

Reconciling similarity across models of continuous selections

Peter D. Kvam
University of Florida

Brandon M. Turner
The Ohio State University

Recently-developed models of decision making have provided accounts of the cognitive processes underlying choice on tasks where responses can fall along a continuum, such as identifying the color or orientation of a stimulus. Even though nearly all of these models seek to extend diffusion decision processes to a continuum of response options, they vary in terms of complexity, tractability, and their ability to predict patterns of data such as multimodal distributions of responses. We suggest that these differences are almost entirely due to differences in how these models account for the similarity among response options. In this theoretical note, we reconcile these differences by characterizing the existing models under a common framework, where the assumptions about psychological representations of similarity, and their implications for behavioral data (e.g., multimodal responses), are made explicit. Furthermore, we implement a simulation-based approach to computing model likelihoods that allows for greater freedom in constructing and implementing continuous response models. The resulting geometric similarity representation (GSR) can supplement approaches like the circular / spherical diffusion models by allowing them to generate multimodal distributions of responses from a single drift, or simplify models like the spatially continuous diffusion model by condensing their representations of similarity and allowing them to generate simulations more efficiently. To illustrate its utility, we apply this approach to multimodal distributions responses, two-dimensional responses (such as locations on a computer screen), and continua of response options with nontrivial, nonlinear similarity relations between response options.

Keywords: continuous response; decision making; cognitive model; evidence accumulation; response time

A number of recent papers proposing new computational models have targeted a long-standing problem in decision-making, which is that dynamic theories of choice only account for selections between a small number of options (most often, sets of 2-3). Continuous-response paradigms are prevalent across cognitive sciences and psychology, including perception and memory of color (Awh et al., 2007; van den Berg et al., 2014; Zhang & Luck, 2008) and orientation (Rademaker et al., 2012; Kvam, 2019b), simple studies of brightness detection (Ratcliff, 2018), and more applied cognitive tasks like spatial navigation and wayfinding (Blanchard et al., 1987; Montello & Pick Jr, 1993), pricing (Mellers et al., 1992; Kvam & Busemeyer, 2018, 2020) and preference or confidence rating tasks (Snizek & Henry, 1989; Harvey, 1997). Despite this, theories of the cognitive processes underlying selections in continuous-report paradigms were underdeveloped or non-existent until the recent advent of the cir-

cular diffusion model (Smith, 2016), the spatially continuous diffusion model (Ratcliff, 2018), the more general geometric framework for constructing models (Kvam, 2019a), and extensions and generalizations of these approaches (Kvam, 2019b; Kvam & Busemeyer, 2020; Smith & Corbett, 2018; Smith, 2019; Smith et al., 2020; Ratcliff & McKoon, 2020). Like preceding evidence accumulation approaches such as diffusion models (Ratcliff et al., 2016) or accumulator models (Brown & Heathcote, 2008; Smith & Vickers, 1988; Usher & McClelland, 2001; Smith & Van Zandt, 2000), these paradigms allow for new insights into the cognitive processes underlying performance by quantifying responses and response times in terms of meaningful cognitive variables like signal strength, response caution, bias, and non-decision time.

The first of these models, the circular diffusion model (CDM; Smith, 2016), provided a tractable approach to modeling continuously-distributed responses on a circle. It was particularly aimed at accounting for behavior on tasks like the color wheels used in studies on visual working memory (van den Berg et al., 2014) and accounts for behavior by representing decision making as a two-dimensional diffusion process bounded within a circle. The left panel of Figure 1 illustrates the model: A decision maker's evidence state

This work was supported by a grant from the Air Force Office of Scientific Research, Grant FA8650-16-1-6770. Some of the material in this manuscript was previously presented at the 2019 meeting of the Society for Mathematical Psychology.

starts somewhere within the circle, typically at the center, and moves around as the decision maker gathers information from the stimulus. Once the evidence state crosses the edge of the bounding circle, a response corresponding to that location on the circle is generated. The CDM predicts a right-skewed distribution of response times and a von Mises distribution of responses (similar to a normal distribution wrapped around a circle). The biggest advantage of the CDM is that it has an analytically tractable likelihood, giving the likelihood of the data directly as a function of its parameters. Such tractability makes the CDM easy to use, and the von Mises distribution of responses that results from this type of diffusion process provides a natural connection to existing work that has examined the distribution of responses on a circle (Awth et al., 2007; Zhang & Luck, 2008).

Another approach to modeling continuously-distributed responses is the spatially continuous diffusion model (SCDM) developed by Ratcliff (2018). Instead of the circular distributions typically encountered in color and orientation tasks, the SCDM was designed to predict performance in tasks where responses could fall along a line or plane, such as in function learning tasks (DeLosh et al., 1997) or spatial location tasks (Naveh-Benjamin, 1987). This was particularly important because it provided a method for predicting responses in a bounded 1-dimensional or 2-dimensional space, rather than the wrapped / unbounded circle. As a result, the SCDM provided an account of responses on a scale with endpoints, where different distributions of response times and accuracy arise in responses near the edges of the response space (consistent with work on absolute judgment; Luce et al., 1982; Marley & Cook, 1984). In addition, the SCDM provides a natural way of accounting for multimodal distributions of responses, which can occur when there are multiple equally-good options that correspond to spatially distinct regions of the response space. However, as is often the case, accounting for more complex patterns in behavioral data often comes at the cost of a more complex theory, and the SCDM is no exception. Of note, the SCDM can be difficult to use because it must approximate the continuous set of response options discretely, and generating fits to data requires simulation of a large number of accumulators. Both the quality of fits and the computational complexity of this model depend on the resolution of the discrete approximation of the continuous space.

Although Ratcliff (2018) noted some differences between the models in terms of what they could predict, the mathematical structure of these models has not been thoroughly compared. One reason for this gap in comparison relates to purposes for which the models have been designed: whereas the CDM was developed primarily for tasks with a circular / wrapped response space, the SCDM was developed with an eye toward cases where responses fall within a fixed range on a line or plane. In this theoretical note, we examine the differ-

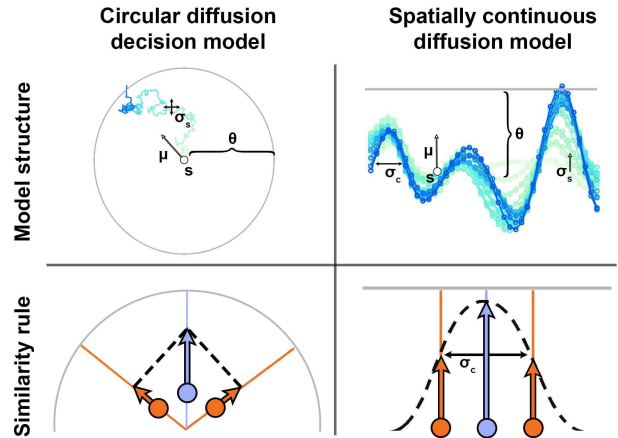


Figure 1

Top row: Structure of the circular diffusion decision model (left) and the spatially continuous diffusion model (right), including parameters for drift μ , starting point / state s , diffusion rate σ_s , threshold θ , and kernel width specifying evidence covariance σ_c (specific to the SCDM). Bottom row: Similarity rules for mapping evidence for a target alternative (blue) onto other nearby alternatives (orange) for each model.

