#### THEORETICAL REVIEW



# Speeded multielement decision-making as diffusion in a hypersphere: Theory and application to double-target detection

Philip L. Smith<sup>1</sup> · Elaine A. Corbett<sup>1,2</sup>

Published online: 2 July 2018 © Psychonomic Society, Inc. 2018

#### **Abstract**

We generalize the circular 2D diffusion model of Smith (*Psychological Review*, *123*, 425–451: 2016) to provide a new model of speeded decision-making in multielement visual displays. We model decision-making in tasks with multielement displays as evidence accumulation by a vector-valued diffusion process in a hypersphere, whose radius represents the decision criterion for the task. We show that the methods used to derive response time and accuracy predictions for the 2D model can be applied, with only minor changes, to predict performance in higher-dimensional spaces as well. We apply the model to the double-target deficit paradigm of Duncan (*Psychological Review*, *87*, 272–300: 1980) in which participants judge whether briefly presented four-element displays contain one- or two-digit targets among letter distractors. A 4D version of the hyperspherical diffusion model correctly predicted distributions of response times and response accuracy as a function of task difficulty in single-target and double-target versions of the task. The estimated drift rate parameters from the model imply that the mental representation of the decision alternatives, which we term the "decision template" for the task, encodes configural stimulus properties that reflect the number of targets in the display. Along with its application to multielement decision-making, the model has the potential to characterize the speed and accuracy of multiattribute decisions in studies of cognitive categorization, visual attention, and other areas.

Keywords Decision-making · Diffusion model · Response times · Double-target deficit · Categorization

A long-standing question in cognitive psychology is how people make decisions about stimulus arrays composed of multiple elements. Such decisions are central to many studies of cued visual attention (Dosher & Lu, 2000), near-threshold visual search (Eckstein et al., 2000), visual working memory (Vogel et al., 2006), and are important in other areas as well. In statistical decision terms, an array composed of multiple elements can be thought of as a single multidimensional, multiattribute stimulus, and the question then arises as to how people combine information across the elements, dimensions, or attributes

✓ Philip L. Smith philipls@unimelb.edu.au
 Elaine A. Corbett elainecorbett@gmail.com

- Melbourne School of Psychological Sciences, The University of Melbourne, Victoria, 3010, Australia
- Trinity College Institute of Neuroscience, Trinity College, Dublin, Ireland

of such a stimulus to make a decision. Decision accuracy in brief, near-threshold visual displays has been modeled successfully using multichannel generalizations of signal detection theory (see Palmer et al. (2000), for a review), but this work leaves open the question of whether the decision model can be extended to account for response time RT as well. There have been some recent successes in modeling RT in suprathreshold visual search tasks using sequential-sampling decision models (Moran et al., 2013; Schwarz & Miller, 2016), but the models involve sequential attention shifts, so they are not just models of the decision process and involve the additional complexity of distinguishing between parallel and serial processes (Dosher et al., 2010; Townsend & Nozawa, 1995; Thornton & Gilden, 2007).

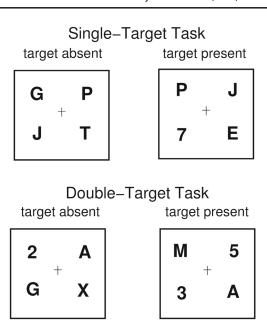
In this article, we describe a new approach to modeling these kinds of decisions that was motivated by the perspective above—that a multielement stimulus array can be thought of as a single multidimensional stimulus whose dimensionality equals the number of elements in the array. To this end, we generalize the circular diffusion model of Smith (2016)—in a way foreshadowed in that



article—and represent the decision process as a diffusion process within a hypersphere imbedded in an n-dimensional evidence accumulation space, whose dimensionality equals the number of elements in the display. The circular diffusion model can itself be thought of as a generalization of the diffusion decision model of Ratcliff (1978), which represents two-choice decision-making as diffusion on a line (i.e., a one-dimensional evidence space). Smith introduced the model to try to account for RTs and decision outcomes in continuous-report tasks, which are widely used in studies of visual working memory (van den Berg et al., 2014; Zhang & Luck, 2008). In continuous report tasks, people make decisions about continuously distributed stimuli that are defined in a two-dimensional (2D) feature space (Prinzmetal et al., 1998), and express their decisions using an analog device like a mouse or a trackball. Although Smith's model was developed to account for decisions that have a continuous range of outcomes (i.e., an infinite choice set), the model can account for categorical decisions as well, by partitioning the outcome space into discrete categories with decision bounds. When equipped with decision bounds in this way, the model can be viewed as a stochastic form of Ashby and Townsend's (1986) general recognition theory. It is this form of the circular diffusion model that serves as the basis for the models we consider in this article. We do not attempt to elaborate the relationship between the circular diffusion model and general recognition theory here, but the interested reader should be aware of the theoretical links, both with other forms of stochastic general recognition theory (Ashby, 1989, 2000; Townsend et al., 2012) and with work that has extended general recognition theory into more complex spaces (Townsend et al., 2001, 2006).

#### The double-target detection task

The hyperspherical diffusion model is a general theoretical framework with potential applications in a variety of cognitive settings, several of which we consider subsequently. However, our immediate reason for developing the model was to try to characterize decision-making in tasks like the one shown in Fig. 1. The figure shows stimuli from a recent study of the double-target deficit by Corbett and Smith (2017). The task, which used a paradigm devised by Duncan (1980), required participants to detect digit targets among letter distractors in four-element arrays. The actual stimuli were rendered in a simplified stroke font to minimize orthographic cues that could distinguish between digits and letters and were presented in noise for 200 ms and backwardly masked. The brief exposure duration precluded gaze refixations (Hallett, 1986) and minimized the possibility of any kind of serial search. The digits were drawn from the set 2-9, the letters from the set A, E, G, J, M, P, T, X. In



**Fig. 1** Double-target decision task used by Corbett and Smith (2017, Experiment 4a). In the single-target task, the target was a single digit in an array of letters. On target-absent trials, the array contained four letters. In the double-target task, the target was a pair of digits in an array of letters. On target-absent trials, the array contained three letters and one digit

the single-target task, there was either a single digit or no digit and participants responded "target present" or "target absent," accordingly. In the double-target task, there were either two digits or one digit and participants responded "target present" or "target absent" depending on whether they detected two digits or fewer than two. The doubletarget deficit is the finding that if people are required to detect a pair of simultaneous (auditory or visual) targets, then their performance is appreciably worse than would be predicted from their performance when detecting either of the targets separately (Duncan, 1980, 1985; Moray, 1970a, b; Sorkin & Pohlmann, 1973; Sorkin et al., 1973, 1976). The double-target deficit has been interpreted as evidence for two-stage theories of search, in which preattentive filtering processes impose a selection bottleneck on the stimuli that enter the decision process (Duncan, 1980; Hoffman, 1979; Smith & Sewell, 2013; Wolfe et al., 1989, 2010). Our interest in this article is not in the double-target deficit per se but in the processes involved in making these kinds of decisions. We therefore wanted to focus on the decision processes in the two forms of the task, uncontaminated, as much as possible, by differences in the preattentive selection processes they engage. To that end, we have chosen to focus on the experiment from among the nine in Corbett and Smith's study that yielded the smallest double-target deficit. The reader is referred to the original article for further details.

One way in which one could go about developing a response time model for the task in Fig. 1 is to consider



sequential-sampling generalizations of the signal detection models that have previously been used to characterize accuracy in such tasks. The two main classes of signal detection model that have been influential in this area are independent detectors models and integration models (which includes likelihood ratio models) (Eckstein et al., 2000; Kinchala, 1974; Palmer et al., 1993, 2000; Shaw, 1980, 1982; Shaw et al., 1983; Shimozaki et al., 2003; Smith, 1998; Verghese, 2001). Independent detector models assume that separate decision mechanisms identify the stimulus at each location. The outputs of the mechanisms are then combined by Boolean operators to implement the appropriate decision rule. Integration models assume the contents of each array location are summed or pooled (using either a linear or a nonlinear combination rule) to create a single decision variable that expresses the contents of the array as a whole. 1 Corbett and Smith (2017) found that both an independent detectors model and an integration model (in likelihood ratio form) provided a satisfactory account of response accuracy in the nine experiments in their study.

In an investigation that complements the present one, Corbett and Smith (submitted) considered sequential-sampling generalizations of these two models in which the component decision processes are represented by diffusion processes (Ratcliff, 1978; Ratcliff & McKoon, 2008). They focused specifically on the single-target task and, in agreement with Corbett and Smith (2017), found that either of two models could provide a fairly good account of the distributions of response times and choice probabilities in four different decision tasks investigated by Corbett and Smith (2017). One model assumed that evidence about the contents of the four locations was pooled and accumulated by a single diffusion process. The other model assumed

<sup>1</sup>The distinction between independent detectors models and integration models is blurred by the existence of a third class of model, the maximum-of-outputs or MAX models (Palmer et al., 2000), which share some of the properties of both of the other two classes of model. MAX models assume that the decision variable is the maximum of a set of random variables that individually represent the strength of evidence at each display location. One way of obtaining the value of the maximum is as the limit of a Minkowski sum computed on the values of the random variables in the set. This value is compared to a criterion to make a decision. The MAX model can therefore be thought of as a form of integration model with a nonlinear combination rule. When the same decision criteria are used at every location, independent detectors models and MAX models make identical predictions because then the probability that any of the random variables exceeds its criterion is the same as the probability the *largest* of them exceeds the criterion. Despite the equivalence of their predictions, the two kinds of model make quite different assumptions about the representation of stimuli in the decision process: Independent detectors models assume that n independent random variables enter the decision process; MAXmodels assume there is only one. The usual psychological interpretation of MAX models is that the operation of taking-the-maximum is performed by preattentive filtering processes that select the most "target-like" stimulus in the display (Corbett & Smith, 2017).

that decisions about the contents of the four locations were made by four independent diffusion processes whose outputs are combined to implement the decision rule. We chose to investigate the hyperspherical diffusion model as an alternative decision model for these tasks both because of its general theoretical interest as a model for multiattribute decision-making and because, to date, we have not been able to find a model for response times in the double-target task using either the single or the multiple diffusion model that accounts for the data with a consistent and interpretable set of parameters.

## Decision-making as diffusion in an (n-1)-sphere in $\mathbb{R}^n$

Figure 2 shows the circular diffusion model of Smith (2016), its relationship to the diffusion model of Ratcliff (1978), and the way in which we generalize the circular model to higher dimensions in this article. The model conceptualizes decision-making as a process of noisy evidence accumulation in an n-dimensional Euclidean evidence space,  $\mathbb{R}^n$ , to a decision boundary or response criterion. The evidence accumulation process is represented by an *n*-dimensional (vector-valued) Wiener diffusion process,  $\mathbf{X}_t$ , with independent components,  $(X_t^1, X_t^2, \dots, X_t^n)$ , all with infinitesimal standard deviation,  $\sigma$ . In this notation, the subscript t denotes time and the superscripts 1, 2, ... index the coordinates of the process; they are *not* powers. In the application to the double-target detection task in Fig. 1, the evidence space is four-dimensional, reflecting the number of noisy elements in the stimulus space. The decision criterion or decision boundary, which represents the evidence needed for a response, is an (n-1)-dimensional hypersphere,  $S^{(n-1)}$ , of radius a, imbedded in the n-dimensional Euclidean evidence space,  $\mathbb{R}^n$ . In the circular model, the decision criterion is a circle, or 1-sphere,  $S^1$ . In the 3D generalization of the model shown in Fig. 2c, the decision criterion is an "ordinary" sphere,  $S^2$ , while in Ratcliff's model the decision criteria are an isolated pair of points, a and -a, which may be viewed as a 0-sphere,  $S^0$ . (This parameterization differs from that of Ratcliff who uses the "gambler's ruin" form of the diffusion process (Feller, 1967), in which the upper bound is set to a and the lower bound set to zero.)

Figure 2a shows the main elements of Ratcliff's (1978) diffusion model, which we generalize here. We have used the terminology of signal detection theory in the figure and denoted the two classes of stimuli as "signal" and "noise." The model assumes that the information in the stimulus is encoded in the *drift rate* of the diffusion process, which determines the average rate at which evidence accumulates. The sign of the drift rate depends on the identity of the



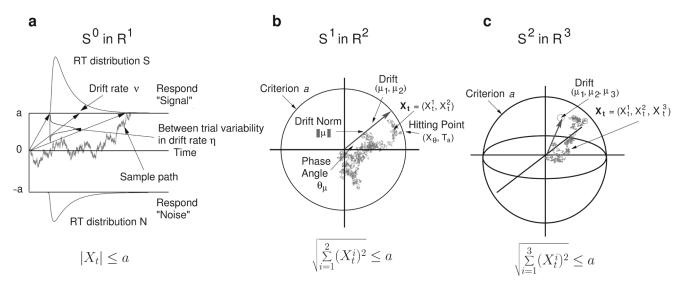


Fig. 2 Decision-making as evidence accumulation in an (n-1)-sphere in  $\mathbb{R}^n$ . a Decision-making as diffusion on a line (Ratcliff, 1978). Evidence accumulation is modeled as a one-dimensional diffusion process on the real line,  $\mathbb{R}^1$  (the vertical axis), between absorbing boundaries (decision criteria) at -a and a. The isolated points a and -a are a zero-sphere,  $S^0$ . The horizontal axis represents time. b Decision-making as diffusion in a 1-sphere,  $S^1$  (Smith, 2016). Evidence accumulation is modeled as a two-dimensional diffusion process on the interior of a disk of radius a, whose bounding circle represents the decision criterion. The process  $\mathbf{X}_t = (X_t^1, X_t^2)$  consists of

two independent components, both with infinitesimal standard deviation  $\sigma$ . The drift rate,  $\mu=(\mu_1,\mu_2)$ , is vector-valued, with length, or norm,  $\|\mu\|$  and phase angle  $\theta_{\mu}$ . The phase angle specifies the direction of the drift vector in polar coordinates. A decision is made when the process  $\mathbf{X}_t$  reaches the criterion. The decision time,  $T_a$ , is the time to reach the boundary and the response,  $X_{\theta}$ , is the hitting point on the boundary.  $\mathbf{c}$  Diffusion in a two-sphere,  $S^2$ , in  $\mathbb{R}^3$ . The process  $\mathbf{X}_t = (X_t^1, X_t^2, X_t^3)$  consists of three independent components, all with infinitesimal standard deviation  $\sigma$ . The drift rate is  $\mu = (\mu_1, \mu_2, \mu_3)$ 

stimulus (positive for signals and negative for noise) and the magnitude of the drift rate determines the quality of the encoded information on a given trial. Evidence accumulation begins at stimulus onset and continues until the process hits one of the two decision boundaries. The first boundary hit determines the response—"signal" at the upper boundary and "noise" at the lower boundary—and the time taken to hit the boundary determines the decision time component of RT. The time taken to first hit a boundary and the probability of hitting the upper boundary before the lower boundary depend on the sign and magnitude of the drift rate and the noise in the evidence accumulation process, which in turn depends on the infinitesimal standard deviation,  $\sigma$ . Because of noise in the accumulation process, on some proportion of trials the process may terminate at the wrong boundary, and an error will be made. As Fig. 2a depicts, the model can predict choice probabilities (response accuracy) and distributions of RTs for correct responses and errors. In addition to the elements shown in Fig. 2a, the standard form of the model contains components of processing that represent various kinds of decision bias, which we discuss later. The properties of the model have been described in many places, including Ratcliff and McKoon (2008), Ratcliff and Smith (2004, 2015), and Ratcliff et al. (2016), and Smith and Ratcliff (2015).

Figure 2b shows the two-dimensional (2D) generalization of the model of Smith (2016). Unlike the standard

model, the circular model assumes that stimuli are defined in a 2D evidence space. The dimensions represent features of a stimulus that can vary independently, such as color, orientation, or spatial frequency and phase, or any pair of psychologically relevant feature dimensions of a decision task. The drift rate is then vector valued,  $\mu$ , with components  $\mu_1$  and  $\mu_2$ , which represent the values of the two features. Although it is not conceptually necessary, as is discussed in the next section there are powerful analytic advantages to be gained by making a strong symmetry assumption and assuming that the evidence accumulation process starts at the center of the circle,  $S^1$ , with radius a.

As in the standard model, evidence accumulation begins at stimulus onset and continues until the process hits a point on the decision boundary. The time taken to reach the boundary determines the decision time and the point at which it hits the boundary determines the response. The latter is continuously valued on the range  $(0, 2\pi)$ . The decision time, which we denote by  $T_a$ , is the time for the Euclidean norm of the process,  $\sqrt{\sum_{i=1}^2 (X_t^i)^2}$ , to first equal a. The time  $T_a$  is referred to as the *first passage time* of the process through the boundary a. Because of the circular symmetry of the process it is more convenient to work in polar coordinates rather than Cartesian coordinates. In polar coordinates, the drift rate is defined by the *drift norm*,  $\|\mu\| = \sqrt{\mu_1^2 + \mu_2^2}$ , and *phase angle*,  $\theta_{\mu} = \arctan(\mu_2/\mu_1)$ .



Psychologically, the drift norm represents the quality of the evidence in the stimulus and the phase angle represents its identity (i.e., the direction in which drift vector points). Likewise, we denote the point at which the process exits the bounding circle, the *hitting point*, in polar coordinates as  $\mathbf{X}_{\theta}$ . If  $X_T^1$  and  $X_T^2$  denote the horizontal and vertical Cartesian coordinates of the process at the point at which it hits the boundary, then the phase angle of the hitting point is  $\cap \Theta_T = \arctan\left(X_T^2/X_T^1\right)$ .

As discussed by Smith (2016), all of the features of the standard diffusion model that have been found to be useful in fitting empirical data carry over to the 2D case. When there is no across-trial variability in drift rates or decision criteria, the model predicts that decision times will be independent of decision outcomes. When there is across-trial variability in drift rates, the model predicts an inverse relationship between RT and decision accuracy. When there is across-trial variability in decision criterion (the value of *a*), the model predicts a direct relationship between RT and decision accuracy. These relationships are continuous counterparts of the slow errors and fast errors that the standard diffusion model predicts when there is across-trial variability in drift rates and starting points, respectively (Ratcliff & McKoon, 2008).

The model also predicts families of RT distributions whose shapes remain invariant as drift rates or decision criteria change. Technically, it predicts RT distributions whose quantiles form affine families. Let  $Q_{i,k}$  denote the kth distribution quantile in the ith condition and let T denote the RT. Symbolically,  $P_i[T \leq Q_{i,k}] = p_k$ , that is,  $Q_{i,k}$  is the value of time that cuts off the fastest proportion  $p_k$  of the RTs in condition i. For any pair of conditions, i and j, the model predicts that  $Q_{i,k} = a_{i,j} Q_{j,k} + b_{i,j}$ , where  $a_{i,j}$ and  $b_{i,j}$  are constants that depend on the pair of conditions being compared. Put informally, the distribution quantiles form a set of straight lines. As we discuss in the section Treatment of Empirical Data, empirical RT distributions often exhibit this property to quite a high degree. In addition to these RT distribution properties, the model predicts that decision outcomes follow a von Mises distribution, which is a circular counterpart of the normal distribution (Fisher, 1993). Studies of visual working memory using continuousreport decision tasks have found that empirical decision outcomes are well described as mixtures of von Mises distributions (van den Berg et al., 2014; Zhang & Luck, 2008).

Figure 2c shows how the circular model can be generalized in a straightforward way to higher dimensions. For a 3D stimulus, the evidence accumulation process becomes three-dimensional, with components,  $\mathbf{X}_t = (X_t^1, X_t^2, X_t^3)$ , each with infinitesimal standard deviation  $\sigma$  and drift rate  $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ . The decision criterion then becomes an ordinary sphere,  $S^2$ , with radius a, in

the 3D evidence space  $\mathbb{R}^3$ . As is Fig. 2b, a decision is made when  $\sqrt{\sum_{i=1}^3 (X_t^i)^2} = a$ , that is, at the first time when the Euclidean norm of the evidence accumulation process equals the decision criterion. Although the process in higher dimensions is hard to visualize, the decision rule is simply defined. In four dimensions, which is the process that concerns us here, the accumulation process is  $\mathbf{X}_t = (X_t^1, X_t^2, X_t^3, X_t^4)$ , the drift rate is  $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)$ , and a decision is made when  $\sqrt{\sum_{i=1}^4 (X_t^i)^2} = a$ .

