2}\mu^2 t\right) dP_t(-b). \tag{15}$$

A comparison of Equations 9 and 11 with Equations 10 and 12 shows they are related in precisely this way. In either case, the drift-dependent exponential terms in Equations 11 and 12 are equal to the exponential martingale in the Girsanov theorem. We can, therefore, obtain the first passage time distributions for the nonzero process via a change of measure on the distributions for a zero-drift process.

Before leaving this example, there is one more important property of the exponential martingale in the Girsanov theorem to highlight. The quantity Z_t in Equation 13 is the product of two terms: The first of these, $\exp(\mu X_t)$, depends on the sign of the drift; the second, $\exp(-\mu^2 t/2)$ does not—it depends on the drift magnitude but not on its sign. This means that, for a given value of drift, the relationship between the predicted distributions of correct responses and errors will depend on the first of these terms but will be independent of the second. Because the first term does not depend on time, when the boundaries are equal the two distributions will differ only by a scale factor, equal to $\exp(\mu a)$ and $\exp(-\mu b)$, respectively. This is the basis of the well-known (or notorious) property of random walk and diffusion models, of predicting equal mean RTs for correct responses and errors when the process is unbiased or the drift rates for the two stimuli are equal in magnitude and opposite in sign (Luce, 1986, pp. 340–344; Townsend & Asbhy, 1983, pp. 294–310).[1] The process $\boldsymbol{Z}_t(\boldsymbol{X})$ that effects the change of measure for the 2D process in Equation 6 has the same properties: Only the first of the terms in the exponential depends on the phase angle of the drift vector; the second depends on the norm of the drift but is independent of its direction.

Change of measure for the 2D model. For the 2D model, the exponential martingale of Equation 6, evaluated at $T = T_a$, the time at which the process hits the boundary at a, is equal to

$$\boldsymbol{Z}_T(\boldsymbol{X}) = \exp\left[(\boldsymbol{\mu} \cdot \boldsymbol{X}_T) - \frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right], \tag{16}$$

where $(\boldsymbol{\mu} \cdot \boldsymbol{X}_T) = \sum_{i=1}^{2} \mu_i X_T^i$ is the dot product of the drift vector and a random vector $\boldsymbol{X}_T$ with norm $\|a\|$ and phase angle θ_T. (In Equation 16 and elsewhere I write T rather than T_a to avoid double subscripts in the notation.) For a process hitting the boundary at θ, the components of X_T at the hitting points are $\boldsymbol{X}_T^{\theta} = (a\cos\theta, a\sin\theta)$, as shown in Figure 2. The norm of the drift vector on the right of Equation 16 is $\|\boldsymbol{\mu}\| = \sqrt{\mu_1^2 + \mu_2^2}$. As in the 1D model, the joint density of the pair (T_a, θ_T) is obtained from Equation 6, as the product of the exponential martingale in Equation 16 and the first passage time density for the Bessel process, which is given in the following section.

As noted previously, $\boldsymbol{Z}_T(\boldsymbol{X})$ is equal to the value of the likelihood ratio at $T = T_a$, the time at which $\boldsymbol{X}_t$ first exits the criterion circle. Evidently, $\boldsymbol{Z}_T(\boldsymbol{X})$ will be maximized when the dot product $\boldsymbol{\mu} \cdot \boldsymbol{X}_T$ is maximized. This will occur when the cosine of the angle between the drift vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and the process vector $\boldsymbol{X}_T = (X_T^1, X_T^2)$ is unity, that is, when the phase angles θ_μ and θ_T of the two vectors are equal. This provides the formal justification for the earlier statement that a maximum likelihood observer reports the hitting point of the criterion circle as the value of the stimulus attribute. I illustrate this property with computational examples in a subsequent section.

An important property of the model, which is implied by the product form of the Girsanov change-of-measure formula, Equation 6, is that the choice probabilities are independent of the conditional RT distributions, and are consequently also independent of the mean decision times. This property was originally proved by Reuter (1959); a simple proof based on the product form of the change in measure in the Girsanov theorem was given by Rogers and Williams (1987/2000, p. 84). A proof based on the one in Rogers and Williams can be found in Appendix B. The standard form of the Girsanov theorem used here assumes that the diffusion coefficient, σ^2, is equal to unity. When the diffusion coefficient is nonunity, the exponential martingale in Equation 16 instead becomes

$$\boldsymbol{Z}_T(\boldsymbol{X}) = \exp\left[\frac{1}{\sigma^2}(\boldsymbol{\mu} \cdot \boldsymbol{X}_T) - \frac{1}{2\sigma^2}\|\boldsymbol{\mu}\|^2 T\right]. \tag{17}$$

This makes the process dimensionally invariant, that is, invariant under a change of measurement scale.

The Bessel Process

The radial, or Euclidean distance, process, R_t, of Equation 5 is a 2D Bessel process. The process is 2D in the sense that it is composed of two independent components but is 1D in the sense that it is scalar-valued rather than vector-valued. Mathematically, a d-dimensional Bessel process satisfies the stochastic differential equation

$$dR_t = \frac{d-1}{2R_t} dt + dB_t, \tag{18}$$

where

[1] The two conditions for equality of mean correct and error RTs reflect two different ways of characterizing correct and error responses in the model. One way involves a comparison of RTs at the upper and lower boundaries, a and $-b$, for a single stimulus and its associated drift rate. The other way involves a comparison of correct and error RTs at a single boundary (either a or $-b$) for a pair of stimuli, each with its own drift rate. Denote by M_{ij}, $i, j \in \{a, b\}$, the mean time to make response j to stimulus i. The model predicts that $M_{aa} = M_{ab}$ for drift rate μ_a and $M_{bb} = M_{ba}$ for drift rate μ_b when $a = b$, that is, when the starting point of the process is equidistant from the boundaries. The process then has an equal distance to travel to reach the correct and error boundaries. As these relationships each involve only a single stimulus, no relationship between the drift rates μ_a and μ_b is implied. The model also predicts $M_{aa} = M_{ba}$ and $M_{bb} = M_{ab}$, when $\mu_a = -\mu_b$, that is, the mean time to make a response at a given boundary correctly is the same as that to make it in error. Equality of means implies that the effects of the two stimuli are equal in magnitude and opposite in sign. As these relationships each involve only a single boundary they continue to hold when the boundaries are unequal. All of these relationships are implied by Equations 9 and 10. Historically, the relationship between mean times for correct responses and errors was first studied, in a quite general way, for discrete-time random walk models. The relevant articles are by Thomas (1975), Swensson and Green (1977), and Laming (1979).

$$dB_t = \sum_{i=1}^{d} \frac{W_t^i}{R_t} dW_t^i \tag{19}$$

is the differential of a standard (zero drift, unit variance) Brownian motion process, B_t, and the terms W_t^i are independent components of a d-dimensional Brownian motion process (Karatzas & Shreve, 1991, p. 159). For the 2D model, the drift rate of the process is simply equal to $1/(2R_t)$. To understand why the sum of terms in Equation 19 should equal dB_t, the differential of a standard Brownian motion, observe that the components of the sum, dW_t^i, are scaled by W_t^i/R_t. For a 2D process, the vertical component of this ratio is equal to a random sine (opposite over hypotenuse) and the horizontal component is equal to a random cosine (adjacent over hypotenuse). As $\sin^2\theta + \cos^2\theta = 1$, the Brownian increments at each instant are weighted by the components of a random vector of norm one. The weighted sum will, therefore, have a mean and variance that are the same as each of its individual components and so will also be a standard Brownian motion process.[2] The transition density of the d-dimensional Bessel process is

$$dP_t(x, y) = \frac{\exp[-(x^2 + y^2)/2t]}{t(xy)^{(d-2)/2}} y^{d-1} I_{(d-2)/2}\left(\frac{xy}{t}\right);\ t > 0;\ x,\ y > 0, \tag{20}$$

(Karlin & Taylor, 1981, p. 238), where

$$I_v(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+v}}{k!\Gamma(k + v + 1)},$$

is the modified Bessel function of the first kind of order v (Abramowitz & Stegun, 1965, p. 375), and $\Gamma(.)$ is the gamma function. The quantity $dP_t(x, y)$ is the probability density that the process starting at x at time 0 will be found at y at time t.

The Laplace transform for the first passage distribution takes the form of a ratio of modified Bessel functions (Getoor & Sharpe, 1979; Kent, 1978) and was inverted by Kent (1980). Kent's results were restated in a more accessible form by Hamana and Matsumoto (2013); related results can be found in Borodin and Salminen (1996, p. 297). For a process starting at the origin, $R_0 = 0$, the probability that the process will hit the boundary a at or before time t is

$$P[T_a \leq t] = 1 - 2\sum_{k=1}^{\infty} \frac{1}{j_{0,k} J_1(j_{0,k})} \exp\left(-\frac{j_{0,k}^2 \sigma^2}{2a^2} t\right). \tag{21}$$

In this equation, $J_1(x)$ is a first-order Bessel function of the first kind and the $j_{0,k}$ terms are the zeros of a zero-order Bessel function, $J_0(x)$, where

$$J_v(z) = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{2k+v}}{k!\Gamma(k + v + 1)},$$

(Abramowitz & Stegun, 1965, p. 360). For each k, $J_1(x)$ must be evaluated at $j_{0,k}$, the k-th zero of $J_0(x)$, as shown in Figure 3. Because there is no analytic expression for the zeros of $J_0(x)$, they must be found numerically, but this is a comparatively straightforward task.[3] In practice, the series must be approximated using a finite number of terms. Because consecutive terms decrease in absolute magnitude exponentially fast as $j_{0,k}^2 t\sigma^2/(2a)^2$, the approximation is good for large values of t with even a small number of terms and improves for small values of t as the number of terms increases. In the applications described here, I truncated the series after 50 terms, which is more than sufficient to ensure convergence. For models like this one, whose predictions are in the form of an infinite series, computational efficiency can be increased by deriving bounds for the truncation error of the series and terminating it appropriately (e.g., Blurton, Kesselmeir, & Gondan, 2012), but my primary concern here is to characterize the model's theoretical properties, not to present a tool for fitting data, so I have not attempted to derive bounds of this kind.

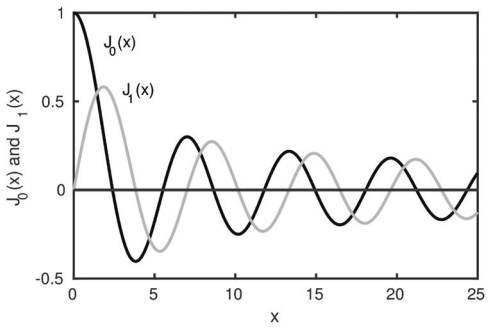

Figure 3. Bessel functions of the first kind, $J_0(x)$ (black) and $J_1(x)$ (gray). To evaluate the infinite series in Equations 21 and 22, the function $J_1(x)$ must be evaluated at consecutive points $j_{0,k}$, the points at which $J_0(x)$ crosses the x-axis.

Differentiating Equation 21 with respect to t gives the first passage time density function,

$$dP_t(a) = \frac{\sigma^2}{a^2} \sum_{k=1}^{\infty} \frac{j_{0,k}}{J_1(j_{0,k})} \exp\left(-\frac{j_{0,k}^2 \sigma^2}{2a^2} t\right). \tag{22}$$

Figure 4 shows a simulation of 50,000 trials of the zero-drift, 2D model, with $\sigma = 1.0$ and $a = 1.0$, on a time scale of $h = .1$ ms. The inset in the figure shows schematically (using four bins as an example) how I performed the simulation. I divided the circular region $\|\boldsymbol{X}_t\| \leq a$ into m bins of width $w = 2\pi/m$ radians, centered on a set of equally spaced phase angles θ_i, $i = 1, \ldots, m$. Processes crossing the boundary in the region $(\theta_i - w/2, \theta_i + w/2)$ in the time interval $(t, t + h)$ were assigned to the i-th bin and the t/h-th time index. The resulting distributions are joint mass-density distributions. They are (approximately) density functions in their time indexes and mass distributions in their phase angles. For plotting purposes, I converted these distributions to joint density distributions in both variables by dividing them by w. For the simulation in Figure 4, 1 set $m = 11$.

The left hand panel of Figure 4 shows simulated joint density functions for the zero-drift process for each phase angle, θ_i. The highly oscillatory behavior of the functions is a result of the fine time scale of the simulation. To reduce these effects, I filtered the functions on the time axis using a cosine-tapered filter. Because the zero-drift process is radially symmetrical, the 11 functions are the same

[2] The drift rate in Equation 18 is undefined for $R_t = 0$. This is a reflection of the fact that, for a 2D process, the origin (0, 0), is an entrance boundary for the process: The process can begin at the origin but cannot attain it again subsequently (Karlin & Taylor, 1981, pp. 232–236).

[3] I obtained zeros of the Bessel function $J_0(x)$ by means of a Matlab routine, besselzero.m, written by G. von Winckel, which is publicly available and downloadable from the MathWorks website.

except for sampling variability. The distributions are unimodal, positively skewed and look much like the predicted RT distributions from the standard diffusion model.

The right hand panels show the mean simulated joint density, averaged over bins, and the theoretical joint density for the Bessel process, $dP_t(a)$, from Equation 22. As can be seen, the predicted and simulated values agree, within the limits of resolution of the simulation.[4] Equation 21, therefore, provides an appropriate tool for obtaining predictions for the zero-drift model. In the following section, I show how these predictions are used with the Girsanov theorem to obtain predictions for the nonzero-drift model.

Decision-Time Predictions of the Nonzero-Drift Model

I denote by $d\tilde{P}_t(\theta_T)$ the joint density function of the nonzero-drift process, written as a function of the phase angle of the hitting point of the boundary. That is, $d\tilde{P}_t(\theta_T) = P[T \in dt, \theta_T \in d\theta]$. This density is the product of the exponential martingale in the Girsanov theorem and the first passage density of the Bessel process through the boundary a. Parameterizing Equation 17 by the phase angle of the hitting point, we obtain

$$d\tilde{P}_t(\theta_T) = \exp\left[\frac{a}{\sigma^2}(\mu_1 \cos\theta_T + \mu_2 \sin\theta_T) - \frac{1}{2\sigma^2}\|\boldsymbol{\mu}\|^2 t\right] dP_t(a), \tag{23}$$

where $dP_t(a)$ is the first passage density for the Bessel process in Equation 22. Figure 5 shows 150,000 simulated trials of the nonzero drift 2D model. The drift vector was $\boldsymbol{\mu} = (1,1)$ (phase angle $\theta = \pi/4$ radians). The infinitesimal *SD* was $\sigma = 1.0$ and the criterion was $a = 1.0$. The top panel shows the simulated distributions; the bottom panel shows the predicted density functions from Equation 23. As before, the predicted and simulated values agree. The methods I have described here therefore provide an effective way of obtaining complete decision-time predictions for the 2D model.

As noted in a previous section, only the first term in the exponential martingale on the right of Equation 23 depends on the phase angle. This means that the distributions in Figure 5 are all scaled copies of each other, with scale factor $K_\theta = \exp(\boldsymbol{\mu} \cdot \boldsymbol{X}_T)$. The conditional distributions, which are obtained by normalizing the masses, will be identical, and so also will be the mean decision times. The property of predicting equal mean RTs for correct responses and errors made by the 1D model therefore carries over to the 2D setting. Subsequently, I show how these predictions are changed when there is variability across trials in drift or criterion.