ences between the models in three important dimensions: (1) how the models represent evidence for different alternatives, (2) how support for one alternative affects support for another (similarity rule), and (3) how evidence is accumulated and mapped to a decision rule. We show that the primary difference between the models is in how similarity among alternatives is represented: whereas the CDM represents evidence as a point within a circle (hypersphere) of possible responses, constraining similarity to a cosine-based rule, the SCDM uses an additional free parameter to represent the relationships among choice options with a similarity kernel. A result from our analysis is that the additional free parameter used in SCDM can be integrated into the CDM using the geometric framework, producing a new geometric representation of similarity [GSR] that merges the benefits of both the CDM (in terms of continuity and tractability) and the SCDM (in terms of its ability to capture multimodal response distributions from a single drift vector).

Models of Continuous Report

Figure 1 illustrates how the two models account for decision dynamics over time (top panel) and how they relate support for one choice alternative (blue) to other alternatives (bottom panel; orange). Both models posit that evidence for a distribution of choice options is represented as an evidence state, and that evidence accumulates over time as a decision-maker gathers information from the stimulus during the de-

liberation period. Both models also agree that a decision is terminated at the moment in time when the amount of evidence for a choice exceeds a threshold. At a high level, their search, stopping, and decision rules are all the same (Weitzman, 1979; Gigerenzer et al., 1999) and so these conventional elements used to characterize discrete choice models cannot sufficiently distinguish between them. Instead, we have to look at characteristics that are unique to multi-alternative and continuous models – namely, how evidence for one response option in the choice set affects evidence for other response options.

Despite the apparent similarities, the models differ in several critical ways that affect both their practical and theoretical utility. The CDM has analytic solutions that allow us to easily compute the likelihood of the data given different values of its parameters (such as drift, diffusion, threshold – outlined in detail below). The likelihood covers a continuous span of both responses on the continuum and their associated response times. However, the CDM cannot predict multimodal distributions of responses, where selections are clustered in more than one main area, from a single drift rate. These types of distributions appear in continuous response tasks when the stimulus consists of a display with multiple targets (Ratcliff, 2018) or when there is conflict between two sources of information such as a stimulus and a predecision cue (Kvam, 2019b) or multiple simultaneous cues (Ratcliff, 2018, Experiment 6).

It is important to note that the CDM can produce multimodal distributions of responses across trials by allowing the drift vector to vary from trial to trial, as in work by Kvam (2019b) and Smith et al. (2020). Its main limitation in this respect is that these distributions cannot be produced by a *single* drift vector. Naturally, a mixture of different drift rates across trials can produce multimodality. If a participant is driven toward responses in one direction (north) for one type of stimulus or on some proportion of trials and driven toward responses in another direction (east) for another type of stimulus / proportion of trials, then the overall distribution of responses can be multimodal (north and east, as opposed to northeast for a single drift). Whether the drift should vary across trials in principle, and what form such variability should take, are broader issues that must be resolved with domain-specific theory (see Jones & Dzhafarov, 2014; Heathcote et al., 2014; Smith et al., 2014). Whether a model with drift variability can accommodate multimodal distributions will depend heavily on the distribution of drifts – for example, a unimodal Gaussian distribution on drift vectors is much less likely to generate multimodality than a mixture distribution of two Gaussians.

For the present purposes, we operate under the assumption that it ought to be possible to generate multimodal distributions of responses from a single drift. Multimodal distributions and the relative concentration of responses at dif-

ferent modes are often theoretically important, as they reflect potentially interesting properties of the generative process. For example, Dowd & Golomb (2019) showed continuous reports of spatial locations in a feature-binding task that were distributed with multiple modes – one at the correct target location and other smaller but substantial modes at important nontarget locations in the display (see Dowd & Golomb, 2019, Figure 5). The concentration of responses around each of the modes reflected different types of binding errors, meaning that a plausible model must reflect not only the multimodality of the response distributions but also the relative prevalence of responses at each mode. The fact that people do not integrate these disparate sources of information, where integration would result in a ‘compromise’ response somewhere in between the conflicting sources, reflects competition between options where one alternative can compete with information favoring another in the set.

Conversely, the SCDM has no trouble predicting multimodal response distributions from a single drift vector (e.g., Ratcliff, 2018, Experiment 6), but doing so comes at a cost. The SCDM models the evidence accumulation process using a large set of accumulators that each correspond to a unique response, discretizing the continuous space into a finite number of separate locations (dots in the top right panel of Figure 1). As opposed to the CDM, which has analytic solutions and a single accumulator, the likelihood of different responses predicted by the SCDM must be computed by simulating the evolution of a pool of accumulators. There is nothing inherently wrong with simulated models or approximations to a continuous process, but there is a computational cost accompanying the increase in model complexity. As a result, a researcher is left with the predicament of determining the degree to which they are willing to compromise between fidelity (high number of accumulators / states) and practicality (reasonable fitting times). Simulating a larger number of states will result in a finer approximation of continuity, but will naturally demand more computational resources. As we develop the geometric similarity representation model, we show how it can alleviate some of the computational demands of the SCDM by condensing the (correlated) information stored across many accumulators into a single state representation of the evidence a decision maker accumulates. This makes simulation less cumbersome, and it builds representations of similarity into the model in terms of converging representations of response options (similar alternatives) as opposed to redundancy between different pools of evidence (correlated accumulators).

The Circular Diffusion Model

In the CDM, the evidence state is represented as a point located on the interior of a circle, where each point on the circle corresponds to a response alternative (e.g., 0-360 degrees for a motion stimulus). The state describes the degree

of support for each option by how close it is to the location of the response option on the circle. A neutral state for a starting point would therefore be in the middle of the circle. As a person gathers information from the stimulus favoring different response options, their evidence state moves toward the response options supported by the incoming evidence. This updating process continues until the state crosses the boundary of the circle at a particular point. The point at which the state crosses determines the outcome of the decision, triggering a response in favor of the response option corresponding to the location on the circle where the state crossed.

The evidence accumulation process is specified by five main parameters as illustrated in Figure 1: (1) the initial state / starting point (s) which indicates what options are favored before the decision making begins gathering evidence; (2) the average direction that the state moves over time (drift direction μ_{dir}), usually given by the properties of the stimulus; (3) the rate at which it moves in this direction over time, indicating the strength of incoming information (drift magnitude $|\mu|$); (4) the noise in the evidence accumulation process, or rate at which random information is sampled (diffusion rate σ_s); and (5) the amount of information needed to make a decision favoring one of the response options, describing the radius of the circle of response options (threshold θ). Together, these parameters shape how the state unfolds over time and, as a consequence, the predicted distribution of responses and response times.

