

Psychological Review

Modeling Continuous Outcome Color Decisions With the Circular Diffusion Model: Metric and Categorical Properties

Philip L. Smith, Saam Saber, Elaine A. Corbett, and Simon D. Lilburn

Online First Publication, March 9, 2020. <http://dx.doi.org/10.1037/rev0000185>

CITATION

Smith, P. L., Saber, S., Corbett, E. A., & Lilburn, S. D. (2020, March 9). Modeling Continuous Outcome Color Decisions With the Circular Diffusion Model: Metric and Categorical Properties. *Psychological Review*. Advance online publication. <http://dx.doi.org/10.1037/rev0000185>

Modeling Continuous Outcome Color Decisions With the Circular Diffusion Model: Metric and Categorical Properties

Philip L. Smith and Saam Saber
The University of Melbourne

Elaine A. Corbett
Trinity College Dublin and The University of Melbourne

Simon D. Lilburn
The University of Melbourne

The circular diffusion model is extended to provide a theory of the speed and accuracy of continuous outcome color decisions and used to characterize eye-movement decisions about the hues of noisy color patches in an isoluminant, equidiscriminability color space. Heavy-tailed distributions of decision outcomes were found with high levels of chromatic noise, similar to those found in visual working memory studies with high memory loads. Decision times were longer for less accurate decisions, in agreement with the slow error property typically found in difficult 2-choice tasks. Decision times were shorter, and responses were more accurate in parts of the space corresponding to nameable color categories, although the number and locations of the categories varied among participants. We show that these findings can be predicted by a theory of across-trial variability in the quality of the evidence entering the decision process, represented mathematically by the drift rate of the diffusion process. The heavy-tailed distributions of decision outcomes and the slow-error pattern can be predicted by either of 2 models of drift rate. One model is based on encoding failures and the other is based on a nonlinear transformation of the stimulus space. Both models predict highly inaccurate stimulus representations on some trials, leading to heavy-tailed distributions and slow errors. The color-category effects were successfully modeled as stimulus biases in a similarity-choice framework, in which the drift rate is the vector sum of the encoded metric and categorical representations of the stimulus.

Keywords: diffusion model, decision making, continuous outcome, response time, visual working memory

"Tis clear enough the elephant is very like a tree.

—*The Blind Men and the Elephant*, John Godfrey Saxe

Continuous outcome decision tasks, in which people express their decisions about a stimulus attribute or property on a continuous scale, have become increasing important in cognitive psy-

Philip L. Smith and Saam Saber, Melbourne School of Psychological Sciences, The University of Melbourne; Elaine A. Corbett, Trinity College Institute of Neuroscience, Trinity College Dublin, and Melbourne School of Psychological Sciences, The University of Melbourne; Simon D. Lilburn, Melbourne School of Psychological Sciences, The University of Melbourne.

This research was supported by Australian Research Council Discovery Grants DP140102970 and DP180101686. A conference paper describing this work was presented at the Australian Mathematical Psychology Conference, Melbourne, Australia in February 2019. We are grateful to David Sewell for his assistance with pilot testing the eye movement decision task. We thank Gordon Logan and Marius Usher for helpful comments on an earlier version of this article. Data and code for the model can be downloaded from <https://github.com/philipls/CircularDiffusion>.

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

chology, particularly in the study of visual working memory. Historically, the continuous outcome decision task can be viewed as a development from the method of adjustment of classical psychophysics (Woodworth & Schlosberg, 1954), in which sensory thresholds were measured by asking people to adjust the intensity of a variable stimulus so that it matched a standard. Prinzmetal, Amiri, Allen, and Edwards (1998) adapted the method to study the effects of attention on perceptual variability and it was subsequently used by Wilken and Ma (2004) to study visual working memory. Since then, it has become the method of choice for many visual working memory researchers and has formed the basis of a number of key studies in the area (e.g., Adam, Vogel, & Awh, 2017; Bays, Catalao, & Husain, 2009; Oberauer & Lin, 2017; van den Berg, Awh, & Ma, 2014; Zhang & Luck, 2008). Wilken and Ma's reason for preferring the task over traditional signal detection methods was that it simplifies the decision task for participants, but for subsequent researchers its main appeal has been that it yields a measure of the precision of responding, which characterizes the variability of the decision outcomes on a continuous scale. How precision varies with the number of items stored in memory has been a dominant theme in contemporary visual working memory research for the past 15 years.

Despite the sophistication of current theories of visual working memory, until recently there has been no formal model of the

decision process that is involved in continuous outcome decision tasks. Instead, theories of visual working memory have tended to treat the precision of memory and the precision of responding as if they were the same thing, with no intervening decision process required to translate one into the other. This “direct read-out” assumption has been made both in studies investigating the effects of memory load on item precision and in those that have used continuous outcome tasks to characterize decisions about single items, either stored in memory or present in the display at the time of the response (e.g., Bae, Olkkonen, Allred, & Flombaum, 2015; Hardman, Vergauwe, & Ricker, 2017; Persaud & Hemmer, 2016). With minor differences, the methods used to characterize continuous outcome performance in these latter studies are the same as those used in studies of visual working memory and, like those studies, they have focused solely on distributions of decision outcomes. There has been nothing comparable to the sequential-sampling models of two-choice decisions that predict the joint probabilities of decision outcomes and decision times, which are widely used in other parts of cognitive psychology (Busemeyer & Townsend, 1993; Ratcliff, 1978; Smith & Ratcliff, 2009; Usher & McClelland, 2001).

The lack of a satisfactory formal model of the decision process represents a significant theoretical and practical limitation. Theoretically, the most commonly made assumption in cognitive modeling is that the variability of observed responding is jointly a function of the variability of the cognitive representation of the stimulus and the variability of the decision process that acts on it to produce a response. If this assumption is correct, then it implies that the variability of the stimulus representation can only properly be characterized when is done in conjunction with a model of the decision process. Practically, as models of visual working memory have become increasingly more complex and flexible it has become increasingly challenging to distinguish between them on the basis of the distributions of decision outcomes alone. Current theories of visual working memory tend to theorize almost exclusively about one dimension of what is in reality a two-dimensional joint distribution of decision outcomes and decision times. Influential theoretical concepts like “variable precision” (Shen & Ma, 2016, 2019; van den Berg et al., 2014), which describes across-trial variability in the memory representations of nominally equivalent stimuli, have been defined and evaluated solely with reference to the distribution of decision outcomes, with no reference to the distribution of decision times. How useful they are in characterizing the full two-dimensional joint distribution is unknown and remains to be ascertained.

In two-choice cognitive tasks, models like the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008) have been successful in predicting distributions of decision outcomes and decision times and characterizing them in terms of the quality of the information in the stimulus and the amount of evidence needed to make a response. The success of these models is dependent on them taking advantage of the additional constraints imposed by distributions of response times (RTs) to provide estimates of parameters of the underlying cognitive processes. Although there have been some applications of RT modeling methods to the study of visual working memory (e.g., Donkin, Nosofsky, Gold, & Shiffrin, 2013; Schneegans & Bays, 2016; Sewell, Lilburn, & Smith, 2016), the lack of a comparable decision model for continuous outcome tasks has hindered attempts to characterize performance on these kinds

of tasks in a similar way. Our aim in this article is to fill this gap, by providing a theoretical and empirical characterization of the processes involved in making continuous outcome decisions. Although much of the recent interest in continuous outcome tasks has been associated with the study of visual working memory, our focus here is on the prior and more fundamental question of understanding the task as a decision task. Accordingly, we restrict ourselves to decisions about single stimuli that remain present in the display until the decision is made. We have adopted a data-driven approach to theory development: First we provide an empirical characterization of the speed and accuracy of continuous outcome decisions and then we develop the theoretical machinery needed to explain their properties.

Recently, two evidence accumulation models of continuous outcome decisions have been proposed, one by Smith (2016) and one by Ratcliff (2018). These models are the circular diffusion model and the spatially continuous diffusion model, respectively. Both models represent evidence accumulation mathematically as a diffusion or a diffusion-like process, in which decisions are made by accumulating noisy evidence to a decision criterion. These models generalize the diffusion model of two-choice decisions (Ratcliff, 1978; Ratcliff & McKoon, 2008) in different ways. The two-choice diffusion model represents evidence accumulation as a one-dimensional (1D) Wiener diffusion process (i.e., diffusion on a line). Smith’s circular diffusion model assumes that the dimensionality of the evidence accumulation process is equal to the dimensionality of the stimulus space. For continuous outcome decisions about color and orientation, which are two of the most widely used decision tasks in the visual working memory literature, the evidence accumulation process is two dimensional (2D). Smith and Corbett (2019) also developed a 4D version of the model to model decisions about the contents of four-item displays in which the items each took on one of two feature values. Ratcliff’s model, which generalizes his earlier model of response confidence (Ratcliff & Sterns, 2013), assumes that the dimensionality of the evidence accumulation process is equal to the number of alternatives in the response space. For continuous outcome decisions, the process is infinite dimensional.

These two models have complementary virtues. Ratcliff’s (2018) model can accommodate different display geometries, including decisions on a circle and decisions on a line, whereas Smith’s (2016) model, which makes strong symmetry assumptions about the evidence accumulation process, applies most naturally to decisions on a circle. It applies to decisions about color, orientation, and direction of motion, but does not obviously apply to decisions on a line, such as number-line judgments (Brezis, Bronfman, & Usher, 2015), for example. The cost of the flexibility of Ratcliff’s model is that there appears to be no method known to generate analytic predictions for it, which means that it must be evaluated by simulation, which increases the challenge of fitting it to data. In contrast, the strong symmetry assumptions in Smith’s model mean that it is possible to derive explicit analytic predictions for the distributions of decision times and decision outcomes. As well as facilitating the empirical evaluation of the model, its analytic structure offers theoretical insights into the cognitive basis of precision that would be difficult to obtain in other ways.

A key tool for studying continuous-outcome tasks on circular domains is the von Mises distribution, which is a circular analogue of the normal distribution. The von Mises distribution is com-

monly used as the meeting point of theory and data in analyzing such tasks: Distributions of responses around the true stimulus value are characterized empirically using the von Mises distribution and theoretical models of visual working memory typically represent the fidelity of memory using finite or continuous mixtures of von Mises components (Bays, Wu, & Husain, 2011; Oberauer & Lin, 2017; van den Berg et al., 2014; Zhang & Luck, 2008). The von Mises distribution is a two-parameter family of the angular variable, θ , with probability density function

$$f(\theta; \kappa, \varphi) = \frac{e^{\kappa \cos(\theta - \varphi)}}{2\pi I_0(\kappa)}. \quad (1)$$

The parameter φ identifies the center of the distribution, κ specifies its precision, and the normalizing constant $I_0(\kappa)$ is a modified Bessel function of the first kind of order zero (Abramowitz & Stegun, 1965, p. 375). Loosely, precision is the reciprocal of variance; low precision represents large variance and vice versa.

Typically, precision is viewed psychologically as a fundamental or irreducible construct in visual working memory models. To the extent that its neural or cognitive basis is specified at all, precision is usually seen as an expression of the kind of tuned-detector model architecture originally outlined by Prinzmetal et al. (1998). In contrast, the circular diffusion model provides an analytic decomposition of precision into the cognitive components of the evidence accumulation process: the drift rate norm $\|\mu\|$, which characterizes the quality of the information in the stimulus; the decision criterion, a , which characterizes the evidence needed for a response; and the diffusion coefficient, σ^2 , which characterizes the noisiness of the decision process. (These three quantities are introduced more formally below.) Smith (2016, Equation 29) showed that the distribution of decision outcomes predicted by the circular diffusion model analytically has the form of a von Mises distribution in which precision admits the decomposition

$$\kappa = \frac{a\|\mu\|}{\sigma^2}. \quad (2)$$

In words, precision is equal to the quality of evidence in the stimulus times the amount of evidence needed for a response divided by the noisiness of the evidence accumulation process. Increasing the quality of the stimulus information or the amount of evidence used to make a response increases precision, while increasing the noisiness of the decision process reduces it. Precision can thus be viewed as an expression of the same components of processing that have been shown to characterize the speed and accuracy of two-choice decisions in a variety of cognitive settings (Ratcliff, Smith, & McKoon, 2015). Indeed, as pointed out by Smith (2016), Equation 2 is a continuous counterpart of the sensitivity index derived by Link (1975) for a two-choice random walk model. The close analytic relationship between the circular diffusion model and earlier models of two-choice decisions is one of its attractive properties.

Our aim in this article is to develop the circular diffusion model to characterize the processes involved in decisions about the hues of color patches, which, as we noted earlier, is one of the most widely used tasks in the visual working memory literature. We focus on the ability of the model to provide a detailed account of the joint distributions of decision times and decision outcomes as a function of stimulus discriminability. To account for the speed and accuracy of continuous hue judg-

ments, we develop the theory presented in Smith (2016) in two main ways. First, we develop and compare two alternative models of across-trial variability in the evidence entering the decision process. One model is based on encoding failures and the other is based on a nonlinear transformation of the stimulus space. The effect of transformation is to increase the proportion of very low precision stimulus representations beyond what would otherwise be found. Second, we develop a diffusion model of the categorical effects of nameable colors on decision times and decision outcomes. Early articles using hue judgments to investigate visual working memory treated the color space as isotropic and considered only the distributions of decision outcomes aggregated across stimulus types. More recently, a number of studies have highlighted the fact that nameable color categories can have large effects on the accuracy with which particular stimuli are reported. Color category effects have been shown in long-term memory (Persaud & Hemmer, 2016), visual working memory (Bae et al., 2015; Hardman et al., 2017), and in immediate decisions with no delay (Bae et al., 2015). These effects are found in both the speed and accuracy of decisions (Ratcliff, 2018). We present a cognitive model of stimulus bias based on similarity-choice theory (Navarro, 2007; Nosofsky, 1984), which assumes that the evidence entering the decision process reflects a combination of the metric and the categorical properties of the stimulus.

Because of the complexity of our model evaluation study and the number of different model variants we considered, we have not attempted to carry out a comparison with Ratcliff's (2018) model, although such a comparison would at some point be useful. Apart from the complexity of the exercise, the lack of analytic predictions for Ratcliff's model means its predictions could not be brought into direct alignment with those of the circular diffusion model in any straightforward way.

Eye Movement Decision Task

The task we used to evaluate the model was an eye-movement decision task in which participants made decisions about the dominant hues of color patches that had been perturbed by chromatic noise and expressed their decisions by moving their eyes from a central fixation point to a point on a surrounding color wheel or annulus. Because our task involved decisions about a single, centrally presented stimulus, the addition of chromatic noise provided a convenient way of manipulating task difficulty. RT was defined as the time to break fixation and the decision outcome was defined as the landing point of the first corrective eye movement after the initial saccade. Figure 1 shows a representation of the stimulus display used in the task. The task is a version of a task reported by Ratcliff (2018) in his article on the spatially continuous diffusion model. The motivation for using an eye movement rather than a hand movement response was to try to minimize possible anisotropies in the time to plan and execute movements directed toward different points on the color wheel—although with hindsight these concerns were probably unwarranted. Ratcliff (2018) reported comparisons of continuous-outcome decision tasks using eye-movement, mouse movement, and touch-screen responses and found no systematic differences between them, and Kvam (2019) obtained well-behaved data from a similar

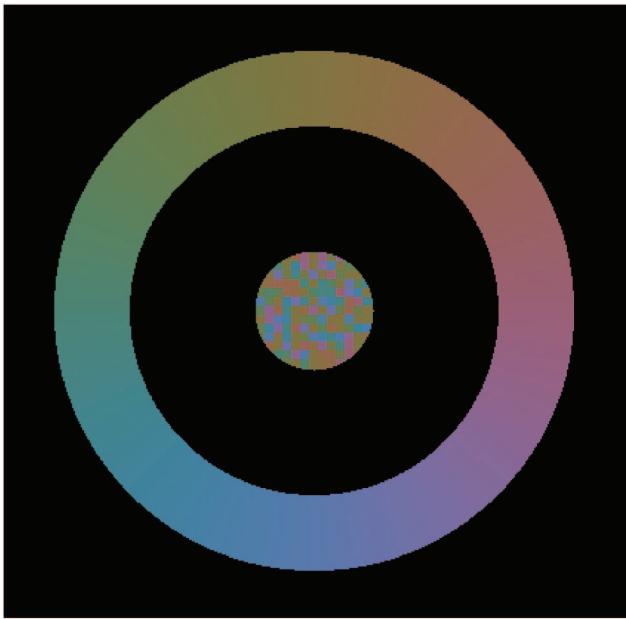


Figure 1. Eye-movement decision task. Participants made decisions about the dominant hues of color patches that were perturbed with chromatic noise and made saccadic eye movements from a central location to points on a surrounding color wheel. Stimuli were defined in an isoluminant, equidiscriminability color space (CIELUV space). The dominant color in this example was teal (green-blue).

task using mouse movements. It therefore seems unlikely that RTs measured in this way are materially affected by differences in response selection times prior to response initiation. Regardless of whether there is substance in these concerns, the eye-movement decision task is a well-established paradigm that is widely used in neuroscience to study decision making in awake behaving monkeys (Smith & Ratcliff, 2004). Our task therefore links directly to standard methods in the primate decision-making literature.

We carried out our experiment using a **small-*N* psychophysical design**, in which data were collected over multiple sessions from four practiced participants. Each participant provided enough data to provide stable estimates of the joint distributions of decision outcomes and RTs at each of three levels of stimulus discriminability defined by three levels of added noise. This allowed us to fit the model on a participant by participant basis. As argued by Smith and Little (2018) and Little and Smith (2018), in small-*N* designs the individual participant becomes the replication unit and the study as a whole is effectively self-replicating. Faced with a putative “replication crisis” in psychology, many authors have advocated the use of larger samples of participants. Smith and Little instead advocated the use of small-*N* designs, individually tailored experimental manipulations, and the use of strong measurement methods and detailed mathematical models that provide rich contact with experimental data at the individual participant level. A similar argument was made by Miller and Schwarz (2018). Such methods are of course widely used in both human psychophysics and animal neuroscience, but they are less common in other parts of psychology.