Mean decision times can be obtained computationally from the density in Equation 23, by conditioning with respect to hitting probability and then taking the expectation with respect to time. An analytic expression for the mean can be obtained from the Laplace transform of the first passage time distribution of the

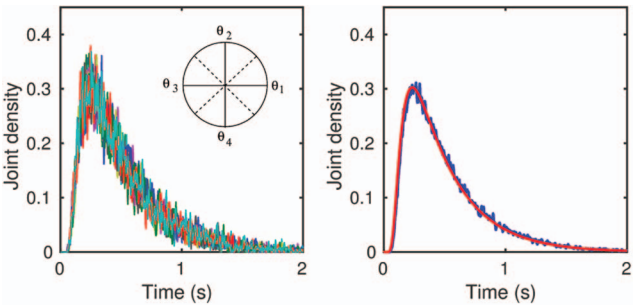

Figure 4. Simulated decision time distributions for the zero-drift process as a function of phase angle and first passage time. The panel on the left shows 11 simulated distributions of a zero-drift process, with $a = 1.0$, $\sigma = 1.0$, for processes terminating in each of 11 bins of width $2\pi/11$. The inset shows schematically how the bins were constructed, using four bins as an example. The panel on the right shows the marginal simulated distribution, averaged over the 11 distributions on the left, together with the predicted distribution for the Bessel process of Equation 22.

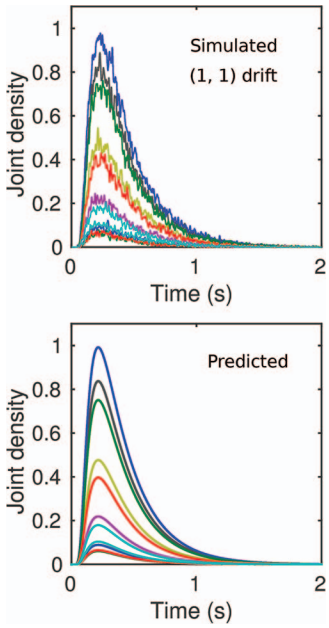

Figure 5. Simulated decision time distributions for a nonzero-drift process as a function of phase angle and hitting time. The top panel shows 11 simulated distributions of a process with $\|\boldsymbol{\mu}\| = \sqrt{2}$, $a = 1.0$ and $\sigma = 1.0$. The bottom panel shows the predicted distributions from Equation 23.

[4] The agreement between the simulated and predicted density functions in Figure 4 is deceptive. Although the simulated and predicted functions coincide to the level of resolution of the figure, the predicted distributions are actually shifted to the left by several milliseconds relative to the simulated distributions. The effects become more pronounced when considering predicted and simulated means, like those in Figures 8 and 9. Even when using a simulation with a small time step (e.g., $h = .001$ s or smaller), predicted mean times can be up to .01 s shorter than simulated times. The reason for the discrepancy is the phenomenon of overshoot of the boundary, discussed by Smith (1990), which arises when approximating a continuous time diffusion process with a discrete time random walk. Whereas a diffusion process terminates immediately on hitting a boundary, a random walk does not terminate until it has taken a step beyond the boundary. The distance traveled by a diffusion to reach a boundary is a, whereas the average distance traveled by a random walk is $a + \zeta/2$, where ζ is the average step size of the walk. Because ζ is of the order $\sqrt{h}\sigma$, surprisingly large discrepancies in mean first passage times can arise, even with small step sizes. For the simulations in Figures 8 and 9, I used a correction factor proposed by Heath and Kelly (1988) and discussed by Smith (1990), in which the criterion in the simulated process is set to $a' = a - \sqrt{h}\sigma/2$. This brings the predicted and simulated means into close alignment.

zero-drift process. In Appendix C, I show that the mean decision time corresponding to Equation 23 has the form

$$\tilde{E}[T_a] = \frac{aI_1(a\|\boldsymbol{\mu}\|/\sigma^2)}{\|\boldsymbol{\mu}\|I_0(a\|\boldsymbol{\mu}\|/\sigma^2)}, \tag{24}$$

where $I_1(.)$ and $I_0(.)$ are modified Bessel functions of the first kind of order one and zero, respectively (Abramowitz & Stegun, 1965, p. 375).

Response Probabilities and the von Mises Distribution of Precision

In empirical studies of continuous report, like the Zhang and Luck (2008) study in Figure 1, the main measure of theoretical interest is the precision of the participants' reports, that is, the reciprocal of the *SD* of the angular error. Predicted precision in the 2D model is given by the marginal distribution of the hitting angle, θ_T, obtaining by integrating over the hitting time. As the previous paragraph noted, predicted precision depends only on the first term in the exponential martingale, and so has the form

$$P[\theta_T \in d\theta] = K_{\|\boldsymbol{\mu}\|}e^{(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)/\sigma^2} = K_{\|\boldsymbol{\mu}\|}\exp\left[\frac{1}{\sigma^2}(a\mu_1\cos\theta + a\mu_2\sin\theta)\right]. \tag{25}$$

The normalizing constant, $K_{\|\boldsymbol{\mu}\|}$, is the integral of the product of the first passage time density for the Bessel process, $dP_t(a)$, from Equation 22 and $\exp[-\|\boldsymbol{\mu}\|^2 t/(2\sigma^2)]$, the phase angle independent component of the exponential martingale from Equation 23. Using Equation A7 in the Appendix, we find the normalizing constant is equal to

$$K_{\|\boldsymbol{\mu}\|} = [2\pi I_0(a\|\boldsymbol{\mu}\|/\sigma^2)]^{-1}. \tag{26}$$

The scaling factor $(2\pi)^{-1}$ arises because the normalizing constant is a probability density on a circle, obtained by distributing the probability mass in Equation C9 uniformly on the range $(0, 2\pi)$.

Figure 6 shows the results of two simulations, each of 50,000 trials of the nonzero-drift process, with drift vectors $\boldsymbol{\mu} = (1, 1)$ and $(-1, 1)$, respectively (i.e., phase angles of $\pi/4$ and $3\pi/4$ radians). To better show the distribution of hitting points, I grouped the phase angles into 100 bins. The figure shows the simulated hitting point densities, obtained by summing across hitting times in each bin, together with the predicted densities from Equation 25, with the normalization constant appropriately chosen. Again, the predicted and simulated results agree. The methods presented here, therefore, also provide a way to obtain exact predictions for the distribution of precision.

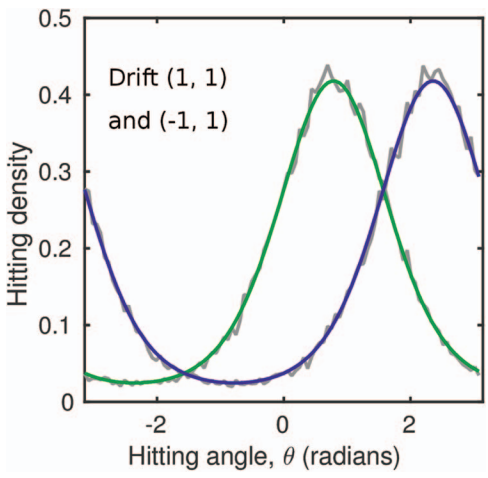

Figure 6. Predicted and simulated distributions of report for a nonzero-drift diffusion process. The distribution of report is the distribution of hitting points, parameterized by phase angle. The parameters of the process were $\|\boldsymbol{\mu}_1\| = \|\boldsymbol{\mu}_2\| = \sqrt{2}$, $\theta_{\mu,1} = \pi/4$ (green), $\theta_{\mu,2} = 3\pi/4$ (purple).

Precision is usually characterized empirically using a von Mises distribution, which approximates a circular normal distribution[5] (Fisher, 1993, p. 48). The von Mises distribution is a two-parameter family, which depends on a concentration or precision parameter, κ, and a mean or location parameter, ϕ. The von Mises probability density function, $f(\theta; \kappa, \phi)$, has the form

$$f(\theta; \kappa, \phi) = \frac{e^{\kappa\cos(\theta-\phi)}}{2\pi I_0(\kappa)} = \frac{e^{\kappa\cos(\theta-\phi)}}{\int_0^{2\pi} e^{\kappa\cos(\vartheta-\phi)}d\vartheta}. \tag{27}$$

The normalizing constant, $I_0(\kappa)$ is a modified Bessel function of the first kind of order zero. The form of the denominator on the right of the second equality emphasizes that the normalizing constant is simply the integral of the exponential term with respect to phase angle, taken around the perimeter of the circle.

Rogers and Williams (1987/2000, p. 85) note that the distribution of hitting points on a circle for a 2D diffusion process follows a von Mises distribution. This result is an important one theoretically, so I show it explicitly. For the diffusion model, the distribution of hitting points is given by Equation 25. We set $\kappa\cos\phi = a\mu_1/\sigma^2$, $\kappa\sin\phi = a\mu_2/\sigma^2$ and obtain

$$\begin{aligned} P[\theta_T \in d\theta] &= K_{\|\boldsymbol{\mu}\|}\exp(\kappa\cos\theta\cos\phi + \kappa\sin\theta\sin\phi) \\ &= K_{\|\boldsymbol{\mu}\|}\exp[\kappa\cos(\theta-\phi)], \end{aligned} \tag{28}$$

where the second equality uses elementary trigonometric identities and where the precision is $\kappa = a\sqrt{\mu_1^2 + \mu_2^2}/\sigma^2$. Substituting the value of $K_{\|\boldsymbol{\mu}\|}$ from Equation 26 gives

$$P[\theta_T \in d\theta] = \frac{1}{2\pi I_0(\kappa)}\exp[\kappa\cos(\theta-\phi)],$$

which is Equation 27. To emphasize the correspondence between the two models of the hitting point distribution, Figure 7 shows the simulated distributions of hitting points from Figure 6, but characterized instead using the von Mises distribution of Equation 27. As expected from the preceding analysis, the figures are indistinguishable.

Theoretical Decomposition of Precision: Quality Times Quantity

The theoretical distribution of report in Equation 28 has two important implications. First, as noted previously, empirical stud-

[5] The von Mises distribution is often referred to as a "circular normal" distribution, but this terminology is misleading. Defining a normal distribution on a circle instead of a line leads to a so-called *wrapped normal* distribution, whose probability density function takes the form of an infinite series (Fisher, 1993). The von Mises is actually an approximation to the wrapped normal, and is usually preferred in applications because of its simpler form. Here, the predicted distribution of hitting points of the 2D model follows a von Mises distribution exactly; it is not an approximation.

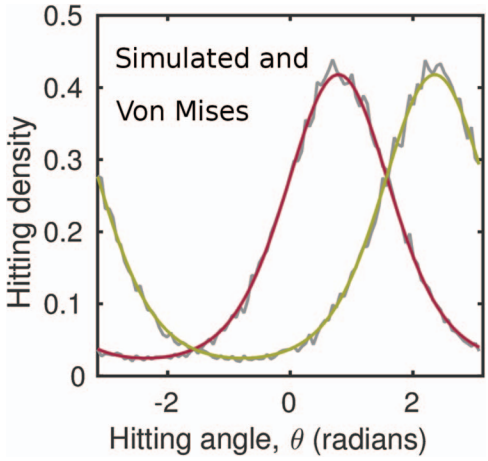

Figure 7. Simulated hitting point (phase angle) distributions for nonzero-drift diffusion processes and von Mises distributions with $\kappa = a\|\boldsymbol{\mu}\|/\sigma^2$ and $\phi = \arctan(\mu_2/\mu_1)$.

ies of continuous report typically characterize report precision using the von Mises distribution. Empirical data are usually well-fitted either by a finite mixture of von Mises distributions and a uniform distribution (e.g., Bays et al., 2011; Zhang & Luck, 2008), or by a continuous mixture of von Mises distributions in which the concentration parameter is treated as a random variable (van den Berg et al., 2012, 2014). The 2D diffusion model provides a theoretical basis for the empirically observed variability in report accuracy. Like the standard model, it attributes variability in accuracy to noise in the evidence accumulation process. According to the model, trial-to-trial variability in report accuracy in the continuous report task is a manifestation of exactly the same kinds of processes as lead to variability in decision outcomes and RT in 2AFC tasks.

The second important implication concerns the form of the von Mises precision parameter predicted by the 2D diffusion model. The preceding analysis showed that, theoretically, precision is equal to

$$\kappa = \frac{a\|\boldsymbol{\mu}\|}{\sigma^2}, \tag{29}$$

that is, precision is the product of the norm of the drift vector and the radius of the criterion circle, scaled in diffusion coefficient units. The norm of the drift vector depends on how well the stimulus is represented in memory while the radius of the criterion circle depends on the participant's speed–accuracy trade-off settings. Rather than being an indivisible parameter of continuous report performance, precision in the diffusion model is the product of two basic cognitive processes: those that determine the quality of the evidence in the stimulus and those that determine the amount of evidence required for a response. This characterization of precision shows, precisely, in what sense the remark by Engen (1971) quoted in the Introduction is true. According to the diffusion model, the amount of time a participant takes to make a judgment will depend on the time required to retrieve a criterion quantity of evidence from memory.

It is instructive to compare the expression for precision for the 2D model in Equation 29 to the corresponding expression for detection sensitivity for a 1D random walk model derived by Link (1975). The measure is the random walk counterpart of the signal detection sensitivity measure d' and has the same form as the one obtained from a Luce choice model with exponentially distributed response strengths, or equivalently, from a signal detection model with logistic distributions of sensory effect (McNicol, 1972/2005, chap. 6; Yellott, 1977). For an unbiased random walk, this measure, which I denote as d'_{RW}, is equal to $\log\{[P(C)/[1 - P(C)]\}$, where $P(C)$ is the proportion of correct responses. When the increments to the walk are Gaussian distributed, with means $\pm\mu$ (depending on the stimulus) and *SD* σ, it reduces to

$$d'_{\mathrm{RW}} = \frac{2a\mu}{\sigma^2}, \tag{30}$$

(Smith, 2015). Link derived his measure for a discrete-time random walk, but it can also be interpreted as a sensitivity measure for a 1D Wiener diffusion process because a Gaussian-increments random walk and a continuous Wiener process have the same first-passage time statistics (Smith, 1990).

Apart from the scaling factor of 2, Equation 30 is the scalar counterpart of Equation 29 and, like Equation 29, it says that sensitivity is equal to the product of the evidence in the stimulus and the decision criterion, scaled in diffusion coefficient units. The additional factor of 2 in Equation 30 arises because the means of the distributions of sensory effect associated with the stimuli are 2μ apart. If the means are located at $\pm\mu/2$, giving an effective stimulus discriminability of $|\mu|$, then Equation 30 becomes an exact scalar counterpart of Equation 29. The equivalence of the expressions for precision in the 2D model and sensitivity in the 1D model highlights the close theoretical relationship between the two models.