Formal specification

Formally, the state is represented as a two-dimensional vector $s(t)$ that describes its position on a plane for any given time $t \geq 0$. At the beginning of a trial, $t_0 = 0$, the state typically begins at the origin, $s(0) = [0, 0]$. One can assume that $s(0)$ is randomly initialized according to a zero-centered distribution, but this makes the analytic likelihood much more difficult to evaluate (see Smith, 2016, Equation 32). Alternatively, to incorporate a starting point bias (i.e., $s(0) \neq 0$), a modeler can use simulation-based methods to compute the likelihood (Turner & Sederberg, 2014; Turner & Van Zandt, 2018). Hence, while the CDM can incorporate both starting point bias and between-trial variability in starting point, additional computational expenses are needed, either in the form of integrating out starting point variability or moving to a simulation-based protocol for fitting the model to data. The latter approach is fortunately not too strenuous, as we show in the comparison section below.

The evidence state is updated over time according to dynamics set out by the drift and diffusion parameters. As with the unidimensional diffusion model, it can be approximated by a discrete-time random walk. In such a case, at each time step, a bivariate normal random variable is drawn, $\delta(t) \sim \mathcal{N}(\mu, \sigma_s^2)$, where the drift μ is a 2-dimensional vector $[\mu_1, \mu_2]$ and σ_s^2 is the noise or diffusion rate. The drift can also

be represented in terms of a *direction* $\mu_{dir} = \tan^{-1}(\mu_2/\mu_1)$ and *magnitude* $|\mu| = \sqrt{\mu_1^2 + \mu_2^2}$ (i.e., ϕ and r in polar coordinates from x and y motion in Cartesian coordinates). Because the CDM does not assume covariance between noise in the x and y directions, a modeler can simply draw a normal random variable $\delta_1(t) \sim \mathcal{N}(\mu_1, \sigma_s^2)$ for the motion in the x direction and another normal random variable $\delta_2(t) \sim \mathcal{N}(\mu_2, \sigma_s^2)$ for motion in the y direction. The state is then updated at each step by simply adding this new evidence:

$$s(t + \Delta t) = s(t) + \delta(\Delta t) \cdot \sqrt{\Delta t} \quad (1)$$

As the step size ($\sqrt{\Delta t}$) and time between steps (Δt) approach zero, this will generate a continuous two-dimensional diffusion process. The decision process terminates when it crosses the circular boundary surrounding the origin – specified with radius θ – and a response corresponding to the location of the evidence state is made.

The biggest advantage of the CDM is in its elegance: Smith (2016) develops the differential equation and solutions for the resulting continuous-time process, providing an analytic solution for the first-passage distribution of hitting times. The resulting dynamics of the CDM yield a joint probability density for a particular response x and response time t as

$$\begin{aligned} Pr(\text{Response} = x, RT = t | \mu, \sigma, \theta, t_{ND}) = \\ \exp\left(\frac{\theta(\mu_1 \cos(x) + \mu_2 \sin(x))}{\sigma^2} - \frac{|\mu|^2(t - t_{ND})}{2\sigma^2}\right) \\ \times \frac{\theta^2}{\sigma^2} \sum_{k=1}^n \frac{j_{0,k}}{J_1(j_{0,k})} \exp\left(-\frac{j_{0,k}^2 \sigma^2(t - t_{ND})}{2\theta^2}\right) \end{aligned} \quad (2)$$

The vector $\mu = [\mu_1, \mu_2]^T$ specifies the direction and magnitude of accumulation, where its length $|\mu|$ describes the overall rate of accumulation and the balance between μ_1 and μ_2 determines the direction as $\mu_\phi = \tan^{-1}(\mu_2/\mu_1)$. Consequently the drift can also be easily represented in terms of polar coordinates, radius $|\mu|$ and angle μ_ϕ .

In the equation above, the function $J_1(\cdot)$ is a first-order Bessel function of the first kind, and the elements $j_{0,k}$ are the zeros of the zero-order Bessel function of the first kind. Bessel functions and their zeros tend to appear in the equations for functions defined on a circle, as they allow the function to decrease from a central location but (like sine and cosine) wrap at 360 degrees around the circle. The ones used here have an integer as an order and are therefore sometimes referred to as cylindrical functions or harmonics; as their name suggests, these are used to describe distributions where at least one of the coordinates is circular as in cylindrical or spherical coordinate systems. The series $\sum_{k=1}^n$ computes and evaluates the function for the first n zeros. The true likelihood is given by an infinite series, $n = \infty$, but this series can be well approximated in practice with $n = 50$ to $n = 150$ (Kvam, 2019b; Smith & Corbett, 2018).

The circular diffusion model has since been extended to modeling responses and response times on a sphere or hypersphere (Corbett & Smith, 2020), where the distribution of responses appears to follow a multidimensional von Mises-Fisher distribution. This development is particularly important for the geometric similarity approach we present here, where evidence accumulation unfolds in a multidimensional space. Indeed, the geometric similarity representation model uses the same state representation and stopping rule as the hyperspherical diffusion model, meaning that the distribution of response times that it predicts will be identical. The main change to the circular / hyperspherical diffusion model is in how response options are represented relative to one another, building representations of similarity into the dynamic decision making process.

Similarity rule

In the case of the circular diffusion model, we can naturally think of the similarity between representations in terms of spatial distances – the closer together two response options are on the circle, the more similar they appear to the decision-maker (Smith, 2019). This applies to the evidence state as well – how well a body of evidence aligns with a particular response alternative can be computed by literally examining the alignment between the state and the response by evaluating the angle between them. Stronger evidence favoring a response alternative will move it farther away from the center, meaning that two main factors determining the degree of support for a response option are the distance of the evidence state from the origin and the alignment of the evidence state with the response alternative. Thus, the support for a particular response at a particular point in time is computed by taking the *component* of the state relative to the response. Similarly, the degree of similarity between two response options can be computed by taking the component of one response relative to the other (Kvam, 2019a).

In accordance with this, the similarity between two alternatives in a continuous set is given by the cosine of the angle between them (dissimilarity can be quantified by the sine of the angle between them). The evidence that information favoring one alternative A shares with another alternative B in the choice set, $Ev(A, B)$, is a function of the alignment between A and B (\mathbf{v}_A and \mathbf{v}_B for alternatives A and B, respectively):

$$Ev(A, B) = \cos(\angle AB) = \mathbf{v}_A \cdot \mathbf{v}_B \quad (3)$$

Here, \mathbf{v}_A is the unit-length vector describing alternative A. For example, motion direction on a circle might assign the vector $\mathbf{v}_A = [1, 0]$ to rightward motion or $\mathbf{v}_A = [\sqrt{5}, \sqrt{5}]$ to motion at 45 degrees. Then \mathbf{v}_B is the unit-length vector describing alternative B, and $\angle AB$ is the angle between them (45 degrees or $\pi/4$ radians, $\cos(\angle AB) \approx 0.71$).