Smith (2016) proposed the circular diffusion model as a model of decisions in continuous report tasks, but it also provides a model for categorical decisions if the outcome space is partitioned into discrete categories with decision bounds. To see how this works, consider, first, a twoelement version of the digits-among-letters task of Fig. 1, in which the display contains a pair of characters, each of which can be a letter or a digit. Following the terminology of Fig. 2, we will refer to digits as "signals" and letters as "noise." Assume that the evidence encoded at each location varies (negative to positive) from strong evidence for a letter (noise) to strong evidence for a digit (signal). For a two-element display, there are four possible stimulus configurations,  $s_1s_2$ ,  $n_1s_2$ ,  $n_1n_2$ , and  $s_1n_2$ , where s and n denote signal and noise, and the subscripts denote the display locations, which we identify with the horizontal and vertical axes of the evidence space,  $\mathbb{R}^2$ , respectively. For specified stimulus discriminability conditions (exposure, contrast, etc.), the drift vectors associated with these four stimulus conditions will be  $(\mu, \mu), (-\mu, \mu), (-\mu, -\mu)$ , and  $(\mu, -\mu)$ , respectively. These coordinate pairs represent drift vectors with phase angles of  $\pi/4$ ,  $3\pi/4$ ,  $5\pi/4$ , and  $7\pi/4$ radians (i.e., oriented at  $45^{\circ}$ ,  $135^{\circ}$ ,  $225^{\circ}$ , and  $315^{\circ}$ ).

As shown in Fig. 3, an appropriate decision model for this task can be obtained by placing decision bounds that align with the cardinal axes, at 0,  $\pi/2$ ,  $\pi$ , and  $3\pi/2$  radians, which partition the continuous outcome space into four equalsize quadrants that define response regions, and which we denote, going anticlockwise from zero, as  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ . Any process terminating in one of the four quadrants is labeled with the associated response. These quadrants can be mapped via Boolean operators to the response rule for the task. For the single-target task of Fig. 1, quadrants  $R_1$ ,  $R_2$ , and  $R_4$  map to the "target present" response and quadrant  $R_3$  maps to the "target absent" response. For the double-target task, quadrants  $R_1$  maps to the "target present" response and quadrants  $R_2$ ,  $R_3$  and  $R_4$  map to the "target absent" response.

For the task in Fig. 1, which used four-element displays, each of which independently provides noisy evidence for the presence or absence of a digit, the natural way to conceptualize the decision process is as a 4D accumulation process in the evidence space  $\mathbb{R}^4$ . As in the preceding



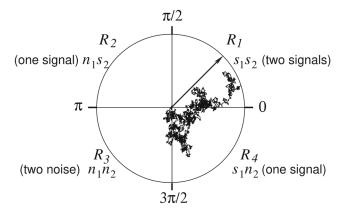
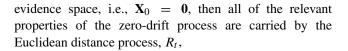


Fig. 3 Circular (2D) diffusion model with decision bounds. The decision bounds are aligned with the cardinal axes and partition the outcome space into four response regions,  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ , each of which is a quadrant of the outcome space. Processes finishing in the first quadrant lead to a "two signals" response; processes finishing in the third quadrant lead to a "two noise" response; and those finishing in the second and fourth quadrants lead to a "single signal" response. Processes that finish in the second quadrant are associated with detecting noise at location 1 and signal at location 2; those that finish in the fourth quadrant are associated with detecting a signal at location 1 and noise at location 2

paragraph, the response region, which is the surface of the hypersphere  $S^3$ , can be partitioned into response regions with decision bounds. If the decision bounds are aligned with the cardinal axes of  $S^3$ , then there are  $2^4 = 16$  such response regions, which are the *orthants* of the hypersphere. ("Orthants" are higher-dimensional analogs of quadrants. For an ordinary sphere, the orthants are the octants of the sphere and can be visualized as eighths of a ball. For the three-sphere, the orthants are mathematically well defined but difficult to visualize.) As in the 2D case, the orthants of the hypersphere map via Boolean operators to the response rule for the task. Before we describe these mappings, we outline the methods used to obtain response time and response outcome predictions for the model.

# The Girsanov theorem and the Bessel process

In order to apply the hyperspherical diffusion model to empirical data we need predicted distributions of decision outcomes and decision times. This apparently formidable task is simplified by means of a theorem in stochastic processes known as the Girsanov theorem (or Cameron–Martin–Girsanov theorem) (Karatzas & Shreve, 1991; Rogers & Williams, 2000). The Girsanov theorem allows us to relate the properties of a diffusion process with nonzero drift rate (which is the process of theoretical interest) to those of the simpler zero-drift process. If we assume that evidence accumulation always begins at the origin of the



$$R_t = \sqrt{\sum_{i=1}^4 (X_t^i)^2}.$$
 (1)

The Euclidean distance process  $R_t$  is known as the *Bessel process* (Borodin & Salminen, 1996, p. 297, Hamana & Matsumoto, 2013) and describes the distance of the 4D diffusion process,  $\mathbf{X}_t$ , from the origin at time t. The advantage of working with  $R_t$  rather than  $\mathbf{X}_t$  is that it allows us to reduce a problem in four dimensions to a problem in one dimension. Smith (2016) describes in detail how the combination of the Girsanov theorem and the Bessel process can be used to derive predictions for the circular diffusion model. Together, they allow us to obtain simple analytic expressions for all of the relevant properties of the model required to fit data. Here we confine ourselves to stating the key results needed to extend the model to the four-dimensional case. To do so, we need some notation.

We denote by  $dP_t(a)$  the probability density function of the Bessel process through the decision boundary a at time t. (We have chosen to use differential notation as the most convenient way to state these results). That is,

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

This relationship states that  $dP_t(a)$  is the time derivative of the *first passage time* probability of  $R_t$  through the decision boundary a. The condition on the right stipulates that  $R_t$  is less than a for all values of time less than t and is needed to ensure that we are dealing with the first boundary crossing rather than any later crossing.

For the 4D Bessel process, the probability density function  $dP_t(a)$  is given by the time derivative of Eq. 2.7 in Hamana and Matsumoto (2013) with v (a parameter in their notation that determines the dimensionality of the process) set to 1.0,

$$dP_t(a) = \frac{\sigma^2}{2a^2} \sum_{k=1}^{\infty} \frac{j_{1,k}^2}{J_2(j_{1,k})} \exp\left(-\frac{j_{1,k}^2 \sigma^2}{2a^2}t\right). \tag{2}$$

In this equation,  $J_2(x)$  is a second-order Bessel function of the first kind (Abramowitz & Stegun, 1965, p. 360) and  $j_{1,k}$ , k = 0, 1, 2, ... are the zeros of the first-order Bessel function  $J_1(x)$  whose properties are readily computed (Appendix A). Like other infinite-series representations associated with diffusion processes, this series must be truncated for use in applications. In our application, we truncated the series at 50 terms, which is more than sufficient for convergence (see Smith (2016), for further



discussion and illustration of the properties of Bessel functions).<sup>2</sup>

For us, the function of theoretical interest is the joint density of the event that the nonzero-drift process hits a designated point on the decision boundary at time t, conditional on this being the first such boundary crossing. In the circular model of Smith (2016), the hitting point on the boundary is specified by its phase angle,  $\theta$ . For the 4D model, we work in hyperspherical coordinates (Misner et al., 1973, pp. 703–704; Weisstein, 2009, pp. 1877–1878), so we need three phase angles,  $\psi$ ,  $\phi$ , and  $\theta$ , to describe the point at which the process hits the bounding 3-sphere. The first coordinate,  $\theta$ ,  $0 \le \theta < 2\pi$ , describes the point (the azimuth) at which the process exits a disk oriented in the horizontal plane, like the one shown in Fig. 2b. The second coordinate,  $\phi$ ,  $0 \le \phi < \pi$ , identifies the elevation of a designated disk within a sphere like the one shown in Fig. 2c and is expressed as a deviation from the north pole. (It is helpful to visualize the volume enclosed by the sphere in Fig. 2c as being approximated by a stack of disks, whose diameters are maximal at the equator,  $\phi = \pi/2$ , and decrease towards either pole. Then  $\theta$  identifies a point on the boundary of a disk and  $\phi$  identifies a particular disk within the stack.) The third coordinate,  $\psi$ ,  $0 < \psi <$  $\pi$ , similarly identifies the "elevation" of a sphere within a hypersphere and is also expressed as a deviation from the point of maximal elevation. Although it is difficult to visualize,  $\psi$  behaves mathematically in the same way as  $\phi$ . Using these three phase angles, the point at which the process hits the bounding 3-sphere in  $\mathbb{R}^4$  can be specified using four Cartesian coordinates:

$$X_T^4 = a \cos \psi$$

$$X_T^3 = a \sin \psi \cos \phi$$

$$X_T^2 = a \sin \psi \sin \phi \cos \theta$$

$$X_T^1 = a \sin \psi \sin \phi \sin \theta.$$
(3)

Unlike the coordinate transformation for the circular model described earlier, the  $X_T^1$  coordinate in Eq. 3 refers to the vertical rather than the horizontal axis of the circle in  $\theta$ . This way of parameterizing coordinates is fairly standard

when working with spherical and hyperspherical coordinate systems.

In order to state our results compactly, we will introduce the additional piece of notation,  $\mathbf{\Theta} = (\psi, \phi, \theta)$ , to denote a vector whose three components are the three phase angles of the hitting point on the boundary. We then denote by  $d\tilde{P}_t(\mathbf{\Theta}_T)$  the probability density that the process hits the bounding sphere at a point  $X_T$  with phase angle  $\mathbf{\Theta}_T$  at time  $T_a$ . We use a capital T to indicate that the hitting point and hitting time are both random variables that depend on the entire history of the evidence accumulation process, but omit the "a" from the notation for the hitting time to avoid double-subscripting. The tilde indicates that  $\tilde{P}_t$  is the probability distribution of the nonzero-drift process.

The Girsanov theorem states that the probability density functions  $dP_t(a)$  and  $d\tilde{P}_t(\Theta_T)$  are related to one another in a simple way, via an exponential martingale,  $\mathbf{Z}_T(\mathbf{X})$ ,

$$d\tilde{P}_t(\mathbf{\Theta}_T) = \mathbf{Z}_T(\mathbf{X})dP_t(a), \tag{4}$$

where

$$\mathbf{Z}_{T}(\mathbf{X}) = \exp\left[\frac{1}{\sigma^{2}}\boldsymbol{\mu} \cdot \mathbf{X}_{T} - \frac{1}{2\sigma^{2}}\|\boldsymbol{\mu}\|^{2}T\right]$$
 (5)

and where  $\mu \cdot \mathbf{X}_T = \sum_{i=1}^4 \mu_i X_T^i$  is the dot product of the drift rate and a random vector  $\mathbf{X}_T$  with norm ||a|| and phase angle  $\mathbf{\Theta}_T$ . For a specified  $\mathbf{\Theta}_T$ , the four components of  $\mathbf{X}_T$  are given by Eq. 3.

Martingales are fundamental to the modern theory of stochastic processes but they will be unfamiliar to many readers. In brief, a martingale is a bounded stochastic process whose expected value remains constant.<sup>3</sup> For readers wishing to learn more about martingales, Karlin and Taylor (1975) and Williams (1991) are recommended starting places. For our purposes here, the significance of the function  $\mathbf{Z}_T(\mathbf{X})$  is that it allows us to derive the statistics of the nonzero-drift process from those of the 4D Bessel process.

The ratio of probability density functions

$$\frac{d\tilde{P}_t(\mathbf{\Theta}_T)}{dP_t(a)} = \mathbf{Z}_T(\mathbf{X}) \tag{6}$$

has an illuminating statistical interpretation. This ratio is known as the *Radon-Nikodym derivative* of the probability measure  $\tilde{P}_t$  with respect to the measure  $P_t$ . The Radon-Nikodym derivative is similar to, but generalizes, the familiar likelihood ratio of statistics from a discrete sequence of observations to a process in continuous time. Equation 6 states that the process  $\mathbf{Z}_T(\mathbf{X})$  can be interpreted as a likelihood ratio; specifically, it is the likelihood of

<sup>&</sup>lt;sup>3</sup>The term "martingale" is of equestrian origin: It refers to a piece of harness designed to keep a horse's head level. Its use to describe a stochastic process whose expectation is constant is a metaphorical appropriation of its equestrian meaning.



<sup>&</sup>lt;sup>2</sup>Recent articles on mathematical methods for the 1D diffusion model have emphasized that there are two complementary infinite-series representations of the first passage time probability density: one which converges rapidly for small values of time, t, and another which converges rapidly for large values of t (Navarro & Fuss, 2009). The first passage time density for the Bessel process is a "large-time" representation. As a result, truncating the series can introduce a negative excursion into the computed density function in the neighborhood of t=0 for some values of the parameters. The duration of such negative excursions is usually short (typically a few milliseconds) and they can easily be identified and filtered out without affecting those parts of the distribution that are important for predicting experimental data.

obtaining the observed cumulative evidence process  $\mathbf{X}_t$ ,  $0 \le t \le T$ , given the drift rate was  $\boldsymbol{\mu}$ , versus the likelihood of obtaining the same process given the drift rate was  $\mathbf{0}$ . Evidently, the process  $\mathbf{Z}_T(\mathbf{X})$  will be maximized when the dot product  $\boldsymbol{\mu} \cdot \mathbf{X}_T$  is maximized, which will occur when the phase angles  $\boldsymbol{\mu}_{\boldsymbol{\Theta}}$  and  $\boldsymbol{\Theta}_T$  are the same, that is, when the process exits the bounding hypersphere at a point that coincides with the orientation of the drift vector. This point is the most likely hitting point for a process with drift vector  $\boldsymbol{\mu}$  and an observer (i.e., a decision maker) who reports this point as the identity of the stimulus is a maximum-likelihood observer.

Because the process  $\mathbf{X}_T$  is guaranteed to terminate at the decision boundary in finite time (Mörters & Peres, 2010), the integral of the probability density  $d\tilde{P}_t(\boldsymbol{\Theta}_T)$  over all possible hitting points and hitting times will be unity. Expanding  $\boldsymbol{\Theta}_T$  in components, this condition states that

$$\int_0^\infty \int_0^\pi \int_0^\pi \int_0^{2\pi} d\tilde{P}_t(\psi,\phi,\theta) V(\psi,\phi,\theta) d\theta d\phi d\psi dt = 1,$$
(7)

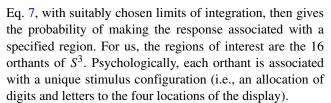
where  $V(\psi, \phi, \theta)$  is the volume element,

$$V(\psi, \phi, \theta) = a^3 \sin \psi^2 \sin \phi$$

associated with the change of variables from Cartesian to hyperspherical coordinates (Misner et al., 1973; Weisstein, 2009). To understand the role of the volume element, it is helpful to visualize the integral over phase angle in Eq. 7 as a summation over a set of small 3D regions on the surface of the hypersphere,  $S^3$ . For the 2D model, each of these regions is just a segment of arc of length  $a d\theta$  that forms the outer boundary of a wedge-shaped region that falls between the phase angles  $\theta$  and  $\theta + d\theta$ . These segments are the same length for all  $\theta$  and the volume element is simply a, the radius of the bounding circle. For the 4D model, the regions are 3D volumes whose size varies with phase angle. They are largest at the "equator" ( $\psi = \pi/2, \phi = \pi/2$ ) and decrease towards the "poles" ( $\psi = 0$ ,  $\phi = 0$  and  $\psi = \pi$ ,  $\phi = \pi$ ). The volume element specifies how the size of these regions changes as the phase angles are varied in equal steps.

### Application to the double-target detection task

Like the circular model of Smith (2016), the 4D model of the preceding section is a model of continuous report. It predicts the distributions of decision times and decision outcomes for reporting the identity of the stimulus as a specific point in 4D space. It becomes a model of categorical decision-making when the surface of  $S^3$  is partitioned into response regions with decision bounds. The integral of



Corbett and Smith's (2017) study of double-target detection used a two-alternative decision task in which participants responded "target present" or "target absent," according to the decision rule for the task. Table 1 shows how the orthants of the decision space for the 4D model map to the two response alternatives for their single-target and double-target tasks. The mappings assume that people respond "signal" (target present) if there is at least one signal associated with the response orthant in the singletarget task and at least two signals in the double-target task; otherwise they respond "noise" (target absent). The decision time and decision outcome predictions are obtained by summing across all of the orthants that map to a given response. The decision time densities are the sums of the hitting time densities, conditioned on the response probabilities. Explicitly, denote by  $d\tilde{P}_{ijk}(t)$  the joint density of a response at time t in orthant ijk. By taking appropriate bounds of integration in Eq. 7 we obtain

$$d\tilde{P}_{ijk}(t) = \int_{\psi_i}^{\psi_{i+1}} \int_{\phi_j}^{\phi_{j+1}} \int_{\theta_k}^{\theta_{k+1}} d\tilde{P}_t(\psi, \phi, \theta) \times V(\psi, \phi, \theta) d\theta d\phi d\psi, \tag{8}$$

where  $\psi_i = i\pi/2$ ; i = 0, 1;  $\phi_j = j\pi/2$ ; j = 0, 1; and  $\theta_k = k\pi/2$ , k = 0, 1, 2, 3. The joint densities of target-present and target-absent responses in the single-target and double-target tasks are obtained by summing the joint densities in Eq. (8) across all of the orthants in Table 1 that map to the given response,  $\mathcal{R}$ , for the particular task,

$$d\tilde{P}_{\mathcal{R}}(t) = \sum_{ijk\in\mathcal{R}} d\tilde{P}_{ijk}(t). \tag{9}$$

The probability of making response  $\mathcal{R}$ ,  $\tilde{P}_{\mathcal{R}}$ , is obtained by integrating across all of the decision times associated with that response,

$$\tilde{P}_{\mathcal{R}} = \int_0^\infty d\tilde{P}_{\mathcal{R}}(t) \, dt. \tag{10}$$

#### **Across-trial variability**

The standard diffusion model (Ratcliff & McKoon, 2008), in addition to within-trial diffusive variability, includes three sources of across-trial variability that have been found to be useful in fitting data: variability in drift rate, variability in starting point, and variability in nondecision time. These sources of variability have been implemented in third-party packages for fitting the model



**Table 1** Mapping of hitting-point phase angles and orthants to responses

Orthant	$\psi$	$\phi$	$\theta$	Stimulus	Single target	Double target
1	$(0, \pi/2)$	$(0, \pi/2)$	$(0, \pi/2)$	ssss	S	S
2	$(0, \pi/2)$	$(0, \pi/2)$	$(\pi/2, \pi)$	ssns	S	S
3	$(0, \pi/2)$	$(0, \pi/2)$	$(\pi, 3\pi/2)$	ssnn	S	S
4	$(0, \pi/2)$	$(0, \pi/2)$	$(3\pi/2, 2\pi)$	sssn	S	S
5	$(0, \pi/2)$	$(\pi/2, \pi)$	$(0, \pi/2)$	snss	S	S
6	$(0, \pi/2)$	$(\pi/2, \pi)$	$(\pi/2, \pi)$	snns	S	S
7	$(0, \pi/2)$	$(\pi/2, \pi)$	$(\pi, 3\pi/2)$	snnn	S	N
8	$(0, \pi/2)$	$(\pi/2, \pi)$	$(3\pi/2, 2\pi)$	snsn	S	S
9	$(\pi/2, \pi)$	$(0, \pi/2)$	$(0, \pi/2)$	nsss	S	S
10	$(\pi/2, \pi)$	$(0, \pi/2)$	$(\pi/2, \pi)$	nsns	S	S
11	$(\pi/2, \pi)$	$(0, \pi/2)$	$(\pi, 3\pi/2)$	nsnn	S	N
12	$(\pi/2, \pi)$	$(0, \pi/2)$	$(3\pi/2, 2\pi)$	nssn	S	S
13	$(\pi/2, \pi)$	$(\pi/2, \pi)$	$(0, \pi/2)$	nnss	S	S
14	$(\pi/2, \pi)$	$(\pi/2, \pi)$	$(\pi/2, \pi)$	nnns	S	N
15	$(\pi/2, \pi)$	$(\pi/2, \pi)$	$(\pi, 3\pi/2)$	nnnn	N	N
16	$(\pi/2, \pi)$	$(\pi/2, \pi)$	$(3\pi/2, 2\pi)$	nnsn	S	N

*Note.* Lower case letters s and n denote signal and noise stimuli; upper case letters S and N denote signal and noise responses. The ordering of orthants is the order obtained as the phase angles,  $\psi$ ,  $\phi$ , and  $\theta$  are progressively varied (the first angle changing most slowly)

to data (Vandekerckhove & Tuerlinckx, 2007; Vogel et al., 2006; Wiecki et al., 2013). For our purposes, the most important of these is variability in drift rate. Variability in drift rate is assumed to arise because of trial-to-trial variability in the quality of the evidence provided by nominally equivalent stimuli. This variability may be due to variability in perception or memory or in the operation of matching the encoded stimulus representation to the mental representations of the decision alternatives. We refer to these mental representations, by analogy with Lu and Dosher's perceptual template model (Dosher & Lu, 2000; Lu & Dosher, 1998), as the *decision template* for the task.