Trial-to-trial variability in stimulus information and decision criterion variability in drift rates. The 1D diffusion model of Equations 8 and 9, like log-likelihood-ratio random walks (Laming, 1968) and Gaussian-increment random walks (Link & Heath, 1975; Smith, 1990), predicts equal mean RTs for correct responses and errors when the decision criteria are located equidistantly from the starting point or, when comparing responses made at the same boundary, when the drift rates for the two stimuli have the same magnitudes but opposite signs (see Footnote 1). Equality is almost never found empirically—one notable exception being a study by Green, D. M. Smith, and von Gierke (1983), in which participants were practiced on the same task for an entire semester. More commonly, errors RTs are either shorter or longer than correct RTs, depending on the task and the experimental instructions. Errors are typically faster than correct responses when the task is easy and speed is emphasized or slower than correct responses when the task is difficult and accuracy is emphasized (Luce, 1986). The standard diffusion model attributes the differences in mean RTs under these conditions to variability in decision criteria (or starting point), and variability in drift rate, respectively.

As discussed previously, in the 2D diffusion model, correct responses and errors are not categorically distinct as they are in the 1D model. Nevertheless, it is possible to look at the relationship between decision time and the distribution of report errors across trials. In the 2D model, both the magnitude and the phase angle of the drift can vary and so could plausibly vary across trials. I restrict my consideration here to variation in magnitude, because of its simple and identifiable effects on precision. Figure 8 shows the effects of trial-to-trial variability in drift magnitude. The top panel

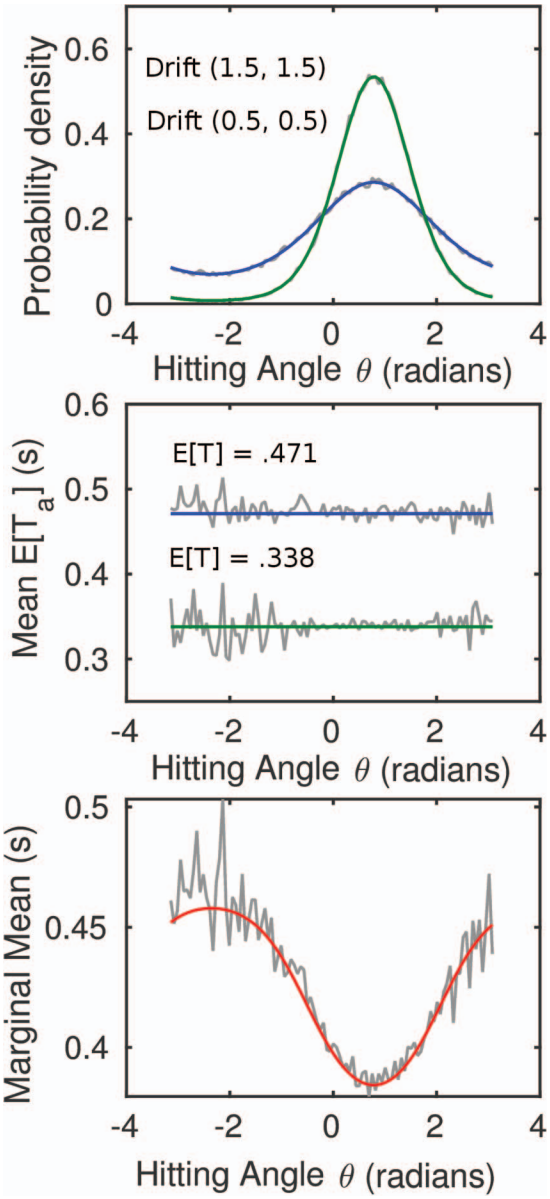

Figure 8. The effect of drift variability on mean decision times as a function of report accuracy. The top panel shows simulated (gray) and predicted (colored) distributions of report for large and small drift magnitudes, $\|\boldsymbol{\mu}_1\| = 2.12$ (green), $\|\boldsymbol{\mu}_2\| = 0.71$ (blue), $a = 1.0$, $\sigma = 1.0$. The middle panel shows simulated mean decision times and predicted times from Equation 24. The bottom panel shows simulated and predicted mean times for a mixture of drift values, in which the means in the middle panel are weighted by the hitting probabilities in the top panel.

shows predicted distributions of hitting points for two values of drift, $\boldsymbol{\mu}_1 = (1.5, 1.5)$ and $\boldsymbol{\mu}_2 = (0.5, 0.5)$. These drift rates have the same phase angle, but different magnitudes, $\|\boldsymbol{\mu}_1\| = 2.12$ and $\|\boldsymbol{\mu}_2\| = 0.71$, respectively. The infinitesimal *SD* and the radius of the criterion circle were held fixed at $\sigma = 1.0$ and $a = 1.0$. Consistent with the fact that precision equals $\kappa = a\|\boldsymbol{\mu}\|/\sigma^2$, with the larger drift magnitude the distribution of hitting points is appreciably more concentrated around the true value of $\pi/4$.

The middle panel of Figure 8 shows predicted mean decision times as a function of phase angle. Mean decision times are faster for larger drift rates but are independent of phase angle, again consistent with the earlier analysis. If $\pi_1(\theta)$ and $\pi_2(\theta)$ denote the hitting probabilities as a function of phase angle for drift vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and $E_1[T(\theta)]$ and $E_2[T(\theta)]$ denote the corresponding mean decision times, then the marginal mean decision times are weighted averages

$$E[T(\theta)] = \frac{\pi_1(\theta)E_1[T(\theta)] + \pi_2(\theta)E_2[T(\theta)]}{\pi_1(\theta) + \pi_2(\theta)}. \tag{31}$$

The bottom panel of Figure 8 shows the predicted mean decision time for a weighted average of the two distributions. As the figure shows, mean decision time and accuracy are inversely related. Small errors of report are associated with fast responses and conversely. This result corresponds to the property of the standard diffusion model with drift variability, in which error RTs are longer than correct RTs. In the 2D model, the same property holds as a continuous function of report error, $\theta_T - \theta_\mu$.

In practice, one would wish to assume a more complex, probably continuous, distribution of drift magnitudes. In the standard diffusion model, drift rates are assumed to be normally distributed with a *SD* that is independent of the mean. In the 2D model, in which the drift rate is vector-valued, the theoretical possibilities are correspondingly richer. A natural generalization of the standard model can be obtained by assuming that the two components of drift are independently normally distributed with means v_1 and v_2 and a common *SD*. The squared drift norm would then follow a noncentral chi-square distribution with noncentrality parameter proportional to $v_1^2 + v_2^2$ (Johnson, Kotz, & Balakrishnan, 1995, p. 433). A direct link to the variable precision models of van den Berg et al. (2014) can be obtained by instead assuming that the magnitude of the drift rate is gamma distributed. I consider an example with gamma-distributed drift magnitudes in the section Quantile Families with Variability in Drift and Decision Criterion.

Variability in criterion. In the standard model, fast errors are predicted by variability in decision criterion, which is implemented in the model as variability in the starting point of the diffusion process. This is equivalent to negatively correlated variability in criteria for a process with a fixed starting point. In the 2D model, it is theoretically possible that the accumulation process could start anywhere inside the criterion circle. However, the methods I have developed here assume radial symmetry of the zero-drift process, which requires that the accumulation process start at the origin. For this reason, instead of variability in starting point, I investigated the effects of trial-to-trial variability in the radius of the criterion circle, a. The value of a characterizes the amount of evidence needed to make a decision, so trial-to-trial variability in a can be thought of as variability in the decision-maker's speed–accuracy trade-off settings.

Figure 9 shows the effects of trial-to-trial variability in decision criterion. The top panel shows the distributions of hitting points for two different values of decision criterion, $a_1 = 0.75$ and $a_2 = 1.25$. The drift vector and infinitesimal *SD* were fixed at $\boldsymbol{\mu} = (1, 1)$ and $\sigma = 1.0$, respectively. Consistent with the fact that predicted precision is $\kappa = a\|\boldsymbol{\mu}\|/\sigma^2$, the distribution of hitting points is narrower when the criterion is larger. The middle panel shows mean decision times as a function of phase angle for the two values of criterion. Again, mean decision time is independent of the phase

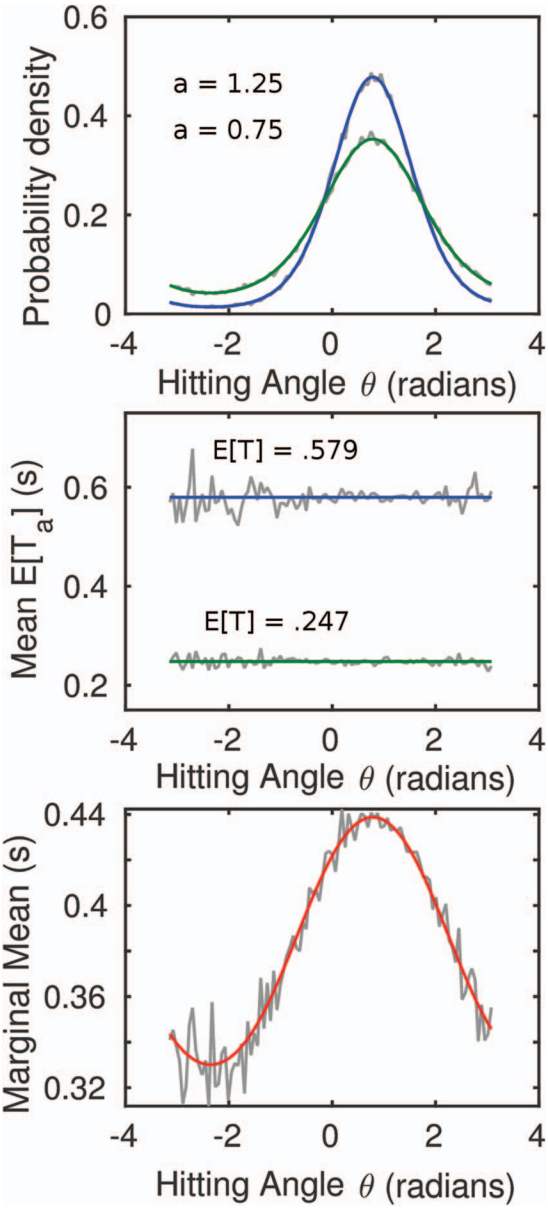

Figure 9. The effect of criterion variability on mean decision times as a function of report accuracy. The top panel shows simulated (gray) and predicted (colored) distributions of report for large and small decision criteria, $a_1 = 1.25$ (blue) and $a_2 = 0.75$ (green), $\|\boldsymbol{\mu}\| = \sqrt{2}$, $\sigma = 1.0$. The middle panel shows simulated mean decision times and predicted times from Equation 24. The bottom panel shows simulated and predicted mean times for a mixture of criteria, in which the means in the middle panel are weighted by the hitting probabilities in the top panel.

angle of the response and is uniformly longer for the larger value of a.

The bottom panel shows the marginal mean decision times obtained from a two-point probability mixture, like the mixture of drift rates in Equation 31. The effect of trial-to-trial variability in criterion is the obverse of that obtained with variability in drift. Mean decision time is directly rather than inversely related to report accuracy. Small errors of report are now associated with long decision times rather than short times, and conversely.

The preceding results have two important theoretical implications. The first is that the qualitative effects of trial-to-trial variability in drift rates and criteria in the 2D model are precisely the same as in the standard model. Although the 2D model does not make a categorical distinction between correct responses and errors, the same relationships are obtained when mean decision times are considered as a continuous function of report error: Variability in drift leads to slow errors and variability in criterion leads to fast errors. The close parallels between the predictions of the two models lends support to the idea that continuous report can be viewed as a decision process with a continuous instead of a categorical outcome set.

The second implication involves the interpretation of precision studies. As discussed in the following section, van den Berg et al. (2014) found that the best model for continuous report was one in which the precision parameter of a von Mises distribution was allowed to vary from trial to trial, which they modeled using a gamma distribution. In the 2D diffusion model, precision is the product of drift magnitude and criterion, scaled in diffusion coefficient units. The standard diffusion model, with variability in drift rates and criterion (starting point), has now accounted successfully for 2AFC performance in numerous experimental tasks. The analysis I have presented here shows that, in the context of continuous report, variability in either drift magnitude or criterion translates into trial-to-trial variability in precision of the kind proposed by van den Berg et al. The analysis therefore provides a way to theoretically unify performance on the two kinds of task and provides a process account of the empirically estimated variability in precision. Moreover, like the standard diffusion model, the 2D model makes strong and testable predictions about the effects on decision time of the two components of variability in precision. Because variability in precision due to variability in drift rate and criterion have opposite effects on the ordering of times for correct responses and errors, the 2D model provides a way to characterize the basis of variable precision via an analysis of RT data.

Jones and Dzhafarov (2014) were critical of the use of trial-to-trial variability in the standard diffusion model because of the amount of flexibility it can introduce into the predictions. They showed that, if all a priori constraints are removed from the distributions of across-trial variability, an unfalsifiable model can be obtained, whose predictions can be made to match any set of empirical data. In response, Smith, Ratcliff, and McKoon (2014) pointed out that the distributions of across-trial trial variability in the standard diffusion model have specific, prescribed forms and that most of the predictive work in the model is done by within-trial diffusive variability. This constrains the model and makes it highly falsifiable. They also noted that the assumptions of across-trial variability in evidence quality and decision criteria are theoretically principled and have well-established precedents in the literature. The assumption that evidence quality varies from trial to trial has its origins in Thurstone's law of comparative judgment (Thurstone, 1927) and signal detection theory (Green & Swets, 1966), while the assumption that decision criteria vary from trial to trial is a feature of adaptive criterion-setting models that seek to explain how people dynamically regulate their speed-accuracy settings (Brewer & G. A. Smith, 1989; Rabbitt & Rodgers, 1977;

Vickers, 1977; Vickers & Lee, 1998, 2000). These models are empirically supported by the literature on sequential effects (Luce, 1986, chap. 6.6).

Distributions of Decision Times

One of the features of empirical 2AFC data that is well-described by the standard diffusion model is the linearity of families of RT distribution quantiles (Ratcliff & Smith, 2010; Ratcliff et al., 2015). Formally, the p-th distribution quantile, Q_p, is defined to be the value of time such that the proportion of RTs in the distribution that are less than or equal to Q_p is equal to p. Many studies of 2AFC studies summarize distributions of RT using five quantiles: usually the .1, .3, .5, .7, and .9 quantiles. The .1 and .9 quantiles represent the lower and upper tails of the distribution, that is, the fastest and slowest responses, respectively. The .5 quantile is the median and represents the distribution's central tendency.

The properties of a family of RT distributions can be characterized in a Q-Q plot. In a Q-Q plot, the RT distribution quantiles for each condition in an experiment are plotted against the corresponding quantiles of a reference condition (usually the fastest). Linearity of the Q-Q plot means that the RT distributions in different conditions are similarly shaped, in the sense that they are related to one another by a change in time scale (a dilation or contraction), possibly also accompanied by a change in the intercept. Formally, the combination of a shift and similarity transformation of the time scale is known as an affine transformation.

Q-Q plots of 2AFC data from many experimental tasks are highly linear. Ratcliff and Smith (2010, Figure 20) showed that the Q-Q plots from 12 perceptual experiments were predominantly linear. Ratcliff et al. (2015) showed that Q-Q linearity holds across a wide range of cognitive tasks. Rouder, Yue, Speckman, Pratte, and Province (2010) reported additional examples, although they also found deviations from strict linearity using a model-based statistical test. Ratcliff and McKoon (2008) showed that Q-Q linearity is predicted by the standard diffusion model and that it is a fairly robust feature of the model that is preserved, approximately if not exactly, under reasonable assumptions about trial-to-trial distributions of drift and criteria (i.e., unimodality). The model therefore predicts the same shaped families of RT distributions that are found empirically and, to a good approximation, this is all it can predict. Given that linearity or near-linearity of quantile families is a feature of so many empirical data and that the ability to predict it is one of the reasons for the success of the standard diffusion model, it is of interest to know whether the 2D model predicts the same properties.