The issue that prevents the standard CDM from generating multimodal distributions of responses with a single drift vector is that any time there are response options located at every point on the circle (or hypersphere), a combination of evidence for two alternatives along the continuum will result in more evidence for another alternative in the set. Figure 2 illustrates how evidence for two options can be combined by adding them together. When we combine evidence for alternatives A (blue) and alternative B (orange), we do not obtain a multimodal distribution, but rather a unimodal distribution with its mean between alternatives A and B. This is because of the cosine similarity rule implemented in the CDM: the sum of evidence for response options between A and B is greater than either the sum of evidence for A or the sum of evidence for B. Therefore, the overall sum of evidence from the two sources winds up being concentrated somewhere in the middle.

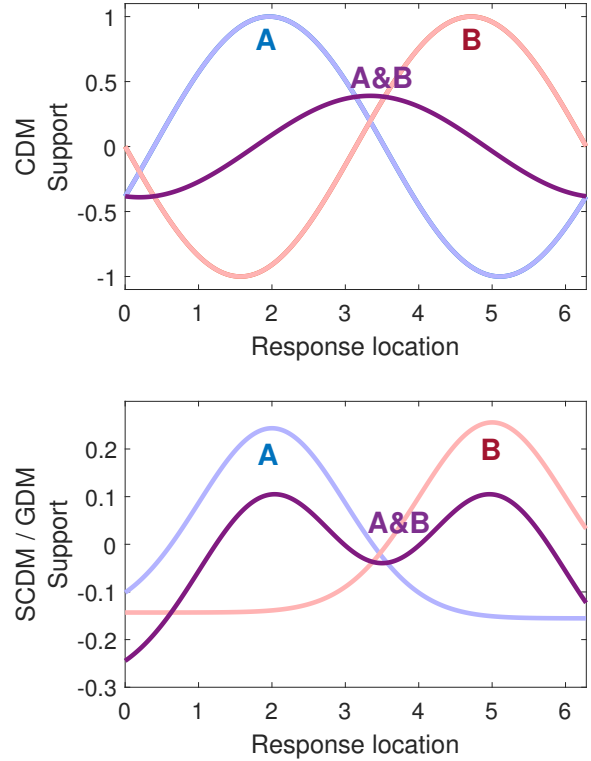


Figure 2

Distributions of evidence arising from combining two separate information sources, A and B, for the different models. This necessarily creates a unimodal distribution in the CDM (top) whereas it can create multimodal distributions in the SCDM or GSR model (bottom). CDM: Circular Diffusion Model. SCDM: Spatially Continuous Diffusion Model. GSR: Geometric Similarity Representation.

For example, take two sources of two-dimensional motion information: one favoring motion at 0 degrees / east, $\mathbf{v}_E = [1, 0]$, and one favoring motion at 135 degrees / north-west, $\mathbf{v}_{NW} = [-\sqrt{5}, \sqrt{5}]$. If these two sources of information are evenly weighted, the vector sum will be approximately $\mathbf{v}_{E,NW} = [0.29, 0.71]$, or motion toward 67.5 degrees / north-northeast. Naturally, this is a compromise between the two input vectors, and feeding $\mathbf{v}_{E,NW}$ into an accumulation process (Equation 1) will result in a unimodal distribution centered on 67.5 degrees rather than a multimodal distribution with modes at 0 and 135 degrees. This will happen with any circular arrangement of alternatives in two dimensions: there is always a ‘compromise’ alternative situated exactly between the two competing sources of information, with the rare exception of cases where they are perfectly opposite one another, in which case there exist two compromises.

As we show below, there is a way to solve the merged evidence issue discussed above by constructing a higher-dimensional representation of alternatives. When a two-dimensional span of alternatives is represented in three or more dimensions, it becomes possible to create a multimodal distribution of evidence, and thus create a multimodal distribution of responses from a single drift. This occurs because there are not any response options “in between” every pair of response options – formally, for any pair of points on the hypersphere, there is not necessarily another point on the geodesic between them that corresponds to another response alternative. We show how to do this in the Model Synthesis section, taking the similarity rule from the SCDM and integrating it into the representation of alternatives in the CDM to form a new geometric model that inherits advantages of both approaches.

Spatially continuous diffusion model

The spatially continuous diffusion model (SCDM) takes a slightly different approach to modeling evidence accumulation in selections along a continuum of response options. Rather than using a single two-dimensional position to represent the evidence for a variety of responses on the circle, it allocates a separate accumulator to each possible response that a decision maker could select. These accumulators increase or decrease over time as new evidence is gathered, and the first accumulator to exceed the threshold θ determines what response is made. Naturally, it is impossible to represent a continuous span of responses in terms of a finite set of accumulators, so this model provides an approximation of the response space. The granularity of the model, in terms of how closely it approximates the continuum, is determined by the number of accumulators. An increase in the number of accumulators results in a more complex and more continuous representation of evidence, along with an increase in the computational demands of the model. Hence, using the SCDM involves an important trade-off: more accumulators

allow better approximations of a continuous representation of evidence, but each additional accumulator creates greater computational complexity and makes the model more difficult to use.

The use of many accumulators also comes with an additional challenge – namely, how to specify how all of the accumulators change over time. In a typical racing accumulator model, each accumulator can have its own drift and diffusion rate, and sometimes its own threshold (Brown et al., 2008). The SCDM sidesteps the problem of specifying a large number of accumulators to some extent by representing response options with a Gaussian process, where the evidence in each accumulator is constrained to follow a multivariate normal distribution. The Gaussian process approach induces correlations between the accumulators such that information favoring one response in the set provides evidence for other nearby alternatives as well. In practice, most stimuli feature only a limited number of targets and distractors, meaning that there are a limited number of drift and diffusion rates that must be parameterized. Although there are exceptions like the sinusoidal threshold used in color identifications (Ratcliff, 2018), each accumulator in the SCDM also typically has the same threshold. In many ways this brings the SCDM into alignment with the CDM, as it has a primary target accumulator (drift direction) with a unique accumulator rate (drift rate), distractor accumulators and noise processes in accumulation (diffusion), and a level of support needed to generate a response for a particular option (threshold).

Despite these overlapping mechanisms, the SCDM adds one critical element that makes it unique from the CDM. Instead of arranging the choice options in a circle and using a cosine distribution to represent support for the response options (Figure 1, left) it uses a Gaussian process to represent the distribution of evidence (Figure 1, right). This means that a kernel is used to relate changes in evidence in one accumulator onto changes in evidence in other accumulators, representing the distribution of evidence across accumulators as a multivariate normal random variable. We detail how this is carried out formally below, but essentially the amount of support that information favoring option A on the continuum provides for another option / accumulator B is based on the height of a normal distribution, with mean centered at A and standard deviation σ_c , at the location of accumulator B. The addition of the parameter σ_c allows the SCDM to produce multimodal distributions of evidence and thus multimodal distributions of responses from a single drift vector. Although this feature makes the SCDM more capable of capturing real empirical data, it increases the model’s complexity, both in terms of the computational cost (i.e., simulation time) and model flexibility (i.e., the range of data the model is able to fit).