In the standard diffusion model, drift rates are assumed to vary normally across trials with standard deviation  $\eta$ . In higher dimensions the possibilities are richer, as discussed by Smith (2016), but here we try to remain close to the spirit of the standard diffusion model and assume that the four components of the drift rate  $\mu$  are independently normally distributed with common standard deviation  $\eta$ . When there is across-trial variability in drift rate (or any other model parameter) it is necessary to integrate (marginalize) across the decision time and decision outcome distributions for different values of the parameter to obtain predictions. To improve computational efficiency, we carried out this integration analytically rather than numerically. The first passage time joint density  $dP_t(\Theta_T)$  in Eq. 4 is the product of the exponential martingale  $\mathbf{Z}_T(\mathbf{X})$  in Eq. 5 and  $dP_t(a)$ , the first passage time density for the Bessel process in Eq. 2. Because only the first of these terms depends on drift rate, the effect of across-trial variability in drift can be completely characterized by its effects on  $\mathbf{Z}_T(\mathbf{X})$ . The value of  $\bar{\mathbf{Z}}_T(\mathbf{X}; \mathbf{v}, \eta)$ , the marginal form of  $\mathbf{Z}_T(\mathbf{X})$ , is a function of  $\mathbf{v} = \{v_1, v_2, v_3, v_4\}$ , the mean of the across-trial distribution of drift rates, and the common standard deviation,  $\eta$ ,

$$\bar{\mathbf{Z}}_{T}(\mathbf{X}; \mathbf{v}, \eta) = \prod_{i=1}^{4} \frac{1}{\sqrt{(\eta/\sigma)^{2}T + 1}} \exp \left\{ -\frac{v_{i}^{2}}{2\eta^{2}} + \frac{[X_{T}^{i}(\eta/\sigma)^{2} + v_{i}]^{2}}{2\eta^{2}[(\eta/\sigma)^{2}T + 1]} \right\}. (11)$$

The details of this computation are described in Appendix A. As in Eq. 5, when carrying out computations using Eq. 11 the values of  $X_T^i$  are given by the coordinate mapping functions in Eq. 3 that relate the phase angles  $\psi$ ,  $\phi$ , and  $\theta$  to the hitting points of the process on the surface of the hypersphere.

The standard diffusion model also includes variability in the starting point of the evidence accumulation process. In Ratcliff's (1978) gambler's ruin notation, the decision boundaries are denoted 0 and a and the starting point is denoted z (0 < z < a) and is assumed to vary uniformly across trials with range  $s_z$ , ( $s_z$  < a). Because of the spatial homogeneity (translation invariance) of the Wiener diffusion process, variability in starting point is equivalent to negatively correlated variability in decision criteria with a fixed starting point. Starting-point variability allows the model to predict the fast errors that are often found when stimulus discriminability is high and instructions emphasize



speed of responding (Luce, 1986; Ratcliff & Smith, 2004). Smith (2016) showed that the circular diffusion model can similarly predict fast errors (i.e., positively correlated RT and decision accuracy) when there is across-trial variability in the radius of the criterion circle, a. However, in near-threshold psychophysical tasks like the ones used by Corbett and Smith (2017) errors are almost always slower than correct responses. For such tasks, we have found little evidence that fits are improved by including variability in starting points or decision criteria (Smith & Ratcliff, 2009; Smith et al., 2004) and, indeed, recovery of other model parameters can be impaired by including it in the model, suggesting that the starting point parameter is not well identified in these kinds of tasks. We therefore chose not to include it in our models here.

The third source of across-trial variability is in nondecision time. The standard diffusion model assumes that RT is the sum of independent decision and nondecision times. The latter is denoted  $T_{\rm er}$  (for "time for encoding and response") and is assumed to be uniformly distributed with range  $s_t$ . This assumption is made mainly for convenience: If nondecision time variability is small relative to decision time variability, then the shape of the RT distribution will be almost completely determined by the distribution of decision times and the shape of the distribution of nondecision times will be largely immaterial. This assumption was questioned recently by Verdonck and Tuerlinckx (2016) who proposed a deconvolutional method for inferring the distribution of nondecision times that yields distributions with appreciable positive skewness. As their method has yet to be widely applied and because we wished to stay as close as possible to the standard diffusion model, we continued to assume uniformly distributed nondecision times. For Corbett and Smith's (2017) task, which was difficult and yielded long and variable RTs, this assumption seems a reasonable one and is not critical to the performance of the model.

#### Response bias and stimulus bias

Two kinds of decision bias have been identified in models of speeded two-choice decision-making: response bias and stimulus bias. Response bias is bias in the amount of evidence needed for a response and is implemented as a difference in the decision criteria for the two responses or, in the standard diffusion model, as asymmetry in the placement of the starting point between the two decision boundaries. Response bias is needed to account for performance in decision tasks in which the prior probabilities of the alternatives are unequal (Diederich & Busemeyer, 2006; Leite & Ratcliff, 2011; Mulder et al., 2012) and in tasks like the lexical decision task, in which there is asymmetry in the amounts of evidence needed to

decide whether a character string is a word or a not (Ratcliff & Smith, 2004). In contrast, many perceptual tasks exhibit a high degree of symmetry in the response probabilities and RT distributions for the two alternatives, especially when the prior probabilities of the stimulus alternatives are equal, and can be modeled without response bias (Ratcliff & Smith, 2004; Smith & Ratcliff, 2009; Smith et al., 2004).

In higher-dimensional spaces like those that arise in the 2D and 4D models, it is not obvious that starting point placement is the most appropriate way to think about response bias-at least not when the prior probabilities of the stimulus alternatives are equal. Pragmatically, the introduction of starting point bias breaks the symmetry on which the analytic solutions via the Girsanov theorem and Bessel process depend. Consequently, for reasons of computational efficiency if for no other reason, we are motivated to seek other ways of representing response bias in these models. A form of bias that seems particularly relevant to the tasks of Fig. 1 is bias in the placement of the decision boundaries on the surface of the hypersphere. An unbiased decision maker will align the boundaries with the cardinal axes of the hypersphere, partitioning it into equalsized orthants. For the single-target task, we found that this assumption was too restrictive. As is apparent from Table 1, there is asymmetry in the decision rule for both forms of the task that is particularly pronounced in the single-target task: For this task, 15 of the 16 orthants are associated with a "signal" response and only one of them is associated with a "noise" response. This response rule strongly biases the model to make "signal" responses, even when the drift rates for signal and noise are the same. In order to allow for this kind of bias, we increased the size of the response region associated with "noise" responses by moving the decision boundaries symmetrically away from the category midpoint.

Stimulus bias is a difference in the rates at which evidence for the decision alternatives accumulates. It is conceptualized as a bias in the process that compares encoded stimulus representations to the decision template. Link and Heath (1975) proposed that sensory evidence is compared to a mental standard or referent, which determines the relative rates at which evidence for the two alternatives accumulates. Ashby (1983) proposed that stimulus bias could arise from bias in a log-likelihood computation. The standard diffusion model conceives of stimulus bias as arising from a drift criterion, which determines the relative drift rates for the two alternatives. Drift rate bias is needed to account for performance in recognition memory tasks, in which drift rates are affected by the prior probability that an item was on a studied list. The changes in drift rates across conditions in recognition memory are well described by changes in drift criterion (Ratcliff & Smith, 2004).



#### The decision template for multiattribute tasks

To characterize drift bias in the 4D model it is useful to have a formal representation of the processes involved in the computation of drift rates. We coined the term "decision template" to refer to the mental representation of the stimulus configurations that are associated with the response alternatives in the task. In the 1D model, the decision template consists of a pair of alternatives; for the 4D model it potentially consists of 16, although only a subset of these alternatives were actually presented in Corbett and Smith's (2017) task. These 16 alternatives correspond to the rows of Table 1 and consist of all possible assignments of signals (digits) and noise (letters) to the four locations of the array. The drift rate is a four-element vector that represents the encoded evidence for one of these 16 configurations on an experimental trial.

A natural way to represent the process of computing drift rates is as a linear operation in the vector space of possible stimulus configurations. Although this approach is too restrictive to accommodate the configuration-dependent biases we found empirically, it is useful to consider it first because it helps motivate a more appropriate representation. To this end, we define the *configuration space* of stimuli by associating with each of the 16 stimulus configurations a basis vector.

$$\mathbf{u}_i = (u_1, u_2, u_3, u_4)';$$
  
 $i = 1, \dots 16; \quad u_1, u_2, u_3, u_4 \in \{0.500, -0.500\}, \quad (12)$ 

where, in keeping with our conventions for stimulus coding, we assume signals (digits) are coded as positive and noise (letters) are coded as negative. In words,  $u_i$  is a four-element vector with components 0.500 or -0.500 and the set of  $2^4$  such vectors enumerates all possible assignments of these values to components. The prime denotes the matrix transpose and the choice of 0.500 sets the norms of the basis vectors to unity,  $\|u_i\| = 1$ . The basis set in Eq. 12 is redundant in that it contains subsets of linearly dependent vectors, including the collinear pairs  $u_i$  and  $-u_i$ , but the redundancy does not affect its algebraic properties in any essential way.<sup>4</sup>

A stimulus, s, of intensity I can be represented as a vector in this 4D space,

$$s = Iu_i + \xi$$
,

for some i, where  $\xi$  is a vector of normally distributed noise, which characterizes the effects of encoding variability on a given trial. The vector space operation

$$\mu = \frac{1}{4} \sum_{i=1}^{16} (s \cdot u_i) u_i = s, \tag{13}$$

yields a veridical representation of the stimulus. In this representation, the dot product,  $(s \cdot u_i) = ||s|| ||u_i|| \cos \alpha$ , computes the orthogonal projection of the stimulus vector on the basis vector  $u_i$ , where  $\alpha$  is the angle between the two vectors. The value of  $\mu$  is a weighted sum of the basis vectors, with weights equal to the orthogonal projections of the stimulus vector on each of the basis vectors. The scaling factor of 1/4 compensates for the redundancy in the basis set and puts the computed drift rates on the same scale as the stimulus. The operation in Eq. 13 can be viewed as a vectorspace generalization of the one assumed to underlie the computation of drift rates in the 1D diffusion model, which reflect the difference in the strength-of-match between a stimulus and the two decision alternatives. The drift rates in the 1D model can be regarded as a weighted sum of two collinear vectors of opposite sign.

In general, the drift rates will not be veridical representations of the stimuli, but may exhibit various forms of stimulus bias. The estimates of drift rates we obtained from fitting the model suggest they depend at least in part on the configural properties of the stimuli, that is, on the particular combination and arrangement of elements in the display. Specifically, our data suggest that participants were searching for particular patterns of stimulus elements that would allow them to classify the trial as "target present" or "target absent." For the double-target task, these patterns consisted of either a pair of digits or a triplet of letters, indicating a target-present trial or a target-absent trial, respectively. For the single-target task the patterns were a single digit or a quartet of letters. The biases we found empirically seem to have been directed at emphasizing the contents of those display locations that carry information about these patterns while de-emphasizing other locations.

These kinds of configural biases imply that the metric and the categorical properties of stimuli are both important for how they are coded cognitively. The metric properties of a stimulus are those that identify it uniquely as a point in a four-dimensional perceptual space; the categorical properties are those that associate this point with one of two response categories. Such interactions between metric and categorical properties of stimuli recall the picture provided by the work of Fific et al. (2010) who showed how



 $<sup>^4</sup>$ We use the term "basis vector" in a psychological rather than a strict mathematical sense. In vector analysis, a set of basis vectors spans a vector space without linear dependencies: Exactly four linearly independent basis vectors are required to span the space  $\mathbb{R}^4$ . Mathematically, our "basis vectors" are a *spanning set* for  $\mathbb{R}^4$ . Psychologically, they function as a cognitive basis for the representation of stimuli in the sense they enumerate all possible stimulus configurations in the task. We stipulate that the basis vectors are normalized for reasons of convenience, to simplify the expression of the model. For unnormalized basis vectors, the expression for the projection of a stimulus on a basis vector is equal to the dot product divided by the norm of the basis vector. The norms of normalized vectors are unity, so they can be omitted from the model.

evidence accumulation rates in serial and parallel processing models depend on category structures defined by logical rules acting upon the metric properties of an underlying stimulus space. Our data suggest that rule-based category structures induce stimulus biases that depend on the way in which particular sets of stimulus elements are mapped to responses. Such biases cannot be represented by applying a single biasing transformation uniformly to the result of the linear operation in Eq. 13, because they depend on the rule that maps a specific configuration of stimulus elements to a response.

We can formalize the ideas in the preceding paragraphs by thinking of the decision template, not as the set of all possible stimulus configurations for the task, but instead, as a set of diagnostic patterns that participants actually use (or we infer they use) to distinguish between target-present and target-absent trials. Components of the decision template that do not represent one or other of these diagnostic patterns are set to zero (for "don't care"). Conceptualized in this way, the set of diagnostic patterns we hypothesize forms the decision template for the double-target task can be represented as a set of ten normalized vectors,  $v_i$ , comprising the six element-wise permutations of

 $v_{\text{Double Present}} = (0.707, 0.707, 0, 0)'$ 

and the four element-wise permutations of

$$v_{\text{Double Absent}} = (-0.577, -0.577, -0.577, 0)'.$$

For the single-target task, the hypothesized decision template is a set of five normalized vectors comprising the four element-wise permutations of

 $\boldsymbol{v}_{\text{Single Present}} = (1, 0, 0, 0)'$ 

together with the single vector

$$v_{\text{Single Absent}} = (-0.500, -0.500, -0.500, -0.500)'.$$

Like the linear-space representation of Eq. 13, we can think of the dot product  $(s \cdot v_i)$  as characterizing the similarity between a stimulus s and a template vector  $v_i$ , and the drift rate,  $\mu$ , as the outcome of a process in which the stimulus is simultaneously matched to every vector in the template. Now, however, instead of depending on the sum of the projections of the stimulus on the basis vectors in configuration space, the drift rate depends on which of the set of template vectors it resembles most closely,

$$\mu = (s \cdot v_i) v_i + (1 - \beta) [s - (s \cdot v_i) v_i]$$

$$i = \underset{j}{\operatorname{argmax}} (s \cdot v_j); \quad j = 1, 2, \dots; \quad 0 \le \beta \le 1.$$
 (14)

In Eq. 14, the "argmax" function picks out the index i at which similarity is maximized, that is, it identifies the template vector that most closely matches the stimulus. A biased drift rate,  $\mu$ , is obtained by orthogonally decomposing the stimulus vector into two components. The

first,  $(s \cdot v_i) v_i$ , is the projection of the stimulus on the template vector  $v_i$  and the second,  $s - (s \cdot v_i) v_i$ , is the orthogonal complement of the projection. The stimulus vector, s, is the sum of these two components. The extent of the drift bias is controlled by the parameter  $\beta$  which varies between zero and one. When  $\beta$  is zero, there is no bias and  $\mu = s$  and the drift rate is a veridical representation of the stimulus. As  $\beta$  increases, the contribution of the orthogonal complement is reduced and the drift vector progressively approaches its projection onto the nearest template vector. The effect of this kind of bias is to de-emphasize those elements of the stimulus that are not part of one or other of the diagnostic patterns for the decision rule while preserving the values of the remaining elements.

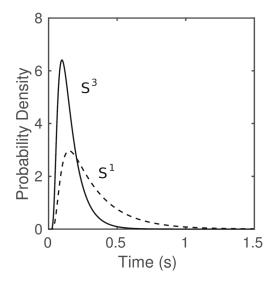
Computationally, the "argmax" function in Eq. 14 could be realized in a straightforward way by reformulating the model as a similarity-choice model, in which similarity is an exponentially decreasing function of the psychological distance between stimuli and similarities are summed across items or exemplars in psychological space (Navarro, 2007; Nosofsky, 1984; Shepard, 1987). In such models, when the exponential decay grows large, the summed similarity becomes dominated by the nearest item and the sum across items approximates the argmax function. For our purposes, there is little to be gained by elaborating the model in this way, but the theoretical connection to the larger class of similarity-choice models is nevertheless illuminating.

#### Does dimensionality matter?

Notwithstanding our earlier argument that a fourdimensional evidence accumulation process is the natural setting in which to represent decisions about four-element displays, the reader may still be wondering whether the dimensionality of the model really matters. Specifically, are the predictions of the 2D and 4D models sufficiently different that it would affect our ability to account for data if we assumed a model with the wrong dimension? Figure 4 shows that the dimensionality of the model does indeed have a significant effect on the RT distributions it predicts. The figure shows predicted response time distributions for zerodrift diffusion processes (Bessel processes) in two  $(S^1)$  and four  $(S^3)$  dimensions. The distribution for the 4D process was generated from Eq. 2 and that for the 2D process from Eq. 22 in Smith (2016) with the decision criterion a = 0.8and infinitesimal standard deviation  $\sigma = 1.0$  in both cases.

In comparison to the distribution for the 2D process, the distribution for the 4D process has a smaller mean and a smaller variance. Processes in a 4D space are likely to finish faster and to be less variable than those in a 2D space. The way to understand this relationship is by analogy with the properties of independent parallel processes. As is well known (e.g., Colonius (1995), Eq. 6),





**Fig. 4** Predicted response time distributions for zero-drift diffusion processes in two dimensions ( $S^1$ , dashed line) and four dimensions ( $S^3$ , solid line) for a = 0.8 and  $\sigma = 1.0$ 

when a decision depends on the fastest-finishing of a set of racing independent processes, RTs become progressively faster and less variable as the number of racing processes increases. This is because the probability that the system as a whole takes longer than a given time t to finish is the product of the probabilities that all of its components take longer than t to finish and this (usually) decreases with the number of components for all t.5 A multidimensional diffusion process with independent components behaves similarly. Increasing the number of dimensions increases the number of different ways in which the process can exit the bounding hypersphere. This increases its opportunities for finishing by time t, or conversely, decreases the probability that it will still be within the hypersphere at time t. This manifests itself as a reduction in the mean and variance of the RT distribution as shown in Fig. 3. So the answer to our question is, yes, dimensionality does matter—in the sense that the relationship between the predicted RT distributions and the underlying model parameters is different in the 2D and 4D cases—which justifies our use of the more complex 4D model for the Corbett and Smith (2017) task.