Method

Stimuli and Apparatus

The experiment was run on a pair of linked computers, one controlling a Cambridge Research Systems ViSaGe framestore that presented the stimuli with 12-bit color resolution and the other controlling an Eyelink II eye movement monitor with head-mounted infrared sensors that recorded responses. The stimuli were presented on a gamma-corrected 21" Sony Trinitron Multi-scan CRT monitor with refresh rate set at 100 Hz. Stimulus presentation and response recording were carried out by custom C++ software. An example of the stimulus display is shown in Figure 1. The stimuli were specified in CIELUV space, which is one of two isoluminant, equidiscriminability color spaces defined by the CIE in 1976 (Wyszecki & Stiles, 1982; pp. 164–169), the other being CIELAB space. Stimuli defined in these spaces have the properties of being isoluminant and, if they are separated by equal distances, equally discriminable. The two spaces are similar but CIELUV space is thought to be superior for luminous devices like monitors, whereas CIELAB space is thought to be superior for reflective devices like offset printers, although the difference is unlikely to have been consequential for our results. We chose to use an isoluminant, equidiscriminability color space to try to attenuate the effects of nameable colors reported by previous authors. As can be seen from Figure 1, the resulting color wheel is more homogeneous and less perceptually banded than color wheels defined in the more commonly used XYZ space. Despite these precautions, our data show evidence of strong, but idiosyncratic, categorical responding, and one of the challenges we faced was how to deal with these effects in a theoretically satisfying way.

Participants viewed the stimulus display from a distance of 100 cm. The stimuli were presented in a central, circular region subtending 2.0° of visual angle. Each stimulus was composed of a dominant background color, specified by one of 253 pen colors on the color circle, some fraction of whose pixels were replaced with randomly chosen chromatic noise pixels. On each trial, the participant was required to judge the dominant hue of the stimulus and to make a saccadic eye movement to the corresponding point on the color wheel. The difficulty of the task could be systematically varied by varying the proportion of noise pixels in the display. The color wheel was an annulus of width 1.5° whose inner circle subtended an angle of 6.1°. Unlike many experiments in the visual working memory literature, we did not rotate the color wheel on each trial. Although this is often seen as a useful way to guard against stereotypical responding, it introduces uncertainty into the stimulus-response mapping that can have a destabilizing effect on performance in RT tasks and is usually avoided in studies of decision processes.

Participants

The participants were four young adults from the University of Melbourne population with normal visual acuity and normal color vision who served in multiple experimental sessions, each of around 50-mins duration, who were paid \$A15/session for their participation. They are referred to in this article as S_1 through S_4 . There are pronounced individual differences in the quality of the eye-movement recordings that can be obtained with the Eyelink II,

so we verified we could obtain reliable eye movements from each of the participants at the time of recruitment before proceeding with the experimental sessions. Our initial experimental design called for each of the participants to serve in 12 experimental sessions following several sessions of familiarization and practice, providing 768 trials of data at each of three levels of stimulus discriminability. S_1 completed only nine of the scheduled 12 sessions, yielding 557/558 trials per condition. S_4 completed the full 12 sessions, but showed poor stability during the first three sessions, so we have analyzed only the last nine sessions for this participant, which yielded 576 trials per condition. S_2 and S_3 completed the full 12 sessions, each yielding the planned 768 trials per condition.

Procedure

The experimental sessions were run individually in a dimly lit booth. The Eyelink II provides automatic calibration and drift-correction procedures and we verified during pilot testing that it reliably recorded the end points of eye movements made to targets on the circumference of a circle of the same diameter as the color wheel. The calibration procedure required participants to fixate on each of nine fiducial markers on the screen in turn and was carried out at the start of the session and was repeated at the start of each 48-trial block. The participant's gaze position was shown on an external monitor outside the experimental booth where it could be assessed for calibration and drift artifacts by the experimenter.

Participants were told that, at the beginning of each trial, they should fixate a central cross on the display and should maintain their fixation until they had made a decision. Once they had achieved fixation, they were to press a button to initiate the trial. The experimental program verified they had achieved fixation before the trial began. Once fixation had been achieved, the stimulus appeared and remained present until they moved their gaze outside a 2.0° region surrounding the fixation point, at which point it was extinguished. We adopted the procedure of extinguishing the stimulus as soon as fixation was broken to discourage the strategy of multiple looks at the stimulus and the color wheel. During pilot testing, in which participants were required to make eye movements to designated points on a target circle, we found that on the majority of trials they made an initial ballistic eye movement that finished on or near the target circle, followed by a small corrective adjustment that brought them to the target. We therefore measured the RT as the onset of the initial ballistic movement and the final decision as the end point of the first corrective eye movement.

Participants performed from three to five familiarization and calibration sessions prior to the collection of eye movement data, during which the experimenter set the levels of noise for the three conditions for each participant individually. In setting the levels of noise we did not attempt to calibrate participants to any prespecified level of accuracy, but simply attempted to ensure that the three levels of noise resulted in distributions of decision outcomes with different precisions. The experiment we report here was a replication of an earlier one in which the obtained precision varied only slightly across the three levels of noise. We thought the data from that experiment were insufficiently constraining to test the model, so we replicated it with different levels of noise to try to increase the differences among conditions.

Theory development and modeling of data. We present the results of the experiment in a later section after first summarizing the circular diffusion model. We report the analysis in a layered way, which seeks to characterize the data at progressively finer levels of granularity. In keeping with our data-driven approach to theory development, we first consider the ability of the model in its current, published form to account for the marginal and joint distributions of decision outcomes and decision times, and then develop new theory to correct for the limitations of the existing model. The first model we consider is the model of Smith (2016), augmented with the results of Smith and Corbett (2019), who derived analytic expressions for a continuous-outcome model in which the evidence entering the decision process is normally distributed across trials. The assumption that stimulus information across trials is normally distributed is a natural one to make theoretically and links the circular diffusion model to the 1D diffusion model of two-choice decisions (Ratcliff, 1978; Ratcliff & McKoon, 2008) and, more generally, to Case V of Thurstone's law of comparative judgment (Thurstone, 1927) and signal detection theory (Green & Swets, 1966). We find, however, that a model with normally distributed stimulus information is unable to capture the fine-grained structure of our distribution data. We then consider two alternative models of across-trial variability: One model is based on encoding failures and the other is based on nonlinear transformation of the stimulus space. Both of these models successfully account for the joint and marginal distributions of decision outcomes and decision times. We then consider color category effects, in which we analyze the data conditional on the location of the stimulus in color space and present a model for categorical effects that characterizes the speed and accuracy of decisions as a function of stimulus location.

The circular diffusion model. The main properties of the circular diffusion model are shown in Figure 2. The model represents evidence accumulation by a Wiener diffusion process, $\mathbf{X}_t = (X_t^1, X_t^2)'$, whose growth is characterized by a vector-valued stochastic differential equation,

$$d\mathbf{X}_t = \boldsymbol{\mu} dt + \boldsymbol{\sigma} d\mathbf{W}_t, \quad (3)$$

or in components,

$$\begin{bmatrix} dX_t^1 \\ dX_t^2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} dt + \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dW_t^1 \\ dW_t^2 \end{bmatrix}, \quad (4)$$

where $d\mathbf{W}_t = (dW_t^1, dW_t^2)'$ is the differential of a two-dimensional Brownian motion, or Wiener, process and where the prime denotes the matrix transpose. (In these expressions the superscripts index the coordinates of the process; they are not powers.) The quantity $d\mathbf{W}_t$ describes the horizontal and vertical components of the random change in the process \mathbf{X}_t during a small interval of length dt . The quality of the information in the stimulus is represented by the drift rate, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, while the rate at which the process diffuses toward the decision boundary is determined by the infinitesimal standard deviation, $\boldsymbol{\sigma}$, which is assumed to be the same in the horizontal and vertical directions. The square of the infinitesimal standard deviation, σ^2 , is the diffusion coefficient.

Psychologically, the model has very similar properties to the 1D diffusion model of two-choice decisions of Ratcliff and colleagues (Ratcliff, 1978; Ratcliff & McKoon, 2008). On stimulus presentation, noisy evidence begins to accumulate,

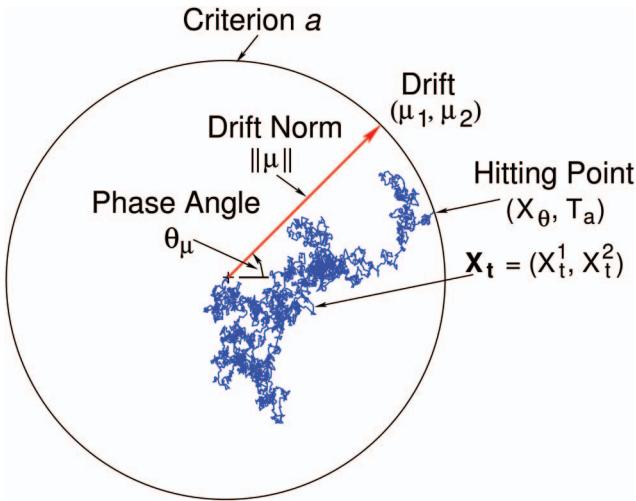


Figure 2. Circular diffusion model of continuous-outcome decisions. Evidence accumulation is modeled as a two-dimensional Wiener diffusion process on the interior of a disk, whose bounding circle, of radius a , represents the decision criterion. The process $X_t = (X_t^1, X_t^2)'$ consists of independent components, with drift rate $\mu = (\mu_1, \mu_2)'$ and infinitesimal standard deviation σ . In polar coordinates, the norm of the drift rate is $\|\mu\|$ and the phase angle is θ_μ . A decision is made when the accumulating evidence reaches the criterion. The hitting point, X_θ , represents the decision outcome and the hitting time, T_a , represents the decision time. See the online article for the color version of this figure.

starting at the origin, $X_0 = \mathbf{0}$, at a rate and in a direction determined by the drift rate vector μ . The irregular sample path in Figure 2 represents the accumulating evidence on a single experimental trial. Because of the circular symmetry of the model, it is convenient to represent the drift rate in polar coordinates. In polar coordinates, the drift rate has a length or norm, $\|\mu\| = \sqrt{\mu_1^2 + \mu_2^2}$, and a phase angle, or direction, $\theta_\mu = \arctan(\mu_2/\mu_1)$. Psychologically, the drift norm represents the quality of the information in the stimulus while the phase angle represents its identity. Evidence accumulation continues until the process hits a point on the bounding circle. The hitting point is denoted X_θ and the hitting time is denoted T_a , where the capitalization in the notation indicates that the hitting point and hitting time are both random variables. In applications of the model to continuous outcome decision tasks, the hitting point corresponds to the participant's report of the stimulus identity and the hitting time corresponds to the decision time.

In fitting the model to experimental data we make the usual assumption that RT is the sum of two independent components, a decision time, T_a , and a nondecision time, T_{er} . The latter represents the sum of all of the nondecision portions of RT. The notation T_{er} is due to Ratcliff (1978) and stands for “time for encoding and responding”—where “responding” is interpreted here as referring to premotor components of RT prior to the saccade. The nondecision time is often treated, on pragmatic grounds, as uniformly distributed with range s_r , because it appears to contribute relatively little variability to the overall RT (although see Verdonck & Tuerlinckx, 2016, for a contrary view). We found that for the majority of our participants nondecision time could be character-

ized as a simple additive constant and for only one of them (S_4) was the model fit improved by making T_{er} variable.

Joint distributions of decision outcomes and decision times.

The main theoretical quantities of interest in the circular diffusion model are the joint distributions of decision times and decision outcomes. Smith (2016), following Kent (1978), showed that the joint distribution can be obtained by applying the Girsanov change-of-measure theorem (Karatzas & Shreve, 1991, pp. 190–193; Rogers & Williams, 2000, pp. 81–85) to the first-passage time distribution of the 2D Bessel process. The Bessel process, R_t , describes the Euclidean distance, $\|W_t\| = \sqrt{(W_t^1)^2 + (W_t^2)^2}$, of a 2D Wiener diffusion process, W_t , with zero drift rate from its starting point. When $W_0 = \mathbf{0}$, all of the relevant information about the zero-drift process is carried by the Euclidean distance process. We denote by $dP_t(a)$ the probability density function of the Bessel process through the decision boundary a at time t . Symbolically,

$$dP_t(a) = \frac{d}{dt} P[R_t \leq a | R_s < a, 0 \leq s < t].$$

The first-passage time density function has an infinite series representation of the form

$$dP_t(a) = \frac{\sigma^2}{a^2} \sum_{k=1}^{\infty} \frac{j_{0,k}}{J_1(j_{0,k})} \exp\left(-\frac{j_{0,k}^2 \sigma^2}{2a^2} t\right), \quad (5)$$

(Borodin & Salminen, 1996, p. 297; Hamana & Matsumoto, 2013). 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 of the first kind, $J_0(x)$, (Abramowitz & Stegun, 1965, p. 360). The term $J_1(j_{0,k})$ in the denominator represents the values of the function $J_1(x)$ evaluated at the zeros of $J_0(x)$, that is, at the points where the function crosses the x -axis. An example of the density $dP_t(a)$ is shown in Figure 3; plots of the functions $J_0(x)$ and $J_1(x)$ are shown in Smith (2016, Figure 3).

We denote by $d\tilde{P}_t(\theta_T)$ the probability density that the nonzero-drift diffusion process hits the bounding circle at a point X_T with phase angle θ_T at time T_a . The Girsanov theorem states that the probability density functions $d\tilde{P}_t(\theta_T)$ and $dP_t(a)$ are related in a simple way, via a function $Z_T(X)$,

$$d\tilde{P}_t(\theta_T) = Z_T(X) dP_t(a), \quad (6)$$

where

$$Z_T(X) = \exp\left[\frac{1}{\sigma^2}(\mu \cdot X_T) - \frac{1}{2\sigma^2}\|\mu\|^2 T\right], \quad (7)$$

and where $(\mu \cdot X_T) = \sum_{i=1}^2 \mu_i X_T^i$ is the dot product of the drift rate vector and a random vector X_T with norm a and phase angle θ_T . In polar coordinates, the components of X_T map out a locus of random hitting points on the bounding circle with $X_T^1 = a \cos \theta_T$ and $X_T^2 = a \sin \theta_T$, so $Z_T(X)$ can be represented in polar form as

$$Z_T(X) = \exp\left[\frac{1}{\sigma^2}(a\mu_1 \cos \theta_T + a\mu_2 \sin \theta_T) - \frac{1}{2\sigma^2}\|\mu\|^2 T\right]. \quad (8)$$

Equations 5 and 8 used in Equation 6 provide explicit expressions for the distributions of decision times and decision outcomes

which can be used to fit the model to data.¹ Figure 3 shows an example of the Girsanov transformation of the first-passage time density of the Bessel process in Equation 5.

The function $Z_T(X) = d\tilde{P}_t(\theta_T)/dP_t(a)$ in Equation 6 is known as the Radon-Nikodym derivative of the probability measure \tilde{P}_t with respect to the measure P_t , and has an illuminating statistical interpretation as a likelihood ratio process. Specifically, it defines the likelihood of obtaining the observed evidence process X_T , given a drift rate of μ , versus the likelihood of obtaining the same process given a drift rate of $\mathbf{0}$. The value of $Z_T(X)$ is maximized when the dot product $(\mu \cdot X_T)$ is maximized, which will occur when the phase angles of μ and X_T coincide and the cosine of the angle between them is unity. This means that the most likely place for the process to hit the bounding circle is at the point corresponding to the phase angle of the drift vector. An observer who reports the hitting point as the decision about the identity of the stimulus is a maximum likelihood observer.

The top two panels of Figure 4 show predicted distributions of decision outcomes and decision times for the model. Figure 4a shows the predicted distribution of decision outcomes for a constant drift rate norm, $\|\mu\|$, for three different phase angles of the stimulus, θ_μ , and Figure 4b shows several joint distributions of decision times and decision outcomes for the $\theta_\mu = 0$ case. The decision outcomes follow a von Mises distribution centered on the stimulus, with precision κ that depends jointly on the decision criterion, the drift norm, and the diffusion coefficient, as specified

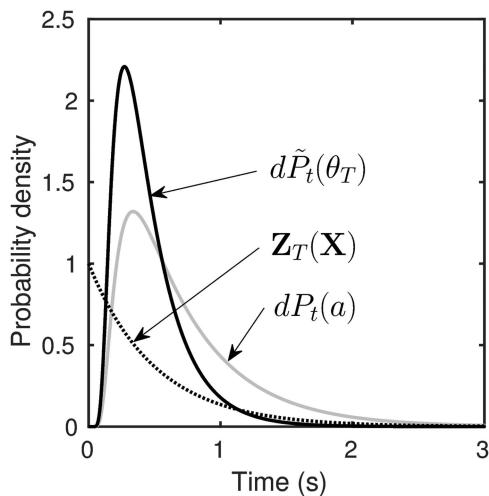


Figure 3. Response time distribution predictions obtained by Girsanov transformation of the Bessel process. The first-passage time density, $d\tilde{P}_t(\theta_T)$, for a process with a nonzero drift rate, μ (continuous black line), is obtained by multiplying $dP_t(a)$, the first-passage time density of the Bessel process of Equation 5 (continuous gray line), by the function $Z_T(X)$ (dotted black line). Only the time-dependent part of $Z_T(X)$ is shown. The product of functions $Z_T(X)dP_t(a)$ represents a slice through the joint density of decision times and decision outcomes, and so is not a proper probability density, but its mass has been normalized to unity to facilitate comparison with the Bessel process. The joint densities for different decision outcomes shown in Figure 4 are all scaled copies of this product of functions, with a scale factor equal to $\exp[(\mu \cdot X_T)/\sigma^2]$, the phase-angle dependent part of $Z_T(X)$. The predictions are for a model with $a = 1.0$, $\sigma = 1.0$, and $\|\mu\| = 2.0$.