Formal specification

The initial state in the SCDM is $s(0)$, an n -dimensional state where n is the number of accumulators that a modeler has chosen to use. As in the CDM, an unbiased initial state is realized by setting all entries of $s(0)$ equal to zero. As the decision-maker accumulates information over time, their state at time t will be reflected by updating $s(0)$ to the n -dimensional vector $s(t)$. The evidence accumulation process unfolds according to the dynamics specified by the drift vector μ and noise / diffusion parameter σ_s , as well as the covariance across the different accumulators, specified by a kernel with free parameter σ_c .

At each time step, a multivariate normal random variable $\delta(t) \sim \mathcal{N}(\mu, K)$ is drawn, where μ is a vector describing the average drift for each accumulator and K is the covariance matrix across accumulators (specified in Equation 5 below) with the main diagonal describing the within-accumulator noise at each time step (σ_s^2). The vector $\delta(t)$ describes how the state changes from one time step to the next, and it is added to the state at time t to compute the state at time $t + \Delta$:

$$s(t + \Delta t) = s(t) + \delta(\Delta t) - \sum_i \left(\frac{s_i(t) + \delta_i(\Delta t)}{n} \right) \quad (4)$$

The far right-hand term is a normalization factor that sets the sum of evidence across all of the states to be equal to zero by subtracting the mean level of evidence across all accumulators from each individual accumulator.

The process of drawing a new multivariate normal random variable, adding it to the current state, and normalizing the state is repeated to calculate $s(t + 2\Delta t)$, $s(t + 3\Delta t)$, and so on until one of the accumulators exceeds the threshold θ . Once one of the accumulators reaches the threshold, the corresponding alternative is chosen and the response time is calculated as a linear function of the number of steps it took to finish. For example, if it took 75 steps to finish, assuming each step took 10 ms, then the response time would be 750 ms plus the non-decision time.

Similarity rule

As we suggested, the main difference between the CDM and SCDM is in the similarity rule, depicted in the bottom panels of Figure 1. Whereas the CDM uses cosine similarity to map evidence for one response option onto the others along the circle, the SCDM (typically) uses a Gaussian distribution to map evidence stored in one accumulator onto support for the other accumulators and response options. When combining two of these distributions, the SCDM can have multiple modes to the aggregate distribution of evidence as long as the two distributions are sufficiently far apart, but this cannot occur for the CDM because the cosine rule does not allow for the distributions to be sufficiently far apart (because of the wrapping incurred from the circular representation of

responses). Figure 2 illustrates this difference by comparing the aggregation process of these two approaches.

Formally, similarity in the SCDM is computed using a Gaussian / squared exponential kernel function to compute the degree of evidence sharing between options $Ev(A, B)$ from the distance between them:

$$Ev(A, B) = \exp \left(- \frac{(L_A - L_B)^2}{\sigma_c^2} \right) \quad (5)$$

where L_A and L_B denote the *locations* of alternatives A and B. These could be locations along a line, or a two-dimensional grid, or three dimensional psychological space generated from multidimensional scaling. The key here is that the kernel takes the Euclidean distance between two alternatives and turns them into a similarity relation between 0 and 1.

The bottom right panel of Figure 1 illustrates the similarity rule of SCDM. Critically, the SCDM incorporates an additional free parameter $\sigma_c > 0$ into its similarity metric, giving it greater flexibility in how evidence should be mapped from one alternative to another. A smaller value for σ_c results in a stricter similarity relation, making nearby alternatives appear more different from one another, whereas a larger value for σ_c results in a more lenient rule, making alternatives appear more similar and thus causing evidence for one alternative to affect a wider spread of nearby alternatives. As a result, greater σ_c creates a smoother pattern of tightly related accumulators in the SCDM, whereas smaller σ_c creates a rougher pattern of more independent accumulators. The covariance matrix K is determined by calculating all pairwise distances between the accumulators, and evaluating their psychological similarity by calculating Equation 5 for each accumulator. If we denote the set of all pairwise (Euclidean) distances between the accumulators \mathbf{D} , then the covariance matrix K is calculated as

$$K = \exp \left(- \frac{\mathbf{D}^2}{\sigma_c^2} \right). \quad (6)$$

The relation of the value of σ_c to the width of the span of alternatives is determined by the degree of multimodality in the distribution of evidence and thus the distribution of responses. In the Appendix, we provide a formula specifying how many possible modes the model can produce based on the values of σ_c .

Notably, parameter estimates of σ_c (covariance between adjacent states) and σ_s (diffusion / noise parameter) are highly correlated in the SCDM and they tend to produce similar changes in the distributions of responses, as shown in Figure A2 of Ratcliff (2018). Thus, greater values of σ_s will decrease the prevalence of multimodality in the response distributions.

The Geometric Similarity Representation Model

The CDM and the SCDM each provide unique benefits to a researcher seeking to account for behavior in continu-

ous response tasks, with the former bringing simplicity and tractability and the latter bringing greater flexibility (such as accounting for non-circular responses) and a capacity to account for multimodality with a single drift vector. In this section, our goal is to provide a bridge between the two models by generalizing the CDM to allow for psychological similarity among sets of response options. Our generalization enables the model to parsimoniously account for multimodal response patterns, while also condensing the correlation structure across accumulators in the SCDM, making it more wieldy and more efficient to simulate. We can accomplish these goals by putting the models together in a common framework. The geometric approach proposed by Kvam (2019a) offers the capacity to evaluate the two models through a single lens, evaluating how each model represents similarity in order to construct a tractable but effective approach to modeling continuous responses and associated response times.

The geometric approach evaluates models in terms of how they represent information for different options, how incoming information is mapped onto those options, and what rules trigger a response in favor of an alternative. The CDM and SCDM models overlap almost entirely in terms of how they account for the accumulation and decision making steps – evidence is mapped onto alternatives in the choice set as new information is gathered from the stimulus, and once one of the response alternatives reaches a critical level of support (threshold), it is chosen. The key difference, as we have highlighted, comes from the similarity rule that relates evidence for one choice option to evidence for the others.

In the next sections, we leverage the geometric similarity approach to develop a new model – which we call the Geometric Similarity Representation (GSR) – that puts together the benefits of the CDM and the SCDM. First, we use the geometric approach to synthesize the similarity rules used by the CDM and SCDM, and show that the similarity structure can be expressed more generally to allow flexible interplay between the two models of continuous report. This model can be viewed as an extension of the hyperspherical diffusion model (Smith & Corbett, 2018) where responses lie on only a subset of the surface of the hypersphere. Second, we discuss the evidence accumulation dynamics in the GSR model, and make the mathematical details of the model explicit. Finally, we illustrate the model in three simple examples.