Figure 10a shows a Q-Q plot for five different values of drift magnitude and shows that this is indeed the case. The symbols are the predicted quantiles of the model; the straight lines are fitted regression lines. Because the model predicts invariant distribution shapes as a function of the phase angle of the response, I have marginalized the distributions across the phase angle of the hitting point, θ_T. The resulting Q-Q plot is highly linear. This means that the 2D model produces families of decision time distributions whose members are related to one another by a combination of a shift and a similarity transformation (i.e., an affine transformation). For comparison purposes, I have shown a Q-Q plot for the 1D model of Equations 11 and 12 with

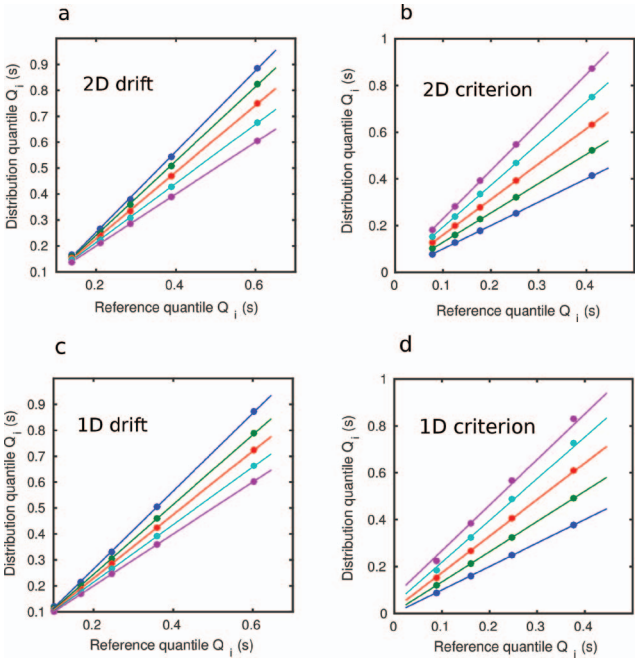

Figure 10. Predicted distribution quantiles obtained by varying drift rates and decision criteria for two-dimensional (2D) and one-dimensional (1D) models. The symbols in each panel (bottom left to top right) are the .1, .3, .5, .7, and .9 distribution quantiles predicted by the model and the lines are fitted regression lines. (a) 2D model, drift magnitude varying, $\|\boldsymbol{\mu}\| = \{0.71, 1.06, 1.41, 1.77, 2.12\}$, $a = 1.0$; (b), 2D model, criterion varying, $a = \{0.7, 0.8, 0.9, 1.0, 1.1\}$, $\|\boldsymbol{\mu}\| = \sqrt{2}$; (c) 1D model, drift varying, $\mu = \{.90, 1.25, 1.5, 1.75, 2.0\}$, $a = 1.4$; (d) 1D model, criterion varying, $a = \{0.6, 0.7, 0.8, 0.9, 1.0\}$, $\mu = 1.0$. For the 2D model, the infinitesimal *SD* was $\sigma_i = 1.0$, $i = 1, 2$, for both of its components; for the 1D model it was $\sigma = 1.0$. Colors (dark blue, green, red, light blue, and purple) represent increasing values of drift rate and criterion.

different drift rates, marginalized across correct responses and errors, in Figure 10c. These quantiles are for a pure diffusion process with no across-trial variability, and equal decision criteria for the two responses. The plot for the 1D model exhibits the same high degree of linearity as the one for the 2D model. Linear Q-Q plots are obtained for both models despite the fact that the first passage time densities do not have simple algebraic representations as shift-scale families (cf., Thomas & Ross, 1980, for a discussion of such families in the context of distribution averaging).

Another facet of the Q-Q families predicted by the 2D model is shown in Figure 10b. The distributions in this figure were obtained by holding the drift rates constant and varying the criterion, a. Such a family would be obtained experimentally by parametrically varying the speed-accuracy instructions given to participants. The figure shows that the same Q-Q linearity that is found across conditions when drift rates are varied is also found when drift rate is held constant and criteria are varied. Figure 10d shows the corresponding plot for the 1D model, with fixed drift rate and variable criteria, which is also highly linear. Like the 1D model then, linearity of Q-Q plots for the 2D model is a signature of the distribution families it predicts, and may afford the basis for strong experimental tests in future.

Quantile Families With Variability in Drift Rate and Decision Criterion

Figure 11 shows the effects of trial-to-trial variability in drift magnitude and decision criterion on report accuracy and on RT distributions. The left-hand panels show the effects of drift variability; the right-hand panels show the effects of criterion variability. Following van den Berg et al.'s (2014) characterization of variable precision models, I assumed that drift magnitude and criterion were gamma distributed. The predictions for drift variability are based on a two-stage gamma distribution with rate constant 1.5. The predictions for criterion variability are based on a two-stage gamma distribution with rate constant 0.35 and an offset of 0.35. For criterion, an offset was needed to obtain well-behaved RT distributions, because a drift rate of zero yields well-behaved distributions but a criterion of zero does not. The phase angle of the stimulus was 0 radians, so a hitting angle of $\theta = 0$ represents the most accurate responses.

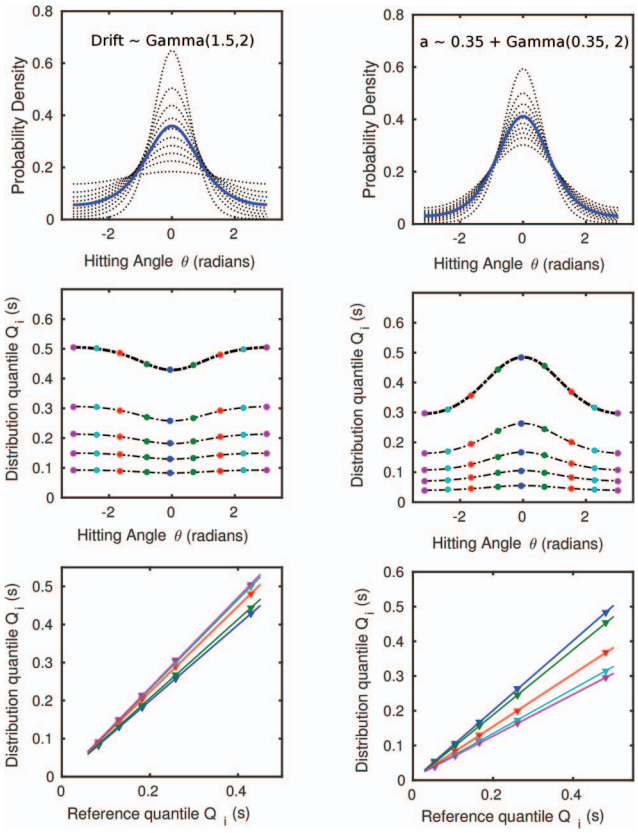

Figure 11. Effects of trial-to-trial variability in drift magnitude (left panels) and decision criterion (right panels) on distributions of report and response time (RT). In the panels on the left, the decision criterion was held fixed at $a = 0.75$; in the panels on the right, the drift magnitude was held fixed at $\|\boldsymbol{\mu}\| = \sqrt{2}$. The phase angle of the drift rate was zero for both cases and the diffusion coefficient was $\sigma^2 = 1.0$. The top row shows the distributions of report accuracy for 10 of the components of the mixture distributions (dotted lines) and the distributions marginalized across components (heavy blue lines). The middle row shows RT distribution quantiles for the mixture distribution. The bottom row shows quantile-quantile (Q-Q) plots of RT distributions for selected hitting angles (shown colored in the middle row). The symbols are the distribution quantiles; the lines are fitted regression lines. The colors in the bottom row correspond to those of the selected distributions in the middle row.

The top two panels show distributions of report accuracy for the components of the mixture distributions, together with the marginal distributions obtained by integrating across them. For both drift magnitude and criterion, a peaked, high-tailed distribution of report accuracy is obtained, which is qualitatively like the distributions found empirically in continuous reports tasks. The Zhang and Luck (2008) study reproduced in Figure 1 is typical of this pattern of data. The effects of these two kinds of variability on the distribution of report accuracy are indistinguishable, but they can be distinguished, both quantitatively and qualitatively, in the distributions of RT.

This point is illustrated graphically in the middle panels of Figure 11, which show the quantiles of the RT distributions as a function of report accuracy (hitting angle). The lines in the figure are, from bottom to top, the .1, .3, .5, .7, and .9 RT distribution quantiles. The distribution quantiles exhibit the same properties as the mean RTs in the middle panels of Figures 8 and 9: When drift rate is variable, inaccurate responses are slower than accurate responses; when the criterion is variable, inaccurate responses are faster than accurate responses. As in the standard diffusion model, the effects of trial-to-trial variability are most apparent in the upper quantiles of the RT distribution. These patterns of accurate and inaccurate responding are the empirical signatures of drift variability and criterion variability, respectively, in the 2D diffusion model. Working memory researchers regard a peaked, high-tailed distribution of report accuracy (relative to the standard von Mises distribution) as indicative of trial-to-trial variability in the cognitive processes that determine report accuracy. The 2D diffusion model provides a way to distinguish the components of this variability in data.

Because variability in drift rates and criteria can reduce Q-Q linearity, in the bottom panels of Figure 11 I have shown Q-Q plots of selected members of the families of distributions from the two panels above. To construct this plot, I selected five equally spaced phase angles from mirror-symmetric points of the continuous families of RT, and plotted their quantiles (shown as colored dots) against those for the distribution for $\theta = 0$. As in Figure 10, the symbols are the predicted distribution quantiles and the lines are fitted regression lines. The purpose of these plots is to ascertain whether across-trial variability in drift rate and criterion affects the linearity of a family of distributions, when the members of a family come from a single experimental condition, parameterized by report accuracy. This differs from the Q-Q plots reported by Ratcliff and McKoon (2008); Ratcliff and Smith (2010), and Rouder et al. (2010) for the 1D model, in which the members of a family were from different experimental conditions. The Q-Q plots in the bottom panels of Figure 11 are again highly linear, showing that, within a family of distributions, Q-Q linearity appears to a robust feature of the 2D model under mixing.

In summary then, linearity or near-linearity of Q-Q plots is a feature of models that represent evidence accumulation as a diffusion process, but it is not a property shared by other sequential-sampling decision models. Notably, it is not a property of accumulator models or Poisson counter models (Ratcliff & Smith, 2004; Smith & Vickers, 1988; Smith & Van Zandt, 2000). The fact that Q-Q linearity is frequently found empirically and is predicted

by some, but not all, models, is important in attempting to distinguish between model classes. Discussions of the relationships among the various model classes can be found in Bogacz et al. (2006) and Smith and Ratcliff (2015b).[6]

Finite Mixtures and Guessing

In the previous section I showed that variable precision models, like those proposed by van den Berg et al. (2012, 2014), are a natural consequence of the assumption that drift rates and decision criteria vary across trials. In variable precision models, the predicted distribution of report is a continuous mixture of von Mises distributions with different concentration parameters. However, finite mixture models, like the ones proposed by Zhang and Luck (2008), Bays et al. (2009, 2011), and others can also be represented using the 2D diffusion model. In finite mixture models the predicted distribution of report is a mixture of a small number (often two) of von Mises distributions. These models are an expression of the assumption that VSTM has an item capacity limit, usually estimated to be around four items (Cowan, 2001). Items in memory are represented with some precision, which may depend on the number of stored items and the allocation of resources among them (Bays, 2014; Bays et al., 2011; Sewell et al., 2014; Smith, 2015; Zhang & Luck, 2008); if an item is not in memory, the participant is forced to guess. The guessing process has a straightforward representation in the 2D diffusion model as a zero-drift diffusion process. When the retrieval cue is presented, a noisy retrieval process is initiated to determine whether the item is in memory and, if so, to identify it. Items in memory are represented with nonzero precision; items not in memory are represented with zero precision and the process retrieves only noise. The distribution of report will then be uniformly distributed on the range $(0, 2\pi)$.

Figure 12 shows the predicted distribution of report and associated RT distribution quantiles for a two-state mixture of 2D diffusion processes, in which items are either in memory, in which case their precision is proportional to the drift norm, $\|\boldsymbol{\mu}\|$, or not in memory, in which case both the drift rate and the associated precision is zero. To generate Figure 12, I followed the example of Zhang and Luck (2008) in Figure 1 and assumed a four item memory and a six item display, so there was a one third probability of a probed item not being in memory. The figure shows that the resulting two-state mixture model produces a peaked, high-tailed distribution of report, similar to those shown in Figure 11, and a family of distribution quantiles that exhibit the slow-error pattern associated with across-trial variability in drift rate. This pattern of RTs would be expected if the same noisy retrieval process governs performance for items in memory and items not in memory. Slow errors are an empirical signature of a single-process retrieval mechanism and serve to distinguish it from errors initiated from outside the retrieval process, such as fast guesses (Yellott, 1971). The 2D diffusion model can also predict similar performance to a pure fast-guess process if some proportion of responses are made using a very low decision criterion (i.e., a small value of a). Performance on such trials will be dominated by retrieval noise and will exhibit the fast-error pattern of RTs shown on the right in Figure 11.

The examples I have given above are intended to illustrate that both finite-mixture and variable precision models have natural representations using the 2D diffusion model. Of course, any empirical distribution of report that can be characterized as a mixture of von Mises distributions can also be characterized by an appropriately chosen combination of parameters of the 2D diffusion model. My aim in showing these predictions is not to reproduce any particular set of experimental data, or to adjudicate between models that assume discrete and continuous distributions of precision. It is, rather, to show how across-trial variability in precision can arise either as the result of variability in the quality of the memory representation of the stimulus or the amount of evidence needed to make a response. The effects of these two kinds of variability on the distribution of report accuracy are indistinguishable, but they can be distinguished, both quantitatively and qualitatively, in the distributions of RT.

As noted previously, much of the debate in the recent VSTM literature has focused on whether or not VSTM has a fixed item capacity. Finite mixture models like the one in Figure 12 are a natural expression of an item-capacity limit. The simplest finite mixture models are two-component models, like that of Zhang and Luck (2008), which assume an item is either in memory or not. More elaborate finite mixture models (Bays et al., 2011; Fougnie & Alvarez, 2011) also assume intrusion errors from items in memory but not at the report location. Variable precision models like those advocated by Ma, van den Berg, and colleagues (van den Berg et al., 2012, 2014), assume that the number of items in memory and the precision with which individual items are represented may both vary randomly from trial to trial. Like finite mixture models, these models allow for the possibility that some items will be represented with zero precision and will produce responding that is indistinguishable from guessing.