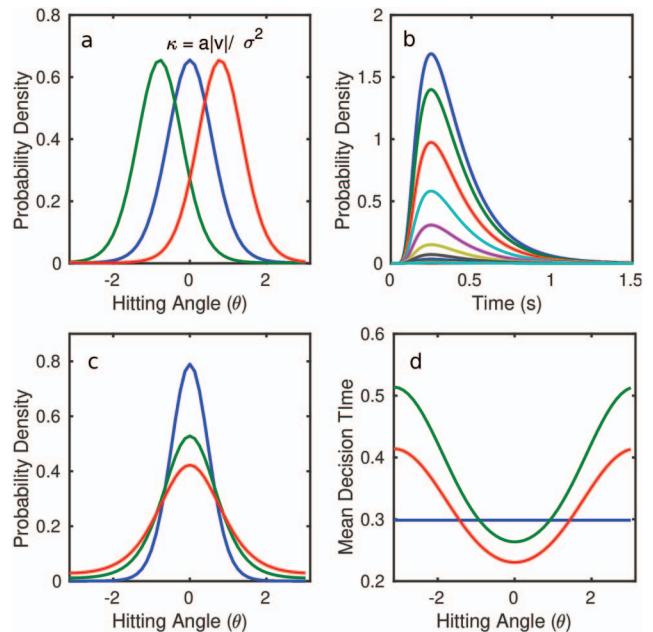


Figure 4. Predicted distributions of decision outcomes and decision times for the circular diffusion model. (a) Predicted decision outcomes for stimuli with drift norm $\|\mu\| = 2.0$ and phase angles (left to right) of $\theta_\mu = (-\pi/4, 0, \pi/4)$ radians. The other model parameters were $a = 1.2$ and $\sigma^2 = 1.0$. The decision outcomes follow a von Mises distribution with precision $\kappa = a\|\mu\|/\sigma^2$. (b) Distributions of decision times for selected response angles for the stimuli in Panel a. (c) Distributions of decision outcomes with normally distributed across-trial variability in drift rates. The mean drift rate was $\nu = (2.5, 2.5)$ and the standard deviations, $\eta = (\eta_1, \eta_2)$; most to least peaked), were 0.0, 2.0 and 3.0. The other parameters were as in Panels a and b. (d) Mean decision times with across-trial variability in drift rates for the model parameters in Panel c. When there is no across-trial variability in drift rates mean decision times are constant (≈ 0.3 s). When there is across-trial variability in drift rates responses in the tails of the distribution of decision outcomes are slowed. See the online article for the color version of this figure.

in Equation 2. The distributions of decision times are obtained by applying the function $Z_T(X)$ in Equation 8 to the distribution of first-passage times of the Bessel process in Equation 5. Because Equation 8 can be factored into a product of terms, one dependent on response angle and independent of time and the other dependent on time and independent of response angle, it follows that the joint distributions of decision times are scaled copies of one another. The conditional distributions of decision times, obtained by dividing the joint distributions by the associated response probabilities will be identical and the mean decision times will be constant for all decision outcomes, as shown in Figure 4d.

Across-trial variability in drift rates. In applications of the model to data, hypotheses about the stimulus representations that

¹ Time is capitalized as T in Equations 6 through 8 to indicate that decision time is a random variable. The fact that Equation 6 holds at both fixed and random times is what allows us to substitute points on the boundary $(\cos\theta, \sin\theta)$ for X_T to obtain the explicit representation Equation 8. When $Z_T(X)$ is applied to $dP_t(a)$ in (7), T takes on specific values $T = t$ and the time-dependent part of Equation 6 becomes $\exp[-\|\mu\|^2 t/(2\sigma^2)]dP_t(a)$.

drive the decision process can be expressed as assumptions about the distribution of drift rates across trials. As shown in Figures 4c and 4d, when there is across-trial variability in drift rates, decision times and decision outcomes will be negatively correlated: The least accurate responses tend also to be the slowest. Increasing the drift rate standard deviation decreases the precision of the distribution of decision outcomes, increasing the number of inaccurate responses and the number of slow responses. This property is a continuous counterpart of the slow errors predicted by the standard form of the 1D diffusion model with drift rate variability (Ratcliff & McKoon, 2008). Across-trial variability in drift rate can also be viewed as a diffusion model counterpart of the concept of variable precision, which van den Berg et al. (2014) proposed to explain the distribution of decision outcomes in their model of visual working memory.

Studies of visual working memory using continuous-report decision tasks often yield peaked, heavy-tailed distributions of decision outcomes (van den Berg et al., 2014). Such distributions have a bell-shaped central region flanked by flat or near-flat tails. Contrasting accounts of these distributions have been offered by different theorists. Zhang and Luck (2008) modeled them as two-component mixtures of a von Mises distribution and a uniform distribution, which they interpreted as evidence of the joint action of a memory retrieval process and a guessing process. van den Berg et al. (2014) subsequently showed they were better fit by a variable precision model, consisting of a continuous mixture of von Mises components with differing precisions. They interpreted the variability of the components as evidence of trial-to-trial variability in the fidelity with which items are represented in memory, and argued that the heavy tails of the distributions could be attributed to a combination of variable precision item representations and variability in the number of items entering memory. Other theorists, such as Bays et al. (2011) and Oberauer and Lin (2017), have proposed models consisting of finite mixtures of von Mises components that are intermediate between the two-component mixture model of Zhang and Luck and the continuous mixture model of van den Berg et al.

In our study, we also found peaked, heavy-tailed distributions of decision outcomes when there were high levels of chromatic noise added to the stimuli, which we attributed to encoding failures, as we discuss subsequently. A significant theoretical challenge for us was how to account for the joint distributions of decision times and decision outcomes obtained under these conditions in a consistent way. We sought to do so via an appropriate model of across-trial variability in drift rates. This model can be viewed, on the one hand, as a continuous counterpart of the across-trial variability in drift rates in the diffusion model of two-choice decision making (Ratcliff & McKoon, 2008) and, on the other hand, as a diffusion model counterpart of the variable precision model of van den Berg et al. (2014). We considered a number of different ways of representing across-trial variability in drift rates in the model, with contrasting theoretical properties. We describe these in the following sections.

Bivariate-normally distributed drift rates. The assumption that links most directly with the 1D diffusion model of two-choice decisions is that drift rates across trials follow a bivariate normal distribution. Equation 6 shows that the effect of variability in drift rate in the circular diffusion model is completely characterized by its effect on the function $Z_T(X)$. Consequently, the joint density of

decision times and decision outcomes in the presence of drift rate variability can be obtained by integrating, or marginalizing, $Z_T(X)$ across the distribution of drift rates. When the components of drift rate variability are independent, with mean $\nu = (\nu_1, \nu_2)'$ and standard deviation $\eta = (\eta_1, \eta_2)'$, the resulting function has the form

$$\bar{Z}_T(X) = \prod_{i=1}^2 \frac{1}{\sqrt{(\eta_i/\sigma)^2 T + 1}} \exp \left\{ -\frac{\nu_i^2}{2\eta_i^2} + \frac{[X_T^i (\eta_i/\sigma)^2 + \nu_i]^2}{2\eta_i^2[(\eta_i/\sigma)^2 T + 1]} \right\}, \quad (9)$$

(See Appendix A in Smith & Corbett, 2019; see also Tuerlinckx, 2004), where the overbar notation implies averaging or marginalization. When the components are correlated, $\bar{Z}_T(X)$ has a more complex form (Smith, 2019) that can be linked theoretically with the general recognition theory of Ashby and Townsend (1986). Correlation induces small left-right asymmetries in the distributions of decision outcomes and large asymmetries in the distributions of decision times. We saw little evidence of these kinds of effects in our data, so we do not consider models with correlated drift rates further. The predictions for the model with drift rate variability in Figures 4c and 4d were obtained with $\eta_1 = \eta_2$ in Equation 9, making the effects of drift rate variability the same for all stimuli (i.e., the model is isotropic).

When the mean drift rate vector, ν , aligns with one of the cardinal axes of the (X^1, X^2) space, then the two components of the drift-rate distribution can be interpreted as the radial and tangential components of drift rate, respectively. The radial component describes the variability along the mean drift rate vector and the tangential component describes the variability at right angles to it. Psychologically, the radial component describes variability in the quality of the encoded stimulus information and the tangential component describes variability in its identity. Variability in the radial component affects both the distribution of decision times and the distribution of decision outcomes while variability in the tangential component mainly affects the distribution of decision outcomes. Radial variability leads the model to predict slow errors and reduces the precision of responding while tangential variability leads mainly to a reduction in precision.²

Fitting the model to data. We fit the model to data by maximum likelihood, where the likelihood, $L(\theta, T)$, associated with decision outcome θ and response time T is $L(\theta, T) = d\tilde{P}_t(\theta_T)$, as given by Equation 6 for specified values of the model parameters. To do so, we minimized minus the sum of the log-likelihoods ($-LL$) where,

$$LL = \sum_i \sum_j \log L(\theta_{ij}, T_{ij}),$$

for each participant individually using the Nelder-Mead simplex algorithm (Nelder & Mead, 1965) in Matlab (fminsearch). In this

² If the mean drift rate vector does not align with a cardinal axis then the components of η no longer describe the radial and tangential components of drift rate variability. It is possible to write an expression for $\bar{Z}_T(X)$ for arbitrary phase angles of the mean drift rate vector, θ_ν (Smith, 2019), but the predictions for a model with independent radial and tangential components of drift rate variability can equally well be obtained by circularly shifting the predictions for the $\nu = (\|\nu\|, 0)$ case, in which the phase angle of the mean drift rate vector is zero, which was the way we implemented the model in our fitting routines.

expression θ_{ij} and T_{ij} are the decision outcome and RT for the j th trial in the i th condition, respectively. A problem with maximum likelihood fitting of RT data, as discussed by Ratcliff and Tuerlinckx (2002), is the presence of outlier observations to which the model assigns zero likelihood, which can create convergence problems and distort the fit. We dealt with this problem in an ad hoc but effective way, by assigning a value of $L = 0.05$ to any observation to which the model assigned zero likelihood. The effect of this procedure was to downweight any observation that fell too far from the rest of the data. This stabilized the fitting procedure and eliminated the convergence problems to which maximum likelihood is otherwise prone.

To compare the fits of alternative models, we used versions of the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978), adjusted for statistical overdispersion. The AIC and the BIC penalize the log-likelihoods of the fitted models for their numbers of free parameters; the BIC does so in a sample-size dependent way that makes it less prone than the AIC to favor complex models with large sample sizes. This endows the BIC with an inherent parsimony, but it can sometimes fail to capture apparently real features of empirical data (Smith & Corbett, 2019) and it has been shown to perform more poorly than the AIC in simulated model recovery studies (Oberauer & Lin, 2017). We use the BIC and AIC in the same spirit as they are used in other studies, as more or less strict criteria, respectively, for comparing models of differing complexity (Voss, Lerche, Mertens, & Voss, 2019).

The versions of the AIC and the BIC we used are called the QAIC (Burnham & Anderson, 2002) and the QBIC (Hastie & Tibshirani, 1990), respectively. Behavioral data are often more variable than predicted by the mathematical models used to describe them because of factors like learning, fatigue, inattention, and across-trial autocorrelation in performance that are not explicitly represented in the model, and which introduce additional, extraneous sources of variability across experimental blocks and sessions. This unmodeled variability is termed *overdispersion*. If there is overdispersion in the data, then the penalty terms used in model selection will be incorrect, because the fit statistics will not be distributed as true log-likelihoods. A simple, if approximate, way of dealing with this problem is to adjust the likelihood by a factor q , that quantifies the overdispersion. The QAIC and QBIC are defined as

$$\text{QAIC} = -2LL_{\max}/q + 2m \quad (10)$$

$$\text{QBIC} = -2LL_{\max}/q + m\log N, \quad (11)$$

respectively, where LL_{\max} is the maximized log-likelihood of the fitted model, m is the number of free parameters, and N is the total sample size. We used a chi-square method described by Cox and Snell (1989) and McCullagh and Nelder (1989, pp. 174–175) to estimate q . In the method, the N observations are grouped into an $a \times b$ contingency table and the bin bounds and the proportion of observations in each bin are used to calculate expected frequencies in a chi-square test. The data are then partitioned into c subsets on the basis of a relevant experimental grouping variable like sessions or blocks, and the observed frequencies in each subset that fall into each bin are tallied. If there is no overdispersion across subsets, then each subset should be an independent random sample from a joint distribution with expected frequencies approximately equal to

those computed from the marginal table. A chi-square statistic computed on the ab bins of each subset and summed across c subsets should be distributed as a chi-square random variable with $df = c(a - 1)(b - 1)$ degrees of freedom. To the extent that the chi-square is greater than df , which is its expected value under the null hypothesis, overdispersion may be inferred. Formally, the overdispersion factor is defined as χ^2/df , where $\chi^2[c(a - 1)(b - 1)]$ is the summed Pearson chi-square across subsets.

In our application of the method, we formed a contingency table on the joint distributions of decision outcomes (a) and decision times (b) and used sessions (c) as the partition variable, so $c = 9$ or $c = 12$, depending on the participant. In order to satisfy the continuity requirement of the chi-square test that no more than 20% of expected cell frequencies should be less than five we used $a = 5$ and $b = 6$ to represent the joint distribution. For q calculated in this way, we obtained values of 1.38, 4.01, 1.75, and 1.81 for participants S_1 through S_4 , respectively. These values of overdispersion are broadly consistent with typical estimates of overdispersion in the data from decision tasks reported by Ratcliff and Childers (2015), Smith (1998), and Smith and Corbett (2019).

Results

Figure 5 shows the fit of a version of the circular diffusion model with bivariate-normally distributed drift rates to the marginal distributions of decision outcomes and RTs for the three levels of chromatic noise to each of the four participants. In fitting the model we assumed that the parameters that characterized the mean and across-trial variability in drift rate varied as a function of the stimulus condition, whereas the decision criterion, a , and the nondecision time, T_{er} , were the same for all conditions. For each participant, the panels on the left (the symmetrical distributions) are the distributions of decision outcomes, expressed as a function of the angular error of the response, and the panels on the right (the skewed distributions) are the distributions of RT. Overall—with some important qualifications we have yet to make—the model fits well, capturing the shapes of both the distributions of decision outcomes and RT as a function of the stimulus. (The qualifications, as we discuss subsequently, are that the model shown in Figure 5 includes an encoding-failure process and a color-category process. The former allows it to account for the heavy-tailed distributions of decision outcomes at high noise levels; the latter allows it to account for the effects of specific stimulus colors.) There is a slight tendency to underestimate the peaks of the decision outcome distribution at high levels of stimulus discriminability (low levels of noise), but, otherwise the model is capturing the main features of the data in a fairly satisfying way. In evaluating the fits in Figure 5 and comparing them to fits of other models of the continuous report task in the literature it is important to bear in mind that most other models only characterize the distributions of decision outcomes on the left; they are silent on the distributions of RT on the right. In requiring our model to account for both decision outcomes and decision times, we are setting it an appreciably more complex challenge than is usually faced by models of continuous outcome decisions.

As well as requiring the model to account for the marginal distributions of decision outcomes and RT we also require it to account for their joint distributions. The latter characterize how accuracy and RT covary across the space of decision outcomes.

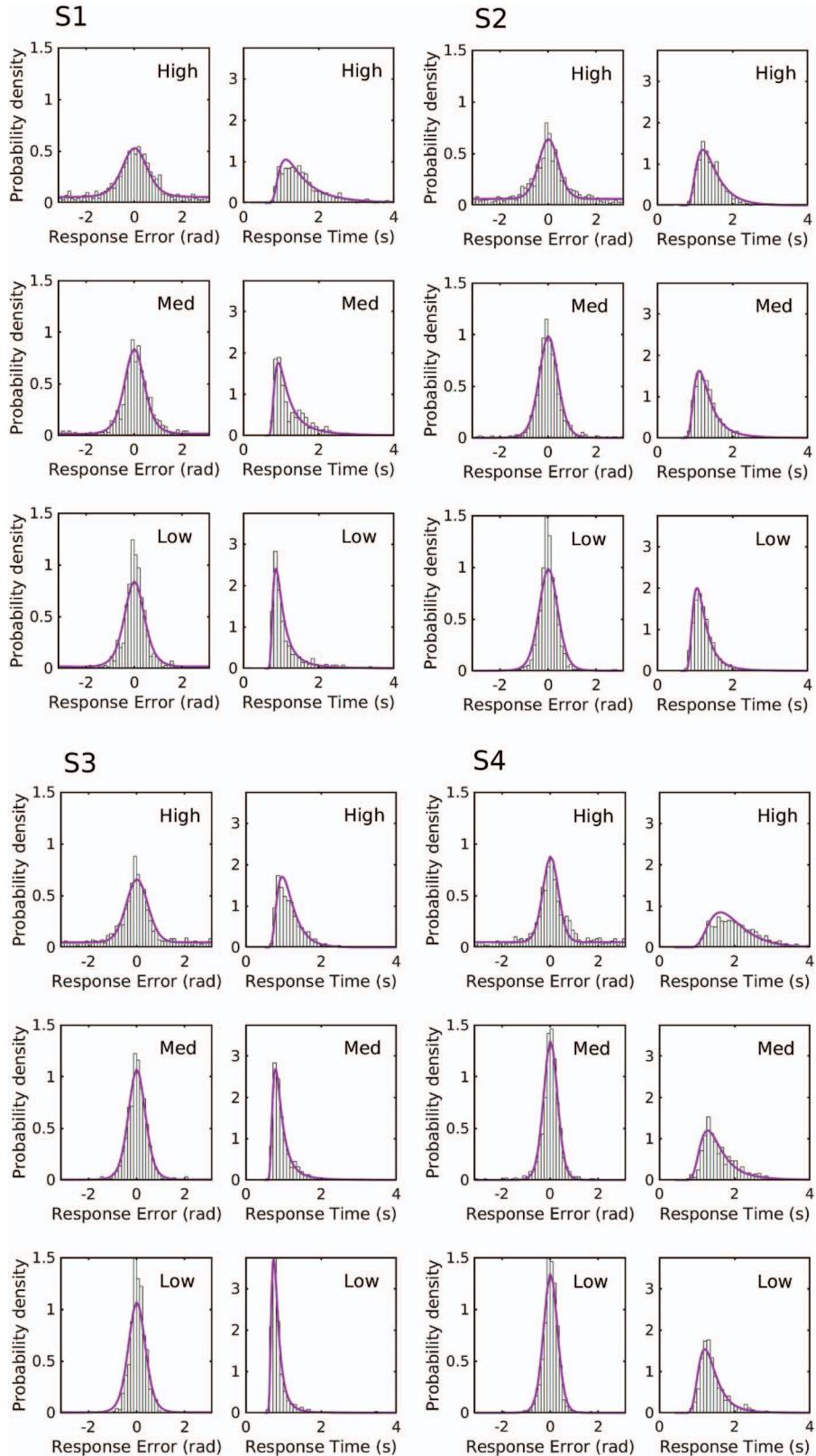


Figure 5. Fits of the bivariate normal model with encoding failures to the distributions of decision outcomes and reaction times (RTs) for individual participants for high, medium, and low levels of chromatic noise. See the online article for the color version of this figure.