Similarity of Response Options: A Synthesis

In a sense, we can view the CDM as a two-dimensional (or multidimensional, as in Smith & Corbett, 2018) model where all the response options are located on the surface of a circle (hypersphere), and the support different responses provide for one another is related with a cosine rule as in Equation 3. Conversely, the SCDM is an n -dimensional model where all of the n response options are represented sepa-

ately, but incoming evidence is correlated across dimensions. In essence, what we need to do to go from one option to the other is (1) dimension reduction, and (2) turn correlated evidence (across accumulators) into correlated representations (across alternatives) by arranging them at non-orthogonal angles relative to one another. The fields of psychology and data science more generally offer a variety of tools for dimension reduction, which allows us to perform this mapping. We pursue a basic multidimensional scaling approach here (Shepard, 1962), but spherical multidimensional scaling (Cox & Cox, 1991), factor analysis, or other decompositions could certainly be implemented for the dimension reduction steps discussed below.

First, we focus on how to turn the similarity rule from the SCDM, which creates correlated accumulators, to the similarity rule that can be implemented in fewer dimensions as part of the CDM. Because both models provide a similarity metric that explains how evidence for one alternative affects evidence for other alternatives in the continuum, we can connect them together by setting the right-hand terms from Equations 3 and 5 equal to one another. Given the angles corresponding to option A (a) and option B (b) on a circle, or their locations in a Cartesian coordinate space, the amount of evidence that information favoring B provides for A is

$$\cos(\angle AB) = \exp\left(-\frac{|a-b|^2}{\sigma_c^2}\right) \quad (7)$$

Inverting the cosine transformation on both sides of the equation produces the formula for going from the kernel-based similarity rule to an angle-based similarity:

$$\angle AB = \cos^{-1}\left(\exp\left[-\frac{|a-b|^2}{\sigma_c^2}\right]\right) \quad (8)$$

This simple transformation allows a researcher to map the similarity relations specified by the free parameter σ_c in the SCDM onto cosine similarity relations that can be incorporated within a model like the CDM.

The relation between the squared exponential similarity metric in the SCDM and the cosine similarity metric in Equation 8 allows us to connect the structures of the two models. Even if another kernel is used in the SCDM – substituted into Equation 5 in place of the squared exponential / Gaussian similarity rule – it can be translated into the cosine similarity relation by inserting the kernel function into the right side of Equation 8.

The key difference between the CDM and models constructed by translating the similarity rule is that the set of angles generated from Equation 8 can almost never be perfectly represented in just two dimensions. Instead, the set of angles between alternatives must be translated into a multidimensional representation of the choice options. Fortunately, the tools to create a GSR model from dimension reduction of

the SCDM are already well-established in cognitive science and psychology. Because of its simplicity and wide use, we implement this construction using multidimensional scaling (Kruskal, 1964), but in principle many algorithms for turning pairwise differences into vector space relations – such as other approaches relying on singular value decomposition or machine learning clustering methods – could be applied to produce similar results. Specifically, we use non-classical multidimensional scaling that minimizes the stress metric reflecting the difference between a nonparametric, monotonic transformation of the pairwise differences between points (locations of accumulators in the SCDM) and the Euclidean distances between their multidimensional representations.¹

The goal of the multidimensional scaling procedure is to take the set of pairwise similarities among the response options, and use these pairwise similarities to arrange them spatially. Response options that are more similar should be closer together, and ones that are more dissimilar should be farther apart. For example, the multidimensional scaling representation of color hues often follows a circle or sphere, where purple/red hues are close to one another, green/blue hues are close together, and opposing hues like blue/yellow and red/green are on opposing sides of the circle. The arrangement of options in multidimensional scaling is determined by first transforming the similarities between options into distances, where the (Euclidean) distance between two options is inversely proportional to their related similarity. The algorithm then takes these distances and tries to “match” them as closely as possible while situating all of the response options in a d -dimensional space. More dimensions (higher d) allows the multidimensional scaling procedure to more closely represent these distances: imagine trying to represent colors on a line / one dimension as opposed to a circle / two dimensions – the two dimensional representation is “closer” to what we might imagine the subjective representation is. A very high-dimensional representation is likely to reproduce the subjective similarities between options very well, but it means that our model is more complex. We can use lower-dimensional representations to simplify the model and generate a more parsimonious account of how the choice options are represented psychologically. The goal of the GSR model is to condense the similarities among all response options into a simple psychological representation that is likely to correspond to our subjective experiences.

Formally, we take a set of locations $X = \{x_1, x_2, \dots\}$ of alternatives / accumulators from the SCDM and compute all pairwise distances among them (just as we did to compute \mathbf{D} in the SCDM above). Applying the kernel function gives us the covariance matrix K using Equation 6, again the same as in the SCDM. This gives us the angle cosines. For multidimensional scaling, we want the actual metric distances between points. This is given by computing the *chord* between each pair of points, which is given as twice the sine of half

the angle. The metric distance m between two points i and j describing choice options is thus given as

$$m_{i,j} = 2 \sin \left\{ \cos^{-1} \left[\exp \left(- \frac{|x_i - x_j|}{\sigma_c^2} \right) \right] / 2 \right\} \quad (9)$$

$$= 2 \sin(\cos^{-1}(K(i, j))/2)$$

Multidimensional scaling can then take a full dissimilarity matrix M (with entries $m_{i,j}$) and transform it into a set of points $Y = \{y_1, y_2, \dots\}$ in d dimensions. We can arbitrate between different numbers of dimensions either a priori or by model comparison, if necessary. To set d a priori, the number of dimensions required can be inferred from the value of σ_c relative to the span of the alternatives, as described in the Appendix. The new representation Y of alternatives leads us to a novel extension of the circular diffusion model based on the similarity rule of the SCDM. Because the covariance structure of the SCDM is integrated into the representation of alternatives, rather than the evidence accumulation process, the dynamics of the GSR model are vastly simpler and more computationally tractable. It also uses d_c dimensions, which in practice will always be fewer than the number of accumulators (and therefore dimensions) required to implement the SCDM.

Conceptually, what the GSR does is takes information that is represented in correlations between accumulators and transforms it into similarity in the representations of response options. If we have a priori similarities between response options, such as those generated from confusion matrices or similarity ratings (Shepard, 1962), then they can be built into the model through the multidimensional scaling procedure as well. Another advantage of this approach is that the resulting GSR model produces a truly continuous distribution of responses, meaning that we no longer have to worry about whether the number of accumulators is sufficiently granular to approximate the continuous space of responses, as in the SCDM.

Representation and dynamics

The representation of alternatives Y is given by the result of the multidimensional scaling procedure, giving us a condensed representation of the information that is contained in the accumulators of the SCDM. This also allows us to go back to the continuum of response options: we can compute the support for any alternative in the set by seeing where it would be represented relative to the response options we rescaled in the multidimensional scaling procedure. For example, motion at 34.5 degrees – if the set of points was 0, 1, ..., 360 degrees – can be computed as a weighted sum of vectors

¹In MATLAB, this can be done using the `mdscale` function and in R, it can be implemented using `isomds` in the MASS package. Classical multidimensional scaling can be implemented using `cmdscale` in both languages.

for 34 and 35 degrees. This allows the decision maker to gather support for options that are not in the finite set of options used in the multidimensional scaling procedure, instead representing evidence for the complete continuum of options they have available to them.