One of the methodological challenges associated with complex mixture models of this kind is that their parameters can be difficult to estimate reliably from experimental data. Van den Berg et al. (2014) compared, among other models, variable precision models with fixed and variable item capacities. In the latter, the number of items in memory was Poisson distributed across trials. They found that the fits of the doubly stochastic Poisson item-encoding models were, on average, better than those of fixed item-capacity models, but that the estimates of the average number of items stored in memory were highly variable and the distributions of estimates

[6] The term "accumulator model" is ambiguous, because different investigators have used it to refer to models with different properties. Here I use it in the sense it was used originally by Audley and Pike (1965) and Vickers and colleagues (Smith & Vickers, 1988), to describe a model in which evidence for competing responses accrues in separate totals and the totals are only incremented and never decremented. Usher and McClelland (2001) used it subsequently to refer to a model in which the evidence totals are modeled as racing, mutually inhibitory, diffusion processes, in which the evidence totals increase and decrease. Ratcliff and Smith (2004) showed that the stochastic process used to model evidence accrual is a more important determinant of model performance than is model architecture, that is, whether the model assumes one evidence total or two. Because the Usher and McClelland model represents evidence accrual as a diffusion process, its properties are much more similar to those of the standard diffusion model than they are to the accumulator models of Audley, Pike, Vickers, and others.

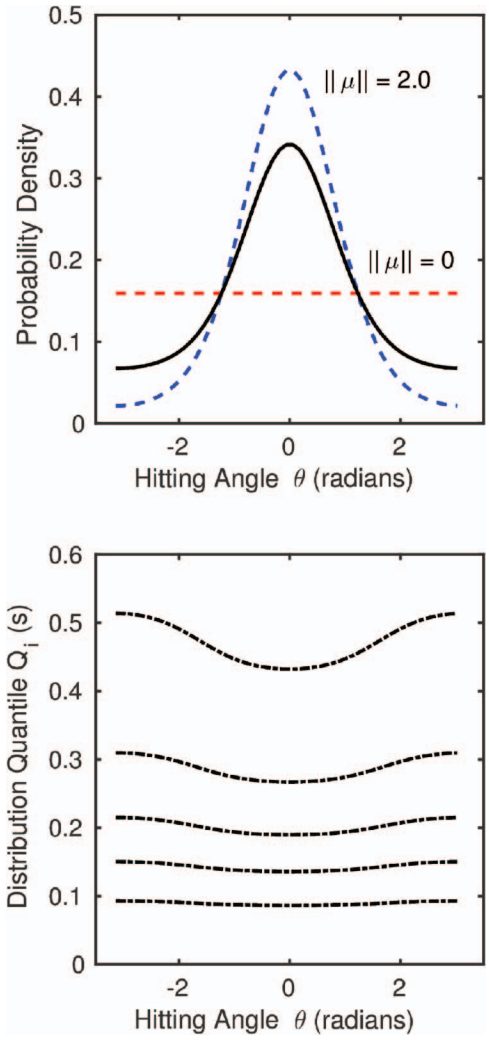

Figure 12. Distribution of report accuracy and response time (RT) distribution quantiles for a two-component, two dimensional (2D) diffusion, mixture model. The model assumes that items not represented in memory initiate a zero-drift retrieval process. The probability of an item not being in memory was one-third. Items in memory were represented with precision $\kappa = a\|\boldsymbol{\mu}\|/\sigma^2$, $a = 1.0$, $\|\boldsymbol{\mu}\| = 2.0$, $\sigma^2 = 1.0$ (blue dashed line); items not in memory were represented with precision zero (red dashed line). The black continuous line is the marginal distribution of report outcomes averaged across both kinds of items. The bottom panel shows the quantiles of the marginal distribution of RT as a function of report outcome (phase angle). The phase angle of the stimulus was $\theta_\mu = 0$, so a report of $\theta_a = 0$ represents the most accurate responses.

were extremely positively skewed (see their Table A1).[7] The estimation problem that arises in fitting these models is a common one in psychology, namely, how to identify the parameters of a complex process model from comparatively limited experimental data — in this case, the distributions of report from a number of experimental conditions. In studies of decision making, the use of a combination of RT and accuracy data has been helpful in identifying the components of the decision process from data and in allowing criterion effects to be distinguished from stimulus quality effects. It seems probable that some of the problems that arise in attempting to fit complex process models to distributions of report might be alleviated by simultaneously considering distributions of RT because of the additional constraints they impose on fits.

Extensions and Generalizations of the Model

The theoretical results in previous sections rely on strong symmetry assumptions that allow the properties of the 2D diffusion model to be characterized using the Girsanov theorem and the 2D Bessel process. The assumption that evidence accumulation begins at the origin of a circular domain makes it possible to derive simple analytic solutions for the RT distributions and the distribution of report. The key theoretical insights concerning the von Mises form of the distribution of report, the theoretical decomposition of precision, and the relationship between the 1D and 2D models all depend on these analytic solutions. However, these assumptions may be too restrictive when attempting to fit the model to data. In this section, I consider ways of relaxing them and extending the model to other domains. One of the costs of relaxing the assumptions is that analytic results are more difficult to obtain. Under these circumstances, predictions from the model can be obtained either by Monte Carlo simulation or using the matrix method of Diederich and Busemeyer (2003), which approximates the diffusion process using a finite-state Markov chain with the same infinitesimal moments (Stroock & Varadhan, 1979, pp. 266–272).

Complete versus incomplete stimulus domains. One extension of the model is to tasks with incomplete stimulus domains, in which only a subset of the full range of decision outcomes in the circular model are possible responses. I restrict my consideration to tasks in which the 2D vector-space stimulus representation of the circular model continues to hold, but only a subset of the full range of responses is used, either because of the way in which the experiment is set up or because of the properties of the stimulus domain. For example, van den Berg et al. (2012) used a task requiring continuous report of color, in which the range of colors was restricted to half the color wheel, which maps to the range $(0, \pi)$. Rademaker et al. (2012) used a task requiring orientation judgments of Gabor patch stimuli in which the set of unique stimuli was restricted to the range $(0, \pi)$ because of their rotational symmetry.

A natural model for decisions in tasks of this kind is 2D diffusion on a semicircular region, $X_t^1 \in \mathbb{R}$, $X_t^2 \in \mathbb{R}^+$, as shown in Figure 13. Like the model of Figure 2, the bounding semicircle is an absorbing barrier (or boundary) for the process and represents the set of possible decision outcomes. Its radius, a, represents the amount of evidence required for a decision. Evidence is accumulated by the process $\boldsymbol{X}_t$, starting at the origin and continuing until the process reaches the criterion a. The response that is made is the one corresponding to the hitting point, $\boldsymbol{X}_\theta$, and the decision time is the hitting time T_a.

To ensure the process does not stray outside the semicircular region $X_t^2 \geq 0$, $(\boldsymbol{X}_t)^2 \leq a^2$, and, in particular, to ensure that it cannot

[7] I thank Weiji Ma and Ronald van den Berg for confirming my interpretation of the parameter estimates of the Poisson encoding models in Table A1 of van den Berg et al. (2014).

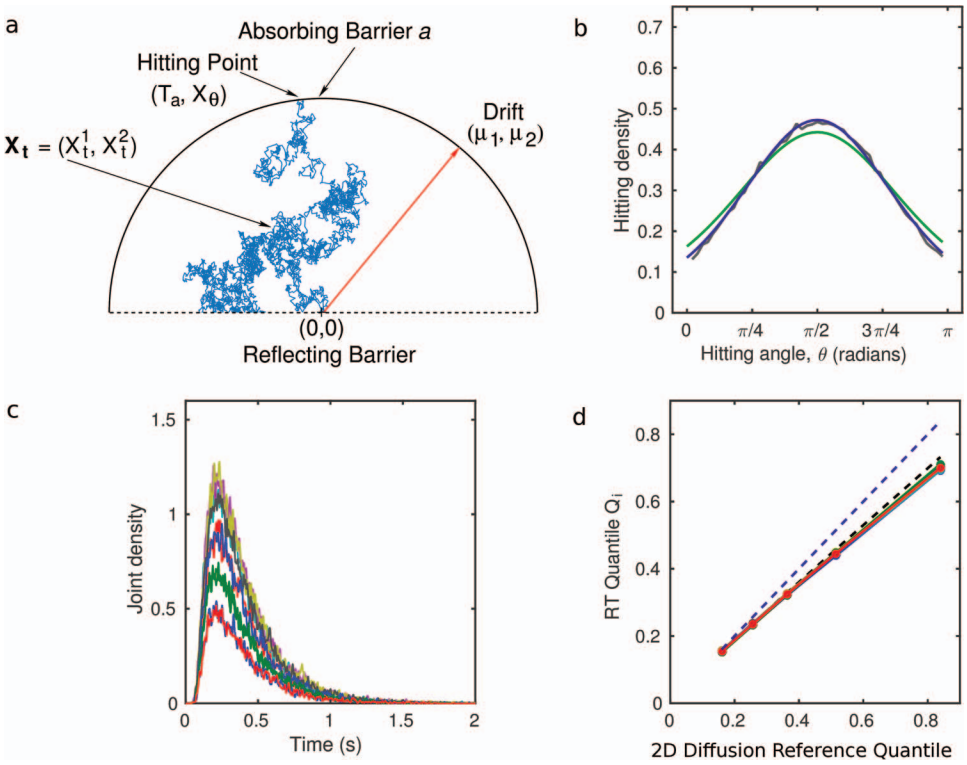

Figure 13. Diffusion on a semicircular region with reflecting barrier. (a) A vector-valued process $\boldsymbol{X}_t$ with drift $\boldsymbol{\mu} = (\mu_1, \mu_2)$ diffuses until it hits a criterion semicircle, a. The hitting point, $\boldsymbol{X}_\theta$, and hitting time, T_a, determine the decision outcome and decision time, respectively. The accumulating evidence is constrained by a reflecting barrier, $X_t^2 \geq 0$. (b) Distribution of report outcomes. The irregular gray line is the simulated distribution of report outcomes for the semicircular model with drift rate $\boldsymbol{\mu} = (0, 1)$. The green and purple lines are the predicted distributions for a circular model with $\boldsymbol{\mu} = (0, 1)$ and $\boldsymbol{\mu} = (0, \sqrt{2})$, respectively. (c) Simulated response time (RT) density functions for 11 decision outcomes for the semicircular model. (d) Distribution quantiles for the semicircular model (solid, colored lines) plotted against quantiles for the circular model with $\boldsymbol{\mu} = (0, 1)$ (purple dashed line) and $\boldsymbol{\mu} = (0, \sqrt{2})$ (black dashed line).

cross the criterion from the wrong side, it is bounded from below by a reflecting barrier. Whenever the process reaches this barrier it is instantaneously reflected back into the upper half-space. Reflecting barriers have been used to constrain accumulating evidence in models proposed by Diederich (1995); Smith and Ratcliff (2009), and Usher and McClelland (2001).[8] Smith and Ratcliff (2009) used the matrix method (Diederich & Busemeyer, 2003) to obtain predicted RT distributions for a model comprised of a pair of racing diffusion processes, each with an absorbing upper barrier and a reflecting lower barrier. Like the semicircular model, in their model the absorbing barriers represent decision criteria and the reflecting barriers constrain the racing processes to the positive half-line. Figure 13a depicts one sample path of the reflecting process in the semicircular model.

Although the model of Figure 13 is a straightforward generalization of the model of Figure 2, it is less tractable analytically, because the accumulating evidence consists of a reflecting Brownian motion with drift in the vertical direction and an ordinary (i.e., nonreflecting) Brownian motion with drift in the horizontal direction. The transition probability of a reflecting Brownian motion with drift is more complex than that of an ordinary Brownian motion process and does not admit of a simple change of measure of the form $d\tilde{P}_t = \boldsymbol{Z}_t(\boldsymbol{X})dP_t$ of Equation 6 (Harrison, 1985, p. 49, Equation 1). However, it can be approximated by such a transformation by choosing the value of the drift rate appropriately. I do not attempt to characterize this approximation here, but merely remark on its consequences. Assuming a positive drift rate in the reflecting component, the approximating drift rate will be larger than the true drift rate because the reflecting process cannot accumulate evidence for a negatively signed response. Consequently, the proportion of its sample paths that are found more than a given distance above the reflecting barrier at any time will exceed the corresponding proportion for an ordinary process. This property suggests that the semicircular model will have similar properties to a circular model with a larger vertical component of drift. Figure 13b shows that this is the case.

[8] Usher and McClelland (2001) formulated their leaky competing accumulator model mathematically as a discrete approximation to a continuous diffusion process and described the mechanism that stopped the accumulating evidence totals becoming negative as "truncation." The continuous (diffusion) equivalent of this process is reflection. A discussion of the possible forms of boundary behavior of diffusion processes can be found in Karlin and Taylor (1981, p. 251).

The gray line in Figure 13b shows the distribution of report for 50,000 simulated trials of the semicircular model of Figure 13a, with drift rate $\boldsymbol{\mu} = (0, 1)$ (i.e., the phase angle of the drift was $\pi/2$), decision criterion $a = 1.0$, and infinitesimal SD $\sigma = 1.0$. The colored lines are truncated distributions of report for the circular model, constrained to the range $(0, \pi)$, and normalized. The green line is the predicted distribution of report for a model with the same parameters as used in the simulation; the purple line is the predicted distribution of report for a model with a drift rate $\boldsymbol{\mu} = (0,\sqrt{2})$ and the other parameters the same. As the figure shows, the predictions of the semicircular model agree fairly well with those of a circular model with a larger drift rate.

Figure 13c and 13d show the corresponding predictions for RT. Figure 13c shows 11 simulated distributions of RT for the semicircular model and Figure 13d shows the quantiles of the corresponding conditional RT distributions plotted against the quantiles of the marginal RT distribution predicted by the circular model (purple dashed line). Figure 13c shows that the distributions are all scaled copies of each other, as they are in the circular model. This means that, like the circular model, the model with no across-trial variability predicts that decision times and decision outcomes will be independent of one another. This property is confirmed in Figure 13d: The distribution quantiles for the semicircular model are highly linear and all lie on top of one another. However, they all lie below the corresponding quantiles for the circular model. This means that the semicircular model predicts faster and more accurate decisions than does a circular model with the same parameters (cf. Figure 5). This difference is because of the effects of the reflecting barrier, which constrains the accumulating evidence to the upper half plane. Psychologically, decisions are faster and more accurate when there is less uncertainty about the range of possible decision outcomes. The representation in Figure 13 has an obvious generalization to tasks in which the range of decision outcomes is less than $(0,\pi)$. These can be represented as diffusion on a wedge bounded by an absorbing arc and two reflecting radiuses.

Stimulus bias and response bias. The assumption that evidence accumulation in the circular model begins at the origin presupposes an unbiased accumulation process, which is likely to be too restrictive in some settings. Two kinds of bias have been identified in sequential-sampling models: response bias and stimulus bias. Response bias is an inequality or asymmetry in the amount of evidence needed for different responses; stimulus bias is an asymmetry in the rates at which evidence for different responses accumulates. In models like the standard model, which represent evidence accumulation using a Wiener process, response bias can be characterized either by asymmetric placement of the decision criteria relative to a starting point of zero, or by asymmetry in the starting point for evidence accumulation relative to criteria placed at zero and a. Asymmetry in starting point is the usual way of parameterizing the model (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2015a).[9]

In the 1D setting, models with stimulus bias were first analyzed by Ashby (1983), who characterized it as bias in the computation of the log-likelihood in a discrete-time random walk. In the standard diffusion model, stimulus bias is modeled using a *drift criterion*, which is used to classify incoming stimulus information and which can result in unequal rates of evidence accumulation for the two responses. The same idea was embodied in Link and Heath's (1975) relative judgment theory, in which new stimulus information is classified at each step of a random walk against a *sensory referent*. Unequal rates of evidence accumulation for the decision alternatives are useful in modeling tasks like recognition memory, which require judgments about whether or not an item was on a previously studied list (Ratcliff & Smith, 2004).