Figure 6 shows fits of the model to the joint distribution in the form of a bivariate quantile ($Q \times Q$) plot. These plots can be viewed as a generalization of the quantile-probability plots that are widely used to show fits of models of two-choice decisions (e.g., Ratcliff & Smith, 2004). In a $Q \times Q$ plot, the horizontal axis shows quantiles of the distribution of decision outcomes and the vertical axis shows quantiles of the distribution of RT. In constructing these plots we have used seven quantiles on the horizontal axes and five on the vertical axes. The quantiles on the horizontal axis partition the marginal distributions of decision outcomes into seven bins, each containing 14.3% of the mass of the distribution. The quantiles on the vertical axis partition the marginal distributions of RT into six bins. The interior bins each contain 20% of the mass of the distribution while the two end bins each contain 10% of the mass. This structure suffices to characterize the main properties of the joint distributions while not being too sensitive to outliers.

As with the marginal distributions, the model does a fairly good job of characterizing the joint distributions for all four participants. All of them show, at least to some degree, the slow error pattern described previously, which is manifested in the dish-shaped profiles of the $Q \times Q$ plot, in which low-accuracy responses (those toward the left and right extremes of the plot) are made more slowly. Like the results of difficult two-choice decision tasks, the slow-error property is most apparent in the upper quantiles of the RT distribution (i.e., the .7 and .9 quantiles), which characterize the slowest 30% and slowest 10% of responses, respectively (Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Smith, Ratcliff, & Wolfgang, 2004; Smith & Ratcliff, 2009). The .1 quantile, which characterizes the leading edge of the RT distribution (i.e., the fastest 10% of responses), shows relatively little change. The pattern of responding in Figure 6 is therefore a close analogue of the typical pattern of data found in two-choice decision tasks when discrimination is difficult.

We noted earlier that studies of continuous report decisions about the hues of color patches have shown strong effects of nameable colors on speed and accuracy and in a later section we present a model of such effects. In fact, the model we used to fit the data in Figures 5 and 6 incorporated a model of categorical effects and we fit the model to our data at the level of the conditional likelihoods, as a function of the locations of the stimuli in color space, as we describe subsequently. (We use the term *conditional likelihood* to refer to a model in which the mean drift rate, v_0 , in $d\tilde{P}_t(\theta_T)$, the predicted joint density of decision times and decision outcomes, varies as a function of the location of the stimulus in color space. The *marginal likelihood* is obtained by integrating the conditional likelihood function over $1/(2\pi)$, the probability density of the stimuli in color space, which was uniform by virtue of the design of the experiment.)

The fitted values shown in Figures 5 and 6 are the marginal predictions of the conditional model, averaged (pooled) over stimulus identity. We fit both conditional and marginal versions of our models; the former included color category effects whereas the latter did not. The fits of the conditional model were numerically better than those of the marginal model because it incorporated stimulus-specific effects that the marginal model lacked, and all of the individual fits of the conditional model were significantly

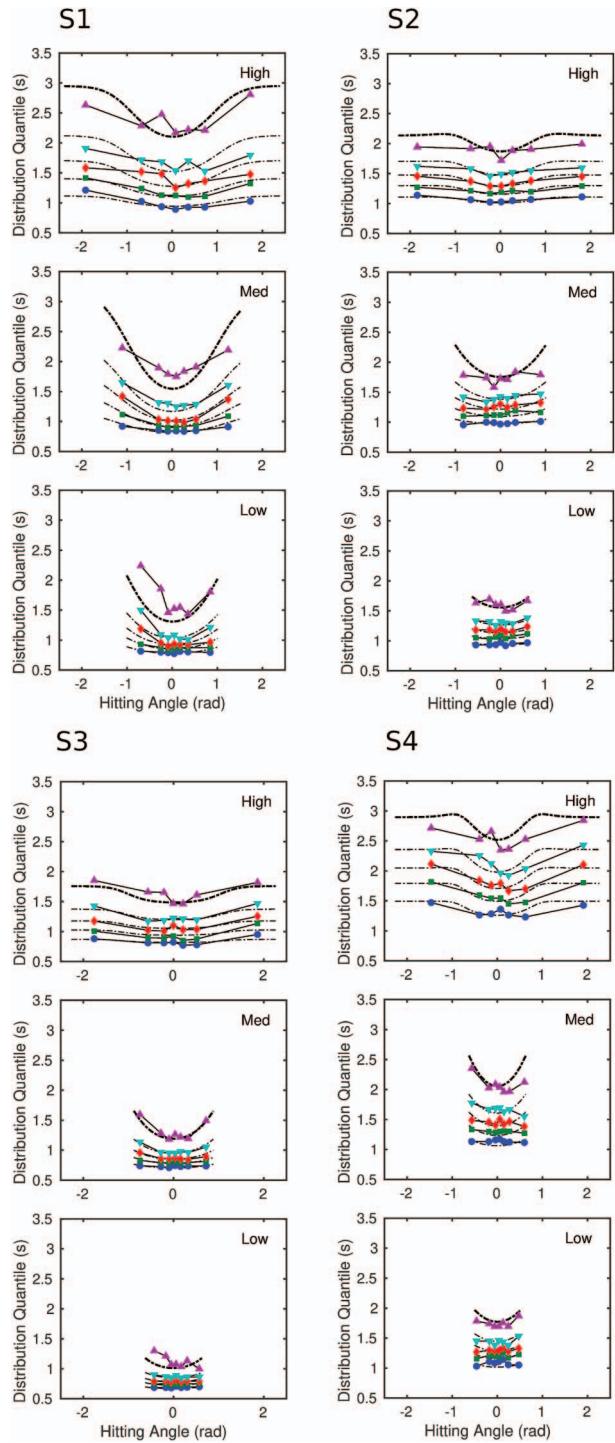


Figure 6. Bivariate quantile ($Q \times Q$) plots of joint distributions of decision outcomes and RTs for high, medium, and low levels of chromatic noise for individual participants. The fitted model had bivariate-normal drift rates and encoding failures. The horizontal axes show the observed and predicted quantiles (symbols and lines, respectively) of the distribution of decision outcomes; the vertical axes show the quantiles of the reaction time (RT) distributions. The circles, squares, diamonds, inverted triangles, and upright triangles are the .1, .3, .5, .7, and .9 quantiles of the RT distributions, respectively. See the online article for the color version of this figure.

better according to a classical nested likelihood ratio test by a large margin, $\chi^2(7)$ or $\chi^2(11)$, all $p < 1.0 \times 10^{-10}$. Because the effects of color categories only appear when performance is conditioned on the location of the stimulus in color space, however, plots of the fits of the two versions of the model to the distributions of decision outcomes and RTs and the $Q \times Q$ plots were virtually indistinguishable. We have therefore chosen to present the predictions of the conditional model, marginalized across stimulus locations, in Figures 5 and 6 to conserve space.

Table 1 compares the fits of the models with and without color category effects and also, as discussed in the next section, with and without encoding failures. A model with color category effects was preferred by the QAIC for all participants and for three of the four participants by the more conservative QBIC. The exception was participant S_2 , for whom the overdispersion estimate and the penalty term in the QBIC were atypically large. For this participant, the QBIC preferred a model without color category effects, although a standard BIC, with no correction for overdispersion, preferred a model with color category effects for all participants. The parameters of the color category model are discussed subsequently. In addition to these effects, as we discuss in the following section, the table also compares the performance of models with and without the effects of an additional, encoding failure, process, which we discuss below.

Heavy Tails and Encoding Failures

There is an important caveat to the apparently good fits in Figures 5 and 6, to which we now turn. As we noted earlier, in the visual working memory literature, heavy-tailed distributions of decision outcomes are commonly found. Zhang and Luck's (2008) slots-plus-averaging model, which assumes memory is item-capacity limited, attributes them to the action of a **guessing process**, whereas van den Berg et al.'s (2014) variable precision model attributes them to **heterogeneity in encoded item representations**. Harlow and Donaldson (2013) found similar heavy-tailed distributions in a study of source memory, in which people were required to recall the location at which a previously studied item had been presented, where the locations were points on the circumference of a circle. Harlow and Donaldson attributed the heavy

tails of the distributions to **a threshold process**: On a proportion of trials the memory trace representing the association between the item and the source is too weak to recall and people are obliged to guess. Essentially, their account of the heavy distribution tails is the same as that of Zhang and Luck and is assumed to arise for similar reasons.

Like those studies, **all four of our participants showed heavy-tailed distributions of decision outcomes at the highest level of chromatic noise**, when the decision task was most difficult (see Figure 5). There is little evidence of similar effects at either medium or low levels of noise. Indeed, these effects were not apparent in the first version of the experiment we ran, and we only obtained them when we replicated the experiment with higher levels of noise. As in the visual working memory and source memory literatures, a significant theoretical question is why do these effects occur. The additional constraints imposed by fits of the circular diffusion model to the joint distributions of decision outcomes and RT can potentially shed light on this situation.

van den Berg et al. (2014) showed that their variable precision model provided a better fit to the distribution of decision outcomes than a variety of other models, including the two-state mixture model of Zhang and Luck (2008). **Any distribution of decision outcomes that can be fit with a variable precision model can be equally well fit by some version of a circular diffusion model with a suitably chosen distribution of drift rates**, because the variable precision model consists of a mixture of von Mises components and the circular diffusion model predicts a von Mises distribution of decision outcomes for any given values of drift rate, decision criterion, and diffusion coefficient. Consequently, it will always be possible to find a set of parameters of the circular diffusion model with across-trial variability that matches the predictions of the variable precision model exactly. However, in the circular diffusion model, the parameters of the model must account for the entire joint distribution of decision times and decision outcomes, not just one margin of it.

Figure 7 shows a failure of what we can regard as a diffusion model counterpart of a variable precision model to participant S_3 . As is apparent in the figure, this participant showed a heavy-tailed distribution of decision outcomes at high levels of chromatic noise

Table 1
Fit Statistics of Models With and Without Encoding Failures

Participant	q	With encoding failures				No encoding failures			
		$-LL_{\max}$	m	QAIC	QBIC	$-LL_{\max}$	m	QAIC	QBIC
Color category effects (conditional model)									
S_1	1.40	2,084	16	3,009	3,096	2,105	15	3,037	3,119
S_2	4.01	2,310	16	1,184	1,276	2,407	15	1,231	1,317
S_3	1.75	1,206	20	1,418	1,533	1,369	19	1,602	1,711
S_4	1.85	1,546	20	1,711	1,820	1,733	19	1,912	2,015
No color category effects (marginal model)									
S_1	1.40	2,357	9	3,383	3,419	2,381	8	3,416	3,459
S_2	4.01	2,398	9	1,214	1,265	2,492	8	1,258	1,304
S_3	1.75	1,319	9	1,576	1,524	1,474	8	1,670	1,746
S_4	1.85	1,775	9	1,935	1,984	1,895	8	2,063	2,106

Note. QAIC = Akaike information criterion with overdispersion; QBIC = Bayesian information criterion with overdispersion; $-LL_{\max}$ = maximized log-likelihood. Bolded entries are best models in a QAIC or QBIC sense; m is the number of free parameters and q is the overdispersion.

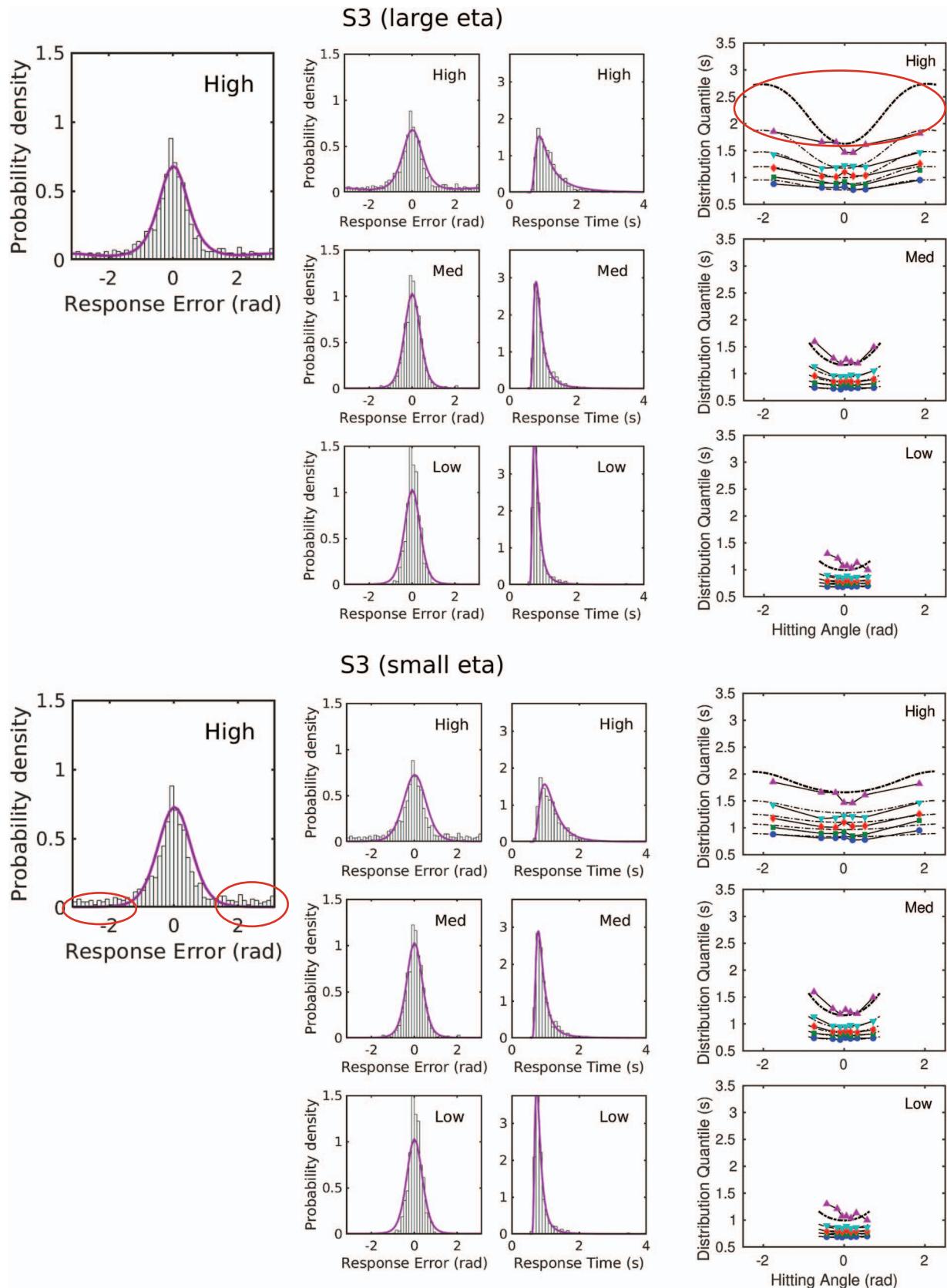


Figure 7 (opposite)

(the top left-hand panel, shown enlarged on the far left for better resolution). The data were fit with the bivariate normal model with a fairly large value of η_1 ($\eta_1 \approx 1.5$), the radial component of drift variability. The resulting model is a mixture of high precision and low precision components and satisfactorily characterizes the tails of the distribution of decision outcomes. However, the effect of the large η_1 is also to create a large slow-error effect, in which the tails of the RT distribution, especially the .9 quantile, are disproportionately long. Clearly, the model does not match the data. Constraining across-trial variability to a smaller value, $\eta_1 \approx 0.5$, allows the model to capture the tails of the RT distributions but it then misses the shoulder of the distribution of decision outcomes. The model with bivariate normal drift rates, viewed as a form of variable precision model, evidently does not capture these data satisfactorily. This problem was apparent to a greater or lesser degree with all our participants. We considered a number of different solutions to this problem.

The simplest solution, and the one we implemented in the fitted model in Figures 5 and 6, we term an *encoding failure model*. The model assumes that, on some trials, at high levels of chromatic noise, the encoded stimulus information is essentially random. We implemented this as a mixture model with a drift norm, $\|\nu\|$, that was the same on encoding failure trials as on regular trials but with a uniformly distributed phase angle (i.e., $\kappa_\mu = 0$). This model was motivated, in part, by the perceptual properties of chromatic noise. Perceptually, the dominant components of chromatic noise are those that contrast most highly with the stimulus and these components tend to be widely separated from the stimulus hue on the color wheel. It thus seemed plausible to us that, at high levels of noise, the system would encode stimulus information of some kind, but this information might be highly inaccurate on some trials. Our encoding failure model tried to formalize this intuition. In a diffusion model setting, the encoding failure model differs from a guessing model in that the latter would assume the decision process on guessing trials is driven by solely by noise. Such a process would predict a large slow-error effect, which is contrary to our data. We also considered a guessing model in which the decision criterion for guesses was lower than for regular responses: Such a model could be viewed as a form of “low threshold” guessing model. Allowing the criteria for guesses and regular responses to differ reduced the slow error effect, but the model was unsuccessful in capturing the fine-grained structure of the joint distributions than the encoding failure model, so we do not consider it further.

The model we have shown in Figures 5 and 6 is the bivariate-normal drift rate model augmented with the assumption that, on a proportion of trials, at high noise levels only, there are encoding failures. On encoding failure trials, the phase angle of the stimulus representation is uniformly distributed around the circle. The en-

coding failure assumption allows the model to predict both the heavy-tailed distributions of decision outcomes in the top row of Figure 5 and the tail quantiles of the joint distributions in Figure 6. Table 1 summarizes the comparative fits of the bivariate normal model with and without encoding failures and Table 2 shows the estimated model parameters. All of the model parameters are scaled relative to a diffusion coefficient of $\sigma^2 = 1.0$. The addition of encoding failures improved the model fits for all participants according to both the QAIC and the QBIC. Qualitatively, the improvements were strongly apparent in the plots of the fitted models for two participants (S_2 and S_3), marginal for one (S_4), and not appreciable for the fourth (S_1). For all participants, the estimated rate of encoding failures on high-noise trials (denoted π_{fail} in the table) was around one third.

We considered two versions of each of the two models shown in Table 1. In one version, both the radial and tangential components of drift rate mean and standard deviation were permitted to vary; in the other, only the radial components varied and the means and standard deviations of the tangential component were set to zero. For three of the four participants (S_2 , S_3 , and S_4), the model with only radial components free to vary was preferred by the QBIC and for the fourth (S_1) the difference between the QBICs for the two versions of the model was small. The models we have reported in Tables 1 and 2 are therefore those in which only the radial components varied and the values of ν and η in the table are the means and standard deviations of this component, respectively. The implication of these fits is that the means and the standard deviations of the radial component of drift rate can account for the shapes of the distributions of decision outcomes and RTs in Figures 5 and 6 with little or no additional contribution from the tangential component and, in particular, with little or no tangential variability. Psychologically, the fits imply that there is appreciable variability across trials in the quality of the encoded stimulus representations, but—aside from the mechanism that produces encoding failures or the appearance of them on some trials—there is comparatively little variability in the encoded stimulus identity.