#### Method

The full details of the experimental method can be found in Corbett and Smith (2017, General Method and Experiment 4a). The stimuli were presented on a linearized (gammacorrected) 21" CRT monitor driven by a Cambridge Research Systems ViSaGe framestore, which controlled all aspects of stimulus presentation and response timing. Responses were made on a Cambridge Research Systems infrared button box and were timed by the ViSaGe's hardware clock. The ViSaGe runs asynchronously of the host computer and is non-interruptible, providing a high degree of timing accuracy. Six highly practiced participants each completed seven 480-trial experimental sessions, yielding 3,360 trials per participant in a 2 (task) × 5 (stimulus contrast) factorial design. The two tasks (single target and double target) were alternated in four consecutive 120-trial blocks in each session. The task order was pseudorandomized across sessions to ensure that no more than four sessions began with the same task. Stimuli were presented in Gaussian noise for 200 ms and backwardly masked with high-contrast checkerboards. Stimulus contrasts and the location of digit targets were randomized across trials within a block. Five levels of stimulus contrast were selected for each participant individually during practice to produce a range of accuracies in the single-target task that varied from just above chance to near-ceiling. The same set of contrasts was used in the single-target and double-target tasks. Participants were encouraged to respond on each trial as accurately as possible and as soon as they had made a decision but there was no emphasis on speed of responding. Trial-by-trial accuracy feedback was given via distinctive auditory tone pairs.

#### Treatment of experimental data

The primary experimental data we wish to explain are the proportions of correct responses and the RT distributions for correct responses and errors as functions of stimulus contrast and the decision task. We evaluated our models using methods similar to those described in Ratcliff and Smith (2004), which we have used in many other studies (e.g., Smith et al. (2004), Smith and Ratcliff (2009), and Smith et al. (2016)). We summarized the information in the RT distributions using five quantiles: the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles. Five quantiles suffice



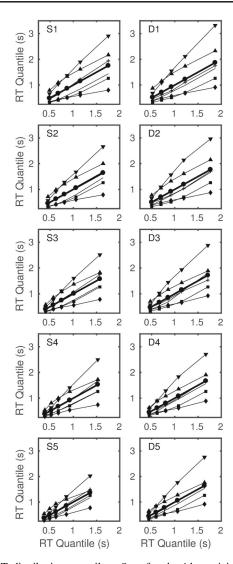
<sup>&</sup>lt;sup>5</sup>We make the qualification "usually" because counterexamples are known. Townsend and Colonius (2005, p. 768) gave the example of a reflected exponential distribution, whose probability mass is concentrated on the negative real axis, with a lower bound of  $-\infty$ and an upper bound of 0. Contrary to intuition, the dispersion of the minimum of n such random variables increases rather than decreases with n. Townsend and Colonius characterized the properties of such distributions using a statistic they called the quantile spread function, which they defined as the difference between the pth and (1 - p)th quantiles for a given p. They showed that the dispersion of the minimum and maximum statistics of sets of identically distributed random variables depends on a technical property they called global positive min-skewness, which is similar to other more familiar measures of positive skewness. For globally positive minskewed distributions (a class that appears to include most distributions encountered in practice in RT modeling), the mean and variance both decrease under the operation of taking the minimum, as stated in the

to characterize the shape of the distribution while being relatively insensitive to contaminants and outliers (Ratcliff & Tuerlinckx, 2002). The 0.1 quantile characterizes the fastest responses in the distribution (the leading edge); the 0.5 quantile characterizes its central tendency (the median), and the 0.9 quantile characterizes the slowest responses (the tail).

We fitted our models both to the individual participant data and to group data obtained by averaging corresponding quantiles for the distributions of correct responses and errors across participants:  $\bar{Q}_k = (\sum_{i=1}^6 Q_{i,k})/6$ . The conditions under which distributions can be safely quantile-averaged without distortion were identified by Thomas and Ross (1980). A distribution obtained from quantile-averaging a set of component distributions will belong to the same family as its components when the quantiles of the components form an affine family, that is, when they fall on a set of straight lines.<sup>6</sup>

Figure 5 shows quantile-quantile (Q-Q) plots for individual participants in each of the ten conditions of Corbett and Smith's (2017) experiment. The columns are the experimental tasks and the rows, top to bottom, represent increasing levels of stimulus contrast. To construct these plots, we plotted the quantiles of the marginal RT distributions for the individual participants,  $Q_{i,k}$ ,  $i=1,\ldots 6$ , against the group average quantiles,  $\bar{Q}_k$ . The heavy solid line is a plot of the values of  $\bar{Q}_k$  against itself, which is a straight line with unity slope. Although there are some individual deviations from linearity discernible in the figure they are small and unsystematic and the plot overall is remarkably linear.

Table 2 represents the same data in another way. The table shows the squared Pearson correlations  $(r^2)$  between the individual participant RT quantiles,  $Q_{i,k}$ , and the average quantiles for the other participants,  $(\sum_{j=1,j\neq i}^{6} Q_{j,k})/5$ , with participant i omitted. The smallest  $r^2$  in the table is .946 and most of the values are around 0.99. These figures imply that the conditions for valid



**Fig. 5** RT distribution quantiles,  $Q_{i,k}$ , for the *i*th participant, k = (.1, .3, .5, .7, .9), plotted against the group average quantile,  $\bar{Q}_k$  across participants for single (S) and double (D) target conditions for five levels of stimulus contrast (1, ... 5). The *plotting symbols* denote the individual participants; the *heavy line* is the group average

quantile-averaging identified by Thomas and Ross (1980) are satisfied. The pattern of data in Fig. 5 and Table 2 is not an exceptional or remarkable one but is fairly typical of data from this kind of task (e.g., Ratcliff and McKoon (2008), Fig. 8; Ratcliff and Smith (2010, Figure 20).

One of the reasons for the success of the diffusion model as a model of two-choice decisions is that the quantiles of its predicted RT distributions also form affine families (Ratcliff & McKoon, 2008; Smith, 2016, Fig. 10). This is true for the 1D model and for models in higher dimensions as well. Consequently, the distributions predicted by the diffusion model have just the right shape to account for empirical data. Because empirical data and the model predictions have the same affine structure, model parameters estimated from group data and the averages of parameters estimated from



<sup>&</sup>lt;sup>6</sup>Thomas and Ross did not characterize their results directly in terms of affine quantile families but in terms of shift-scale distribution families. Their actual result was to show that a sufficient condition for valid quantile averaging was that the component distributions were members of a shift-scale family, F[(t - b)/a], where b and a are shift and scale parameters, respectively. The functional consequences of this representation are that the distribution quantiles will form sets of straight lines as a and b are varied. We refer to such sets, which are closed under averaging, as "affine families." The shiftscale representation of the distribution family is a sufficient but not necessary condition for them to possess this property: Response time distributions for the diffusion model do not have a simple shiftscale algebraic representation but nevertheless generate affine-family quantiles. When there is across-trial variability in drift rates this property will hold only approximately rather than exactly (Ratcliff & McKoon, 2008, Fig. 8) so it is good practice to verify that inferences made on quantile-averaged group data also hold at the individual level.

**Table 2** Individual-group quantile correlation  $(r^2)$  statistics

Participant	<b>S</b> 1	D1	S2	D2	S3	D3	S4	D4	S5	D5
1	0.993	0.982	0.987	0.996	0.985	0.996	0.995	0.983	0.999	0.998
2	0.999	0.998	0.998	0.990	0.997	0.993	0.980	0.997	0.988	0.998
3	0.992	0.991	0.994	0.992	0.999	0.997	0.998	0.998	0.996	0.991
4	0.964	0.978	0.983	0.988	0.991	0.993	0.970	0.979	0.989	0.984
5	0.977	0.982	0.969	0.955	0.946	0.987	0.989	0.989	0.992	0.995
6	0.996	0.996	0.998	0.997	0.999	0.999	0.975	0.990	0.990	0.998

*Note.*  $r^2$  between  $Q_{i,k}$  and  $(\sum_{j,j\neq i}^6 Q_{j,k})/5$  in single (S) and double (D) target conditions

individual data are usually in good agreement (Ratcliff et al., 2003, 2004, 2004).

We fitted our models to the data from the single-target and double-target tasks separately, because there were no strong a priori reasons to assume that any of the model parameters would be the same across tasks. To do so, we fitted the quantile-averaged group data and individual data by minimizing the likelihood-ratio Chi-square statistic  $(G^2)$ ,

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

using the Matlab implementation of the Nelder-Mead Simplex algorithm (fminsearch). In this equation,  $p_{ij}$  and  $\pi_{ij}$  are, respectively, the observed and predicted probabilities (proportions) in the bins bounded by the quantiles.<sup>7</sup> The inner summation over j extends over the 12 bins formed by each pair of joint distributions of correct responses and errors. (There are five quantiles per distribution, resulting in six bins per distribution, or 12 bins in total for each distribution pair.) The outer summation over i extends across the five contrast levels and the two stimulus configurations (target-present and target-absent). The quantity  $n_i$  is the number of experimental trials in each

Burnham and Anderson (2002) recommend increasing the value of m in the QAIC by one, because the dispersion estimate, q, is a parameter that must be estimated from the data. The effect of this adjustment is to add 2.0 to all of the QAIC estimates. As the QAIC is an interval rather than a ratio scale quantity, whose interpretation is unaffected addition or subtraction of a constant, we omitted this adjustment.

stimulus condition (here  $n_i = 168$ ). Because  $G^2$  computed on the joint distributions depends on the relative proportions of correct responses and errors, it characterizes goodness-of-fit to the distribution shapes and the choice probabilities simultaneously.

We performed model selection using the Akaike Information Criterion (AIC), which penalizes models for complexity based on their number of free parameters. To allow for overdispersion (or underdispersion) in the data, we used a variant of the AIC called the QAIC (Burnham & Anderson, 2002, pp. 67–70), which adjusts the likelihood for dispersion prior to applying the penalty,

$$QAIC = G^2/q + 2m,$$

where q is the dispersion adjustment (see Results) and m is the number of free parameters in the model. Many cognitive researchers prefer the alternative Bayesian Information Criterion (BIC) for model selection, which uses a sample-size dependent complexity penalty, but Burnham and Anderson claim that BIC often leads to underfitting. Consistent with their view, Oberauer and Lin (2017) recently reported that the AIC had better model-recovery properties than the BIC in a study of visual working memory. We also computed the BIC, but in agreement with Burnham and Anderson, we found it tended to prefer underfit models.

#### Parameterizing the 4D model

Mean drift rates Because the 4D model is new and because there have been relatively few applications of diffusion models to tasks with multielement displays, we considered several highly parameterized versions of the model to try to understand its empirical properties. Our ultimate goal was to find a parsimonious parameterization that provided insight into the underlying psychological processes, particularly those associated with the computation of drift rates. We do not attempt to report all of the model variants we considered, but focus on the theoretically interesting ones. The most highly parameterized model has a total of 20 mean drift



<sup>&</sup>lt;sup>7</sup>Heathcote et al. (2002) called the maximization of a binned likelihood using bins formed from distribution quantiles "quantile maximum likelihood." Their assumption of a multinomial likelihood function was criticized by Speckman and Rouder (2004) because it of its use data-dependent bins, which is contrary to the classical definition of multinomial likelihood, although Speckman and Rouder also noted that the method had good parameter recovery properties. In fact, as discussed by Moore (1986, pp. 75-76), data-dependent bins lead to valid multinomial likelihoods and asymptotic Chi-square distributions in Cressie–Read divergence statistics (a class that includes  $G^2$  and the Pearson  $X^2$ ), providing the random cell boundaries converge in probability to a set with fixed boundaries. Distribution quantiles are order statistics and so possess this property.

rates: one for each level of stimulus contrast for signal and noise stimuli (digits and letters), which differ for target-present and target-absent trials. The possibility that drift rates for signal and noise stimuli might be different accords with previous treatments of stimulus bias in the literature and can be characterized using a variable drift criterion. However, the idea that drift rates might differ on target-present and target-absent trials has some deep theoretical implications that we consider at length subsequently.

The most striking of these, as was anticipated in the section on the decision template, is that the decision template is not local but configural. By a "local" decision template, we mean the components of the drift rate arise as a result of performing a perceptual matching operation simultaneously and separately at each location in the display. The result of this matching operation is a vectorvalued quantity that characterizes the strength of evidence for a signal at each location. The drift criterion is a local computation of this kind. By a "configural" template, we mean a matching operation in which the strength of evidence for a signal at a given location also depends on the stimuli that are present at other locations. This is, to our knowledge, a new idea in the realm of diffusion process models and one that on first encounter may strike readers as counterintuitive.

Although it is novel in diffusion process settings, the idea that the global features of a display can influence how its elements are processed is a fairly common one in perceptual psychology and mental architecture studies. In perceptual psychology, this kind of nonlocality is expressed in the idea of "gist" processing, which proposes that the global properties of an object or a scene can affect the manner and efficiency with which its components are processed (Palmer, 1975). Mental architecture studies using systems factorial technology (Townsend & Nozawa, 1995) have shown that the configural properties of stimuli affect the speed with which their components are processed and the degree to which they can be processed in parallel (Fific et al., 2010; Moneer et al., 2016; Wenger & Townsend, 2006). In a similar vein, Brady and Tenenbaum (2012) found that visual working memory was better characterized by a model in which top-down configural structure affects the processing of individual stimulus elements than by one in which the elements were processed independently.

In mental architecture models, these kinds of nonlocal processing effects are represented mathematically by crosstalk between channels that process individual stimulus elements (Townsend et al., 2012). Such crosstalk constitutes a form of *coactivation* between channels (Miller, 1982; Townsend & Nozawa, 1995). Our estimated drift rates imply a similar kind of nonlocality in the computation of letter and digit drift rates on double-target trials. Parenthetically,

we should note that while we have termed this "configural" processing, some authors prefer to reserve this term for stimuli that have recognizable higher-order or Gestalt structure, like features in a face (Wenger & Townsend, 2006) and use the term "contextual processing" (Palmer, 1975), for the more general situation in which processing of a stimulus element is affected by elements elsewhere in the display. We do not maintain this distinction rigidly and treat the terms as interchangeable.

**Drift rate variability** Most applications of the 1D diffusion model assume that across-trial variability in drift rates,  $\eta$ , is constant for different levels of stimulus difficulty within an experimental block when stimuli are viewed under identical conditions. Values of  $\eta$  may also vary within blocks if stimuli are viewed under different conditions, such as occurs when there is an across-trial manipulation of attention (e.g., Ratcliff (2014), Ratcliff and Smith (2004), Smith et al. (2016), and Smith and Ratcliff (2009)). When brief nearthreshold stimuli are used, however, like those used by Corbett and Smith (2017), there is a possibility that drift rate means and variances might covary with stimulus difficulty. This possibility arises from the idea that the drift rate mean and variance might both depend on Poisson-like neural coding of stimulus information in perception and visual working memory (Bays, 2014; Smith, 2010, 2015; Smith & McKenzie, 2011). We considered this possibility by allowing mean drift rate and variance to both vary with stimulus contrast: Specifically, we allowed  $\eta$  to vary with contrast and stimulus condition (single- versus double-target trials). Allowing  $\eta$  to vary in this way did improve model fit, but the improvements were smaller than were those associated with the configural encoding of drift rates.

**Nondecision times** We also considered the possibility that nondecision times might vary as a function stimulus difficulty. This possibility has been considered by a number of authors, including Ratcliff and Smith (2010) who studied two-choice decisions about letters presented in dynamic noise. They found large changes in the leading edge (the 0.1 quantile) of the RT distribution with changes in noise density. Donkin et al. (2009) reanalyzed data from Gould et al. (2007), who studied two-choice decisions about lowcontrast grating patches and found similar changes in the leading edge. Smith et al. (2014) showed that the RT distributions from these kinds of tasks could be successfully modeled by the integrated system model of Smith and Ratcliff (2009), which attributes changes in the leading edge of the RT distributions to the process of visual short-term memory formation that drives the decision process. Corbett and Smith's (2017) study also used low-contrast stimuli presented in noise, so we considered models in which both the mean nondecision time,  $T_{er}$ , and its variability,  $s_t$ ,



varied with stimulus condition. We found no evidence that freeing up the fits in this way improved them, so we do not consider models with stimulus-dependent nondecision times any further.

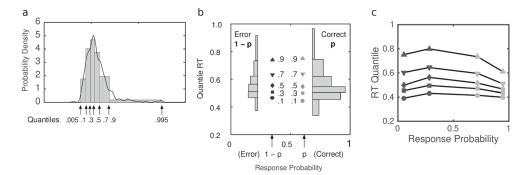
#### Results

## Quantile-probability representation of experimental data

A compact way to represent the effects of experimental manipulations on the RT distributions for correct responses and errors is in a quantile-probability plot. These plots have been widely used in the literature since they were introduced by Ratcliff et al. (2001), but they may be unfamiliar to some readers, so Fig. 6 summarizes how such a plot is constructed. The quantile-probability plot shows how the distributions of RT for correct responses and errors and the associated accuracy varies as a function of stimulus difficulty, which in our task was a function of contrast. For each difficulty level, performance is summarized by the .1, .3, .5, .7, and .9 quantiles of the distributions of correct responses and errors and the associated accuracy, p. Figure 6a shows a comparison of a kernel-density estimate of an RT distribution with an equal-area histogram in which the bin boundaries are formed by the distribution quantiles. Each of the interior bins contains .2 of the probability mass; the remaining .2 of the mass is distributed equally between the two end bins, whose lower and upper boundaries are the .005 and .995 quantiles, respectively. Evidently, the distribution quantiles provide a good summary of the shape of the distribution.

To construct a quantile probability plot, the quantiles of the distribution of correct responses are plotted against the probability of a correct response, p, and the quantiles of the distribution of errors are plotted against the probability of an error, 1-p, as shown in Fig. 6b. Each stimulus condition contributes one pair of distributions to the plot; in general there will be as many pairs of distributions in the plot as there are stimulus conditions. Figure 6c shows a quantile-probability plot from an experiment with two stimulus conditions and four distributions. The distributions on the right-hand side of the plot (light gray) are the distributions of correct responses and the distributions on the left (dark gray) are the distributions of errors. The innermost pair of distributions is from the hardest difficulty level (lowest contrast) and the outermost pair is from the easiest level (highest contrast).

The quantile-probability plot shows in a compact way how distribution shape and accuracy both change as stimulus difficulty is parametrically varied. Most of the changes in RT with changing stimulus difficulty are in the upper quantiles of the distribution (i.e., the median, .7, and .9 quantiles), as are most of the differences between correct responses and errors. The leading edge (the .1 quantile) shows comparatively little change with changing difficulty and comparatively little difference between correct responses and errors. The plot in Fig. 6c is canted upwards towards the upper left hand side. This is the typical pattern of slow errors that is found in difficult tasks in which accurate responding is stressed. If there were no difference between correct responses and errors then the plot would be symmetrical about its vertical midline. Performance in tasks in which speed and accuracy for the two decision alternatives are similar can be summarized using a single quantile-plot by combining correct responses and errors for the two alternatives in each stimulus condition. In tasks in which there are differences in RT or accuracy for the two alternatives, like those we consider here, separate plots are required. Performance under these circumstances can be characterized either in terms of the stimulus or the response; we have to chosen to characterize it in terms of the stimulus.