The more complex geometry of the 2D model means that the options for representing bias are correspondingly richer. Response bias can be represented either by assuming that the accumulation process starts somewhere other than at the origin or by deforming the criterion boundary. The criterion can be viewed as a locus in the plane, $\mathbb{R}^2$, parameterized by phase angle, $a(\cos\theta, \sin\theta)$, and could potentially be any closed, convex curve without violating the fundamental psychological properties of the model. A nonzero starting point and noncircularity of the criterion boundary both imply that some responses are made with less evidence than others, but the kinds of trade-offs they predict will be different.

Stimulus bias can be represented in the 2D model using a bias vector, $\boldsymbol{\xi} = (\xi_1, \xi_2)$, which is added to the drift vector associated with the stimulus by vector summation, such that the effective drift becomes the resultant vector $\boldsymbol{\mu}' = \boldsymbol{\xi} + \boldsymbol{\mu}$. This biases both the magnitude and the phase angle of the drift rate: Evidence will accumulate more rapidly for stimuli that have a drift component in the direction of $\boldsymbol{\xi}$ than for those that have a component in the opposite direction. Like the drift criterion in the 1D model, the bias vector represents a preexisting propensity for the cognitive system to classify stimulus information in a particular way. This propensity may vary randomly from trial to trial or be fixed across trials. The basic idea can be elaborated in different ways depending on the representational assumptions in the stimulus domain.

Computationally, the assumption of either a nonzero starting point or a noncircular decision boundary breaks the symmetry needed to characterize the properties of the 2D Wiener process using the Bessel process. However, as I noted, predictions for these more general models can be obtained using Monte Carlo methods or the matrix method. Yin and Wang (2009) reported an analytic expression for the first passage distribution of a 2D Brownian motion process with drift and arbitrary starting point on a circular domain. Their result can be obtained by applying a Girsanov change-of-measure to the first passage distribution for a zero-drift process with arbitrary starting point derived by Hsu (1986). I quote their result for completeness.[10]

As in Equation 21, let $J_m(x)$ be a Bessel function of the first kind of order m, and $J'_m(x)$ its derivative. Let $j_{m,n}$ denote the zeros of $J_m(x)$ arranged in ascending order and let the vectors $\boldsymbol{z}$ and $\boldsymbol{x}$ denote, respectively, the starting point of the process and the point at which it exits the criterion circle, a, and let θ denote the angle between them. Then the joint density of $(T_a, \boldsymbol{X}_T)$ for a 2D Wiener process with drift $\boldsymbol{\mu}$ is

[9] These two ways of parameterizing bias are not, in general, equivalent; they differ for spatially inhomogeneous diffusions like the Ornstein-Uhlenbeck process in which the drift rate is a function of the accumulated evidence (Busemeyer & Townsend, 1993), but they are the same for the Wiener process.

[10] Hsu (1986) reported an explicit expression for the first passage distribution only for a process of dimension $d \geq 3$, but noted that the solution for the $d = 2$ case can be obtained by the same methods as he used, although its form is slightly different. Yin and Wang (2009) reported explicit expressions for both the $d = 2$ and $d \geq 3$ cases.

$$P[T_a \in dt,\ \boldsymbol{X}_T \in d\boldsymbol{x}] = -\exp\left[\boldsymbol{\mu} \cdot (\boldsymbol{x} - \boldsymbol{z}) - \frac{\|\boldsymbol{\mu}\|^2 t}{2}\right] \times \left\{ \sum_{n=1}^{\infty} \frac{j_{0,n} J_0(\|\boldsymbol{z}\| j_{0,n}/a)}{2\pi a^3 J_0'(j_{0,n})} e^{-j_{0,n}^2 t/(2a^2)} + \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{j_{m,n}\ \cos(m\theta) J_m(\|\boldsymbol{z}\| j_{m,n}/a)}{\pi a^3 J_m'(j_{m,n})} e^{-j_{m,n}^2 t/(2a^2)} \right\}. \quad (32)$$

The hitting point, $\boldsymbol{x}$, is expressed in Cartesian instead of polar coordinates as done elsewhere in the article. The terms in braces are the joint density of the hitting time and hitting point for a zero-drift process starting at $\boldsymbol{z}$, where $\|\boldsymbol{z}\| < a$, and the exponential term on the left is the result of applying a Girsanov change-of-measure to the zero-drift process. Equation 32 is more complex than the corresponding expression for a process starting at the origin because it involves a doubly infinite series and, to my knowledge, its computational properties have not been studied.

Categorical decision boundaries. In the circular diffusion model, the outcomes of the process are continuously distributed on the range $(0, 2\pi)$ and the decision is determined by the hitting point $\boldsymbol{X}_\theta$. However, in some situations, the decision outcome might be expressed using a smaller number of response categories. This situation is depicted in Figure 14a, where the continuous set of decision outcomes is mapped to five unequally sized response categories, R_1 through R_5. The marginal probabilities of each response are then the integrals with respect to phase angle of the hitting point densities across the set of outcomes associated with the response. The response times are similarly obtained by integrating the joint density of $(T_a, \boldsymbol{X}_T)$ over hitting points and conditioning on the response probabilities.

One area in which a model like the one in Figure 14a may be useful is in episodic memory for color. Recent work by Persaud and Hemmer (2014) on memory for color showed that people's color reports are not uniformly distributed around the circle, but are biased toward a cardinal set of hue values associated with Berlin and Kay's (1969) universal color names. Potentially, these biases could be stimulus biases, response biases, or some combination of the two. If they are stimulus biases, then they could be represented as biases in the drift rates. If they are response biases, then one way to represent them would be to assume that the outcome of the decision process is mapped onto one of a smaller number of response categories.[11] Use of a formal decision model like the one in Figure 14a may be helpful in allowing these alternatives to be distinguished.

The introduction of categorical decision bounds implies a theoretical distinction between the outcome of the decision process as a cognitive state and the behavioral response that expresses that state. This distinction has typically not been made in the recent literature on sequential-sampling models, which has tended to characterize decision making as a process of accumulating evidence for the response. This is particularly so in applications that have used the models to characterize neural firing rates in the oculomotor control systems of monkeys performing saccade-to-target decision tasks (e.g., Purcell et al., 2010; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007). In these applications, a close coupling between the processes of evidence accumulation and saccade generation has been assumed. A contrasting interpretation can be found elsewhere in the work of Ratcliff and colleagues

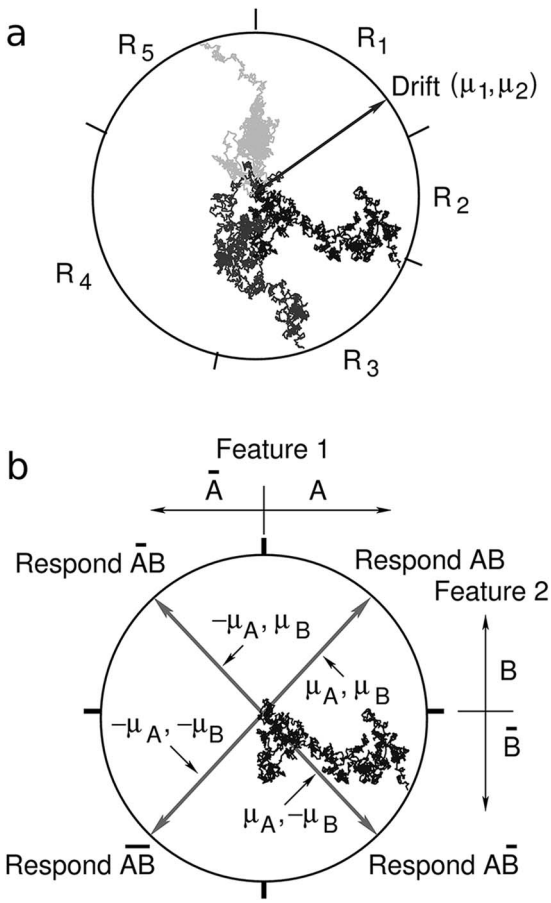

Figure 14. Models with categorical decision boundaries. (a) The circular outcome space is partitioned into five unequally sized response categories, R_1 through R_5. The response that is made is the one associated with the category corresponding to the hitting point, $\boldsymbol{X}_\theta$. Three sample paths are shown (different shades of gray), two terminating with response R_3 and one terminating with response R_5. (b) Model for multiattribute decision making. The stimulus space is defined by a pair of features: Feature 1 with levels A and $\bar{A}$, and Feature 2 with levels B and $\bar{B}$. Factorial combination of these features leads to four values of drift rate, (μ_A, μ_B), $(-\mu_A, \mu_B)$, $(\mu_A, -\mu_B)$, $(-\mu_A, -\mu_B)$. The boundary of the criterion circle is partitioned into four equally sized response categories with boundaries 0, $\pi/2$, π, and $3\pi/2$. The observer responds with the category that corresponds to the hitting point on the criterion circle.

(Gomez, Ratcliff, & Perea, 2007; Ratcliff, 2006), in which a diffusion model with an implicit boundary has been used to model decisions that are made but withheld. More generally, the distinction between stimulus identification and response selection is a fundamental one in the Donders-Sternberg stages approach to RT modeling, which has been a cornerstone of cognitive psychology since the end of the 1960s (Donders, 1869/1969; Sternberg, 1969; Townsend & Asbhy, 1983). The model with categorical bounds highlights the theoretical importance of this distinction, and of the need to account for response selection processes, particularly in

[11] I am grateful to Jerome Busemeyer for suggesting the idea of augmenting the circular model with categorical decision boundaries.

tasks in which there may be a complex (e.g., many-to-one) mapping between stimuli and responses.

Multiattribute decision making. One potential application of the model with categorical bounds is to decisions about multiattribute stimuli. Multidimensional signal detection theory (Macmillan & Creelman, 2005) and general recognition theory (Ashby & Townsend, 1986) provide theoretical accounts of accuracy judgments about such stimuli, but there is as yet no general theory of RT judgments about them—although Ashby (2000) developed a stochastic version of general recognition theory to model two-choice RT. In his model stimuli are represented statistically as points in a multidimensional space and evidence about them is accumulated by a multidimensional diffusion process. Decisions are made by computing a linear discriminant function on the values of this process. The circular diffusion model with categorical response boundaries makes similar representational assumptions to those in Ashby's model but allows for a greater range of responses.

The general set-up is depicted in Figure 14b. The figure shows a decision space for stimuli defined by a pair of binary features or attributes, which have values A and $\overline{A}$, and B and $\overline{B}$, respectively. Associated with each of the four combinations of these features is a vector-valued drift rate, (μ_A, μ_B), $(\mu_A, -\mu_B)$, $(-\mu_A, \mu_B)$, and $(-\mu_A, -\mu_B)$, and an associated partitioning of the criterion boundary into response categories, AA, $A\overline{B}$, $\overline{A}B$, and $\overline{AB}$. Each of the response categories is centered on the phase angle defined by the corresponding drift rates. When $\mu_A = \mu_B$, the category centers are at $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$, as shown. For a stimulus with drift rate (μ_i, μ_j), the most likely hitting point of the boundary, $\boldsymbol{X}_\theta$, is $\arctan(\mu_j/\mu_i)$, $i \in \{A, \overline{A}\}$, $j \in \{B, \overline{B}\}$, so a maximum-likelihood observer will respond with the category that corresponds to the hitting point. The model of Figure 14b represents decisions about stimuli comprised of pairs of features, but it generalizes in a natural way to higher dimensions. For d-feature stimuli, the zero-drift process would be modeled as a d-dimensional Bessel process (Equation 20, Hamana & Matsumoto, 2013) and the points on the boundary of the decision space in the Girsanov theorem would be parameterized in spherical (or hyperspherical) coordinates (Arkfen & Weber, 2001, pp. 121–124).

The domain in which the model of Figure 14b is most likely to apply is to tasks in which the stimulus is processed as a single, unitary entity, for example, to judgments about the color and orientation of a Gabor patch, when the decision is based on the joint likelihood of its two components. An alternative way to model multiattribute decision making is with latent state models, in which decisions are made by combining the outputs of separate detectors that each make a decision about an individual feature. In psychophysics, the latent state approach has led to the widely studied independent detectors model and the closely related *MAX* model (Eckstein, Thomas, Palmer, & Shimozaki, 2000; Graham, 1989; Shaw, 1982; Smith, 1998; Palmer, Verghese, & Pavel, 2000). These models assume that a target stimulus is detected if any one, or any appropriate combination, of a set of component detectors responds.

In RT, the latent state approach has let to logical rule models (Fific, Little, & Nosofsky, 2010), in which both RT and the decision outcome are determined by the responses of a set of parallel detectors, each of which reports a categorical outcome. In psychophysical decisions, such as the detection of the presence or absence of a target stimulus in an m-item display, likelihood based models and independent detectors models often make predictions that are essentially indistinguishable (Nolte, & Jaarsma, 1967), but in more complex tasks their predictions can differ. Ma, Shen, Dziugaite, and van den Berg (2015) were critical of *MAX* models and documented a number of settings in which they perform more poorly than models based on full likelihoods. In comparison to latent state models, the circular diffusion model with categorical bounds more closely resembles the likelihood-based models advocated by Ma and colleagues.

Discussion

In this article, I have presented a new theory of decision making in the continuous report task and obtained a complete analytic characterization of its properties. The principal assumption of the theory is that memory retrieval in continuous report involves a decision process, but one with a continuous rather than a discrete outcome set. I have modeled this process as a two-dimensional diffusion process with independent components on a disk, whose bounding circle represents the decision criterion. The model, purposely, shares many properties with the standard diffusion model for 2AFC decisions. Most fundamentally, it assumes that trial-to-trial variability in decision outcome and decision time are because of moment-by-moment noise in the evidence accumulation process. Despite the comparative complexity of the model, I have been able to obtain analytic predictions for all of its relevant properties: decision time distributions, mean decision times, and distributions of decision outcomes. Given the difficulties associated with generalizing the standard diffusion model in a tractable way to decisions with more than two alternatives, the relative ease with which the model can generalize to decision tasks with an infinite number of outcomes is striking.

The theoretical benefit of treating continuous report as a 2D diffusion process is that it provides an explicit link between the literatures on speeded 2AFC decision making, on the one hand, and working memory retrieval, on the other. The standard diffusion model has provided a successful account of two-choice decision-making in a range of cognitive tasks and estimates of its parameters have provided theoretically based, interpretable characterizations of performance in such tasks, including the individual differences factors that distinguish participant populations. The continuous report task plays a central role in contemporary working memory research and its proponents have argued that it allows tests between alternative models of working memory capacity that are difficult to carry out using conventional 2AFC tasks. For a variety of working memory precision studies, the best model appears to be one in which the precision parameter, which determines variability in report accuracy across trials, is itself subject to trial-to-trial variability.

The 2D diffusion model connects these two bodies of research, and its properties mirror and extend those of models that have been successful in both of these domains. As shown in previous sections, all of the features of the standard diffusion model that are important in fitting data are reproduced, to a quite remarkable degree, by the 2D model: The model predicts unimodal, positively skewed distributions of decision times whose shapes remain invariant across changes in stimulus quality and the amount of evidence needed for a response. The model also predicts a continuous counterpart of the fast and slow errors that are found in

speed-emphasis and accuracy-emphasis conditions, respectively, as functions of variability in criterion or drift magnitude. In addition, the model also predicts a von Mises distribution of decision outcomes, consistent with the way in which precision is modeled in the working memory literature.