Nonlinear Scaling of the Phase-Angle Distribution

The encoding failure model, although it is consistent with the perceptual properties of chromatic noise stimuli and accords with the subjective experience of doing the task, may strike some readers as ad hoc, so we looked for another way to account for the heavy-tailed distributions of decision outcomes found under high-noise conditions. An alternative account of these kinds of distributions in visual working memory was given by Schurgin, Wixted, and Brady (2018) who proposed that the heavy tails of the distribution of decision outcomes are a reflection of nonlinear scaling of the underlying stimulus space. This makes pairs of stimuli that are

Figure 7 (opposite). Failure of the bivariate normal model for participant S_3 . The panels at the top show marginal and joint distributions of decision outcomes for high levels of radial drift rate variability (η_1); the panels at the bottom show the distributions for low levels of radial drift rate variability. With a large η_1 the model captures the heavy tail of the distribution of decision outcomes in the high noise condition but overpredicts the tail quantiles of the RT distributions. With a small η_1 the model captures the tail quantiles of the RT distributions but misses the shoulder of the distribution of decision outcomes. The differences in fit to the distributions of decision outcomes in the high noise condition are highlighted in the enlarged panels on the far left. See the online article for the color version of this figure.

Table 2
Parameters of Models With Multivariate Normal Drift Rates

Participant	ν_1	ν_2	ν_3	η_1	η_2	η_3	a	T_{er}	π_{fail}
With encoding failures									
S_1	1.017	1.572	2.137	0.674	1.602	1.727	1.972	0.571	0.277
S_2	1.929	2.131	2.603	0.825	1.018	1.244	2.120	0.660	0.344
S_3	2.086	3.233	4.238	0.000	1.623	1.748	1.872	0.497	0.302
S_4	2.209	2.864	3.313	0.538	0.954	0.783	4.078	0.408	0.317
No encoding failures									
S_1	0.728	1.459	2.010	1.051	1.528	1.638	1.878	0.583	0
S_2	1.265	1.952	2.379	1.232	0.800	0.921	1.918	0.661	0
S_3	1.332	2.840	3.828	0.000	1.444	1.643	1.575	0.523	0
S_4	1.853	2.546	2.948	0.672	0.787	0.604	3.751	0.408	0

farther from a given reference stimulus appear closer than they would otherwise appear. Schurgin et al. carried out a study of judgments of stimuli defined on circular domains using a scaling method developed by Maloney and Yang (2003) and found evidence of an exponential-like compression of the similarities among stimuli as a function of the distance between them in circular space. They showed that a combination of nonlinear scaling and a signal detection theory model of stimulus noise was able to reproduce the distributions of decision outcomes reported by Zhang and Luck (2008) and others, in which the distribution tails become heavier as the number of items in memory increases.

The model proposed by Schurgin et al. (2018) is essentially a form of population-coding model (Bays, 2014; Prinzmetal et al., 1998; Seung & Sompolinsky, 1993), in which the decision depends on the most active among a set of detectors that code values of a continuous stimulus attribute. Like most other models in the visual working memory literature, it is a model of the decision outcome only and is silent on the subject of decision times. However, it is possible to incorporate similar ideas into a diffusion model framework by assuming that the nonlinear scaling of stimuli is expressed in the across-trial distribution of drift rates. In Schurgin et al.'s model, the decision depends on the most active member in a set of noisy, labeled detectors, whose mean activity strength decreases nonlinearly with the detector's distance from the true stimulus value. In a diffusion model, the drift rate is a random draw from a distribution that represents possible values of the encoded stimuli across trials. Translating Schurgin et al.'s idea into a diffusion model context leads to the idea that the distribution of phase angles of the drift rate, rather than being normally distributed, might instead have extended tails because of the scaling properties of the underlying space. Although the results for the bivariate normal model showed only a weak contribution of tangential drift rate variability to the overall fit, the results do not speak to the question of whether an alternative representation of drift rate variability could more successfully capture the heavy tails of the distributions of decision outcomes.

To implement a model of this kind we looked for a distribution defined on a circular space that was able to flexibly represent the distribution of phase angles across trials. A distribution with these properties was proposed by Jones and Pewsey (2005). Their distribution is a three-parameter family with probability density

$$f(\theta; \kappa_\mu, \varphi, \psi) = \frac{[\cosh(\kappa_\mu\psi) + \sinh(\kappa_\mu\psi)\cos(\theta - \varphi)]^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa_\mu\psi))}, \quad (12)$$

examples of which are shown in Figure 8A. In this equation, θ is the phase angle of the stimulus, φ is its center, κ_μ is the precision, and ψ is a parameter that controls the distribution shape. The term $P_{1/\psi}(\cdot)$ in the denominator, which makes the function a probability distribution, is an associated Legendre function of the first kind of degree $1/\psi$ and order 0 (Gradshteyn & Ryzhik, 2007, 8.7–8.8).

The Jones-Pewsey distribution of Equation 12 is able to represent a wide variety of functional forms (see their article for examples), but the ones of particular concern to us arise when $\psi \leq 0$. As ψ approaches 0, Equation 12 approaches the von Mises distribution of Equation 1, while when $\psi = -1$ the distribution becomes a wrapped Cauchy distribution, which is a circular analogue of the Cauchy distribution on the real line. The latter is widely used as a model of physical and biological phenomena with heavy-tailed distributions. The special case in which the phase angle is von Mises distributed leads to an analytically tractable model we refer to as the “doubly von Mises” model. This model, which was discussed briefly by Smith (2016), arises when variability in the encoded stimulus identity in memory, represented by the drift rate phase angle, θ_μ , is itself also the outcome of a diffusion process. The resulting diffusion can be thought of as taking place on a circle (technically, a circular manifold), rather than on the interior of a disk, and was shown by Carverhill (1985) to follow a von Mises distribution (Rogers & Williams, 2000, pp. 141–144). We investigated versions of the doubly von Mises model and found that its predictions were similar to those of the bivariate normal model. This is unsurprising as the von Mises and normal distributions have similar bell-shaped density functions. In light of the similarities between the two models, we do not report fits of the doubly von Mises model and consider instead a more general Jones-Pewsey model in which ψ is allowed to freely vary.

Models With Jones-Pewsey Distributed Phase Angles

We investigated models in which the across-trial variability in drift rate phase angle followed a Jones-Pewsey distribution and in which the drift rate norm also varied across trials. In polar coordinates, these components are represented by the tangential and radial components of drift rate, respectively. We do not have a

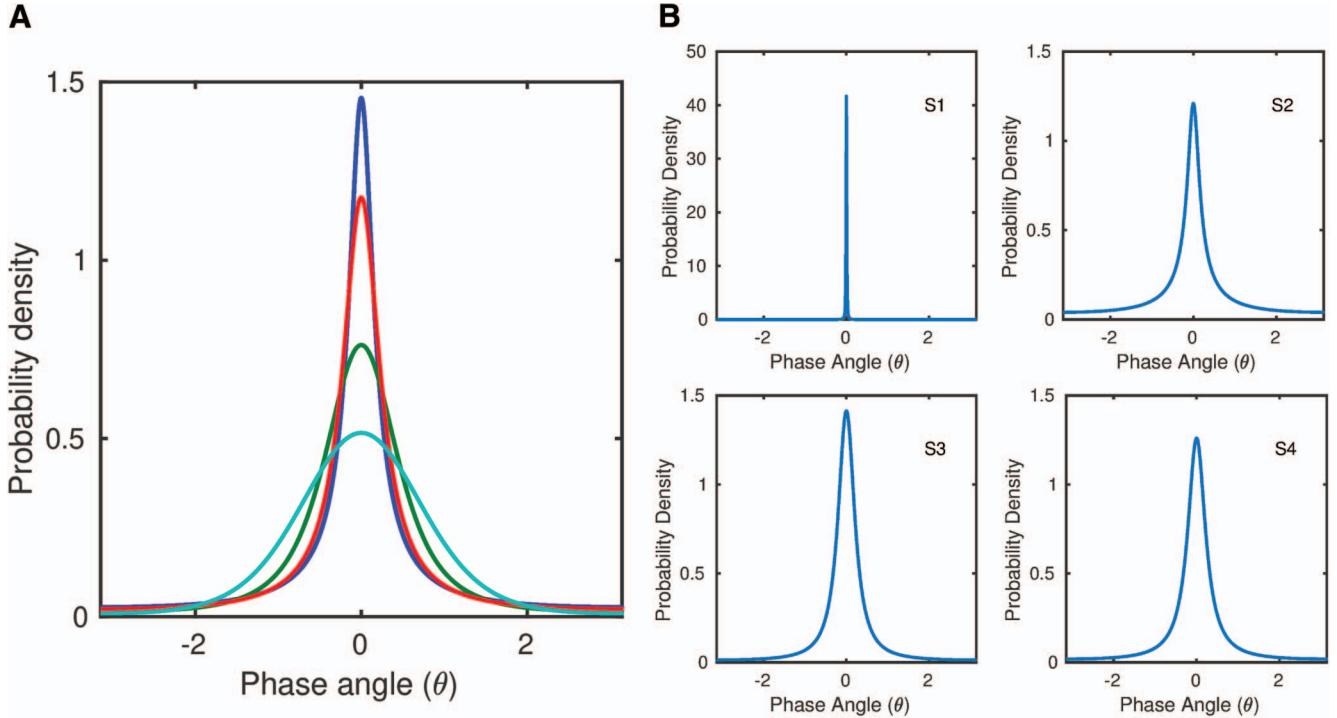


Figure 8. (A) Jones-Pewsey distribution of across-trial variability in drift rates. The curves in the figure, from least to most peaked, are for $\psi = 0, -0.5, -1.0, -1.25$, all with $\kappa_\mu = 2$. The $\psi = 0$ case is a von Mises distribution. (B) Estimated distributions of drift rate phase angle distribution for individual participants in the high-noise condition. Note the differences in y-axis scaling in the four panels on the right. See the online article for the color version of this figure.

theory of how drift rate norms should vary, except that norms should be positive and the distribution that characterizes them should be defined on the positive real line. We implemented a marginal model, averaging over locations in stimulus space, in which the drift rate phase angle and norm were Jones-Pewsey-distributed and Weibull-distributed, respectively, but the conditional model implemented in this way was too expensive computationally because of the number of levels of numerical integration involved. To circumvent this problem, we implemented a conditional model using the analytic predictions of the bivariate normal model of Equation 9, in which the radial component of drift rate varied normally across trials and the tangential component followed a Jones-Pewsey distribution, which we evaluated numerically. We found that the marginal predictions of this model were in fairly close agreement with the model with Weibull-distributed norms, so we again reproduce only the fits of the conditional model to conserve space.

Figure 9 shows the marginal distributions of decision outcomes and RTs for the model with Jones-Pewsey distributed phase angles and Figure 10 shows the $Q \times Q$ plot of the joint distributions. Like the bivariate normal model with encoding failures, the model captures both the shapes of the distributions of decision outcomes and RTs, including the heavy-tailed distributions of decision outcomes in the high-noise condition, and also the shapes of the RT quantiles as a function of response accuracy in the $Q \times Q$ plot. A question that arose when implementing the Jones-Pewsey model was whether the drift rate norm, $\|\boldsymbol{\mu}\|$, should depend on the drift

rate phase angle, θ_μ . The simplest assumption is that they are independent of each other but a plausible alternative assumption is that they are related, as we might expect that on trials on which the identity of the stimulus is encoded poorly, the quality of the associated representation might also be poor. We considered a simple form of the dependent-norm model, in which the mean of the distribution of drift rate norms, $\|\boldsymbol{v}\|$, decreased linearly with $|\theta_\mu|$ on the range $(-\pi, \pi)$.

Table 3 shows the fit statistics of models with varying phase angles and Table 4 shows estimated model parameters. As in Table 1, we have shown fits of versions of the models with and without color category effects. Like the models in Table 1, the QAIC prefers a model with color categories for all four participants, as does the standard BIC; the QBIC prefers a model with color categories for two of the four participants and a model without categories for the other two. The fits of the model with color categories were significantly better than those of the model without categories by a large margin for all four participants according to a nested likelihood ratio test ($\chi^2(7)$ or $\chi^2(11)$, all $p < 1.0 \times 10^{-10}$). These results are similar to those in Table 1.

There is no agreement on whether the model with independent drift rate norms or phase angle dependent norms is the better of the two models according to either the QAIC or QBIC. A comparison of the maximized log-likelihoods of the encoding failure model in Table 1 and the variable phase angle model in Table 3 show that the maximized log-likelihoods of the two models were similar. The QAIC and QBIC model selection statistics were lower for the

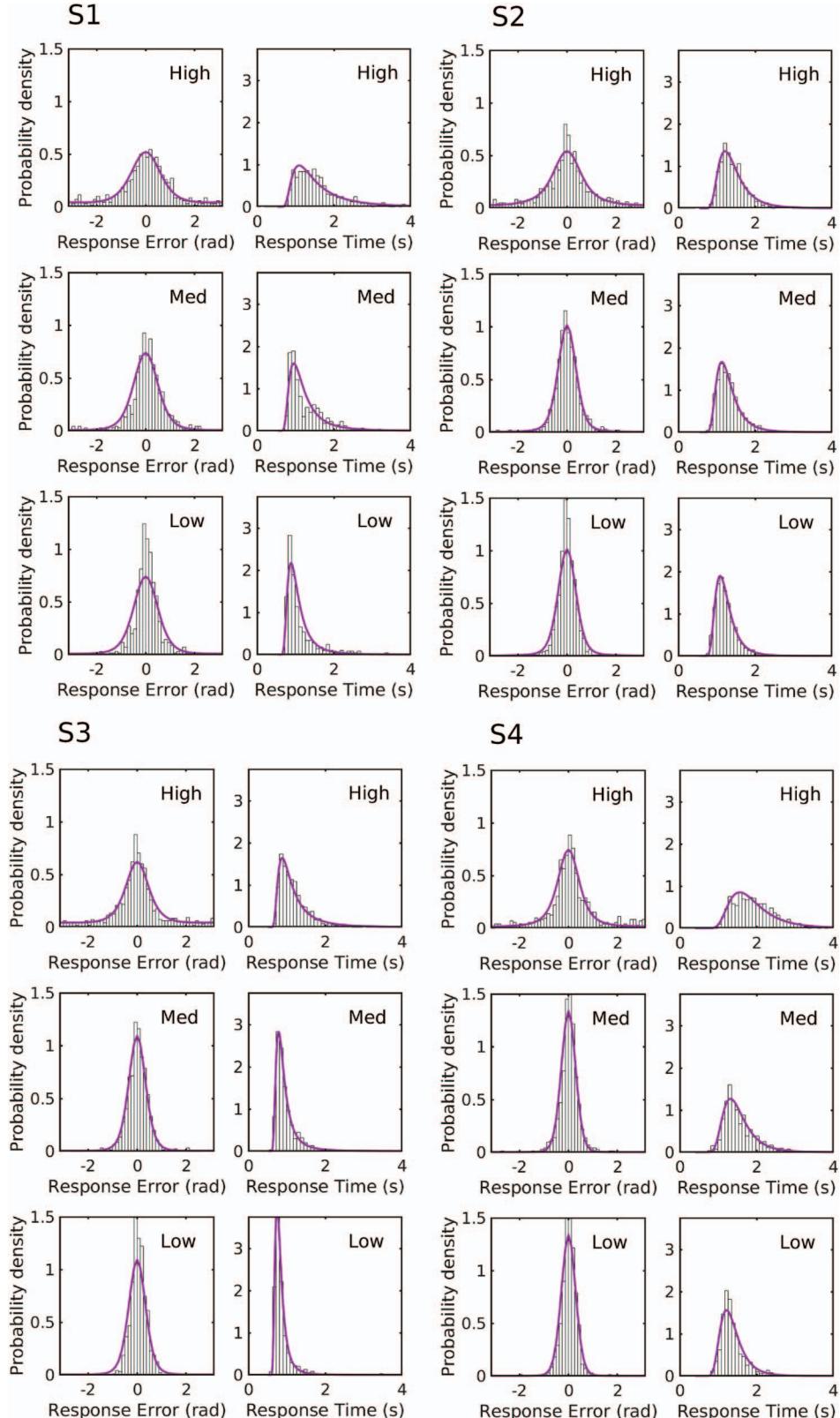


Figure 9. Marginal distributions of decision outcomes and reaction times (RTs) for the model with variable (Jones-Pewsey distributed) phase angles. The drift rate phase angle, θ_μ , was distributed as Equation 12 and the mean drift norm, $\|\nu\|$, was a linear function of the phase angle. See the online article for the color version of this figure.