For cases where multiple sources of information drive the evidence accumulation process, the overall accumulation process can be computed by simply summing all of the sources of evidence together. Formally, this is calculated as the sum of vectors specifying the drifts for all of the cues and items in a stimulus display – everything that a participant considers on a single trial. This yields a single drift vector describing the average direction and rate of evidence accumulation for a given trial.

The drift vector μ and the diffusion / noise σ_s specify how a decision maker’s state changes over time. They accumulate information until their state crosses the surface of a hypersphere, exactly as in the hyperspherical diffusion model (Smith & Corbett, 2018). The resulting sum of multiple processes for the GSR model is shown in the bottom panel of Figure 2 – it mimics perfectly the multimodal shape of combined distribution of two sources of evidence from the SCDM. As illustrated, this approach can thus create multimodal distributions of responses from a single drift rate as long as there are sufficient dimensions in the representation space and as long as the two sources of information do not indicate response options that are too close together (which corresponds to the empirical observation that multimodality disappears when the locations driving accumulation are close together; see Kvam (2019b), Figure 5).

To assist in understanding how the GSR works in different types of choices scenarios, we first provide its formal specification and then walk through three example implementations. For readers who learn better by following along with model scripts, we provide MATLAB code for each of these examples on the Open Science Framework at osf.io/zbyyps.

Formal specification of the GSR model

The evidence state at time t is represented as a d -dimensional vector $s(t)$. At time $t = 0$ it might typically begin at point $s(0) = [0, 0, \dots, 0]$, though it is possible to insert bias toward certain alternatives by setting $s(0)$ to some value other than zero.

As before, the dynamics of the evidence accumulation process in the GSR model are given by the drift μ and diffusion σ_s . The drift μ is a d -dimensional vector that can be specified either by d Cartesian coordinates or by $d - 1$ polar angles and a magnitude (as in Smith & Corbett, 2018, Equation 3). For a particular stimulus, the drift direction can be specified by taking the stimulus and interpolating its location on the hypersphere from vectors describing neighboring directions. This interpolation process is demonstrated in the example code provided at osf.io/zbyyps. Putting the stimulus

direction d_{stim} (in Cartesian coordinates) together with a drift rate $|\mu|$ gives us the overall drift, $\mu = |\mu| \cdot d_{stim}$.

At each time step, a multivariate normal random variable is drawn based on these parameters, $\delta(\Delta t) \sim \mathcal{N}(\mu, \sigma_s^2)$ and added to the state to calculate the evidence distribution at the next time point $s(t + \Delta)$ as in Equation 1. Note that this evidence distribution is equivalent to drawing d normal random variables with mean specified by μ and variance σ_s^2 , as there is no covariance between evidence accumulated along the various dimensions. Naturally, as with the CDM, the evidence accumulation process will approach a continuous-time multidimensional Wiener process as $\Delta t \rightarrow 0$. As a result, the GSR model removes the main source of computational complexity from the SCDM (i.e., correlated accumulators) by building the evidence covariance structure entirely into the representation of alternatives.

The evidence accumulation process halts and an alternative is chosen as soon as the state $s(t)$ is sufficiently far from the origin, as in the CDM. In the original case, this condition is met when the squared x and y components of the state exceed the squared threshold ($s_x(t)^2 + s_y(t)^2 \geq \theta^2$). The stopping rule is the same as before – once the state is at least θ away from the origin, the process halts. In a larger number of dimensions, this stopping rule will be $\sum_j s_j(t)^2 \geq \theta^2$ or simply $\|s(t)\| \geq \theta$.

In many ways, the GSR model’s stopping rule is similar to the rule in the CDM and SCDM in that it will halt as soon as evidence for any of the alternatives in the choice set exceeds the threshold. In fact, the distribution of response times in the GSR model will follow exactly the first-passage distribution for hitting times on the hypersphere in d dimensions derived by Smith & Corbett (2018), and the raw hitting locations will follow a von Mises-Fisher distribution (Kvam, 2019a). The key extension provided by the GSR model is that not every hitting point on the hypersphere corresponds to a unique response option. Instead, the distribution of hitting points is re-mapped onto the set of response options by interpolating the location of a response from the locations of a more limited set of responses along the hypersphere.

As a consequence, the GSR model produces a continuum of responses, including ones that are not included in the discrete set of accumulators used in the SCDM. For example, there could be one alternative A at $[\sqrt{.8}, \sqrt{.1}, \sqrt{.1}]$ and another alternative B at $[\sqrt{.6}, \sqrt{.2}, \sqrt{.2}]$, and the state $s(t)$ would halt when it crossed the sphere at $[\sqrt{.7}, \sqrt{.15}, \sqrt{.15}]$ if $\theta = 1$. Even though it does not exactly match response A or response B, this response can still be mapped onto a continuous response by gauging its distance to each one and interpolating to obtain the value of the response.² Assum-

²More sophisticated methods of interpolation can be used to do this, but simple linear interpolation tends to work suitably well as long as a reasonably sized set of points is used to compute the multidimensional scaling solution. In our experience, 50 points for a

ing A and B are the nearest two points in the set, the point $s(t) = [\sqrt{.7}, \sqrt{.15}, \sqrt{.15}]$ should be mapped onto a response exactly between them. In this case, if A corresponds to a response at 6 degrees and B a response at 10 degrees, then $s(t)$ would naturally generate a response at 8 degrees. It is critical to note that this interpolation only results in a “compromise” alternative being selected if the two representation points are adjacent – it does not remove the model’s ability to generate multimodal distributions of responses.

Example Implementations

With technical models like these, it can be helpful to provide some example applications. We now provide three worked examples of the GSR model to illustrate its generality. The first example focuses on situations in which the response options are wrapped along a disc, much like the most-studied applications of the CDM. The second example focuses on situations where choices are made in a bounded, two-dimensional plane, as in the Gaussian field implementation of the SCDM. The third example looks at a case where the GSR is uniquely well-suited, where there is a nonlinear relationship between the physical stimuli and their subjective psychological representations. Code in both MATLAB and R are provided in an online repository: osf.io/zbyyps.

Example 1: Continuous choice among options wrapped along a disc

Selection among color options has rapidly become one of the most common applications of continuous response models (Awh et al., 2007; Ratcliff, 2018; Smith et al., 2020), and the hue color wheel provides a natural testbed for illustrating the advantages of the GSR model. Suppose that were are determining which hue on the color wheel is most common in a display of colored dots. A participant must look at the stimulus, identify the dot color that is most common, and map the perceptual affect of the colored dot display into a response on the hue color wheel. An example stimulus is shown in the top-left panel of Figure 3. To illustrate the problem of multimodality, suppose a stimulus is shown consisting of two prominent dot colors – violet and green. We would in an ordinary experiment expect responses to be concentrated with two modes, one centered on violet, and another centered on green. Hence, a desirable property of our model is the production of multimodal response patterns in the color selection for this stimulus.