The idea that working memory precision varies across trials was one of the important conclusions from van den Berg et al.'s (2014) model comparison study. When viewed from a diffusion model perspective, this conclusion is wholly consistent with the idea that drift rates and decision criteria vary from trial to trial, as these two quantities jointly determine the predicted precision. In the working memory literature, precision is treated as an indivisible parameter of performance, but in the 2D model, precision is equal to $\kappa = a\|\boldsymbol{\mu}\|/\sigma^2$, the scaled product of the drift magnitude, which depends on the stimulus representation, and the criterion, which determines the amount of evidence needed for a response.

Perhaps the most important theoretical contribution of the 2D diffusion model is that it predicts distributions of decision times in continuous report. In the 2AFC literature, fits of RT distributions have been invaluable in testing between alternative models of the decision process, and it seems likely that they can contribute in a similar way to working memory research. One immediate payoff of taking decision times into account is that they can help identify the processes that give rise to variations in precision, both within and between conditions. If only response accuracy is taken into account, then the components of precision cannot be independently identified: Providing that κ remains fixed, changing drift magnitude and changing criterion will have identical effects on precision and the distribution of responses. However, they can be distinguished by considering their effect on decision times, where they are predicted to have opposite effects. In a similar way, across-trial variation in precision, as assumed in variable-precision models, can be linked theoretically to trial-to-trial variation in drift magnitude or criterion in the 2D model. Potentially, then, an analysis at the level of decision times can provide a way to link variable precision and its determinants to the relative speed of accurate and inaccurate responses.

For these potential theoretical benefits to be realized, it must be possible to measure RT in continuous report tasks. Although this has not been a feature of contemporary working memory studies, there appears to be no reason in principle why it should not be possible to do so, although there may be methodological hurdles to obtaining good data. The biggest is likely to be the nondecision component of RT, which may include a substantial component of motor variability. The standard diffusion model includes a component of nondecision time variability, $T_{er,}$ and, complementing this, the variable-precision models of van den Berg et al. (2014) included a component that represents imprecision in the motor response. A suitable model for RT in the continuous report task would need to take both kinds of variability into account. Ideally, such a model should be used in conjunction with methods designed to minimize the effects of motor variability experimentally.

One of the most important theoretical properties of the 2D diffusion model is that it provides a process account of the von Mises distribution of precision. Rogers and Williams (1987/2000, pp. 141–142) noted that there is another way to obtain a von Mises distribution from a diffusion process, other than the one I have developed here, namely, from diffusion on a one-dimensional circular manifold. (In differential geometry, a circle in the Euclidean plane can be viewed as an embedded, 1D manifold.) The process described by Rogers and Williams was originally analyzed by Carverhill (1985). Instead of beginning at the center of the circle and terminating when it hits the boundary, Carverhill's process begins at a point on the boundary of the circle and diffuses around the perimeter with the passage of time. Although Carverhill's process provides another way to obtain a von Mises distribution of decision outcomes, it does not provide a way to model decision times. It might, however, be useful as a way of modeling a loss of precision in memory representations during a retention interval.

Conclusion

The capacity of visual working memory is an important determinant of capacity limitations in the human processing system and an understanding of these limitations is fundamental to our understanding of the system as a whole. Since the introduction of the continuous report task around a decade ago, there has been a proliferation of precision studies that have led to the discovery of new phenomena and, arguably, a reconceptualization of how working memory is organized. One of the disadvantages of the continuous report task is that it has not always been clear how to relate the results of continuous report studies theoretically to those of studies using more traditional cognitive methods, such as 2AFC tasks. The 2D model provides a mathematically precise theoretical link between the processes of two-alternative decision making and continuous report. As a result, it provides a way, in future, to allow the powerful tools of RT modeling to be brought to bear in precision studies. The theoretical and mathematical properties of diffusion processes provide a contemporary rationale for the historical practices of classical psychophysics, which assumed that the method of adjustment and the method of constant stimuli were equivalent ways of characterizing the same sensory and cognitive processes. According to the theory presented here, performance on both kinds of tasks is limited by the same noisy decision processes.

References

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

Arkfen, G. B., & Weber, H. J. (2001). *Mathematical methods for physicists* (5th ed.). San Diego, CA: Academic Press.

Ashby, F. G. (1983). A biased random walk for two choice reaction times. *Journal of Mathematical Psychology, 27,* 277–297.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44,* 310–329.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Audley, R. J., & Pike, A. R. (1965). Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology, 18,* 207–225.

Baxter, M., & Rennie, A. (1996). *Financial calculus: An introduction to derivative pricing*. Cambridge, United Kingdom: Cambridge University Press.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience, 34,* 3632–3645.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9,* Article 7.

Bays, P., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision, 11,* Article 6.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.

Blurton, S. P., Kesselmeir, M., & Gondan, M. (2012). Fast and accurate calculations of cumulative first-passage time distributions in Wiener diffusion models. *Journal of Mathematical Psychology, 56,* 470–475.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (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.

Borodin, A. N., & Salminen, P. (1996). *Handbook of Brownian motion—Facts and formulae*. Basel: Birkhäser.

Brewer, N., & Smith, G. A. (1989). Developmental changes in processing speed: Influence of speed-accuracy regulation. *Journal of Experimental Psychology: General, 118,* 298–310.

Busemeyer, J., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459.

Carverhill, A. P. (1985). Flows of stochastic dynamical systems: Ergodic theory. *Stochastics, 14,* 273–318.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

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

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology, 47,* 304–322.

Donders, F. C. (1869/1969). On the speed of mental processes. Translated by W. G. Koster, in W. G. Koster (Ed.), *Attention and Performance II*, *Acta Psychologica, 30,* 412–431.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics, 62,* 425–451.

Engen, T. (1971). Psychophysics I. Discrimination and detection. In J. W. Kling & L. A. Riggs (Eds.) *Woodworth and Schlosberg's experimental psychology* (3rd ed., pp. 11–46). New York, NY: Holt, Rinehart & Winston.

Fan, J. E., & Turk-Brown, N. B. (2013). Internal attention to features in visual short-term memory guides object learning. *Cognition, 129,* 292–308.

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review, 117,* 309–348.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge, United Kingdom: Cambridge University Press.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11,* Article 3.

Getoor, R. K., & Sharpe, M. J. (1979). Excursions of Brownian motion and Bessel processes. *Zeitschrift für Wahrscheinlichkeitstheorie, 47,* 83–106.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136,* 347–369.

Graham, N. V. S. (1989). *Visual pattern analyzers*. New York, NY: Wiley.

Green, D. M., Smith, A. F., & von Gierke, S. M. (1983). Choice reaction time with a random foreperiod. *Perception & Psychophysics, 34,* 195–208.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Gunseli, E., van Moorselaar, D., & Meeter, M., & Olivers, C. N. L. (2015). The reliability of retro-cues determine working memory representations. *Psychonomic Bulletin & Review, 22,* 1334–1341.

Hamana, Y., & Matsumoto, H. (2013). The probability distributions of the first hitting times of Bessel processes. *Transactions of the American Mathematical Society, 365,* 5237–5257.

Harrison, M. (1985). *Brownian motion and stochastic flow systems*. New York: NY: Wiley.

Heath, R. A., & Kelly, L. (1988). An application of a discriminatory index for the assessment of individual differences. In R. A. Heath (Ed.), *Current issues in cognitive development and mathematical psychology*. Newcastle, N. S. W.: University of Newcastle.

Hsu, P. (1986). Brownian exit distributions of a ball. In E. Çinlar, K. L. Chung, R. K. Getoor, & J. Glover (Eds.) *Seminar on stochastic processes, 1985* (pp. 108–116). Boston, MA: Birkhäuser.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (Vol. 2, 2nd ed.) New York, NY: Wiley.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review, 121,* 1–32.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Karatzas, I., & Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. New York, NY: Springer.

Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes*. New York, NY: Academic Press.

Kent, J. T. (1978). Some probabilistic properties of Bessel functions. *Annals of Probability, 6,* 760–770.

Kent, J. T. (1980). Eigenvalue expansions for diffusion hitting times. *Zeitschrift für Wahrscheinlichkeitstheorie, 52,* 309–319.

Kool, W., Conway, A. R. A., & Turk-Browne, N. B. (2014). Sequential dynamics in visual short-term memory. *Attention, Perception, & Psychophysics, 76,* 1885–1901.

Laming, D. (1968). *Information theory of choice reaction times*. New York, NY: Academic Press.

Laming, D. (1979). A critical comparison of two random-walk models for two choice reaction time. *Acta Psychologica, 43,* 431–453.

Link, S. W. (1975). The relative judgment theory of two-choice response time. *Journal of Mathematical Psychology, 12,* 114–135.

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

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17,* 347–356.

Ma, W. J., Shen, S., Dziugaite, G., & van den Berg, R. (2015). Requiem for the max rule? *Vision Research, 116,* 179–193.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision, 13,* Article 21.

McNicol, D. (1972/2005). *A primer of signal detection theory*. London, United Kingdom: George Allen & Unwin. Reprinted 2005 by Erlbaum.

Mörters, P., & Peres, Y. (2010). *Brownian motion*. Cambridge, United Kingdom: Cambridge University Press.

Nolte, L. W., & Jaarsma, D. (1967). More on the detection of one of *M* orthogonal signals. *Journal of the Acoustical Society of America, 41,* 497–505.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40,* 1227–1268.

Pearson, B., Raškevičius, J., Bays, P. M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision, 14,* Article 2.

Persaud, K., & Hemmer, P. (2014). *The influence of knowledge and expectations for color on episodic memory*. Proceedings of the 36th Annual Conference of the Cognitive Science Society. Quebec City, Canada.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review, 117,* 1113–1143.

Rabbitt, P. M. A., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29,* 727–743.

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision, 12,* Article 21.

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

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53,* 195–237.

Ratcliff, R., Hasegawa, Y., Hasegawa, R., Smith, P. L., & Segraves, M. (2007). A dual diffusion model for single cell recording data from the superior colliculus in brightness discrimination task. *Journal of Neurophysiology, 97,* 1756–1797.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential-sampling models for two choice reaction time. *Psychological Review, 111,* 333–367.

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relationship between perceptual encoding and decision making. *Journal of Experimental Psychology: General, 139,* 70–94.

Ratcliff, R., & Smith, P. L. (2015). Modeling simple decisions and applications using a diffusion model. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (pp. 35–62). Oxford, United Kingdom: Oxford University Press.

Ratcliff, R., Smith, P. L., & McKoon, G. (2015). Modeling response time and accuracy data. *Current Directions in Psychological Science, 24,* 458–470.

Reuter, G. E. H. (1959). Denumerable Markov processes, II. *Journal of the London Mathematical Society, 34,* 81–91.

Revus, D., & Yor, M. (1990). *Continuous martingales and Brownian motion*. Berlin: Springer-Verlag.

Rogers, L. C. G., & Williams, D. (1987/2000). *Diffusions, Markov processes and martingales. Vol. 2. Itô calculus*. Chichester, United Kingdom: Wiley, reprinted 2000 New York: Cambridge University Press.

Rouder, J. N., Yue, Y., Speckman, P. L., Pratte, M. S., & Province, J. M. (2010). Gradual growth versus shape invariance in perceptual decision making. *Psychological Review, 117,* 1267–1274.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 2214–2242.

Shaw, M. I. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology, 14,* 353–409.

Smith, P. L. (1990). A note on the distribution of response times for a random walk with Gaussian increments. *Journal of Mathematical Psychology, 34,* 445–459.

Smith, P. L. (1998). Attention and luminance detection: A quantitative analysis. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 105–133.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463.

Smith, P. L. (2015). The Poisson shot noise model of visual short-term memory and choice response time: Normalized coding by neural population size. *Journal of Mathematical Psychology, 66,* 41–52.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27,* 161–168.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116,* 283–317.

Smith, P. L., & Ratcliff, R. (2015a). An introduction to the diffusion model of decision-making. In E. J. Wagenmakers & B. U. Forstmann (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 35–62). New York, NY: Springer.

Smith, P. L., & Ratcliff, R. (2015b). Diffusion and random walk processes. In James D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 6, pp. 395–401). Oxford: Elsevier.

Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review, 120,* 589–627.

Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology, 53,* 293–315.

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

Smith, P. L. Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychological Review, 121,* 679–688.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donders' method. In W. G. Koster (Ed.), *Attention and performance II*, *Acta Psychologica, 30,* 276–315.

Stroock, D. W., & Varadhan, S. R. S. (1979). *Multidimensional diffusion processes*. Berlin, Germany; Springer-Verlag.

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception & Psychophysics, 76,* 2136–2157.

Swensson, R. G., & Green, D. M. (1977). On the relations between random walk models of two-choice response times. *Journal of Mathematical Psychology, 15,* 282–391.

Thomas, E. A. C. (1975). A note on the sequential probability ratio test. *Psychometrika, 40,* 107–111.

Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology, 21,* 136–152.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Townsend, J. T., & Asbhy, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, United Kingdom: Cambridge University Press.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121,* 124–149.

van den Berg, R., Shin, H., Chou, W-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 8780–8785.

Vickers, D. (1977). An adaptive module for simple judgments. In J. Requin (Ed.), *Attention and Performance, VII* (pp. 599–614). Hillsdale, NJ: Erlbaum.

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology and Life Sciences, 2,* 169–194.

Vickers, D., & Lee, M. D. (2000). Dynamic models of simple judgments: II. Properties of a self-regulating PAGAN (Parallel, Adaptive, Generalized, Accumulator Network) model for multi-choice tasks. *Nonlinear Dynamics, Psychology and Life Sciences, 4,* 1–31.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4,* Article 11.

Williams, D. (1991). *Probability with martingales*. Cambridge, United Kingdom: Cambridge University Press.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology* (Rev. ed.). London, United Kingdom: Methuen.

Yellott, J. I., Jr. (1971). Correction for guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology, 8,* 159–199.

Yellott, J. I., Jr. (1977). The relationship between Luce's Choice Axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology, 8,* 159–199.

Yin, C., & Wang, C. (2009). Hitting time and place of Brownian motion with drift. *The Open Statistics and Probability Journal, 1,* 38–42.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235.

Appendix A

The Girsanov Theorem

Rigorous presentations of the Girsanov theorem can be found in Karatzas and Shreve (1991, pp. 190–197), Revus and Yor (1990, pp. 301–313), and Rogers and Williams (1987/2000, pp. 81–89). An elementary treatment can be found in Baxter and Rennie, 1996, pp. 63–75).

Let $\boldsymbol{X}_t$ be a d-dimensional, zero-drift, Brownian motion process with unit infinitesimal *SD* (i.e., $\sigma_i = 1$ for all i). Denote by $P_t(.)$ the probability measure on $\boldsymbol{X}_t$, where

$$P_t(A) = P[\boldsymbol{X}_t \in A] = E_t[1_A]. \tag{A1}$$

The quantity $P_t(A)$ is the probability that the process will be found in a designated set $A \subset \mathbb{R}^d$ at time t. The second equality on the right states that this probability is equal to the expected value at time t of the indicator of the set A, where the indicator function 1_A is equal to unity for all points $(x, y) \in A$, and zero otherwise.