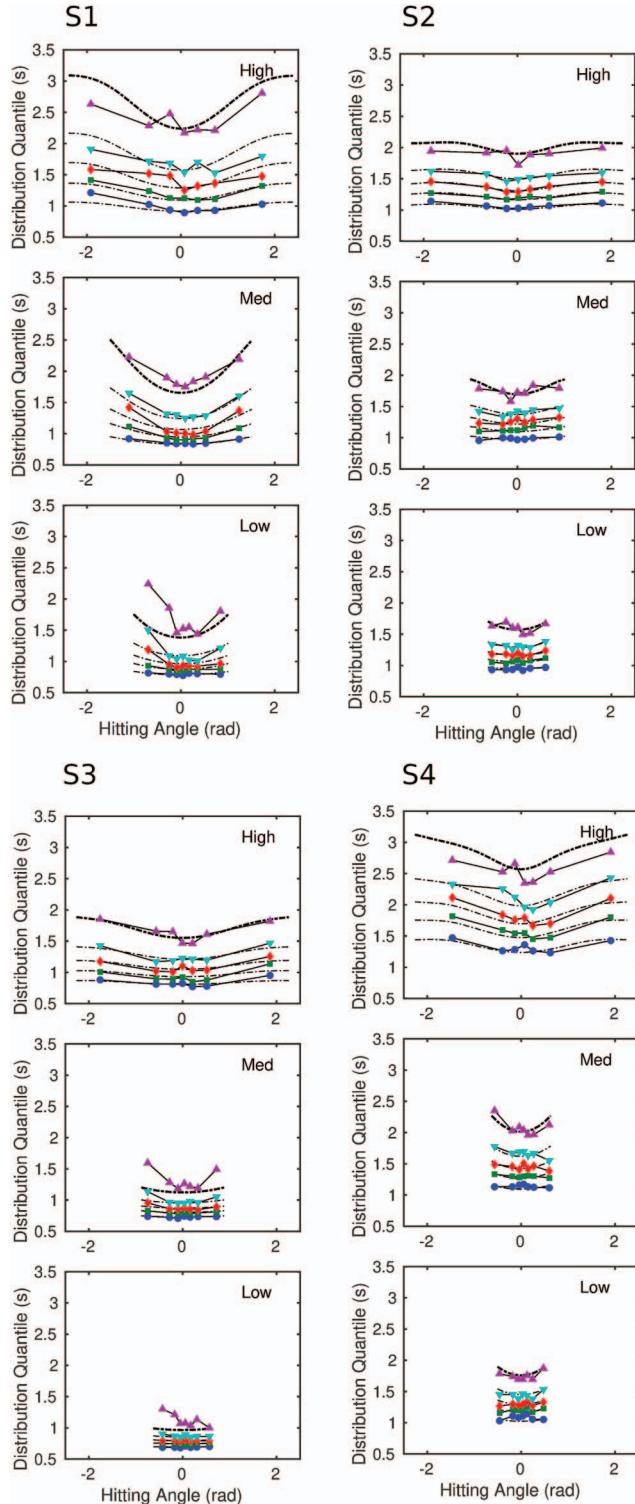


Figure 10. Joint distributions of decision outcomes and reaction times (RTs) for model with Jones-Pewsey distributed phase angles. The horizontal axes show the observed and predicted quantiles (symbols and lines, respectively) of the distribution of decision outcomes; the vertical axes show the quantiles of the RT distributions. The circles, squares, diamonds, inverted triangles, and upright triangles are the .1, .3, .5, .7, and .9 quantiles of the RT distributions, respectively. See the online article for the color version of this figure.

encoding failure model than for the variable phase angle model for most participants, but this is just a reflection of their relative parsimony—as we have parameterized them here—rather than their ability to describe the data. The encoding failure model uses one additional stimulus parameter (π_{fail}), whereas the variable phase angle model uses four ($\kappa_1, \kappa_2, \kappa_3, \psi$). Taken overall, the similarities between the two models outweigh their differences.

Figure 8B shows the estimated Jones-Pewsey distributions of phase angles for the high noise condition for the four participants, based on the parameters in Table 4. For three of the four participants the distribution is a markedly leptokurtic, heavy-tailed distribution, similar to a wrapped Cauchy distribution in form. It is the heavy-tailed nature of this distribution that allows the model to account for the distributions of decision outcomes and the joint distributions of decision outcomes and RT. According to the fitted model, there is a nonnegligible probability that the encoded phase angle of the stimulus will differ from the true angle by amounts as great as $\pm\pi$. It is these extreme phase angles that allow the model to predict the heavy-tailed distributions of decision outcomes.

For the fourth participant, S_1 , the estimated distributions of phase angles resemble a single spike of probability mass located at a phase angle of zero (a Dirac delta function). It is not the case that this distribution has zero probability mass in its tails but the scale of the figure means the tail mass is not visible in the plot. This distribution implies that, for participant S_1 , there was a very low probability that the encoded stimulus phase angle differed appreciably from the true value. Consistent with this interpretation, participant S_1 was the one who showed the smallest difference in the fits of bivariate normal model with and without encoding failures and, as we remarked earlier, the $Q \times Q$ plots of the two models for this participant did not differ in any obvious way.

Overall, the fits of the encoding failure and variable phase angle models provide a highly convergent picture of the underlying process. Both models suggest that the identity of the stimulus (as expressed by its phase angle) is encoded accurately on the majority of trials, but on a small proportion of trials under high noise conditions the encoded identity is wrong. If it is wrong then it is very wrong; rather than being clustered around the true stimulus identity, it is distributed, apparently at random, around the circle. As we commented earlier, this may be a reflection of the perceptual properties of chromatic noise, which increases in contrast and perceptual saliency the further it is from the background color. The encoding failure model characterizes the heavy tails of the distribution of decision outcomes under high noise conditions via an encoding-failure parameter, π_{fail} ; the variable phase angle model characterizes it via the shape parameter of the Jones-Pewsey distribution, ψ . Although the model semantics are different, the psychological process that the models are describing is essentially the same. In either case, the mechanism that produces highly inaccurate encodings only manifests itself under high noise conditions.

Apart from the infrequent trials on which stimulus identity was encoded very inaccurately, we found little evidence for variable precision, in the sense it was defined by van den Berg et al. (2014), as a determining factor in performance. In the bivariate normal model, variable precision is expressed as variability in the tangential component of the distribution of drift rates. In the variable phase angle model, it is expressed in the κ parameter of the Jones-Pewsey distribution, interacting with the shape parameter. In

Table 3
Fit Statistics of Variable Phase Angle Models

Participant	<i>q</i>	Norm-independent phase angle				Norm-dependent phase angle			
		− <i>LL</i> _{max}	<i>m</i>	QAIC	QBIC	− <i>LL</i> _{max}	<i>m</i>	QAIC	QBIC
Categorical effects									
<i>S</i> ₁	1.40	2,160	19	3,120	3,227	2,159	20	3,124	3,235
<i>S</i> ₂	4.01	2,279	19	1,175	1,284	2,275	20	1,174	1,290
<i>S</i> ₃	1.75	1,359	23	1,598	1,730	1,214	24	1,424	1,562
<i>S</i> ₄	1.85	1,603	23	1,778	1,904	1,596	24	1,773	1,904
No categorical effects									
<i>S</i> ₁	1.40	2,419	12	3,422	3,487	2,432	13	3,434	3,504
<i>S</i> ₂	4.01	2,367	12	1,190	1,258	2,355	13	1,186	1,260
<i>S</i> ₃	1.75	1,488	12	1,657	1,726	1,331	13	1,478	1,558
<i>S</i> ₄	1.85	1,817	12	1,934	1,999	1,801	13	1,918	1,989

Note. QAIC = Akaike information criterion with overdispersion; QBIC = Bayesian information criterion with overdispersion; −*LL*_{max} = maximized log-likelihood. Bolded entries are best model in a QAIC or QBIC sense; *m* is the number of free parameters and *q* is the overdispersion.

both models, variability in drift rate norm was needed to account for the joint distributions of decision outcomes and RT, but in neither model was there evidence of significant across-trial variability in encoded stimuli identity, apart from the minority of trials on which encoding comprehensively failed under high noise conditions. In the case of the bivariate normal model, the tangential component of drift rate variability made only a minor contribution to the model's account of the data. In the case of the variable phase angle model, the estimated parameters of the distributions of phase angles in Table 4 and Figure 8B suggest that the majority of the encoded stimulus representations clustered close to the true ones. In both models, the majority of the variance in the distributions of decision outcome is attributable to diffusion noise in the evidence accumulation process. As we comment in the Discussion, the picture that emerges from these fits is in broad agreement with a recent study by Shen and Ma (2019), who found little evidence for variable precision in visual perception once other sources of variability were controlled for.

Our variable phase angle model is related to a model used by Marshall and Bays (2013) to characterize visual working memory in a study investigating interference by irrelevant features. They compared a two-component, memory-plus-guessing model with a continuous model based on a wrapped stable distribution (Pewsey, 2008). Like the Jones-Pewsey model of Equation 12, the wrapped stable distribution can represent a variety of shapes, including distributions with varying skewness and kurtosis. The wrapped stable distribution is a four-parameter rather than a three-parameter family, which allows it to represent distributions that are not symmetrical around their midpoints, but, unlike Equation 12, it does not have a simple analytic form. Marshall and Bays found

that the two models provided similar accounts of the distributions of decision outcomes in one of their experiments but that the wrapped stable distribution fit better in a second. Some of the difference between their results and ours may be due to differences between their experimental task and ours, but it may simply be because our encoding failure model included a component of variable precision due to drift-rate norm variability whereas their two-component mixture model did not.

The Effects of Color Categories

As noted previously, early studies using the continuous outcome hue task treated the color space as isotropic, but more recent studies have found strong effects of nameable colors on the speed and accuracy with which these kinds of decisions are made (Bae et al., 2015; Hardman et al., 2017; Persaud & Hemmer, 2016; Ratcliff, 2018). These effects have been found in both short- and long-term memory and in immediate decisions with no delay prior to the decision. We used an isoluminant, equidiscriminable color space to try to minimize color category effects of this kind, but this manipulation was ineffective.

Figure 11 shows distributions of responses around the color circle for each of the four participants at each of the three levels of stimulus noise. The stimuli were uniformly distributed around the color wheel in our experimental design, so any departures from uniformity in responding are evidence for response biases of some kind. The figure shows that all of our participants responded in a categorical way, but the numbers of categories and the locations of their centers were idiosyncratic. Participant *S*₁ shows a clear and stable three-category structure, centered on the locations 1.9, −0.8,

Table 4
Parameters of Models With Variable Phase Angles

Participant	$\ \nu_1\ $	$\ \nu_2\ $	$\ \nu_3\ $	κ_1	κ_2	κ_3	η_1	η_2	η_3	Ψ	a	T_{er}
<i>S</i> ₁	0.586	1.140	1.556	6.551	7.459	7.810	0.900	0.100	0.206	−0.814	1.844	0.543
<i>S</i> ₂	1.680	2.018	2.261	1.641	3.172	7.877	0.883	0.336	0.194	−0.959	2.696	0.548
<i>S</i> ₃	1.146	1.772	2.286	2.306	7.167	10.009	0.750	0.156	0.062	−0.891	1.812	0.466
<i>S</i> ₄	1.708	2.333	2.717	2.096	4.249	8.048	0.681	0.462	0.359	−0.971	4.012	0.397

Note. Parameters for best-fitting models with color category effects.

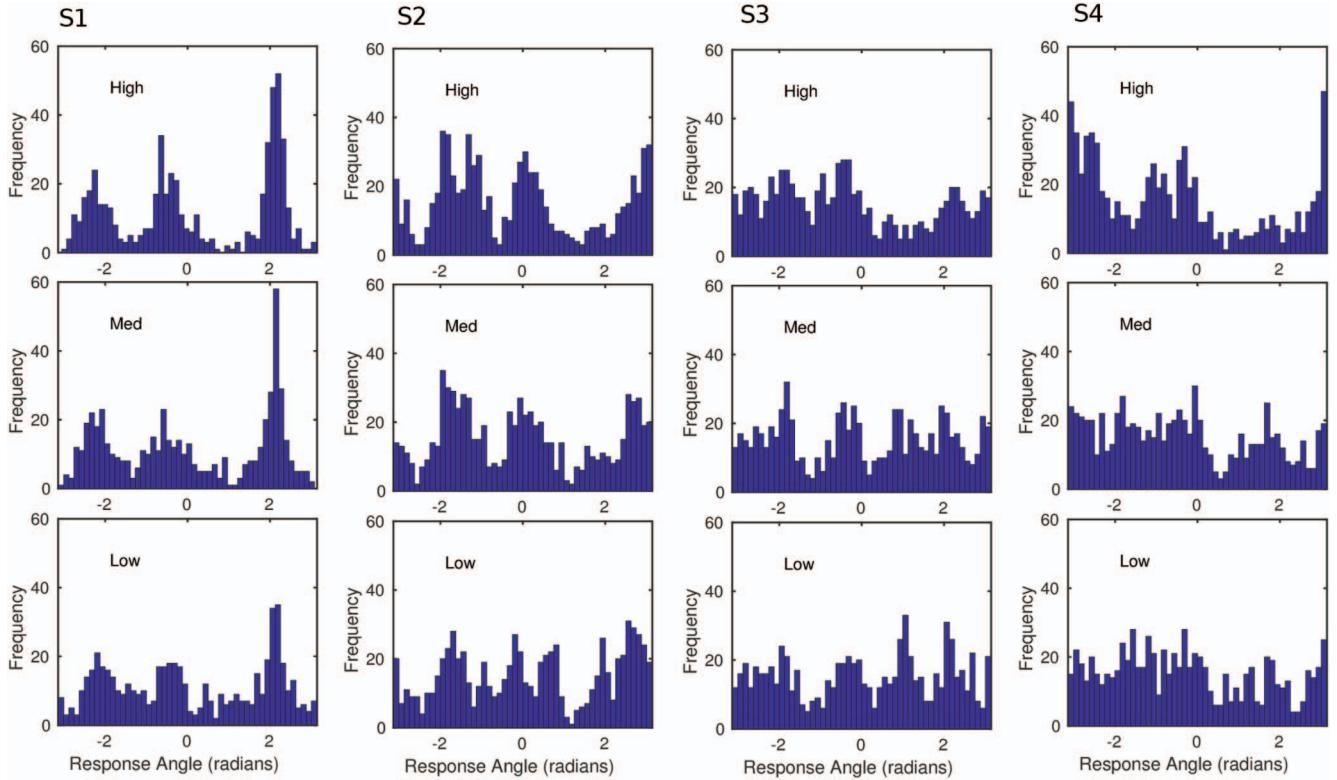


Figure 11. Distributions of responses around the color circle for high, medium, and low noise stimuli. The location of the response is expressed in radians with zero corresponding to the positive real axis and the positive direction taken as anticlockwise. See the online article for the color version of this figure.

and -2.6 radians, which correspond roughly to the transition between the yellows and greens, the center of the teals, and the center of the pinks in Figure 1, respectively. Participant S_2 also shows a fairly clear three-category structure, while the numbers and locations of the categories for the remaining two participants are less clear. Participant S_3 shows evidence of four categories and possibly a fifth (somewhere in the vicinity of $\pm\pi$), while the number for participant S_4 is moot. Because we did not rotate the color wheel from trial to trial, for reasons we discussed earlier, there is a possibility that the biases in Figure 11 are actually positional biases rather than category biases but, in the light of the strong category effects in hue judgments reported by other researchers, this seems unlikely.

Several proposals have been made for how to incorporate category effects into models of continuous outcome color decisions. Bae et al. (2015) and Persaud and Hemmer (2016) proposed Bayesian models, in which the spread of hues within each nameable color category induces a distribution of prior probabilities that biases the distribution of naming responses via Bayes' rule. Hardman et al. (2017) proposed a multinomial processing tree model, in which participants respond on the basis of the metric (hue identity) properties of the stimulus on some trials and on the basis of its color category on others. Ratcliff (2018) proposed a model with a category-dependent decision criterion, in which the criterion for making a response is set lower at the centers of nameable color categories than elsewhere.

The model we propose here is a **drift bias model**, in which participants' cognitive representations of color categories bias the drift rate of the evidence accumulation process. The specific form of the model was guided by both pragmatic and theoretical considerations. Pragmatically, we sought to develop a model that was compatible with the geometric and mathematical structure of the circular diffusion model. Theoretically, it was motivated by our belief that the dominant kinds of biases in decision tasks in which accuracy is stressed over speed are stimulus biases rather than response biases. Our drift bias model has some affinity with the models of both Hardman et al. and Bae et al., although we have developed it in a different way. Parenthetically, we note that we also considered a diffusion analogue of the multinomial mixture model of Hardman et al., but it did not perform as well as the model we report here. In the two-choice RT literature, stimulus bias models have been proposed by Ashby (1983) and Ratcliff and colleagues (Ratcliff, 1985; Ratcliff, Van Zandt, & McKoon, 1999; Ratcliff & Smith, 2004). In the 1D diffusion model, such biases are expressed via the idea of a "drift criterion," which induces asymmetries in the stimulus representations for the decision alternatives and affects the rates at which evidence for them accumulates. Our drift bias model can usefully be seen as a higher-dimensional analogue of a drift criterion model.

Figure 12 summarizes the properties of the drift bias model. As shown in Figure 12A, a stimulus is represented by a drift rate vector, \mathbf{v}_0 , which has a norm and phase angle, as described

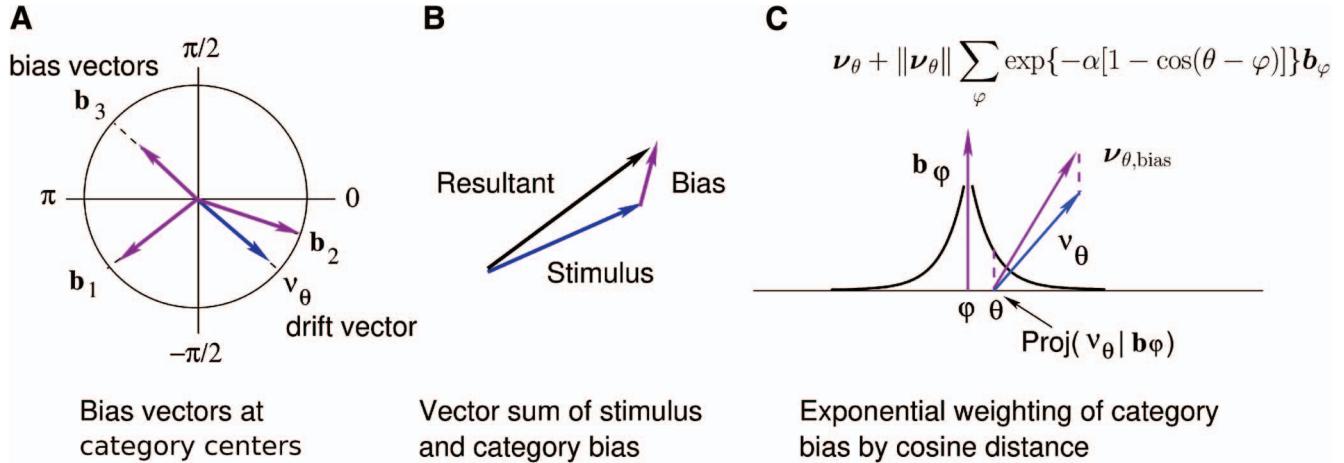


Figure 12. Drift rate bias model. (A) Stimuli presented at location θ are represented by a drift rate vector \mathbf{v}_θ . The colors categories are represented by a set of bias vectors, \mathbf{b}_i . (B) The resultant drift rate is the vector sum of the drift rate associated with the stimulus and the bias induced by the nearest color category. (C) The drift rate biases are exponentially decaying functions of the distance between the stimulus at θ and the bias vector at φ , in the $1 - \cos(\cdot)$ circular distance metric. The total bias the sum of the biases associated with each of the color categories. See the online article for the color version of this figure.

previously. The decision maker's category structure is represented by a set of bias vectors, \mathbf{b}_i , $i = 1, 2, \dots$, each of which has a phase angle and a norm. The bias vectors \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 in the figure correspond roughly to the vectors inferred for participant S_1 from Figure 11. The drift vector on any trial is the vector sum of the stimulus vector and its nearest neighboring bias vector. As shown in Figure 12B, the vectors add nose to tail in the usual manner of vector addition. This relationship has two important consequences. First, it means that the evidence accumulation process, instead of being centered on the stimulus, will be biased in either a clockwise or anticlockwise direction toward the nearest bias vector. Second, when the stimulus vector aligns with a bias vector, the vectors will be collinear and the length of the resultant vector, $\|\mathbf{v}_\theta + \mathbf{b}_i\|$, will be maximized. Evidence will accumulate more rapidly for such stimuli than would otherwise be the case and the associated RTs will be shorter. Such responses will also be made with greater precision, on account of the decomposition in Equation 2. Both of these relationships were reported by Ratcliff (2018).