**Fig. 6** Quantile-probability plot. **a** RT distributions are summarized by equal-area histograms using the .1, .3, .5, .6, and .9 distribution quantiles as bin boundaries. The *continuous curve* is a kernel density estimate of the same distribution. **b** For each stimulus condition, the quantiles of the distribution of correct responses are plotted against the

probability of a correct response, p, and the quantiles of the distribution of errors are plotted against the probability of an error, 1-p.  $\mathbf{c}$  Quantile-probability plot from an experiment with two stimulus conditions showing a slow-error structure. Each condition contributes a pair of distributions to the plot



**Table 3** Models for the double-target task

Model	m	df	$G^2$	QAIC	$w_i(\text{QAIC})$	Model P	Model Parameters			
Group Fits (q	= 0.583)									
Model 1	33	77	44.9	143.0	0.401	$20 \nu$	$10  \eta$	a	$T_{\rm er}$	$s_t$
Model 2	23	87	57.8	145.3	0.127	$10 \nu$	$10  \eta$	a	$T_{\rm er}$	$s_t$
Model 3	25	85	54.0	142.7	0.446	$20 \nu$	$2 \eta$	a	$T_{\rm er}$	$s_t$
Model 4	15	95	71.1	151.7	0.005	$10 \nu$	$2 \eta$	a	$T_{\rm er}$	$s_t$
Averaged Indi	ividual Fits	(variable q)								
Model 1	33	77	178.4	143.2	0.165	$20 \nu$	$10  \eta$	a	$T_{\rm er}$	$s_t$
Model 2	23	87	208.8	137.8	0.010	$10 \nu$	$10  \eta$	a	$T_{\rm er}$	$s_t$
Model 3	25	85	200.6	137.3	0.167	$20 \nu$	$2 \eta$	a	$T_{\rm er}$	$s_t$
Model 4	15	95	228.4	130.8	0.657	$10 \nu$	$2 \eta$	a	$T_{\rm er}$	$s_t$

*Note.* m = number of free parameters

#### The double-target task

The double-target task is more complex psychologically than the single-target task because it shows a greater asymmetry in performance across target-present and target-absent trials, but it was simpler to model, because it did not require response bias, so we consider it first. We begin by discussing the quantile-averaged group data and then report confirmatory fits to the individual participants. Table 3 summarizes the main models of interest for the double-target task and the associated  $G^2$  fit statistics and BICs. Table 4 gives a detailed breakdown of the parameterization of mean drift rates and drift rate variability for all of the models for the double-target and single-target tasks.

Theoretically,  $G^2$  is distributed as Chi-square with degrees of freedom equal to the degrees of freedom in the data minus the number of free parameters in the model (110 - m in the notation of Table 3). We do not assign p-values to  $G^2$ , but use it just as a rough guide to the quality of the fit. For  $G^2$  to be tested against the percentage points of the Chi-square distribution, the quantity  $\sum_i n_i \sum_j p_{ij} \log \pi_{ij}$  (the model-dependent part of  $G^2$ ) must be a likelihood based on independent and identically distributed observations. Individual data from cognitive tasks are typically statistically *overdispersed* (i.e., more variable than predicted) relative to the distributions ordinarily used to describe them (McCullagh & Nelder, 1989; Smith, 1998) because of parameter variability due to

learning, fluctuations in attention, and across-trial sequential dependencies (Luce, 1986). As a result, fit statistics computed at the individual level are often larger than would be expected from an otherwise well-fitting model, even when it seems to capture the essential features of the data. Quantile-averaging of individual data can produce the opposite effect and yield fit statistics that are underdispersed relative to the theoretical Chi-square. In the absence of overdispersion or underdispersion,  $G^2$  or Pearson Chi-square for a well-fitting model should be fairly close to its residual degrees of freedom.

**Group fits** Figure 7 shows plots for two of the models from Table 3 (Model 1 and Model 2). The left-hand panels of the figure show RTs and accuracy for target-present (two-digit) trials; the right-hand panels show target-absent (single-digit) trials. The errors in the panels on the left are the target-absent responses on double-target trials (misses) and the errors in the panels on the right are the target-present responses on single-digit trials (false alarms). (The errors are the symbols to the left of the 0.5 point in each panel.) There is a marked asymmetry in the pattern of RTs and accuracy for the two kinds of trial and the overall shape of the quantile-probability functions is appreciably more complex than is typically found for decisions based on only a single stimulus (e.g., Ratcliff and Smith (2004)).

Model 1, shown in Fig. 7a, was the most complex of the models we considered. It had a total of 20 mean drift rates,  $\nu$ , ten drift standard deviations,  $\eta$ , a single decision criterion, a, and a single pair of nondecision time parameters,  $T_{\rm er}$ , and  $s_t$ . The idea that the mean drift rates for individual digits and letters might vary both as a function of the elements themselves and of the number of similar elements in the display is an expression of the assumption that processing



<sup>&</sup>lt;sup>8</sup>We investigated whether our fits to the double-target task could be improved by introducing orthant bias of the kind we needed to fit the single-target task, but there was no evidence of any appreciable improvement. We omit the details of these fits.

Table 4 Drift Rate Parameters

	Mea	ın drift	rate (v)	)					Drif	t rate v	ariabili	ty (η)				
	Con	trast 1			 Cont	rast 5			Con	trast 1			 Con	trast 5		
Double target (D)																
	2 <i>s</i>	2n	1 <i>s</i>	1n	 2 <i>s</i>	2n	1 <i>s</i>	1 <i>n</i>	2 <i>s</i>	2n	1 <i>s</i>	1 <i>n</i>	 2 <i>s</i>	2n	1 <i>s</i>	1 <i>n</i>
Model 1	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	 $v_{17}$	$\nu_{18}$	v <sub>19</sub>	$\nu_{20}$	$\eta_1$	$\eta_1$	$\eta_2$	$\eta_2$	 $\eta_9$	$\eta_9$	$\eta_{10}$	$\eta_{10}$
Model 2	$\nu_1$	0	0	$\nu_2$	 ν9	0	0	$\nu_{10}$	$\eta_1$	$\eta_1$	$\eta_2$	$\eta_2$	 $\eta_9$	$\eta_9$	$\eta_{10}$	$\eta_{10}$
Model 3	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	 $v_{17}$	$\nu_{18}$	v <sub>19</sub>	$\nu_{20}$	$\eta_1$	$\eta_1$	$\eta_2$	$\eta_2$	 $\eta_1$	$\eta_1$	$\eta_2$	$\eta_2$
Model 4	$\nu_1$	0	0	$\nu_2$	 ν9	0	0	$\nu_{10}$	$\eta_1$	$\eta_1$	$\eta_2$	$\eta_2$	 $\eta_1$	$\eta_1$	$\eta_2$	$\eta_2$
Single target (S)																
	1 <i>s</i>	1 <i>n</i>	0n		 1 <i>s</i>	1n	0n		1 <i>s</i>	1 <i>n</i>	0n		 1 <i>s</i>	1n	0n	
Model 5	$\nu_1$	$\nu_2$	$\nu_3$		 $\nu_{13}$	$v_{14}$	$\nu_{15}$		$\eta_1$	$\eta_1$	$\eta_2$		 $\eta_9$	$\eta_9$	$\eta_{10}$	
Model 6	$\nu_1$	0	$\nu_2$		 ν9	0	$\nu_{10}$		$\eta_1$	$\eta_1$	$\eta_2$		 $\eta_9$	$\eta_9$	$\eta_{10}$	
Model 7	$\nu_1$	$\nu_2$	$\nu_3$		 $\nu_{13}$	$v_{14}$	$v_{15}$		$\eta_1$	$\eta_1$	$\eta_2$		 $\eta_1$	$\eta_1$	$\eta_2$	
Model 8	$\nu_1$	0	$\nu_2$		 ν9	0	$\nu_{10}$		$\eta_1$	$\eta_1$	$\eta_2$		 $\eta_1$	$\eta_1$	$\eta_2$	

Note. 2s = 2 targets, signal; 2n = 2 targets, noise; 1s = 1 target, signal; 1n = 1 target, noise; 0n = 0 targets, noise

is inherently configural. As we foreshadowed earlier and discuss in detail later, we interpret the estimated drift rates from the model as evidence for a biased-template model, in which participants search for particular, diagnostic patterns of stimulus elements.

Overall, the model fits well and captures the main regularities in what is a challenging set of data. The only obvious misses are in the tail quantiles (.7 and .9 quantiles) for errors, but empirical estimates of tail quantiles have high variance because they are based on comparatively few trials and one must be careful to avoid overinterpreting them when assessing model fit (Ratcliff & Tuerlinckx, 2002). Burnham and Anderson (2002) recommend using the goodness-of-fit of the most complex structural model in a set of candidate models to estimate the QAIC dispersion parameter q; we used Model 1 for this purpose. They also recommend not adjusting the likelihood when the estimated q is less than unity, but underdispersion is expected with quantileaveraged data, so use of an adjusted likelihood seems appropriate. We obtained an estimate of q = 44.9/77 =0.583. which we used to calculate the QAICs for all of the fits to group data in Table 1.

In addition to Model 1, we considered Model 2, which had ten fewer mean drift parameters. Our reasons for considering this model were both pragmatic and theoretical: Pragmatically, we were trying to reduce the number of free parameters, but the model is also of theoretical interest because it is the purest expression of the biased-template model, in which all of the stimulus information on target-present trials is carried by digits (signals) and all of the

information on target-absent trials is carried by letters (noise). We were motivated to consider this model by the finding that mean drift rates for letters on target-present trials and for digits on target-absent trials in Model 1 were smaller and less systematic than those for the other two stimulus types (see Mean Drift Rates below). We therefore considered a model in which these ten drift rates were all set to zero (see Table 4 for details). The fit of Model 2 is shown in Fig. 7b. The  $G^2$  is some 29% worse than that of Model 1, but the QAIC differed by only a small amount,  $\Delta_i = 2.3$ . Burnham and Anderson (2002, p. 70) state that a model whose  $\Delta_i$  is within two units of the bestfitting model has "substantial" support. When applied to our data, their criterion implies that a model with 20 free drift rates and one with ten free drift rates perform similarly. In comparison to Model 1, Model 2 is somewhat "stiffer" and consequently does not capture some of the fine-grained details of the data as well. Nevertheless, it captures most of the main qualitative features, with the exception of the concavity in the quantiles of the target-absent error distributions.

In Models 3 and 4 we attempted to reduce the number of free  $\eta$  parameters from ten to two using the same set of mean drift rates as in Models 1 and 2. Estimates of  $\eta$  for noise (letter) stimulus were appreciably smaller than those for signal (digit) stimuli, implying a need for separate parameters for these two classes of stimuli, but the changes in estimated  $\eta$  across stimulus contrasts were not large, so we considered models in which there were only two such parameters. As we noted previously, the majority of



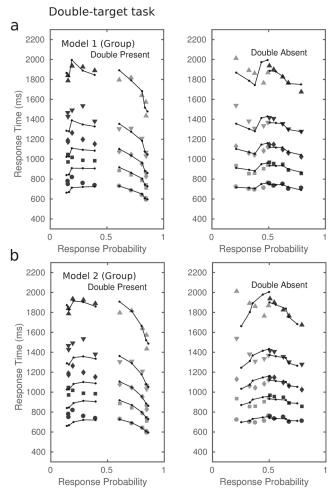
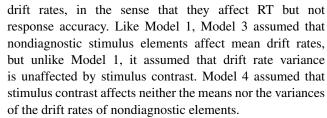


Fig. 7 Fits to double-target task (Group data). The *left-hand panels* are target-present (two-digit) trials; the *right-hand panels* are target-absent (one-digit) trials. a Model with 20 mean drift rates (Model 1); **b** model with ten mean drift rates (Model 2). The *symbols* denote distribution quantiles: circle = .1; square = .3; diamond = .5; inverted triangle = .7, and upright triangle = .9. Light gray symbols are "target-present" responses; dark gray symbols are "target-absent" responses. The points on the x-axis are choice probabilities for the five levels of stimulus contrast. Choice probabilities increase with contrast for correct responses (to the right of 0.5 point) and decrease with contrast for errors (to the left of the 0.5 point)

fits of the 1D diffusion model in the literature assume a single  $\eta$  value for different values of mean drift rate. Again, there was a theoretical as well as a pragmatic reason for considering these two models.

Model 2 assumed that letters on target-present trials and digits on target-absent trials do not contribute discriminative information to the decision process but only add contrast-dependent noise. This assumption is expressed in the combination of zero mean drift rates and contrast-dependent drift-rate variance for these two classes of stimulus elements. The picture implied by Model 2 is that these elements have a measurable, but unsystematic effect on



The  $G^2$  for Model 3 was similar to that of Model 2 and its QAIC was (by a small margin) the best of the three models and within two QAIC units of both of them, implying by Burnham and Anderson's (2002) criterion, "substantial support." In contrast, the fit of Model 4 was poorer, and its QAIC difference from the best model was  $\Delta_i = 9.0$ . Burnham and Anderson (p. 70) classify the support for the associated model as somewhere between "considerably less" and "essentially none."

Another way of viewing the QAIC values is to convert the differences between models to Akaike weights (Burnham & Anderson, 2002, p. 75; Wagenmakers & Farrell, 2004),

$$w_i(\text{QAIC}) = \frac{\exp\left[-\frac{1}{2}\Delta_i(\text{QAIC})\right]}{\sum_{i=1}^k \exp\left[-\frac{1}{2}\Delta_i(\text{QAIC})\right]},$$

where the summation in the denominator extends over the k models in the candidate set. The Akaike weights can be interpreted (with caution) as the probability that Model i is the best model in the set. Table 1 shows the Akaike weights for the four models. The weights suggest that Models 1 and 3 are the most probable models, Model 2 is somewhat less probable, and that Model 4 is highly improbable.

Individual fits When the conditions for valid quantileaveraging are satisfied, the resulting group distributions have some of the nice properties of robust estimators. They identify essential structural regularities in the data while de-emphasizing the influence of outliers and other aberrations, and serve as a corrective to any tendency to overfit and overinterpret the idiosyncrasies of individual participants. Nevertheless, it is important to assess whether a model can account for performance at the individual level and whether inferences made at the group level continue to hold. Individual participant data, as noted earlier, tend to be overdispersed relative to the associated theoretical sampling distributions. Smith (1998) used a method from McCullagh and Nelder (1989) and found that individual two-choice psychophysical data were overdispersed by a factor of 1.2 to 2.5 relative to the theoretical variance of the binomial distribution. Ratcliff and Childers (2015) reported that Chi-square fits of the 1D diffusion model to individual participants tend to vary between the critical value of Chisquare (roughly  $df + 2\sqrt{2 df}$ ) and twice the critical value,

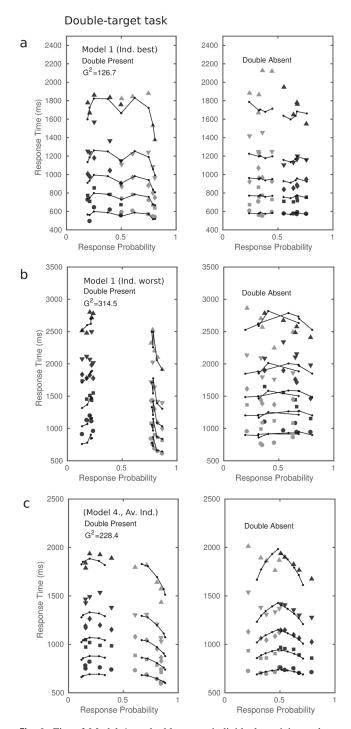


implying overdispersion of around 1.3 to 2.6, in agreement with Smith's estimates.

Using again the fit of Model 1 as a criterion for overdispersion, the estimates of q suggest that the individual double-target data were more variable than is usually found in tasks involving only a single stimulus. The overdispersion estimates, which are given with individual fit statistics in Table 6, ranged from 1.6 to 4.1. The lower part of Table 3 shows the average of the fit statistics for each of the four models. The average  $w_i(QAIC)$  weights are averages from computing Akaike weights for each participant individually. The individual weights are shown in Table 6. These weights suggest that, contrary to the group fit, the best fits at the individual level were provided by Model 4, which had ten free mean drift rates but only two drift variability parameters. Table 6 shows that the QAIC preferred Model 4 for five of the six participants and that in all of these cases, the evidence was overwhelmingly in favor of Model 4. The estimated parameters for the group and individual participants for all four model are shown in Table 7.

To illustrate the range of variability among individuals, Fig. 8 shows the fits of Model 1 for the best-fit (S1) and worst-fit (S6) participants (Fig. 8a and b, respectively). These plots show the differences among participants and highlight the presence of structure in the RT distributions that is more complex than is usually found in decision tasks of this kind. Like the group data in Fig. 7, there are large differences between target-present and target-absent trials, which are particularly apparent on the target-absent trials for S6, whose false alarm rate at the lowest stimulus contrast exceeds 60%. The model is able to predict this pattern of performance because of the differences in the drift rates for noise and signal stimuli. Many large-sample cognitive studies would be inclined to exclude participants like S6 on the grounds of excessive bias, but we believe it is more principled to try to characterize challenging data of this kind rather than to exclude them.

Despite some obvious places where the model misfits the data, it nevertheless captures much of the essential structure fairly well. It fails to capture the tail quantiles for S1 on target-absent trials and the leading edge of S6's target-absent error distributions (false alarms). The shape of the leading edge may be due to fast guess responses, which are outside the scope of model. The model does capture the atypical double-concave "bird-in-flight" pattern of S1's performance on target-present trials; this pattern of RTs indicates that the slowest correct and error responses occurred at intermediate levels of stimulus contrast. The model also captures the unusual convex bowing of the RT distribution quantiles for S6 on target-present trials and the marked asymmetry in S6's performance across the two stimulus types. In sum then, Model 1 appears to be



**Fig. 8** Fits of Model 1 to double-target individual participant data. Best (**a**) and worst (**b**) individual fits of Model 1. Best: Participant S1,  $G^2 = 126.7$ ; worst: Participant S6,  $G^2 = 314.5$ . **c** Quantile-averaged individual fits of Model 4;  $G^2$  is mean across participants. Other details as for Fig. 6

performing fairly well qualitatively and to capture the main features of the individual participant data.

As a contrast to Model 1, Fig. 8c shows the quantile-averaged fitted values of Model 4, which was the best individual model, together the quantile-averaged group



data. Although this model was preferred by the QAIC for most individual participants and again captures the main regularities of the data, qualitatively it appears somewhat underfit, in the sense that there are some aspects of the data that it not accounting for. It overpredicts accuracy at the lowest contrast on target-present trials; it underpredicts it at the highest contrast for on target-absent trials, and it fails to capture the shapes of the upper RT distribution quantiles for these stimuli.

Summary of double-target fits Although there are points of disagreement between the fits to the group data and the individual data, modeling of the data at either level provides strong evidence of a pronounced asymmetry in drift rates on target-present and target-absent trials. Model 1 allowed all 20 mean drift rates to vary; Model 2 allowed digit drift rates on target-present trials and letter drift rates on target-absent trials to vary and set all other drift rates to zero. For the group data, the QAICs for these two models were very similar and for the individual data, on average, the QAICs preferred Model 2. We also considered models in which the letter and digit drift rates had equal but opposite signs but we have not reported them because the fits were so poor. At the individual level, a model of this kind provided a credible account of the data for one participant only.

We consider the magnitude of the estimated drift rates in a later section, but, as we noted, asymmetry in the drift rates on target-present and target-absent trials imply the decision template is configural rather than local; it suggests that participants are searching for particular diagnostic patterns of stimulus elements. Our biased-template formulation represents an attempt to capture these regularities theoretically.

At the group level, similar fits were obtained from Model 2 and Model 3. Model 2 forced all of the mean drift rates for letter stimuli on target-present trials and digit stimuli on target-absent trials to zero, but allowed the drift rate variance to change with contrast. Model 3 allowed all 20 mean drift rates to vary but forced the drift rate variances of letters and digits to be the same for all contrasts. The fact that these two models performed similarly at the group level and also, on average, at the individual level, imply that either of these two representations can capture the main features of the data about as well as each other. That is, for the nondiagnostic parts of the decision template, contrast dependencies in drift-rate mean and in drift-rate variance have similar effects. This suggests that those components of the decision template that are not essential to distinguishing between target-present and target-absent trials are coded psychologically as "don't care." The speed and accuracy of responding is affected by across-trial variance in these parts of the decision template, but the value of the mean does not seem to be critical for performance. This contrasts with formal likelihood-ratio models of the double-target task Corbett and Smith (2017), in which positive and negative instances contribute equally to the joint likelihood.