Define the d-dimensional exponential martingale

$$\boldsymbol{Z}_t(\boldsymbol{X}) = \exp\left[\sum_{i=1}^{d}\int_0^t \gamma_s^i \, dX_s^i - \frac{1}{2}\int_0^t \|\gamma_s\|^2 ds\right], \tag{A2}$$

where γ_t is a progressively measurable, previsible process. A "previsible" process is a technical concept from the theory of stochastic processes, and refers to a process whose value γ_t at time t can be inferred from knowledge of the values of γ_s, $s < t$. The class of previsible processes includes continuous processes but excludes processes with jumps. "Progressive measurability" is, loosely, the requirement that γ_t should be defined on the same probability space as the process $\boldsymbol{X}_t$ and that its properties should be fully characterized there.

Define a new probability measure

$$\tilde{P}_t(A) = E_t[1_A \boldsymbol{Z}_t(\boldsymbol{X})].$$

The Girsanov theorem states that, under $\tilde{P}_t(.)$, the process

$$\tilde{X}_t^i = X_t^i - \int_0^t \gamma_s^i ds;\ 1 \le i \le d, \tag{A3}$$

is a d-dimensional Brownian motion. This means that the probability distribution of the nonzero-drift process, $\tilde{P}_t(A)$, can be written in terms of the distribution of the zero-drift process $P_t(A)$, using the relationship

$$d\tilde{P}_t(A) = \boldsymbol{Z}_t(\boldsymbol{X}) dP_t(A), \tag{A4}$$

where $d\tilde{P}_t(A)$ and $dP_t(A)$ are probability density functions, expressed in differential notation. The statement that, under $\tilde{P}_t(.)$, $\tilde{X}_t^i$ is a Brownian motion (i.e., a zero drift process), means that the original process

$$X_t^i = \tilde{X}_t^i + \int_0^t \gamma_s^i \, ds;\ 1 \le i \le d,$$

is a Brownian motion with drift γ_t^i. Because of this fact, the Girsanov theorem can be characterized as "removing the drift" from the process $\boldsymbol{X}_t$. What this means in practice is that the probability distribution of the nonzero-drift process can be expressed as a function jointly of the distribution of the zero-drift process and the exponential martingale $\boldsymbol{Z}_t(\boldsymbol{X})$ in Equation A2 via Equation A4.

In many applications of the Girsanov theorem, γ_t is a stochastic process. In the 2D model, however, γ_t is simply equal to the drift vector, $\boldsymbol{\mu}$, which is constant, and so is trivially a previsible process. Under these conditions, the Itô (stochastic) integrals with respect to dX_s^i on the right of Equation A2 are simply equal to $\mu_i X_t^i$, $1 \le i \le d$. Under these circumstances, Equation A2 reduces to the exponential martingale of Equation 16.

(*Appendices continue*)

Appendix B

Independence of Decision Times and Decision Outcomes

The proof of the independence of decision times and decision outcomes given below is based on the proof in Rogers and Williams (1987/2000, p. 84). Let $T = \inf\{t : |\boldsymbol{X}_t| = a\}$ be the time at which the process first hits the boundary a and let $\boldsymbol{X}_T$ be its value at the hitting time. Elsewhere I denote the hitting time as T_a but I omit the subscript here to simplify the notation. Let P_t and $\tilde{P}_t$ denote the probability measures for the zero-drift and nonzero-drift processes, respectively, and let E and $\tilde{E}$ denote the corresponding expectations with respect to these measures. The random variables T and $\boldsymbol{X}_T$ are independent under the probability measure $\tilde{P}_t$.

Proof

Let $f(\boldsymbol{X}_T)$ and $g(T)$ be functions of $\boldsymbol{X}_T$ and T. It is clear, by symmetry, that when $\boldsymbol{\mu} = 0$, $\boldsymbol{X}_T$ and T are independent under the probability measure P_t, so

$$E[f(\boldsymbol{X}_T)g(T)] = E[f(\boldsymbol{X}_T)]E[g(T)]. \tag{B1}$$

For $\boldsymbol{\mu} \neq 0$, under the change in measure $\tilde{P}_t$,

$$d\tilde{P}_T(\boldsymbol{X}) = \boldsymbol{Z}_T(\boldsymbol{X})dP_T(\boldsymbol{X}),$$

where

$$\boldsymbol{Z}_T(\boldsymbol{X}) = \exp\left[(\boldsymbol{\mu}\cdot\boldsymbol{X}_T) - \frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right]$$

is the exponential martingale of Equation 16. Taking expectations with respect to $\tilde{P}_t$,

$$\begin{aligned}\tilde{E}[f(\boldsymbol{X}_T)g(T)] &= E\left\{f(\boldsymbol{X}_T)g(T)\exp\left[(\boldsymbol{\mu}\cdot\boldsymbol{X}_T) - \frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right]\right\}\\ &= E\{f(\boldsymbol{X}_T)\exp(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)\}E\left\{g(T)\exp\left[-\frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right]\right\},\end{aligned} \tag{B2}$$

where Equation B1 is used to obtain the product of expectations in the second equality. Multiplying the right hand side of Equation B2 by

$$E[\exp(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)]E\left[-\frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right] / E[\exp(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)]E\left[-\frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right]$$

yields

$$\begin{aligned}&\tilde{E}[f(\mathbf{X}_T)g(T)] =\\ &E\left[f(\boldsymbol{X}_T)\exp(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)\right]E\left[-\frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right]\\ &\quad\times E\left[g(T)\exp\left(-\frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right)\right]E\left[\exp(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)\right]\\ &\frac{}{E\left[\exp(\boldsymbol{\mu}\cdot\boldsymbol{X}_T)\right]E\left[-\frac{1}{2}\|\boldsymbol{\mu}\|^2 T\right]}.\end{aligned} \tag{B3}$$

The denominator of Equation B3 is the expectation of $\boldsymbol{Z}_T(\boldsymbol{X})$ with respect to P_t. The process $\boldsymbol{Z}_T(\boldsymbol{X})$ is a martingale of a 2D Brownian motion starting at the origin at time $t = 0$, so $E[\boldsymbol{Z}_T(\boldsymbol{X})] = \boldsymbol{Z}_0(\boldsymbol{0}) = 1$, that is, its expectation is equal to unity. Using Equation B1 to combine terms in the numerator on the right of Equation B3 then yields

$$\tilde{E}[f(\boldsymbol{X}_T)g(T)] = \tilde{E}[f(\boldsymbol{X}_T)]\tilde{E}[g(T)]. \tag{B4}$$

This shows the independence of functions of $\boldsymbol{X}_T$ and T under $\tilde{P}_t$. In particular, let $f(\boldsymbol{X}_T) = I_{\{\boldsymbol{X}_T \in d\theta\}}$ be the indicator function of the event that the process hits the bounding circle, a, in a small region of arc $d\theta$ centered on θ and let $g(T) = I_{\{T\in dt\}}$ be the indicator function of the event that it hits the bounding circle in a small time interval $(t, t + dt)$. Then Equation B4 yields

$$\tilde{E}[I_{\{\boldsymbol{X}_T\in d\theta, T\in dt\}}] = \tilde{E}[I_{\{\boldsymbol{X}_T\in d\theta\}}]\tilde{E}[I_{\{T\in dt\}}], \tag{B5}$$

that is, the decision time and the decision outcome are independent of each other. Expressed in terms of probability measures, Equation B5 states that

$$d\tilde{P}[T \in dt, \theta_T \in d\theta] = d\tilde{P}_t(a)d\tilde{P}_\theta(a) \tag{B6}$$

This completes the proof.

Appendix C

Mean Decision Time

The mean decision time is the mean first passage time, or hitting time, for the nonzero-drift process through or of the criterion boundary, a. As a first step to obtaining the mean, we note that Equation B6 gives a decomposition of the joint probability measure over hitting times and hitting points into a product of measures on these variables. Because the process is guaranteed to terminate at the boundary a in finite time with probability one, integrating this product over hitting points and hitting times gives

$$\left(\int_0^{2\pi} d\tilde{P}_\theta(a)\right)\left(\int_0^{\infty} d\tilde{P}_t(a)\right) = 1. \tag{C1}$$

The zero-drift process is guaranteed to terminate with probability one because each of its components is almost surely unbounded (Karatzas & Shreve, 1991, p. 240). The nonzero-drift process is guaranteed to terminate with probability one because its probability measure and that of the zero-drift process are absolutely continuous

(*Appendices continue*)

with respect to one another, so it inherits all of the zero-drift process's almost-sure properties (Mörters & Peres, 2010, p. 25).

The densities $d\tilde{P}_\theta(a)$ and $d\tilde{P}_t(a)$ in Equation C1 are obtained by transforming $dP_\theta(a)$ and $dP_t(a)$, the corresponding densities for the zero-drift process, by $\exp[\boldsymbol{\mu} \cdot \boldsymbol{X}_T(\theta)]$ and $\exp(-\|\boldsymbol{\mu}\|t/2)$, the phase-angle dependent and time-dependent parts of the exponential martingale, $\boldsymbol{Z}_t(\boldsymbol{X})$, respectively. The density $dP_t(a)$ is the first passage time density of the Bessel process of Equation 22, while $dP_\theta(a) = (2\pi)^{-1}$, the uniform density on the unit circle.

The mean decision time is obtained by integrating the first passage time with respect to $d\tilde{P}_t(a)$,

$$\tilde{E}[T] = \frac{1}{\tilde{P}(a)}\int_0^\infty t d\tilde{P}_t(a), \tag{C2}$$

where

$$\tilde{P}(a) = \int_0^\infty d\tilde{P}_t(a)$$

makes $d\tilde{P}_t(a)$ into a proper density function. An expression for $\tilde{P}(a)$ is derived below. Its reciprocal is the normalization term which makes the distribution of hitting points, $\exp[\boldsymbol{\mu} \cdot \boldsymbol{X}_T(\theta)]/(2\pi)$, into a von Mises distribution (see the discussion around Equations 27 and 28 in the text for details).

We evaluate the expected value of the first passage time with respect to $d\tilde{P}_t(a)$ in Equation C2 via its moment generating function, considering first the notationally simpler $\sigma = 1.0$ case. The moment generating function (Laplace transform) associated with the second integral on the right of Equation C2 is

$$\begin{aligned}\tilde{E}[e^{-\lambda T}] &= \int_0^\infty e^{-\lambda t} d\tilde{P}_t(a) \\ &= \int_0^\infty \exp\left(-\lambda t - \frac{1}{2}\|\boldsymbol{\mu}\|^2 t\right) dP_t(a) \\ &= E\left[\exp\left\{-\left(\lambda + \frac{1}{2}\|\boldsymbol{\mu}\|^2\right)T\right\}\right].\end{aligned} \tag{C3}$$

The expectations on the left and right hand sides of this equation are taken with respect to the probability measures of the nonzero-drift process and zero-drift process, respectively. We relate Equation C3 to the moment generating function of the zero-drift process.

The moment generating function of the first passage time for the Bessel process was given by Borodin and Salminen (1996, Equation 2.0.1) and Hamana and Matsumoto (2013, Equation 2.2). The process associated with Euclidean distance in the plane corresponds, in the notation of Hamana and Matsumoto's article, to the case $\nu = 0$, although I do not use that notation here. For a process with $\sigma = 1$, starting point z, and criterion boundary a, the moment generating function is

$$E[e^{-\lambda T}] = \frac{I_0(z\sqrt{2\lambda})}{I_0(a\sqrt{2\lambda})}, \tag{C4}$$

where $I_0(.)$ is a modified Bessel function of the first kind of order zero. Making use of the fact that $I_0(0) = 1$, we take the limit $z \to 0$ to obtain the moment generating function for a process starting at zero,

$$E[e^{-\lambda T}] = [I_0(a\sqrt{2\lambda})]^{-1}. \tag{C5}$$

The probability mass in $d\tilde{P}_t(a)$, denoted $\tilde{P}(a)$, is equal to $\tilde{E}[e^{-\lambda T}]|_{\lambda=0}$, the value of the generating function in Equation C3 evaluated at $\lambda = 0$. We write

$$\lambda' + \lambda = \frac{1}{2}\|\boldsymbol{\mu}\|^2,$$

and substitute this value in equation C5, which yields

$$\tilde{E}[e^{-\lambda T}] = E[e^{-\lambda' T}] = \left[I_0(a\sqrt{2\lambda + \|\boldsymbol{\mu}\|^2})\right]^{-1}, \tag{C6}$$

from which we obtain the probability mass $\tilde{P}(a)$,

$$\tilde{P}(a) = \tilde{E}[e^{-\lambda T}]|_{\lambda=0} = [I_0(a\|\boldsymbol{\mu}\|)]^{-1}. \tag{C7}$$

To obtain the mean first passage time, we recall that $E[T] = -(d/d\lambda)E[e^{-\lambda T}]|_{\lambda=0}$, that is, the mean is minus the derivative of the moment generating function evaluated at zero. Differentiating Equation C6 with respect to λ, making use the fact that $(d/dx)I_0(x) = I_1(x)$, yields

$$\frac{d}{d\lambda}\tilde{E}[e^{-\lambda T}] = -\frac{aI_1(a\sqrt{2\lambda + \|\boldsymbol{\mu}\|^2})}{\sqrt{2\lambda + \|\boldsymbol{\mu}\|^2}\left[I_0(a\sqrt{2\lambda + \|\boldsymbol{\mu}\|^2})\right]^2}.$$

We take minus the value of this expression, evaluated at $\lambda = 0$, and divide by the probability mass in Equation C7 to obtain the mean first passage time,

$$E[T] = \frac{aI_1(a\|\boldsymbol{\mu}\|)}{\|\boldsymbol{\mu}\|I_0(a\|\boldsymbol{\mu}\|)}. \tag{C8}$$

The mean for the $\sigma \neq 1$ case is obtained by performing the same calculation using $\boldsymbol{Z}_T(\boldsymbol{X})$ as defined in Equation 17, with a change of variable, $s = \sigma^2 t$. We obtain

$$\tilde{P}(a) = \left[I_0(a\|\boldsymbol{\mu}\|/\sigma^2)\right]^{-1} \tag{C9}$$

for the probability mass and

$$E[T] = \frac{aI_1(a\|\boldsymbol{\mu}\|/\sigma^2)}{\|\boldsymbol{\mu}\|I_0(a\|\boldsymbol{\mu}\|/\sigma^2)} \tag{C10}$$

for the mean first passage time. The reader should note that Equation C10 is dimensionally correct: Its value is unaltered by a change in units for a, $\|\boldsymbol{\mu}\|$, and σ. The decision criterion is dimensioned in units of accumulated evidence, u; the drift norm is dimensioned in units of evidence per unit time, $u \cdot t^{-1}$, and the diffusion coefficient is dimensioned in squared evidence units per unit time, $u^2 \cdot t^{-1}$. Consequently, the arguments to the functions $I_1(.)$ and $I_0(.)$ are dimensionless while the ratio $a/\|\boldsymbol{\mu}\|$ has dimension $u/(u \cdot t^{-1}) = t$. The left and right hand side of Equation C10 are therefore both dimensioned in units of time, as required. The predicted mean decision times in Figures 7 and 8 were obtained using Equation C10.

Received March 15, 2015
Revision received December 17, 2015
Accepted December 22, 2015 ■