To embody these ideas in a working model we borrowed an idea from the similarity-choice literature (Navarro, 2007; Nosofsky, 1984), and assumed that the effects of bias, rather than being uniform or homogeneous, were an exponentially decreasing function of the distance between the stimulus and the bias in circular space. We expressed this distance in the $1 - \cos(\cdot)$ metric that is commonly used in circular statistics (Mardia & Jupp, 1999). We implemented this in the model by assuming that the effect of a bias vector located at φ on a stimulus located at θ is given by the projection of the stimulus norm, $\|\mathbf{v}_\theta\|$, on the bias vector, \mathbf{b}_φ , exponentially attenuated by the circular distance, $1 - \cos(\theta - \varphi)$, between them. The attenuation is of the form $\|\mathbf{v}_\theta\| \exp\{-\alpha[1 - \cos(\theta - \varphi)]\} \mathbf{b}_\varphi$, where α is a generalization parameter that controls the spatial extent of the bias. We assume that the resulting distance-dependent biases are summed over all bias vectors in the color space, such that the resulting biased drift vector is

$$\mathbf{v}_{\theta,\text{bias}} = \mathbf{v}_\theta + \|\mathbf{v}_\theta\| \sum_{\varphi} \exp\{-\alpha[1 - \cos(\theta - \varphi)]\} \mathbf{b}_\varphi. \quad (13)$$

Because the effects of bias decrease exponentially with distance, for reasonable values of the generalization parameter the sum in Equation 13 will be dominated by the bias vector that is closest to the stimulus. Equation 13 embodies the property that bias is dependent on stimulus strength: When the drift norm goes to zero, the associated bias also goes to zero.

The models whose fits we showed in Figures 5, 6, 9, and 10 and reported in Tables 1 to 4 incorporated the category model of Equation 13 in the likelihood equations and predicted performance as a function of the locations of the stimulus and the response in color space, not just of the angular error. The data and the fits shown in the figures are the marginal performance and predictions of the models, aggregating over the color angles of the stimuli. Along with the parameters in Tables 2 and 4, the models had additional parameters to represent the category structure of Equation 13. Each bias vector, \mathbf{b}_i , was represented by a phase angle φ_i and a norm, $\|\mathbf{b}_i\|$, and there was a single, additional generalization parameter, α , that was common to all categories. We allowed the norms to vary across bias vectors based on the evidence of Figure 11, which suggests that some category biases are stronger than others. We compared models with differing numbers of categories for each participant. As expected, S_1 and S_2 were best described by a three-category model. Less obviously, both S_3 and S_4 were best described by a five-category model. Table 5 shows the estimated parameters of the categorical model for the bivariate normal and variable phase angle models.

Figure 13 shows the performance of the categorical model as a function of the phase angle of the stimulus. The drift rate model in this case was the encoding failure model; the performance of the variable phase angle model was similar, so we have not reproduced it. The panels on the left show the signed angular error of the response; the panels on the right show median RTs. The signed

Table 5
Parameters of the Color Category Models

Participant	φ_1	φ_2	φ_3	φ_4	φ_5	$\ b_1\ $	$\ b_2\ $	$\ b_3\ $	$\ b_4\ $	$\ b_5\ $	α
Encoding failure model											
S_1	1.419	1.386	2.173	—	—	2.013	3.924	5.740	—	—	3.142
S_2	0.633	0.776	0.662	—	—	0	2.530	4.525	—	—	1.110
S_3	0.527	0.381	0.085	0.563	0.319	0.924	1.955	3.142	4.287	5.642	4.492
S_4	1.138	0.630	1.621	0.329	1.664	0.175	2.384	4.859	6.277	3.927	0.109
Variable phase angle model											
S_1	1.590	1.654	2.645	—	—	1.986	3.880	5.743	—	—	2.869
S_2	1.393	1.609	1.540	—	—	0	2.310	4.346	—	—	0.490
S_3	1.292	1.155	0.812	1.066	0.839	0.987	2.095	3.136	4.329	5.704	3.141
S_4	0.751	1.292	0.784	0.161	1.023	1.364	2.075	3.127	4.393	5.877	0.349

errors show a roughly sinusoidal pattern, with zero-crossings that correspond to the center of the categories. The existence of zero-crossings does not mean there were no decision errors at these points; rather, they indicate that at the category centers clockwise and anticlockwise errors were equally likely and roughly equal in magnitude and so tended to cancel each other out. The parts of the plot in which the errors are mainly positive are regions of the color space in which the errors were predominantly in an anticlockwise direction and parts where they are negative are regions in which the errors were predominantly clockwise. The sinusoidal pattern of errors around the category centers is a reflection of the vector summation of drift vectors and bias vectors in Figure 12B, which biases the decision in the direction of the nearest neighboring category.

The median RTs also show a roughly sinusoidal pattern, with troughs at the category centers and peaks located between them. This pattern is again a reflection of the vector summation of the drift norm and bias vectors. As discussed above, when the stimulus vector aligns with a category center the vectors will be collinear and short RTs will be found. Overall then, the category model, while it does not capture all of the fine detail of the pattern of performance in Figure 13, captures much of its essential structure. The correspondence between data and model is particularly good for S_1 , intermediate for S_2 and S_3 , and comparatively poor for S_4 . This ordering corresponds to the visual appearance of the category structures in Figure 12. Participant S_1 has a very sharp and stable category structure; S_2 and S_3 have somewhat less well-defined structures, and S_4 has a rather diffuse structure. Within the limits of resolution of the data, the biased drift rate model appears to provide a plausible account of the effects of color category on decision outcomes and RT.

Discussion

In the 19th century, Donders (1869) had what would prove to be one of the deepest and most enduring insights in psychology, namely, that decisions, even simple decisions, take time, and that the systematic study of decision time can provide important insights into the processes involved in making them. Donders' insight was neglected for 100 years until it was rediscovered and reworked during the cognitive revolution of the 1960s by Sternberg (1969). The development of sequential sampling decision models in the 1960s and 1970s by researchers such as Audley and Pike (1965); Laming (1968); Link and Heath (1975); Ratcliff

(1978), and Vickers (1979) married Donders' insight with the other great insight of 19th century psychology, that of Fechner (1860), who realized that simple decisions are inherently probabilistic in nature and that their probabilistic character can be viewed as the signature of the underlying psychological processes. The key insight of the sequential sampling decision literature was that decision times and decision outcomes are not separate and dissociable aspects of decision making that can be studied in isolation of each other but are, rather, closely coupled expressions of a single process. The success of models like the diffusion model is largely attributable to this insight.

In contrast, the recent surge of interest in continuous outcome decision tasks, although it too is built on the foundations of classical psychophysics, has largely neglected the dimension of time. The epigraph for this article from John Godfrey Saxe's *The Blind Men and the Elephant* (Saxe, 1881), is intended to highlight the dangers of this neglect. Concepts such as "precision" and "variable precision" that have dominated theorizing about the continuous outcome decision task in recent applications in the visual working memory literature have been developed and tested solely in relation to the distribution of decision outcomes and without reference to decision times. The term *precision*, in particular, is often used to refer simultaneously to the dispersion of the distribution of outcomes and the cognitive processes that give rise to this distribution with no clearly articulated theoretical account of how one translates into the other. Often the two are treated, either tacitly or explicitly, as isomorphic: The cognitive precursor of the distribution of decision outcomes is assumed to be a neural "population code," whose statistical properties mirror those of that distribution more or less exactly. The probability of a given response on a given trial is assumed to be the probability that the corresponding element of the population code is maximally active at that moment. This identification is convenient because it appears to obviate the need for a formal model of the decision process and licenses the view that the distribution of decision outcomes is a more or less direct readout of the population code.

Such ideas are neither unreasonable nor implausible, but they are inherently limited in their explanatory scope and, if interpreted too literally, come with risks. The limitations, as John Godfrey Saxe astutely reminded us, are that they tell us nothing about decision times (Saxe, 1881). The risks are subtler, but they boil down to the uncomfortable possibility that a "neural population code," as a cognitive explanation of behavior, often may be noth-

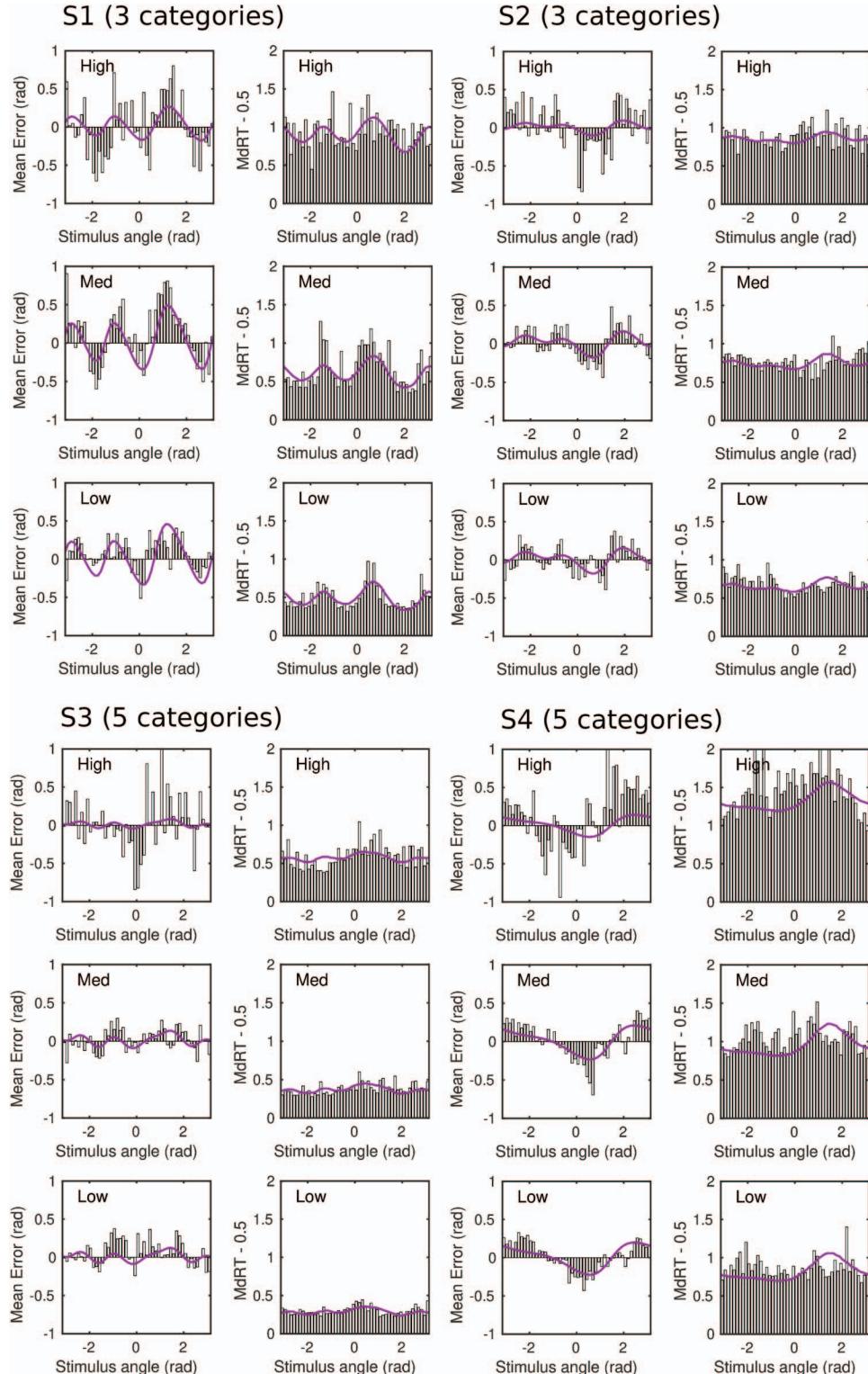


Figure 13. Predictions of the color category model. For each participant, the left-hand panels are the signed angular error of the response as a function of the position of the stimulus on the color wheel. The right-hand panels are the median RT. The histograms are the data and the continuous curves are the predictions the encoding failure model of Table 2 augmented with the category model of Figure 12 using the estimated parameters of Table 5. See the online article for the color version of this figure.

ing more than a redescription of the data in other language. In its malleability the idea of a population code resembles an earlier, similarly malleable, concept—namely, “capacity,” which was proposed during the 1970s as an explanation of attentional limitations in dual tasks (Kahneman, 1973). Capacity was conceived by its proponents as a global cognitive resource that sets limits on the brain’s ability to activate neural structure and to do cognitive work. Initially, capacity theory was seen as a promising alternative to the prevailing bottleneck theories of attention, which could account for the flexible patterns of performance found in dual-task studies. But within a decade the promise had faded (Navon, 1984) as researchers came to realize that a capacity-theory explanation could be constructed post hoc for any pattern of data whatsoever, making the theory unfalsifiable. The neural population code, as a presumed precursor of the distribution of decision outcomes, strikes us as being similarly malleable, and susceptible to the same philosophical objections of unfalsifiability as capacity.

To anticipate likely objections to this analogy, we emphasize that there are ways in which capacity and population coding can both be made precise. Capacity can be made precise if it is interpreted in one of a number of well-defined, quantitative ways, such as the workload capacity coefficient (Townsend & Nozawa, 1995; Townsend & Wenger, 2004) or the sample-size model of visual working memory (Palmer, 1990; Sewell, Lilburn, & Smith, 2014; Smith, Corbett, Lilburn, & Kyllingsbæk, 2018). Population code models can similarly be made quantitatively precise, as in the case of the normalization model of Bays (2014). The normalization model seeks to provide a neural foundation for the empirical model of Bays and colleagues (Bays et al., 2011), which characterized visual working memory using a mixture of von Mises components but provided no theoretical account of why retrieval should be distributed in this way. Our objection to the idea of a neural population code is, rather, when it is invoked post hoc as a piece of putative neural reductionism in ways that are neither quantitatively precise nor empirically falsifiable. Moreover, even a well-specified neural model like that of Bays is silent on the question of RT. Smith (2015) derived a neurally-inspired normalization model of visual working memory capacity, which does predict distributions of RTs, but, in its current form, only in two-choice tasks.

Precision as an Expression of Evidence Accumulation

Here we have taken a very different approach to the continuous outcome decision task to the one described above and the one prevailing in the visual working memory literature. We do not treat the von Mises distribution as an irreducible explanatory construct in the manner of previous researchers. Instead, we claim that the von Mises distribution is an expression of evidence accumulation by a diffusion process. Our theory embodies a transparent and interpretable set of model semantics for how this occurs. Unlike population code models, which need to postulate distributed stimulus representations to account for the dispersion of decision outcomes, the circular diffusion model predicts the dispersion as an expression of the process of evidence accumulation. The stimulus representation on any trial, conceptualized as a vector-valued quantity with a magnitude and a direction, predicts a von Mises distribution of decision outcomes. Repeated presentation of exactly the same stimulus, represented by exactly the same drift rate vector, will yield a von Mises distribution of outcomes across

trials. A corollary of our approach is that “precision” ceases to be an explanatory construct in the theory and instead becomes a dependent variable, along with RT. In the circular diffusion model, the key constructs that account for precision and RT are drift rate, decision criterion, and decision noise. Precision is jointly a function of these three quantities, via Equation 2.

Framed in these terms, the theoretical question then becomes whether the joint distributions of decision outcomes and decision times reflect additional sources of variability beyond what is explicable by the evidence accumulation process itself. In the visual working memory literature, this extra-von Mises variability has been characterized as variable precision by van den Berg et al. (2014). In the circular diffusion model, additional variability of this kind comes from across-trial variability in drift rates, which characterizes trial-to-trial variation in the encoded stimulus representations. Either the phase angle or the norm of the drift rate can vary, or both, and the empirical question is then which of these alternatives provides a better account of performance. Crucially, variability in phase angle affects precision but has no effect on RT; variability in norm affects them both.

An intriguing finding from our study was that, at high levels of chromatic noise, the distributions of decision outcomes all had heavy tails, similar to those found in visual working memory studies. In visual working memory, these kinds of distributions have been associated mainly with high memory loads and have been attributed both to guessing (Zhang & Luck, 2008) and to variable precision memory representations (van den Berg et al., 2014). The novelty of our finding is that we obtained these kinds of distributions for decisions about a single, centrally presented stimulus. We conjectured that such distributions likely reflected encoding failures under high noise conditions and proposed two alternative mechanisms that could give rise to them. The additional theoretical challenge for us was that any proposed mechanism had to account for the entire joint distribution of decision outcomes and RTs, not just one margin of it.

The circular diffusion model with bivariate-normally distributed drift rates can predict the kinds of heavy-tailed distributions that are found under high noise conditions, but overpredicts the tail quantiles of the RT distributions associated with low-accuracy responses. All of our participants showed a continuous form of the slow-error property, manifested graphically as a dish-shaped $Q \times Q$ plot. The model predicts this pattern, but if the drift rate variability is sufficient to account for the tails of the accuracy distribution, then it overpredicts the tail quantiles. The implication of this finding is that there is across-trial variability in the encoded stimuli that cannot be captured by the bivariate normal model.