The tradeoff in drift rate mean and variance seen at the group level was not as apparent at the individual level. On average, the QAICs for Models 2 and 3 were almost the same, but the QAIC-preferred model at the individual level was Model 4, in which the drift-rate means for nondiagnostic stimulus elements was zero and the drift-rate variances for letters and digits were constant. The fact that Model 4 was the model preferred by the QAIC suggest that across-trial variance in drift rates for the nondiagnostic parts of the decision template is not needed to account for the data—although, as Fig. 8c shows, the resulting model appears qualitatively as underfit. In view of the different results for the group and individual fits, we draw no strong inferences about the need for contrast-dependent changes in  $\eta$  in the model. However, the fits were clear in the need for different values of  $\eta$  for letter and digit stimuli.

The single-target task As we noted earlier, the singletarget task was more complex to model than the doubletarget task because of the need to account for response bias, which we conceptualized as a change in the placement of the decision boundaries on the surface of the hypersphere. We can think of the associated decision regions, in which the boundaries no longer align with the cardinal axes, as "generalized orthants." Corbett and Smith (submitted) found a need for a similar kind of bias in modeling the single-target conditions from Corbett and Smith's (2017) study using independent parallel 1D diffusion processes. The need for bias arises in independent channels models for the same reason as it does in the hyperspherical model, because there are many more ways in which processing can terminate with a target-present response than a target-absent response. As we noted in the "Does Dimensionality Matter?" section, the hyperspherical diffusion model and parallel independent-channels models share similar properties, because both are comprised of independent evidence-accumulation components.

**Group fits** We considered four models for the single-target task which paralleled those for the double-target task, which varied in the number of drift rate means and variances. To accommodate the structural asymmetry of the decision rule, we assumed that the response regions were the generalized orthants of the hypersphere, in which the decision bounds need not align with the cardinal axes. In a 2D version of the task, like the one shown in Fig. 3, the target-absent trials are associated with the stimulus configuration nn and the drift vector  $\{-\mu, -\mu\}$ . The target-absent responses are responses made in quadrant  $R_3$ , with decision bounds  $\pi$  and  $3\pi/2$ . We can introduce orthant bias into this model



by allowing the decision bounds to vary between  $3\pi/4$  and  $7\pi/4$  as  $\pi - c\pi/4$  and  $3\pi/2 + c\pi/4$ , with -1 < c < 1. In other words, positive and negative values of c symmetrically increase or decrease the size of the target-absent response region, respectively. In the 4D model, we parameterized the other two phase angles in a similar way. The region corresponding to target-absent responses is orthant 15 of Table 1. To facilitate estimation performance we treated c as a cumulative logistic function of an unbounded parameter, λ, that could vary over the entire real line. Parameterized in this way, unbiasedness is represented by  $\lambda = 0$  and an orthant bias towards noise responses is associated with  $\lambda > 0$ . When the drift rates for signal and noise stimuli are equal in magnitude and opposite in sign, an orthant bias of around  $\lambda = 0.5$  leads to similar frequencies of target-present and target-absent responses.

Table 5 summarizes the four models for the task with their associated parameters and Table 4 gives the details of the mapping of parameters to conditions. Models 5 to 8 correspond to Models 1 to 4 of Table 3, but unlike those models, the maximum number of mean drift rates in the most highly parameterized model was 15 rather than 20. These were the mean drift rates for signals on target-present trials, noise on target-present trials, and noise on target-absent trials, one for each level of contrast. As for the double-target task, we considered models with ten and two drift rate standard deviations. With the exception of the orthant bias,  $\lambda$ , the other parameters are as for the double-target task.

The pattern of model fits for the single-target task is somewhat clearer than the corresponding pattern for the double-target task. There is an overwhelming preference for the simplest of the four models, which had ten  $\nu$  parameters and two  $\eta$  parameters. The model assumes separate mean drift rates for signal stimuli on target-present trials and for noise stimuli on target-absent trials, but all of the mean drift rates for noise stimuli on target-present trials were zero. Allowing these mean drift rates to vary freely produced no improvement in fit. As with the double-target

task, allowing the drift rate standard deviation to vary with stimulus contrast produced no improvement over models in which there were just two drift rate standard deviations, one for signal stimuli and one for noise stimuli. Figure 9 shows quantile-probability plots of the fit of Model 8 to the group data, together with the best and worst individual fits.

The estimated model parameters show evidence of the same kind of asymmetry in mean drift rates that we found previously. The asymmetry can be characterized by assuming that it reflects the properties of the decision template for the task: on target-present trials, it consists of a single digit, with rest of the slots in the template coded as "don't care," and on target-absent trials it consists of a quartet of letters.

Individual fits The individual participant fits show the same pattern as the group fits. The average  $G^2$ , QAIC, and Akaike weights are shown in Table 5; the individual values of these statistics are shown in Table 6, and the associated parameter estimates are given in Table 8. For all six participants the preferred model was Model 8, with individual  $w_i$ 's ranging from .592 to .987. Both at the group and at the individual level, the preferred model was one with ten mean drift rates and two drift-rate standard deviations. We turn to the implications of the single-target and double-target fits in the following section.

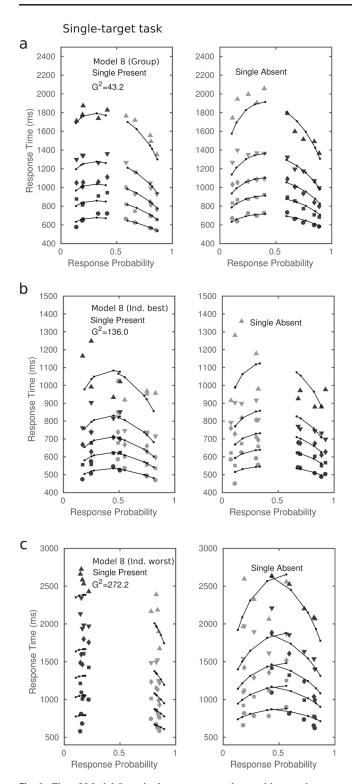
#### Mean drift rates

Figure 10 shows the estimated mean drift rates from the individual fits, averaged across participants for the double-target and single-target tasks. For each task we have shown the estimates for the most and the least highly parameterized of the four models. The estimates for the remaining models, which we have not shown, were similar. The plots emphasize the asymmetry in stimulus information that drives the decision process on target-present and target-absent trials. On target-present trials (two-digit trials in the double-target task and single-digit trials in the single-target

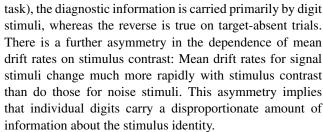
Table 5 Models for the single-target task

Model	m	df	$G^2$	QAIC	$w_i(\text{QAIC})$	Model F	Model Parameters						
Group Fits (q	Group Fits $(q = 0.485)$												
Model 5	29	81	39.3	139.0	0.00	15 ν	$10  \eta$	a	$T_{\rm er}$	$s_t$	λ		
Model 6	24	86	38.5	127.3	0.04	$10 \nu$	$10  \eta$	a	$T_{\rm er}$	$s_t$	λ		
Model 7	21	89	41.4	127.3	0.04	15 ν	$2 \eta$	а	$T_{\rm er}$	$s_t$	λ		
Model 8	16	94	43.2	120.8	0.92	$10 \nu$	$2\eta$	a	$T_{\rm er}$	$s_t$	λ		
Averaged Ind	ividual Fits	s (variable	q)										
Model 5	29	81	188.7	138.4	0.00	15 v	$10  \eta$	a	$T_{\rm er}$	$s_t$	λ		
Model 6	24	86	191.0	128.0	0.01	10 ν	$10  \eta$	a	$T_{\rm er}$	$s_t$	λ		
Model 7	21	89	197.3	124.5	0.06	15 ν	$2 \eta$	a	$T_{\rm er}$	$s_t$	λ		
Model 8	16	94	208.7	118.2	0.93	$10 \nu$	$2 \eta$	а	$T_{\rm er}$	$s_t$	λ		





**Fig. 9** Fits of Model 8 to single-target group data and best and worst individual participants. **a** Quantile-averaged group data. Best (**b**) and worst (**c**) individual fits. Best: Participant S4,  $G^2 = 136.0$ ; worst: Participant S1,  $G^2 = 272.2$ . Light gray symbols are target-present responses; dark gray symbols are target-absent responses



The one exception to this general observation is the nonmonotonic contrast dependence for signal stimuli (1s) for Model 1 in the double-target task in Fig. 10a. The fit of Model 1 to the group data in Fig. 7 showed the same nonmonotonicities as those in Fig. 10a, as did the fits of Model 3  $(20\nu, 2\eta)$  to the individual and group data, so it appears to be fairly robust and not an artifact of this particular fit. What these estimates appear to be showing is that, at low contrasts, single digits provide weak evidence for a targetpresent response, but at high contrast, the sign of the mean drift rate reverses and a single digit instead begins to provide evidence for a target-absent response. At the highest level of contrast, the magnitude of the 1s drift rate exceeds the 1ndrift rate, implying that the single digit target carries more evidence that the trial is a target-absent trial than do any of the three letter distractors individually. This kind of reversal seems paradoxical if drift rates are interpreted as depending only on the local properties of individual stimulus elements, but becomes understandable if they also depend on their configuration: At high contrast, the best evidence for a trial being a single-target trial may be a single, clearly perceived digit. Gleitman and Jonides (1976, 1978) and Jonides and Gleitman (1972, 1976) reported evidence of categorical pop-out effects of this kind in visual search.<sup>9</sup>

Another way to depict the mean drift rates, as shown in Fig. 11, is in a polar plot. Although only two of the four components of the drift rates can be shown in such a plot, they suffice to characterize how the drift rates behave across stimulus conditions in the two tasks. In generating the plot, we have assumed that when there were two digits presented they appeared in the first two positions (quadrant  $R_1$ ) and when there was a single digit it appeared at the first position (quadrant  $R_2$ ). The drift rates are shown as vectors whose lengths represent the norms of the mean drift rates,  $\nu$ , and whose directions represent their phase angles. The fact that the drift vectors for target-present stimuli in the double-target



<sup>&</sup>lt;sup>9</sup>Jonides and Gleitman's (1972) claim of a category pop-out effect in the complete absence of orthographic cues remains controversial. Duncan (1983) reported a failure to replicate their findings with the character "O," when it was described as the letter "oh" in one condition and the digit zero in another. He did, however, replicate their finding of more efficient search for across-category targets than within-category targets. The category pop-out effect reported by Jonides and Gleitman likely underlies the acquisition of automaticity in letter-digit search under consistent mapping conditions reported by Shiffrin and Schneider (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Table 6 Individual model fits

Part	icipant	Double	Target	t					Single	Target				
	q	Model	m	df	$G^2$	QAIC	$w_i$	$\overline{q}$	Model	m	df	$G^2$	QAIC	$w_i$
$S_1$	1.64	1	33	77	126.7	143.1	.000	3.06	5	29	81	230.6	136.4	.000
		2	23	87	149.9	137.3	.000		6	24	86	234.5	125.2	.072
		3	25	85	131.5	130.2	.013		7	21	89	245.1	122.2	.334
		4	15	95	149.9	121.4	.987		8	16	94	272.2	121.1	.592
$S_2$	2.01	1	33	77	155.2	143.1	000	2.90	5	29	81	232.6	138.6	.000
		2	23	87	162.3	126.7	.037		6	24	86	234.7	129.2	.002
		3	25	85	178.1	138.7	.000		7	21	89	240.2	124.2	.018
		4	15	95	181.4	120.2	.963		8	16	94	245.9	116.8	.980
$S_3$	1.71	1	33	77	132.2	143.1	.985	2.07	5	29	81	167.3	139.0	.000
		2	23	87	181.2	151.9	.013		6	24	86	169.1	129.8	.001
		3	25	85	183.7	157.4	.001		7	21	89	169.9	124.2	.012
		4	15	95	221.6	159.6	.000		8	16	94	172.3	115.4	.987
$S_4$	2.23	1	33	77	171.9	143.1	.000	1.68	5	29	81	130.4	138.9	.000
		2	23	87	186.2	129.5	.000		6	24	86	132.6	127.0	.063
		3	25	85	176.0	128.9	.000		7	21	89	139.6	125.0	.170
		4	15	95	183.4	112.2	1.0		8	16	94	151.3	122.0	.767
$S_5$	2.20	1	33	77	169.6	143.1	.002	2.89	5	29	81	228.4	139.0	.000
		2	23	87	231.9	151.4	.000		6	24	86	231.5	128.1	.020
		3	25	85	177.7	130.8	.998		7	21	89	239.8	124.9	.096
		4	15	95	267.5	151.6	.000		8	16	94	255.8	120.5	.884
$S_6$	4.08	1	33	77	314.5	143.1	.000	1.78	5	29	81	142.3	138.7	.000
		2	23	87	341.3	129.6	.007		6	24	86	143.7	128.1	.006
		3	25	85	356.4	137.4	.000		7	21	89	149.1	125.7	.032
		4	15	95	366.6	119.8	.993		8	16	94	154.8	119.0	.962

task appear as collinear is a function of the polar plot, which does not show the third and fourth components of the drift rates, but the collinearity of the drift vectors for target-absent stimuli in the single-target task is an accurate reflection of the way in which we paramaterized the drift rates (Table 4), using a single parameter at each level of contrast.

If the mean drift rates were veridical representations of stimulus contrast, then the drift vectors for target-absent stimuli in the double-target task and for target-present stimuli in the single-target task would have phase angles of  $3\pi/4$ . In both tasks, however, these vectors are rotated away from veridical in a direction that increases the Euclidean distance between the two classes of stimuli. For the doubletarget task, this difference increases with stimulus contrast, but in the single-target task it is present at all levels of contrast and its magnitude is almost constant. Models 2 and 4 for the double-target task, and Models 6 and 8 for the single-target task, provide a way of characterizing the magnitude of the observed rotation theoretically. These four models assume that the mean drift rates for letters on double-target trials and digits on single-target trials are zero. They predict that the phase angles of target-present drift rate vectors in the single-target task will be  $\pi/2$  and of target-absent drift rate vectors in the double-target task will be  $\pi$ . The drift-rate vectors for the single-target task in Fig. 11 show that this property is close to being satisfied at all levels of stimulus contrast. This finding agrees with the strong and consistent support found for Model 8 in both the group and individual fits. Relative to Model 6, Model 8 makes the extra assumption that drift-rate variance is independent of stimulus contrast, which was also supported by the model fits.

In comparison to the single-target task, the mean drift rate vectors for the double-target task are more variable. The phase angles for single-digit stimuli are less than  $\pi$  at the lowest level of contrast, almost exactly equal to  $\pi$  at the three intermediate levels of contrast, and greater than  $\pi$  at the highest level of contrast. The substantive hypothesis embodied in Models 2 and 4 is that the fit is not improved by allowing these phase angles to differ from  $\pi$ . As we noted previously, the pictures obtained from the group and individual fits disagree on this particular point. The group fits imply that these phase angles are contrast dependent, but the individual fits imply that the phase angles are uniformly equal to  $\pi$ .



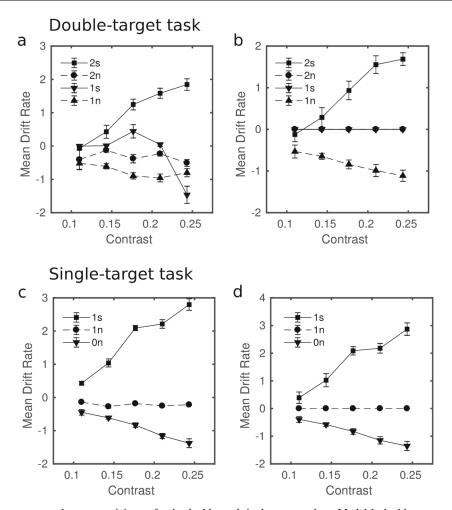
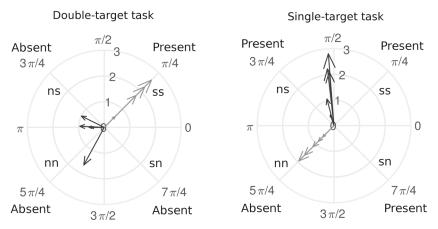


Fig. 10 Mean drift rates,  $\nu$ , averaged across participants for the double- and single-target tasks. **a** Model 1, double-target task; **b** Model 4, double-target task; **c** Model 5, single-target task; **d** Model 8, single-target task. **a** and **b** 2s = signal stimuli, target-present trials; 2n = noise stimuli, target-present trials; n = noise stimuli, target-absent trials; n = noise stimuli, target-present trials; n = noise stimuli, target-absent trials; n = noise stimuli, target-present trials; n = noise stimuli, target-absent trials. n = noise stimuli, target-absent trials. n = noise stimuli, target-absent trials.



**Fig. 11** Polar plot of the first two components of the mean drift vectors for double-target and single-target tasks. For both tasks, the collinear *light gray vectors* represent drift rates in displays in which the first two components were the same: two digits in the double-target task and two letters in the single-target task. The variable *dark gray vectors* 

represent drift rates for displays containing one digit and one letter. The figure assumes that in mixed displays of this kind, the digit is in the first position and the drift vector lies in quadrant  $R_2$ . (In the mapping from polar to Cartesian coordinates of Eq. 3,  $X^1$  is the vertical axis and  $X^2$  is the horizontal axis.)



The estimated phase angles in Fig. 11 help shed light on why we had difficulty in distinguishing between alternative models for the double-target task: Except at the highest level of contrast, where the phase angles is around  $5\pi/4$ , all of the other phase angles are either close to, or equal to  $\pi$ . Figure 11 suggests that the lack of agreement between the group and individual fits largely comes down to what happens at high contrast, where perceptual clarity was high. As we suggested above, these conditions may have produced a categorical pop-out effect, leading to the otherwise paradoxical finding that a single, clearly perceived digit can provide more evidence for the trial being a target-absent trial than do any of the three letters individually. Apart from this point of disagreement between models, our results provide consistent support for the idea that nondiagnostic stimulus elements make either a small or negligible contribution to mean drift rates.

Had it been the case that the drift vectors for targetpresent and target-absent of stimuli in the two tasks were oriented at ±180° to each other, then this could have been interpreted as showing that the decision process projects the stimulus information onto a lower-dimensional subspace and that a simpler 1D diffusion model might be an appropriate one for the task. However, the drift vectors in Fig. 11 suggest that any dimension reduction is at best partial rather than complete and that the stimulus information continues to be represented psychologically in a higher-dimensional space. The drift vectors for the single-target task are more regular than those for the doubletarget task and are consistently oriented at around  $3\pi/4$ (135°) apart, suggesting that the dominant component of the discriminative information used to decide between the two alternatives is encoded in a single stimulus dimension. That this is so is consistent with Corbett and Smith's (submitted) finding that a 1D decision model can provide a satisfactory model for the single-target task.

The greater complexity of the drift rate vectors for the double-target task suggest that any dimension reduction performed by the decision process, if it occurs, happens only at the highest level of stimulus contrast, where accuracy is around 85-90%. At lower levels of contrast, the drift vectors show evidence of being higher dimensional. The greater complexity of the drift rate vectors for the double-target task is consistent with the greater complexity of the associated RT distributions in Figs. 7 and 8, and may explain why we have not been able to find a satisfactory parameterization of the 1D diffusion model for these data.

#### **Discussion**

The theory presented in this article seeks to provide a new way to conceptualize decisions about the contents of multi-element arrays. Our orientation is to view a stimulus array comprised of multiple elements as a single multidimensional stimulus, whose properties are represented mathematically by a point in a vector space. The hyperspherical model represents the process of making a decision about such an array as a single, multidimensional evidence accumulation process with independent components, each of which is a diffusion process. The outcome of the decision process is a description of the contents of the display as a configuration of digits and letters. We assume that the reduction in dimensionality needed to express the decision overtly as a two-choice response involves a response selection process (Sternberg, 1969) that begins once the evidence accumulation process reaches the decision boundary. The response selection process carries out the mapping from the orthants of the decision space to overt responses depicted in Table 1. The fits reported in the previous section show that the hyperspherical model provides a good account of the distributions of RT and the choice probabilities from both the double-target and single-target versions of the task. The fact that the model provides a good fit to our data implies that letter-digit discrimination in four-item arrays can be performed in parallel, without attention switching or any other kind of serial processing—a finding that agrees with the conclusions of Dosher et al. (2010) who studied parallel and serial processing in an orientation discrimination task.