To circumvent the limitations of that model, we considered two alternative models of across-trial variability, one loosely inspired by Zhang and Luck’s (2008) memory-plus-guessing model and the other inspired by van den Berg et al.’s (2014) variable precision model. “Guessing,” as a cognitive process, makes no predictions about RT, but a natural way to represent guessing in a diffusion model is to assume that the drift rate is zero and the process is driven solely by noise. When the drift rate is zero, the circular diffusion model predicts a uniform distribution of decision outcomes, which is simply a von Mises distribution with zero precision. Such a mechanism is unable to account for our data, because it predicts very slow errors and exacerbates the overprediction of tail quantiles associated with the bivariate normal model. How-

ever, a related model, which we called an *encoding failure model*, predicted both the shapes of the distributions of decision outcomes and the tail quantiles of the RT distributions. The encoding failure model assumes that, on a proportion of trials, the phase angle of the drift rate is random but the drift norm is the same as on other trials. As we characterized this model earlier, it assumes that something is encoded on every trial but on a proportion of trials the encoded something is very inaccurate. We attributed these inaccurate encodings to the perceptual properties of chromatic noise.

Our second model assumed that the phase angle of the drift rate varies randomly across trials and follows a continuous, nonnormal distribution. We used the Jones-Pewsey distribution for this purpose because it is defined on a circular domain and characterizes a variety of distribution forms, including distributions with heavy tails. This model also provided a successful account of performance in our task. Although the Jones-Pewsey representation embodies different model semantics to the encoding failure model—one is based on a continuum and the other is based on a two-component mixture—the picture that emerges from the best-fitting variable phase angle model was essentially the same as the one emerging from the encoding failure model. As shown in Figure 8B, the estimated parameters of the Jones-Pewsey distribution for the high-noise condition imply that on the majority of trials there is comparatively little variability in the encoded stimulus identity, except on the small proportion of trials on which the encoded identity is very inaccurate. This is essentially the picture provided by the encoding failure model but expressed in a different formalism.

Both the encoding failure model and the variable phase angle model attribute the heavy-tailed distributions of decision outcomes, the covariation of outcomes and RT, and the color category effects to across-trial variability in drift rates. The construct of drift-rate variability has been a source of some controversy in relation to the two-choice diffusion model, because of the concern that any relaxation of the constraints on the drift-rate distribution may open the door to an unfalsifiable model (Jones & Dzhafarov, 2014; Smith, Ratcliff, & McKoon, 2014). In response, Ratcliff, Voskuilen, and McKoon (2018) carried out a study using a double-pass technique that was able to identify a component of stimulus-specific variability across trials and to show it had a substantial effect on performance. Their results show that across-trial variability in drift rates is both theoretically principled and needed to account for data.

As we noted earlier, our results bear comparison with those of Shen and Ma (2019) who looked for evidence of variable precision in perceptual tasks in which there was no significant memory component. They found that once they had controlled for decision noise and guessing, there was little evidence for variable precision, that is, for across-trial variability in encoded stimulus identity. With two caveats, their conclusions are in agreement with our own. One is that “guessing” as a construct provides no account of RT. In contrast, our encoding failure mechanism predicts zero-precision decision outcomes and plausible distributions of RT, so we prefer to conceive of low-accuracy responding in these terms rather than as guessing in its classical sense. The second caveat is that variable precision in the circular diffusion model can arise either because of variability in the phase angle or the norm of the drift rate. Although we found little evidence of the former, there

was consistent evidence of the latter, manifested as a slow-error pattern of RTs.

Our results also bear comparison with those of van den Berg et al.’s (2014) earlier study of visual working memory, which investigated the construct of variable precision in a systematic way. One puzzling feature of their results is that the estimated parameters of their best-fitting model (their Figure 10) imply there are items in memory that are represented there with zero precision. Put colloquially, they seem to imply that I remember it, but I have absolutely no information about its properties. If viewed as a memory phenomenon this seems paradoxical, but if it is viewed as an encoding phenomenon then it is much less so. Our data suggest that these kinds of failures occur on a proportion of trials with high noise stimuli and our encoding failure model provides a mechanism for how decisions about such stimuli are made. As we have emphasized, the encoding failure model differs from a classical guessing model, like the one proposed by Zhang and Luck (2008), in which guessing occurs when an item is not in memory. The encoding failure model assumes that a durable, noisy representation of a perceptual object is formed, but the representation is too degraded to carry useful information.

The second substantive question we investigated concerned the effect of color categories on continuous outcome decisions. Like other investigators who have studied continuous outcome hue judgments, we found that responding in our task was strongly, if idiosyncratically, categorical. A significant challenge for us was how to represent these effects in a consistent way within the conceptual and mathematical structure of the model. Based on earlier work using the 1D diffusion model, we proposed that these kinds of biases can be viewed as stimulus biases. We proposed that, rather than encoding stimuli purely in relation to their metric properties—that is, to their position on the continuum—they are encoded, at least in part, relative to a set of reference categories. These categories induce drift rate biases, which bias decisions toward the nearest category center and make decisions to stimuli aligned with category centers faster and more accurate. The mathematical machinery we used to represent these biases was borrowed from similarity-choice theory, which has provided a successful framework for representing psychological similarity in other domains. This representation successfully characterized much of the stimulus-specific variability in RT and accuracy shown in Figure 13. That it did so is consistent with our earlier stated belief that many of the biases found in difficult decision tasks in which accuracy is stressed are stimulus biases rather than response biases.

We do not make the stronger claim that response bias plays no role. Ratcliff (2018) showed that, qualitatively at least, his spatially continuous diffusion model could predict faster, higher precision responses for stimuli aligned with the primary, additive colors if he assumed a sinusoidal decision criterion with troughs at the category centers. This idea could be carried over to the circular diffusion model, at least in principle, by replacing the circular decision boundary with some more complex smooth closed curve like a radial-frequency function (Wilkinson, Wilson, & Habak, 1998), whose troughs align with the categories. This assumption would destroy the analytic simplicity of the model, on which much of its appeal depends, but it could still be investigated by simulation. We have not pursued this line of inquiry, but it is useful to

know that response bias effects of this kind could be incorporated into the model, if desired.

Implications for Visual Working Memory Research

Our work has at least three general implications for the study of visual working memory and for the mathematical modeling of retrieval processes. The first is that it emphasizes the importance of understanding the continuous-outcome task as a decision task, in which decisions take a measurable amount of time to make. The second is that, if the continuous-outcome task is viewed as a decision task, then there is an essential unity in the picture of the underlying psychological processes it provides and the one obtained from more traditional approaches based on signal detection theory, or its close relative, the Luce choice model. A third is that peaked, heavy-tailed distributions of decision outcomes, which have been theoretically important in the recent visual working memory literature, may be produced by quite different kinds of mechanism, and these may be difficult or impossible to distinguish in empirical data, even when the constraints provided by distributions of RT are taken into account. We briefly discuss these points in turn.

From a diffusion model perspective, the key to understanding the continuous outcome task as a decision task is the expression for precision, Equation 2. This equation states that precision depends jointly on the quality of evidence in the stimulus, the amount of evidence needed for a response, and the noisiness of the evidence accumulation process. In a typical visual working memory experiment, in which precision is measured as a function of the number of items in the display, we would normally expect that the quality of the evidence entering the decision process, $\|\mu\|$, would vary with the number of items in memory. However, the multiplicative form of the expression for precision, $\kappa = a\|\mu\|/\sigma^2$, means that evidence quality cannot be inferred independently of the decision criterion, a . Although it may be reasonable to expect that decision criterion will not vary with display size, and hence will merely act as a scale factor for the measured precision, it nevertheless remains an assumption. Assumptions of this kind are referred to as *selective influence* assumptions in the modeling literature (Jones & Dzhabarov, 2014; Sternberg, 1969). Importantly, it is one that is amenable to empirical test, but only when RT data used; it cannot be tested from decision outcomes alone. And unless it is actually tested, the precision of the memory representation cannot be inferred from the precision of the behavioral response with any certainty.

The second implication concerns the equivalence of the pictures provided by continuous-outcome tasks and by more traditional two-alternative forced-choice or change-detection tasks—the latter typically analyzed using the Luce choice model or signal detection theory. As noted in the introduction, Link (1975) derived a sensitivity measure for a random-walk model, which is calculated from the choice probabilities in a two-choice task. The theoretical form of Link's measure, which is identical to the sensitivity measure for a Luce choice theory detection model (Luce, 1963, p. 123; McNicol, 1972, p. 139), parallels the precision measure of Equation 2. For a symmetrical random walk with normally distributed increments with means $\pm\mu/2$, standard deviation σ , and decision criteria at $\pm a$, Link's measure is equal to

$$d'_{\text{random walk}} = \frac{a\mu}{\sigma^2}. \quad (14)$$

Further details can be found in Smith (2015, 2016). Equation 14 is a 1D counterpart of Equation 2 and its components have exactly the same meaning as the components of that equation. Because the first-passage time distributions for a random walk with normally distributed increments are mathematically identical to those for a Wiener diffusion process (Smith, 1990), Equation 14 holds in this more general setting as well. It can therefore be thought of as a two-choice, diffusion model based, measure of precision.

The implication of the equivalence in Equations 2 and 14 is that d' measures derived from two-choice tasks carry just as much information about the quality of the evidence entering the decision process as do estimates of precision derived from continuous-outcome tasks. The equivalence is explicit in Equation 14, which reflects a further equivalence in the sensitivity measures derived from random walk and choice theory decision models. The choice theory detection model is based on logistic rather than normal distributions of perceptual effect (McNicol, 1972), and, like the random walk/diffusion model, predicts psychometric functions that are cumulative logistic rather than cumulative normal in form (Link, 1978). But in view of the virtual indistinguishability of the normal and logistic signal detection models for two-choice decisions (Ingleby, 1973), the equivalence can be expected to hold for the standard d' measure as well.

The third point, which we touched on in the previous section, is that very different psychological processes can predict very similar distributions of decision outcomes. Significant effort has been expended over the last 15 years on the question of whether visual working memory has an item-capacity limit. Following Zhang and Luck (2008), the heavy-tailed distributions of decision outcomes found with large memory loads have been interpreted as evidence of such limits, because item-capacity models assume that people are forced to guess when items are not in memory and this becomes increasingly more probable as display size increases. Zhang and Luck's slots-plus-averaging model is a two-component mixture model, whereas competitor models, like those of Bays and colleagues (Bays et al., 2011), Oberauer and Lin (2017), and van den Berg et al. (2014) assume more complex finite or continuous mixtures of components. We found the same kinds of heavy-tailed distributions of decision outcomes as are found with large memory loads with single items presented at low signal-to-noise ratios. We also found that two very different kinds of encoding model, one based on a two-component mixture and one based on a continuous, heavy-tailed distribution, provided similarly good descriptions of these data. This was so even when the additional constraints provided by distributions of RT were taken into account. The degree of model mimicry implied by these fits leaves us pessimistic about whether the item-capacity question can be settled by reference to empirical data in any decisive way. Decision outcomes in item-capacity limited models are predicted to follow a two-component mixture distribution and we have shown that such distributions can be mimicked by a diffusion model with a suitable continuous distribution of drift rates.

Conclusion

Since it was introduced by Prinzmetal et al. (1998) and taken up by Wilken and Ma (2004) to study visual working memory, there

has been a rapid growth of interest in the continuous outcome task and a number of sophisticated memory models have been developed on its foundation. Although this work has yielded valuable insights, a negative consequence of it is that it has become progressively more divorced from theories of decision making in other parts of psychology, particularly from the tradition that holds that the hallmark of a good model is its ability to account for both accuracy and RT. The circular diffusion model, which provides an analytically tractable account of continuous outcome decisions and the associated RTs, seeks to bridge this gap. Smith (2016), showed that the properties of the model, when augmented with across-trial variability, closely align with those of the two-choice diffusion model, which has successfully characterized decision making in a wide variety of domains.

Our aim in this article was to extend this work theoretically and empirically to the widely studied domain of hue judgments. Although our larger interest is in visual working memory, we thought it was important to investigate the model purely as a model of the decision process, using the same kinds of psychophysical stimuli that have been used to characterize decision processes in other settings. The new theory needed to account for performance in this task is essentially a theory of drift-rate variability and consists of two components. One component is a theory of across-trial variability in encoded stimulus identity, focusing in particular on performance under high-noise conditions. **The other component is a theory of categorical responding, which we developed using a drift-bias model based on similarity-choice theory.** Together, these two aspects of the theory of drift rate provide a detailed characterization of performance on the task. The overall message from our work is very similar to the one from many previous studies of two-choice decisions. It is that variability in performance is jointly a function of variability in the stimulus representation and variability in the decision process that acts on the representation. Our study shows that these principles carry across seamlessly from the two-choice to the continuous outcome setting.

References

- Abramowitz, M., & Stegun, I. (1965). *Handbook of mathematical functions*. New York, NY: Dover.
- Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, 97, 79–97.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC*, 19, 716–723.
- Ashby, F. G. (1983). A biased random walk for two choice reaction times. *Journal of Mathematical Psychology*, 27, 277–297.
- 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.
- Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144, 744–763.
- Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *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(10), 7. <http://dx.doi.org/10.1167/9.10.7>
- Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, 49, 1622–1631.
- Borodin, A. N., & Salminen, P. (1996). *Handbook of Brownian motion—facts and formulae*. Basel, Switzerland: Birkhäuser.
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, 5, 1–11. <http://dx.doi.org/10.1038/srep10415>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference* (2nd ed.). New York, NY: Springer.
- 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.
- Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). New York, NY: Chapman and Hall.
- Donders, F. C. (1869). Over de snelheid van psychische processen [On the speed of mental processes]. Translated by W. G. Koster. *Acta Psychologica*, 30, 412–431.
- Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, 120, 873–902.
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig, Germany: Breitkopf & Härtel.
- Gradshteyn, I. S., & Ryzhik, I. M. (2007). *Tables of integrals, series, and products* (7th ed.). Burlington, MA: Academic Press.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- 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.
- Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, 43, 30–54.
- Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review*, 20, 318–325.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. London, UK: Chapman and Hall.
- Ingleby, J. (1973). The separation of bias and sensitivity in multiple choice tasks. *Perception*, 2, 295–305.
- Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121, 1–32.
- Jones, M. C., & Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100, 1422–1428.
- 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.
- Kent, J. T. (1978). Some probabilistic properties of Bessel functions. *Annals of Probability*, 6, 760–770.
- Kvam, P. D. (2019). Modeling accuracy, response time, and bias in continuous outcome orientation judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 301–318.
- Laming, D. (1968). *Information theory of choice reaction times*. New York, NY: Academic Press.
- Link, S. W. (1975). The relative judgment theory of two-choice response time. *Journal of Mathematical Psychology*, 12, 114–135.
- Link, S. W. (1978). The relative judgment theory of the psychometric

- function. In J. Requin (Ed.), *Attention & performance VII* (pp. 619–630). Hillsdale, NJ: Erlbaum.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77–105.
- Little, D. R., & Smith, P. L. (2018). Replication is already mainstream: Lessons from small-*N* designs. *Behavioral and Brain Sciences*, 41, e141.
- Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York, NY: Wiley.
- Maloney, L. T., & Yang, J. (2003). Maximum likelihood difference scaling. *Journal of Vision*, 3(8), 573–585.
- Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics*. New York, NY: Wiley.
- Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, 13(2), 21. <http://dx.doi.org/10.1167/13.2.21>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, UK: Chapman & Hall.
- McNicol, D. (1972). *A primer of signal detection theory*. London, UK: George Allen & Unwin.
- Miller, J., & Schwarz, W. (2018). Implications of individual differences in on-average null results. *Journal of Experimental Psychology: General*, 147, 377–397.
- Navarro, D. J. (2007). Similarity, distance, and categorization: A discussion of Smith's (2006) warning about "colliding parameters." *Psychonomic Bulletin & Review*, 14, 823–833.
- Navon, D. (1984). Resources—a theoretical soup stone? *Psychological Review*, 91, 216–234.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–313.
- Nosofsky, R. M. (1984). Choice, similarity and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104–114.
- Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, 124, 21–59.
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 332–350.
- Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, 88, 1–21.
- Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. *Computational Statistics & Data Analysis*, 52, 1516–1523.
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 261–282.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, 92, 212–225.
- Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review*, 125, 888–935.
- Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, 2, 237–279.
- 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., & McKoon, G. (2015). Modeling response time and accuracy data. *Current Directions in Psychological Science*, 24, 458–470.
- Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, 120, 697–719.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Ratcliff, R., Voskuilen, C., & McKoon, G. (2018). Internal and external sources of variability in perceptual decision-making. *Psychological Review*, 125, 33–46.
- Rogers, L. C. G., & Williams, D. (2000). *Diffusions, Markov processes and martingales: Vol. 2. Itô calculus*. New York, NY: Cambridge University Press.
- Saxe, J. G. (1881). *The poems of John Godfrey Saxe*. Boston, MA: Houghton, Mifflin and Company.
- Schnegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, 83, 181–193.
- Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). Psychological scaling reveals and single parameter framework for visual working memory. *bioRxiv*. Advance online publication. <http://dx.doi.org/10.1101/325472>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 10749–10753.
- 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.
- Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of Experimental Psychology: Learning Memory & Cognition*, 42, 1673–1693.
- Shen, S., & Ma, W. J. (2016). A detailed comparison of optimality and simplicity in perceptual decision making. *Psychological Review*, 123, 452–480.
- Shen, S., & Ma, W. J. (2019). Variable precision in visual perception. *Psychological Review*, 126, 89–132.
- 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. (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. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123, 425–451.
- Smith, P. L. (2019). Linking the diffusion model and general recognition theory: Circular diffusion with bivariate-normally distributed drift rates. *Journal of Mathematical Psychology*, 91, 145–168.
- Smith, P. L., & Corbett, E. A. (2019). Speeded multielement decision making as diffusion in a hypersphere: Theory and application to double-target detection. *Psychonomic Bulletin & Review*, 26, 127–162.
- Smith, P. L., Corbett, E. A., Lilburn, S. D., & Kyllingsbæk, S. (2018). The power law of visual working memory characterizes attention engagement. *Psychological Review*, 125, 435–451.

- Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defence of the small-*N* design. *Psychonomic Bulletin & Review*, 25, 2083–2101.
- 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., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychological Review*, 121, 679–688.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, 44, 1297–1320.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, 39, 321–360.
- Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, 111, 1003–1035.
- Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instrumentation, & Computers*, 36, 702–716.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108, 550–592.
- van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, 121, 124–149.
- Verdonck, S., & Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review*, 123, 208–218.
- Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.
- Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin & Review*, 26, 813–832.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 1120–1135. <http://dx.doi.org/10.1167/4.12.11>
- Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, 38, 3555–3568.
- Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology*. London, UK: Methuen.
- Wyszekci, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York, NY: Wiley.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235.

Received March 30, 2019

Revision received August 14, 2019

Accepted December 1, 2019 ■