The most striking and unexpected finding from our study was the evidence of what we have termed "configural" processing of four-item displays. We use this term to express the idea that the way in which an item at one location is processed depends on the contents of other locations, that is, on the configuration of the display as a whole. In the hyperspherical diffusion model, the rate of evidence accumulation is given by a vector-valued drift rate,  $\mu$ , whose components represent the evidence for a digit or a letter at each display location. Our fits showed that the components of drift rate depend on the number of similar elements (letters or digits) in the array. These kinds of dependencies could not arise if the components depended only on the local features of the display (i.e., on the contents of individual locations). We considered several models for the two tasks, which differed in the way in which they parameterized mean drift rates and drift rate variance, but all of the viable alternatives assumed that the drift-rate components for digits in the double-target task depend on whether there were one or more digits in the display (Fig. 10a and b). We also considered models in which these components were constrained to be equal, but they performed so badly that we have not presented them.

One plausible source of configurality in our double-target task is the one identified by Corbett and Smith (2017) in their study of the double-target detection deficit. They



ascribed the deficit to the attentional selection processes that control access to visual working memory, interacting with working memory capacity limitations. We noted earlier that letter-digit discrimination yielded one of the smaller double-target deficits found in Corbett and Smith's study, but the effect was nevertheless present, as can be inferred by comparing the estimated target-present drift rates (filled squares) across tasks in Fig. 10. For both of the models shown, the mean drift rates for members of a pair of digits (2s) in the double-target task (Fig. 10a or b) are less than the corresponding drift rates for single digits (1s) in the single-target task (Fig. 10c or d). This difference implies that digits as members of a target pair are less well represented in memory than are single-digit targets in isolation.

The double-target deficit is one of the most compelling pieces of evidence supporting two-stage theories of search, of which the best-known is probably Wolfe's guided search theory (Wolfe et al., 1989, 2010). Two-stage theories assume that attentional selection is carried out by preattentive filtering processes that select stimuli by matching them against a cognitive representation of the search targets. Stimuli that pass the filter have privileged access to later processing stages; nonmatching stimuli are either excluded or are represented in attenuated form. Corbett and Smith (2017) showed that the magnitude and the time course of the double-target deficit in their study were well described by the competitive interaction model of attention and decision-making of Smith and Sewell (2013), which implements preattentive filtering using competitiveinteraction dynamics (Grossberg, 1980, 1988).

Preattentive filtering provides a natural account of the nonlocality in drift rates we estimated in our models. A basic tenet of two-stage theories is that the quality of stimulus representations in the decision process will not only depend on how well the stimuli are encoded perceptually but will also depend on how well they match the cognitive representations of the search targets. Nonlocal interactions arise in models like that of Smith and Sewell (2013) because of competition among stimuli for access to later processing stages. In their model, competition for access to visual working memory leads to normalization of stimulus representations (Carandini & Heeger, 2012). Normalization is a process in which the activity in a mechanism representing a stimulus is reduced in proportion to the sum of the activity in other, related mechanisms. In Smith and Sewell's model, competition among stimuli for access to visual working memory leads to normalization of memory trace strengths. The trace strengths determine the drift rates in a diffusion decision model, which we modeled here as diffusion in a hypersphere. Indeed, one of our reasons for developing the hyperspherical model was to provide a decision model that is compatible with the parallel, interactive view of attentional selection in the Smith and Sewell model.

Our current findings extend the picture from Corbett and Smith (2017) study by showing the configural nature of attentional selection. We characterized the goal of the double-target task as search for a pair of digits or a triplet of letters. These representations, because they involve combinations of stimulus elements rather single elements in isolation, are configural rather than local, in our sense of the term. The fact that performance in our tasks could be characterized by models in which the means of the drift rates for the nondiagnostic parts of the decision template—letters on target-present trials and digits on target-absent trialsare uniformly zero is consistent with the idea that drift rates are a product of attentional filtering by a configural decision template. The means of stimulus elements that do not pass the filter are set to zero: They may contribute noise to the output of the filter but they do not contribute to the information entering the decision process in any systematic way.

We have not attempted to relate the estimates of drift rate from our diffusion-model fits to any particular model of preattentive filtering; instead, we characterized them in general terms as the outputs of a biased decision template. To us, however, the most plausible interpretation of this kind of configuration-dependent bias is that it is a product of preattentive filtering, possibly via a process of competitive interaction like the one in the Smith and Sewell (2013) model. Indeed, as discussed by Corbett and Smith (submitted), the scaling relationships among the drift rates obtained from our model fits can potentially be explained by a process of trace-strength normalization acting on the output of a biased decision template like the one in Eq. (14). Viewed in this light, the "argmax" function in Eq. (14) can be interpreted as selecting the winner of a competition on the basis of a similarity-matching rule.

#### The limits of unitization

A basic tenet of the hyperspherical model is that a display comprised of multiple elements is treated in a unitary way, as a single decision about a vector-valued perceptual object. A question we have not addressed concerns the limit on the number of stimulus items that can be unitized in this way. Although we have no direct evidence that bears on this question, we think the maximum is likely to be around four items, which was, by design, the size of the stimulus displays used in Corbett and Smith's (2017) study. According to many estimates, four items is the item-capacity limit of visual working memory (Cowan, 2001; Luck & Vogel, 1997; Shibuya & Bundesen, 1988). The 4D diffusion model presupposes that items enter the decision process in parallel. In tasks that use brief stimulus displays, they must all



be represented in visual working memory—although, as the preceding discussion highlighted, preattentive filtering processes may affect the means of the individual item representations and may even force them to zero.

Whether or not visual working memory actually has a hard item-capacity limit remains highly controversial (Oberauer & Lin, 2017; van den Berg et al., 2014), but even if no such limit exists, the costs associated with forming and maintaining representations of more than four items may increase sharply once this number is exceeded. These costs have been conceptualized both as resource costs (Donkin et al., 2016) and as interference costs (Endress & Szabó, 2017; Oberauer & Lin, 2017). Perhaps coincidentally, four items is the usual estimate of the *subitizing* limit, that is, the size of the set of items whose cardinality can be apprehended in a single glance (Trick & Pylyshyn, 1994). It is also the size of the candidate set of items that later versions of guided search theory suggest can be searched in parallel (Wolfe et al., 2010).

As well as depending on the number of items in the display, the unitization of stimuli in the decision process may be affected by their spatial arrangement and physical separation. Mental architecture studies using systems factorial technology have found that whether stimuli composed of feature pairs are processed serially or in parallel depends in part on whether the features are spatially coincident or separated, with increased evidence of serial processing when they are more widely separated (Fific et al., 2010; Little et al., 2011; Moneer et al., 2016). Complementing these findings, attempts to model visual search for targets in large displays using the diffusion decision model by Moran et al. (2013) and Schwarz and Miller (2016) have had to postulate additional components of RT attributable to serial attention switching and premature search termination in order to fit data.

We characterized the decision process in the letter-digit task using a 4D diffusion model because a four-dimensional evidence space seems the natural setting in which to model decisions about four-item displays. Conceivably, it might be argued that even four dimensions is insufficient, because there were actually 16 different stimuli that could appear at each location and, if each stimulus were associated with its own evidence accumulator, then there would be 16<sup>4</sup> rather than 2<sup>4</sup> unique stimulus configurations that the decision process would have to identify. Apart from the analytic intractability of this kind of model, which makes it a priori unappealing, we think there are good reasons for regarding it as unlikely. Corbett and Smith's (2017) use of letter-digit discrimination in their double-target deficit study was based on Duncan's (1980) article, which drew on earlier work by Jonides and Gleitman who studied such discriminations in visual search. They found that categorically distinct stimuli like letters and digits need not be identified in order to be discriminated (Gleitman & Jonides, 1976; Jonides & Gleitman, 1976), and can be discriminated on the basis of their category membership alone (Jonides & Gleitman, 1972; Gleitman & Jonides, 1978). We suggested that the sign reversal of the mean drift rates for high-contrast, single-digit stimuli in the double-target task may be due to a pop-out effect of this kind.

Consistent with this interpretation, Corbett and Smith (submitted) found that letter-digit discrimination produced the same kinds of quantile-probability functions as three other decision tasks that used only single pairs of stimulus alternatives. There was no evidence of any difference among the RT distributions from these tasks that might have indicated a change in the dimensionality of the decision process in letter-digit discrimination. Taken together, these results suggest that it is appropriate to view letter-digit discrimination as a unidimensional rather than a multidimensional decision task and imply that a 4D vector space is the appropriate representation of a four-element version of it.

#### Is the hyperspherical model the best model?

It is not our intention at this point to claim that the 4D diffusion model is the best among potential candidate models for the task. As we noted in the Introduction, Corbett and Smith (submitted) successfully fitted a version of the 1D diffusion model to the data from the single-target task, but we have not yet been able to find a model for the doubletarget task that fits the data with a consistent and interpretable set of parameters. We were motivated to develop the 4D model as an alternative to existing models by the conjecture that decisions about multiattribute stimuli may involve multidimensional evidence accumulation processes and the fact that the 4D diffusion model provides a theoretically elegant representation of processes of this kind. We have not yet attempted a systematic empirical comparison of the 4D model and plausible alternative models for all of the data reported in Corbett and Smith (2017) study. While a large-scale evaluation of this kind would be useful, it is beyond the scope of this article and will form the subject of future work. Our immediate purpose here was to present the theory of the 4D model and to show that it provides a plausible account of performance in the double-target task, which is one of its areas of potential application.

The scope of the 4D model is, of course, much greater than we have been able to illustrate here. As we noted in the Introduction, the 2D model of Smith (2016) and its 4D generalization provide a stochastic version of Ashby and Townsend's (1986) general recognition theory, which seeks to explain how people make decisions about stimuli composed of multiple features. Cognitive tasks using these



kinds of stimuli are widely studied (e.g., Soto and Ashby (2015) and Soto et al. (2015), and an analytically tractable model that can make both RT distribution and accuracy predictions is likely to have broad theoretical appeal. Another area of potential application is to studies of visual attention. As we noted earlier, in Wolfe's guided search theory of attention, preattentive filtering processes deliver a subset of candidate items to a decision process, which searches the subset to determine whether or not it contains a target. In early versions of guided search (Wolfe et al., 1989), the search of the candidate set was assumed to be serial, but in later versions (Wolfe et al., 2010), the search is assumed to be parallel and to be performed by multiple independent diffusion processes. Thornton and Gilden (2007) assumed a similar kind of parallel search model in their study of redundancy gains in visual search. The 4D diffusion model instantiates a parallel search mechanism of precisely this kind. It can therefore provide a theoretically principled and analytically tractable characterization of RT and accuracy from visual search tasks at a distributional level.

#### **Conclusions**

In this article, we have extended the circular diffusion model of Smith (2016) to higher dimensions. We have shown that the analytic properties of RT distributions and accuracy for the circular model carry over straightforwardly to higherdimensional spaces. We have shown that the resulting hyperspherical model, when equipped with categorical decision bounds, can provide a model of multiattribute decision-making that possesses a number of nice theoretical and empirical properties. We argued that decisions about the contents of four-item displays can be viewed as unitary, multidimensional decisions about vector-valued stimulus objects and have shown that the RT distributions and accuracy from such a task could be well described by the hyperspherical model. The success of the model in capturing these data is notable, because they exhibit several unusual features not normally found in tasks with single stimuli and, to date, we have been unable to capture them in a theoretically satisfying way using extensions of existing models of two-choice decision-making. Our model fits have highlighted the role of nonlocal, configural stimulus properties in determining the drift-rate vectors that characterize the rates of evidence accumulation. We showed that the resulting configural drift rates could be represented as the outcomes of a matching operation between stimuli and the decision template, in which the latter is conceived of as a set of diagnostic patterns that distinguish between the response categories. As well as providing a new model for decisions about multielement arrays, the hyperspherical model is likely to have a variety of applications in areas such as cognitive categorization and visual attention.

Author Note The research in this article was supported by Australian Research Council Discovery Grants DP140102970 and DP180101686 to Philip Smith and a Government of Ireland Postdoctoral Research Fellowship awarded to Elaine Corbett by the Irish Research Council. We thank James Townsend, Mario Fific, and an anonymous reviewer for helpful comments on an earlier version of the manuscript. Data and Matlab and C code for the model can be downloaded from https://github.com/philipls/4DDiffusion.

### Appendix A: Across-trial variability in drift rates

Across-trial variability in drift rate has been a feature of the 1D diffusion model since it was proposed as a model of decision-making by Ratcliff (1978). Implementations of the model typically integrate across drift rates numerically (e.g., Ratcliff and Smith (2004)), which can be done efficiently by parallelizing the code. Greater efficiency can be achieved by performing the integration analytically, and for some time now we have used an implementation of the 1D model that integrates across drift rates in this way. Other researchers have also been aware of the possibility of performing this kind of calculation analytically (Wiecki et al., 2013), but details were published only recently, by Blurton et al. (2017), who derived an analytic expression for the cumulative distribution function for the 1D diffusion model with normally distributed drift rates. The derivation involves a "completion of the square" calculation similar to that involved in calculating the moment generating function or characteristic function of the normal distribution. Ratcliff (1978) carried out a similar calculation to derive d'(t)for a 1D diffusion process in time-dependent processing tasks. We derived an analytic version of the hyperspherical diffusion model with normally distributed drift rates that we used to fit the models investigated in this article, the details of which are given below.

We rewrite the probability density function of the 4D model of Eq. 4 in the form

$$d\tilde{P}_t(\boldsymbol{\mu}, a, \sigma^2) = d\tilde{P}_t(\boldsymbol{\Theta}_T)$$

in order to make explicit its dependence on the drift rate vector  $\boldsymbol{\mu} = \{\mu_1, \mu_2, \mu_3, \mu_4\}$ , the decision criterion, a, and the diffusion coefficient,  $\sigma^2$ . We assume that the drift rates across trials have a multivariate normal distribution,  $N(\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\Sigma})$ , with mean vector  $\boldsymbol{\nu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . We consider only the simplest case of  $\boldsymbol{\Sigma} = \eta^2 \boldsymbol{I}$ , where  $\boldsymbol{I}$  is the  $4 \times 4$  identity matrix. That is, the four components of the drift rate are independent with equal variance,  $\eta^2$ . We denote by  $d\tilde{Q}_t(\boldsymbol{\nu}, \eta^2, a, \sigma^2)$  the joint density function of the first hitting point and first passage time, marginalized across



variation in drift rate. This density function requires us to evaluate the iterated integral

$$d\tilde{Q}_t(\mathbf{v}, \eta^2, a, \sigma^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\boldsymbol{\mu}; \mathbf{v}, \eta^2 \mathbf{I}) d\tilde{P}_t(\boldsymbol{\mu}, a, \sigma^2) d\mu_1 d\mu_2 d\mu_3 d\mu_4. \tag{A1}$$

When the covariance matrix is diagonal,  $N(\mu; \nu, \eta^2 I)$  factorizes into a product of independent 1D components,

$$N(\mu; \nu, \eta^2 I) = \prod_{i=1}^4 \frac{1}{\sqrt{2\pi \eta^2}} \exp\left[-\frac{(\mu_i - \nu_i)^2}{2\eta^2}\right].$$
 (A2)

Equation 4 states that  $d\tilde{P}_t(\boldsymbol{\mu}, a, \sigma^2) = d\tilde{P}_t(\boldsymbol{\Theta}_T)$  is the product of  $dP_t(a)$ , the hitting time density of the Bessel process of Eq. (2) and  $\mathbf{Z}_T(\mathbf{X})$ , the exponential martingale of Eq. (5). The latter factorizes in the same way as does (A2) into a product of four components,

$$\mathbf{Z}_{T}(\mathbf{X}) = \prod_{i=1}^{4} Z_{T}^{i}(X^{i}) = \prod_{i=1}^{4} \exp\left[\frac{1}{\sigma^{2}} \mu_{i} X_{T}^{i} - \frac{1}{2\sigma^{2}} \mu_{i}^{2} T\right],$$
(A3)

where, as before, the values of  $X_T^i$  are points on the boundary of the hypersphere given by the coordinate transformation of Eq. (3) as the values of  $\Theta = \{\psi, \phi, \theta\}$  are parametrically varied. As a result of this factorization, the iterated integral of (A1) can be reduced to a set of componentwise integrals of the form

$$\bar{Z}_T^i(X^i) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\eta^2}} \exp\left[-\frac{(\mu_i - \nu_i)^2}{2\eta^2}\right] \times \exp\left[\frac{1}{\sigma^2}\mu_i X_T^i - \frac{1}{2\sigma^2}\mu_i^2 T\right] d\mu_i, \tag{A4}$$

where we have used the overbar notation,  $\bar{Z}_T^i(X^i)$ , to denote the *i*th component of  $Z_T^i(X^i)$  marginalized (averaged) over the *i*th component of  $N(\mu; \nu, \eta^2 I)$ . We write the integrand of (A4) in the form

$$\frac{1}{\sqrt{2\pi\eta^2}} \exp\left[\frac{-(c\eta^2 + 1)\mu_i^2 + 2(b\eta^2 + \nu_i)\mu_i}{2\eta^2} - \frac{\nu_i^2}{2\eta^2}\right]$$
(A5)

where we have defined  $b = X_T^i/\sigma^2$  and  $c = T/\sigma^2$ . The square in  $\mu_i$  in the exponent on the left hand side of (A5) can be completed by adding and subtracting appropriately chosen terms. Explicitly,

$$\begin{split} \frac{1}{2\eta^2} \left[ -(c\eta^2+1)\mu_i^2 + 2(b\eta^2+\nu_i) \frac{c\eta^2+1}{c\eta^2+1} \mu_i - \frac{(b\eta^2+\nu_i)^2}{(c\eta^2+1)^2} (c\eta^2+1) \right] \\ + \left[ \frac{(b\eta^2+\nu_i)^2}{2\eta^2 (c\eta^2+1)} - \frac{\nu_i^2}{2\eta^2} \right] \end{split}$$

or after collecting terms,

$$-\frac{(c\eta^2+1)}{2\eta^2}\left[\mu_i - \frac{b\eta^2+\nu_i}{c\eta^2+1}\right]^2 + \left[\frac{(b\eta^2+\nu_i)^2}{2\eta^2(c\eta^2+1)} - \frac{\nu_i^2}{2\eta^2}\right].$$

The term in square brackets on the left of this expression is the exponent of a normal density in  $\mu_i$  with mean  $(b\eta^2 + \nu_i)/(c\eta^2 + 1)$  and variance  $\eta^2/(c\eta^2 + 1)$ . Multiplying the right hand side of (A4) by  $\sqrt{(c\eta^2 + 1)}/\sqrt{(c\eta^2 + 1)}$ , and integrating over  $\mu_i$ , the integral evaluates to unity, which eliminates  $\mu_i$  from the equation to leave

$$\bar{Z}_T^i(X^i) = \frac{1}{\sqrt{(\eta/\sigma)^2 T + 1}} \exp\left\{ -\frac{v_i^2}{2\eta^2} + \frac{[X_T^i(\eta/\sigma)^2 + v_i]^2}{2\eta^2[(\eta/\sigma)^2 T + 1]} \right\},$$
(A6)

after substituting for b and c. The product of the four  $\bar{Z}_T^i(X^i)$  terms gives the marginal form of  $\bar{\mathbf{Z}}_T(\mathbf{X})$  in Eq. 11 that is used to obtain the joint density of the first hitting point and first passage time in the presence of drift variability.

#### Software implementation

We implemented the model in two ways, one in C called from Matlab and the other in Julia. The computational efficiency of the two implementations is almost identical. The C implementation used the GNU scientific library, GSL, to evaluate Bessel functions and to find their roots. We discretized the time axis into 300 equal steps and the circumference of the circle into 36 equal angular steps, each of 10°, resulting in 11,664 points on the surface of the hypersphere, which yields 3,499,200 distinct points of the joint density. This computation took around 0.6-0.8 seconds on Linux computers equipped with 3.10-3.33 GHz Xeon processors, or 6-8 seconds per point in the parameter space, which required evaluation of ten different joint densities. A typical model fit (assuming around 1000 iterations of Simplex) required around 1.5 hours.

## Appendix B: Individual model fits and parameter estimates

In Table 7,  $\nu_1$  to  $\nu_5$  are the mean drift rates for digits on double-target trials;  $\nu_6$  to  $\nu_{10}$  are the mean drift rates for



 Table 7
 Parameter estimates: double-target task (group and averaged individual)

Model	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$\nu_6$	$\nu_7$	$\nu_8$	ν <sub>9</sub>	$v_{10}$
1 (Group)	-0.053	0.578	1.289	1.596	1.814	-0.282	-0.069	-0.192	-0.201	-0.430
1 (Indiv.)	-0.070	0.427	1.245	1.582	1.845	-0.409	-0.130	-0.378	-0.233	-0.509
2 (Group)	-0.352	0.280	1.003	1.435	1.682	0	0	0	0	0
2 (Indiv.)	-0.120	0.326	0.946	1.382	1.633	0	0	0	0	0
3 (Group.)	-0.078	0.403	1.211	1.599	1.826	-0.271	-0.072	-0.198	-0.186	-0.421
3 (Indiv.)	-0.057	0.466	1.126	1.567	1.711	-0.342	-0.181	-0.292	-0.297	-0.420
4 (Group)	0.000	0.276	0.990	1.483	1.670	0	0	0	0	0
4 (Indiv.)	-0.124	0.286	0.933	1.553	1.685	0	0	0	0	0
Model	$v_{11}$	$v_{12}$	$\nu_{13}$	$v_{14}$	$v_{15}$	$v_{16}$	$v_{17}$	$v_{18}$	$v_{19}$	$\nu_{20}$
1 (Group)	0.000	0.016	0.786	0.070	-1.285	-0.519	-0.627	-0.865	-0.872	-0.808
1 (Indiv.)	-0.004	0.012	0.441	0.046	-1.465	-0.520	-0.609	-0.897	-0.956	-0.800
2 (Group)	0	0	0	0	0	-0.525	-0.617	-0.772	-0.850	-1.013
2 (Indiv.)	0	0	0	0	0	-0.493	-0.613	-0.808	-0.902	-1.028
3 (Group)	-0.004	0.015	0.656	0.067	-1.304	-0.518	-0.625	-0.846	-0.897	-0.816
3 (Indiv.)	-0.004	0.019	0.277	0.047	-1.114	-0.478	-0.654	-0.839	-0.965	-0.880
4 (Group)	0	0	0	0	0	-0.496	-0.609	-0.764	-0.860	-1.034
4 (Indiv.)	0	0	0	0	0	-0.532	-0.646	-0.840	-0.988	-1.114
Model	$\eta_1$	$\eta_2$	$\eta_3$	$\eta_4$	$\eta_5$	$\eta_6$	$\eta_7$	$\eta_8$	$\eta_9$	$\eta_{10}$
1 (Group)	1.462	1.605	1.613	1.805	1.780	0.781	0.649	0.290	0.449	0.256
1 (Indiv.)	2.706	1.382	1.594	1.455	1.738	1.574	0.774	0.722	0.543	0.166
2 (Group)	1.431	1.736	1.864	2.034	2.004	0.883	0.694	0.536	0.462	0.280
2 (Indiv.)	1.539	1.570	1.683	1.906	1.901	0.731	0.688	0.414	0.437	0.104
3 (Group)	1.742					0.546				
3 (Indiv.)	1.541					0.569				
4 (Group)	1.770					0.556				
4 (Indiv.)	1.829					0.647				
Model	a	$T_{\rm er}$	$s_t$							
1 (Group)	2.751	0.169	0.050							
1 (Indiv.)	2.706	0.175	0.038							
2 (Group)	2.809	0.177	0.040							
2 (Indiv.)	2.647	0.177	0.043							
3 (Group)	2.800	0.174	0.046							
3 (Indiv.)	2.667	0.161	0.036							
4 (Group)	2.792	0.171	0.040							
4 (Indiv.)	2.756	0.173	0.032							

Note. "Indiv." = average of parameters estimated from individual fits

 Table 8
 Parameter estimates: single-target task (group and averaged individual)

Mo del	$\nu_1$	$\nu_2$	ν <sub>3</sub>	ν <sub>4</sub>	ν <sub>5</sub>	$v_6$	ν <sub>7</sub>	ν <sub>8</sub>	ν9	$\nu_{10}$
5 (Group)	0.493	1.099	2.162	2.289	2.856	-0.174	-0.188	-0.190	-0.193	-0.181
5 (Indiv.)	0.321	0.948	2.091	2.204	2.903	-0.198	-0.168	-0.160	-0.168	-0.171
6 (Group)	0.428	0.957	2.042	2.177	2.771	0	0	0	0	0
6 (Indiv.)	0.316	0.918	2.075	2.151	2.897	0	0	0	0	0
7 (Group.)	0.411	1.098	2.077	2.242	2.768	-0.127	-0.273	-0.199	-0.277	-0.219
7 (Indiv.)	0.310	1.064	2.129	2.339	2.991	-0.121	-0.225	-0.134	-0.223	-0.236



Mo del	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$v_6$	$\nu_7$	$\nu_8$	ν9	$v_{10}$
8 (Group)	0.512	1.105	2.165	2.366	2.956	0	0	0	0	0
8 (Indiv.)	0.390	1.026	2.087	2.179	2.870	0	0	0	0	0
Model	$v_{11}$	$v_{12}$	$\nu_{13}$	$v_{14}$	$v_{15}$					
5 (Group)	-0.376	-0.583	-0.825	-1.140	-1.352					
5 (Indiv.)	-0.479	-0.642	-0.881	-1.195	-1.416					
6 (Group)	-0.318	-0.528	-0.754	-1.079	-1.291					
6 (Indiv.)	-0.419	-0.583	-0.839	-1.150	-1.387					
7 (Group)	-0.367	-0.579	-0.798	-1.106	-1.305					
7 (Indiv.)	-0.464	-0.684	-0.921	-1.264	-1.450					
8 (Group)	-0.314	-0.560	-0.800	-1.132	-1.364					
8 (Indiv.)	-0.392	-0.580	-0.828	-1.151	-1.353					
Model	$\eta_1$	$\eta_2$	$\eta_3$	$\eta_4$	$\eta_5$	$\eta_6$	$\eta_7$	$\eta_8$	$\eta_9$	$\eta_{10}$
5 (Group)	1.635	1.803	1.826	1.806	1.795	1.114	1.174	1.179	1.194	1.165
5 (Indiv.)	1.311	1.747	1.909	1.832	1.840	1.057	1.093	1.042	1.097	1.013
6 (Group)	1.573	1.773	1.860	1.816	1.827	1.027	1.101	1.106	1.132	1.103
6 (Indiv.)	1.354	1.749	1.895	1.787	1.863	1.078	1.127	1.077	1.150	1.037
7 (Group)	1.785					1.055				
7 (Indiv.)	1.797					1.106				
8 (Group)	1.855					1.169				
8 (Indiv.)	1.728					1.088				
Model	а	$T_{ m er}$	$s_t$	λ						
5 (Group)	2.954	0.126	0.015	0.627						
5 (Indiv.)	2.608	0.178	0.012	0.597						
6 (Group)	2.788	0.150	0.011	0.664						
6 (Indiv.)	2.608	0.178	0.013	0.652						
7 (Group)	2.759	0.150	0.010	0.608						
7 (Indiv.)	2.659	0.176	0.011	0.596						
8 (Group)	2.881	0.149	0.016	0.678						
8 (Indiv.)	2.608	0.179	0.012	0.666						

Note. "Indiv." = average of parameters estimated from individual fits

letters on double-target trials;  $\nu_{11}$  to  $\nu_{15}$  are the mean drift rates for digits on single-target trials, and  $\nu_{16}$  to  $\nu_{20}$  are the mean drift rates for letters on single-target trials.

In Table 8,  $\nu_1$  to  $\nu_5$  are the mean drift rates for digits on target-present trials;  $\nu_6$  to  $\nu_{10}$  are the mean drift rates for letters on target-present trials, and  $\nu_{11}$  to  $\nu_{15}$  are the mean drift rates for letters on target-absent trials.

#### References

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

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

Ashby, F. G. (1989). Stochastic general recognition theory. In Vickers, D., & Smith, P. L. (Eds.) *Human information processing:* 

measures, mechanisms, and models, (pp. 435–457). Amsterdam: Elsevier.

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

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

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

Blurton, S. P., Kesselmeir, M., & Gondan (2017). The first-passage time distribution for the diffusion process with variable drift. *Journal of Mathematical Psychology*, 76, 7–12.

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

Brady, T. F., & Tenenbaum, J. B. (2012). A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates. *Psychological Review*, 120, 85–109.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*, (2nd ed.). New York: Springer.



- Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13, 51–62.
- Colonius, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review*, *102*, 744–750.
- Corbett, E. A., & Smith, P. L. (2017). The magical number one-on-square-root-two: The double-target deficit in brief visual displays. Journal of Experimental Psychology: Human Perception and Performance, 43, 1376–1396.
- Corbett, E. A., & Smith, P. L. (submitted). A diffusion model analysis of target detection in near-threshold visual search.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavior and Brain Sciences*, 24, 87–185.
- Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoffs on response bias in a perceptual discrimination task: Threshold bound, drift rate change, or two stage processing hypothesis. *Perception & Psychophysics*, 97, 51–72.
- Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, 16, 1129–1135.
- Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology*, 85, 30–42.
- Dosher, B. A., & Lu, Z.-L. (2000). Mechanisms of perceptual attention in precuing of location. *Vision Research*, 40, 1269–1292.
- Dosher, B. A., Han, S., & Lu, Z.-L. (2010). Information-limited parallel processing in difficult heterogeneous covert vision search. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1128–1144.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. Psychological Review, 87, 272–300.
- Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Perception & Psychophysics*, 34, 221–232.
- Duncan, J. (1985). Visual search and visual attention. In *Attention and performance* (Vol. XI, pp. 85–106). Hillsdale, NJ: Erlbaum.
- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. (2000).
  A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, 62, 425–451.
- Endress, A. D., & Szabó, S. (2017). Interference and memory capacity limitations. *Psychological Review*, 124, 551–571.
- Feller, W. (1967). An introduction to probability theory and its applications, (3rd ed.). New York: Wiley.
- Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mentalarchitecture, random-walk, and decision-bound approaches. *Psychological Review*, 117, 309–348.
- Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
- Gleitman, H., & Jonides, J. (1976). The cost of categorization in visual search: Incomplete processing of targets and field items. *Perception & Psychophysics.*, 20, 281–298.
- Gleitman, H., & Jonides, J. (1978). The effect of set on categorization in visual search. *Perception & Psychophysics.*, 24, 361–368.
- Gould, I. C., Wolfgang, B. J., & Smith, P. L. (2007). Spatial uncertainty explains endogenous and exogenous cuing effects in visual signal detection. *Journal of Vision*, 7, 13(4), 1–17.
- Grossberg, S. (1980). How does a brain build a cognitive code *Psychological Review*, 87, 1–51.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1, 17–61.
- Hallett, P. E. (1986). Eye movements. In Boff, K. R., Kaufman, L., & Thomas, J. P. (Eds.) *Handbook of perception and human performance*, (Vol. I, pp. 78–101). New York: 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.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, 9, 394–401.
- Hoffman, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics*, 25, 319–327.
- Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: O as letter or as digit. *Perception & Psychophysics*, 12, 457–460.
- Jonides, J., & Gleitman, H. (1976). The benefit of categorization in visual search: Target location without identification. *Perception & Psychophysics*, 20, 289–298.
- Karatzas, I., & Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. New York: Springer.
- Karlin, S., & Taylor, H. M. (1975). A first course in stochastic processes, (2nd ed.). New York: Academic Press.
- Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception and Psychophysics*, 15, 149–158.
- Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision-Making*, 6, 651–687.
- Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Responsetime tests of logical rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1–37.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77–105.
- Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. Vision Research, 38, 1183–1198.
- Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization. New York: Oxford University Press.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- McCullagh, P., & Nelder, J. A. (1989). Generalized linear models, (2nd ed.). London: Chapman & Hall.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.
- Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). *Gravitation*. New York: W. H. Freeman.
- Moneer, S., Wang, T., & Little, D. R. (2016). The processing architectures of whole-object features: A logical-rules approach. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1443–1465.
- Moore, D. S. (1986). Tests of chi-squared type. In D'Agostino, R. B., & Stephens, M. A. (Eds.) Goodness-of-fit techniques. New York: Marcel Dekker.
- Moran, R., Zehetleitner, M., Müller, H. J., & Usher, M. (2013). Competitive guided search: Meeting the challenge of benchmark RT distributions. *Journal of Vision*, 13(8), 24.
- Moray, N. (1970a). Introductory experiments in auditory time sharing: Detection of intensity and frequency increments. The Journal of the Acoustical Society of America, 47, 1071–1073.
- Moray, N. (1970b). Time sharing in auditory perception: effect of stimulus duration. The Journal of the Acoustical Society of America, 47, 660–661.
- Mörters, P., & Peres, Y. (2010). *Brownian motion*. Cambridge: Cambridge University Press.
- Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, 32, 2335–2343.
- Navarro, D. J. (2007). Similarity, distance, and categorization: A discussion of Smith's (2006) warning about "colliding parameters." *Psychonomic Bulletin & Review*, 14, 823–833.



- Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, 53, 222–230.
- 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., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 108–130.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. Vision Research, 40, 1227–1268.
- Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In Norman, D. A., & Rumelhart, D. E. (Eds.) *Explorations in cognition (pp. 279–307)*. San Francisco: W. H. Freeman.
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention. 1. Colour, 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. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 870–888.
- 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 sequentialsampling models for two-choice reaction time. *Psychological Review*, 111, 333–367.
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relationship between perceptual encoding and decision-making. *Journal of Experimental Psychology: General*, 139, 70–94.
- Ratcliff, R., & Smith, P. L. (2015). Modeling simple decisions and applications using a diffusion model. In Busemeyer, J. R., Wang, Z., Townsend, J. T., & Eidels, A. (Eds.) The Oxford handbook of computational and mathematical psychology (pp. 35–62). Oxford: Oxford University Press.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 3, 438–481.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, 16, 323–341.
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, 65, 523–535.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, 50, 408–424.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, 19, 278–289.
- Ratcliff, R., Smith, P. L., & McKoon, G. (2016). Modeling response time and accuracy data. Current Directions in Psychological Science, 24, 458–470.
- Rogers, L. C. G., & Williams, D. (2000). Diffusions, Markov processes and martingales. Volume 2. Itô calculus. Chichester, U.K.: Wiley, reprinted Cambridge: Cambridge University Press.

- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Schwarz, W., & Miller, J. (2016). GSDT: An integrative model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1654–1675.
- Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In Nickerson, R. S. (Ed.) Attention and performance VIII (pp. 227–296). Erlbaum: Hillsdale.
- Shaw, M. I. (1982). Attending to multiple sources of information: I. The integration of information in decision-making. *Cognitive Psychology*, 14, 353–409.
- Shaw, M. L., Mulligan, R., & Stone, L. (1983). A test between two-state and continuous-state attention models. *Perception & Psychophysics*, 33, 338–354.
- Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, 237, 1317–1323.
- Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 591–600.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127– 190.
- Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2003). Comparison of two weighted integration models for the cueing task: linear and likelihood. *Journal of Vision*, *3*, 209–229.
- 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. (2010). From Poisson shot noise to the integrated Ornstein-Uhlenbeck process: Neurally principled models of diffusive evidence accumulation in decision-making and response time. *Journal of Mathematical Psychology*, 54, 266–283.
- Smith, P. L. (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., & McKenzie, C. R. L. (2011). Diffusive information accumulation by minimal recurrent neural models of decisionmaking. *Neural Computation*, 23, 2000–2031.
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision-making in visual signal detection. *Psychological Review*, 116, 283–317.
- Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attention and decision-making in brief multielement displays. *Psychological Review*, 120, 589–627.
- Smith, P. L., & Ratcliff, R. (2015). An introduction to the diffusion model of decision-making. In Wagenmakers, E. J., & Forstmann, B. U. (Eds.) An introduction to model-based cognitive neuroscience, (pp. 35–62). Berlin: Springer.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, 44, 1297–1320.
- Smith, P. L., Ratcliff, R., & Sewell, D. K. (2014). Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition. *Journal of Mathematical Psychology*, 59, 95–113.
- Smith, P. L., Lilburn, S. D., Corbett, E. A., Sewell, D. K., & Kyllingsbæk, S. (2016). The attention-weighted sample-size model of visual short-term memory: Attention capture predicts resource allocation and memory load. *Cognitive Psychology*, 89, 71–105.



- Soto, F. A., & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, 139, 105–129.
- Soto, F. A., Vucovich, L., Musgrave, R., & Ashby, F. G. (2015). General recognition theory with individual differences: A new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review*, 22, 88–111.
- Sorkin, R. D., & Pohlmann, L. D. (1973). Some models of observer behavior in two-channel auditory signal detection. *Perception & Psychophysics*, 14, 101–109.
- Sorkin, R. D., Pohlmann, L. D., & Gilliom, J. (1973). Simultaneous two-channel signal detection: III. 603 and 1400 Hz signals. *Journal of the Acoustical Society of America*, 53, 1045–1051.
- Sorkin, R. D., Pohlmann, L. D., & Woods, D. D. (1976). Decision interaction between auditory channels. *Perception & Psychophysics*, 19, 290–295.
- Speckman, P. L., & Rouder, J. N. (2004). A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review*, 11, 574–576.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, 21, 136–152.
- Thornton, T., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, 114, 71–103.
- Townsend, J. T., & Colonius, H. (2005). Variability of the MAX and MIN statistic: A theory of the quantile spread as a function of sample size. *Psychometrika*, 70, 759–772.
- 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– 359.
- Townsend, J. T., Solomon, B., & Spencer-Smith, J. B. (2001). The perfect Gestalt: Infinite dimensional Riemannian face spaces and other aspects of face cognition. In Wenger, M. J., & Townsend, J. T. (Eds.) Computational, geometric and process issues in facial cognition: Progress and challenges (pp. 39–82). Mahwah: Erlbaum.
- Townsend, J. T., Aisbett, J., Assadi, A., & Busemeyer, J. (2006). Colonius, H., & Dzhafarov, E. N. (Eds.) General recognition theory and methodology for dimensional independence on simple cognitive manifolds. Mahwah: Erlbaum.

- Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology*, 56, 476–494.
- Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, 101, 80–102.
- van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, 121, 124–149.
- Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, 14, 1011–1026.
- Verdonck, S., & Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review*, 123, 208–218.
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31, 523–535.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1436– 1451
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192–196.
- Wenger, M. J., & Townsend, J. J. (2006). On the costs and benefits of faces and words: Process characteristics of feature search in highly meaningful stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 755–779.
- Weisstein, E. W. (2009). *CRC encyclopedia of mathematics*, (3rd ed.). Vol. 2. Boca Raton: CRC Press.
- Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. Frontiers in Neuroinformatics, 7, 1–10.
- Williams, D. (1991). *Probability with martingales*. Cambridge: Cambridge University Press.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance, 15, 419–433.
- Wolfe, J. M., Horowitz, T. S., Palmer, E. M., Michod, K. O., & Van Wert, M. J. (2010). Getting into guided search. In Coltheart, V. (Ed.) *Tutorials in visual cognition* (pp. 93–119). New York: Taylor & Francis.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235.