To explain how patterns of continuous choice unfold, the first step is to construct a representation of the choice options. To do so, we take a subset of possible responses – such as the hues located at 0, 6, 12, 18, 24, ..., 354 degrees on the color wheel – and compute the pairwise similarity relations between them. We can take this set of 60 points around the circle, with similarity kernel parameter $\sigma_c = .5$, feed them into Equation 9, and then use an multidimensional

scaling procedure to generate a set of points Y . We carried this process out to create a five-dimensional representation of the color hue wheel. We chose five dimensions because it is at least sufficient to represent four dimensions of color plus the degree of support needed to have appropriate resolution for each color (Ekman, 1954).

The result of this procedure is shown in Figure 3, where we show the first three dimensions along with the behavior we would expect to observe if only 1, 2, or 3 dimensions of the solution were used. In one dimension, the colors can only vary along one axis, and so there will be complete confusion between colors that do not vary on that axis. For example, if the axis of variation is the blue-yellow axis, it will be difficult to differentiate between red and green hues, as in red-green colorblindness. The unidimensional representation must be supplemented with additional dimensions to get meaningful color hue responses.

A two-dimensional representation of the response options, generated by projecting the five-dimensional color representation down into the first two dimensions, is shown in the middle panels of Figure 3. This model follows the physical representation of hue, with green and violet on almost opposite sides of the circular response space. Support for a green response (A) is given by the component of the state along a vector v_A , while support for a violet response (B) is given by the component of the state along a vector v_B . However, because both of these stimuli are present in the display, the vector representing the overall drift for the trial, $v_{AB} = v_A + v_B$ points in between violet and green. There is not a bimodal distribution because v_{AB} points toward another option on the hue color wheel, cyan. As a result, the two-dimensional arrangement of response options cannot accommodate a bimodal distribution of responses.

The key to multi-modality becomes more evident in the three-dimensional case, shown on the right side of Figure 3. In three dimensions, the accumulation paths tend to hit the response surface (i.e., stopping criterion) at a location that is near both A and B. Critically, stopping locations do not need to correspond directly to an alternative in the set; rather, they can fall anywhere on the surface of the sphere and are then mapped onto the nearest response. To determine a response, one must determine which point on the color wheel (represented here as a warped circle wrapped around the sphere) is nearest to the point at which the accumulation path hit the outside of the sphere. In the three dimensional case, the nearest points will tend to be associated with colors that are similar in location to the points A and B, and so the distribution of responses will be bimodal. Hence, as the complexity of the representation increases, so too may the complexity of the distribution of responses increase.

One way to grasp this is to literally draw an arc on the circle between the two response options (referred to as circle is usually more than enough.

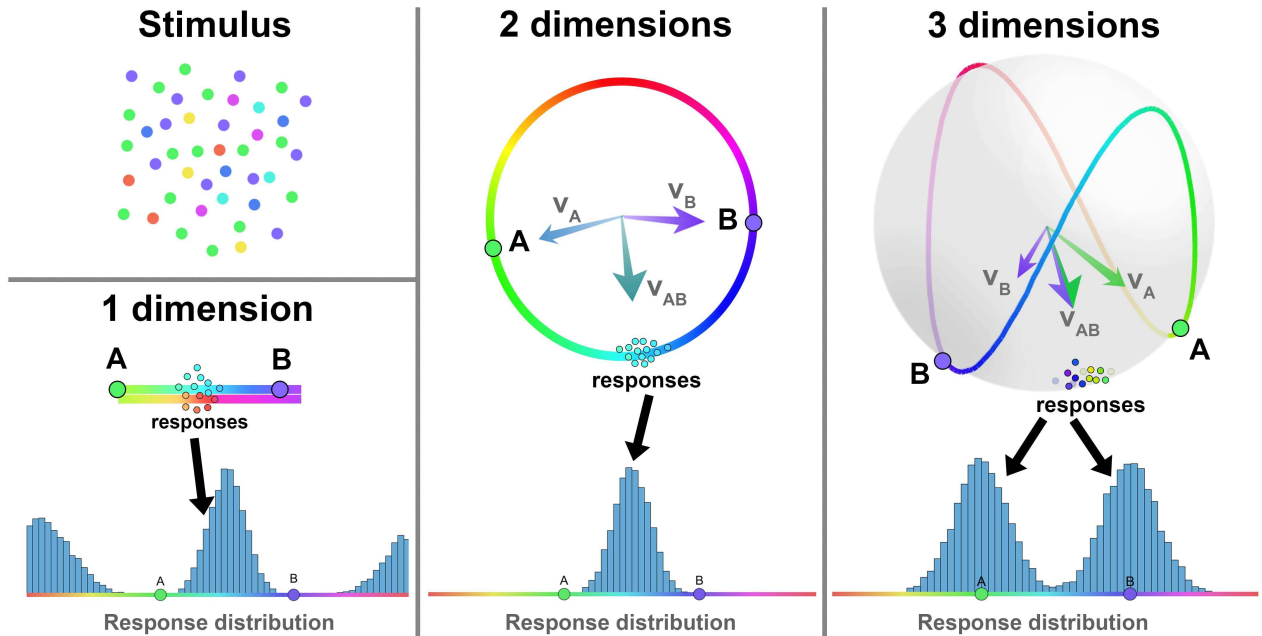
**Figure 3**

Diagram of a simple color task. The stimulus is shown at the top left, along with projections of a five-dimensional geometric similarity representation into one (left), two (middle), and three (right) dimensions. The three dimensional model can create multimodal responses for a single drift (violet / green), whereas lower dimensional models can only produce responses that are a compromise between the two underlying stimulus colors (cyan / light blue).

a geodesic). In two dimensions, this arc passes directly through many other response options. But in three dimensions, the arc does not intersect any other responses, meaning that one can gather evidence for a green and violet response without providing support for a response at say, cyan. Hence, the restriction that constrained the CDM to produce only unimodal distributions of responses is removed. Naturally, responses should be grouped with modes near green and cyan, but the drift describing evidence accumulation should reflect the true information that is gathered from the stimulus and thus generate support for these two options. As with the CDM, the simplest way to compute the evidence accumulation dynamics is to compute the sum of vectors by adding v_A (toward the green response) and v_B (toward the violet response) to compute v_{AB} . Most responses driven by this accumulation process will fall near A and B, or along an arc between them. We can observe this property by drawing an arc on the sphere from one location to the other. The nearest point to any location along this arc will be typically be close to A or close to B. Therefore, slight deviations from the mid-point of the geodesic (around v_{AB}) will result in responses that are closer to A (green) or B (violet) than to most of the points between them. As a result, the distribution of support generated by the stimulus will be a bimodal distribution of responses as shown in the bottom right panel of Figure 3.

For simplicity, the mapping from hitting point to response

is done via linear interpolation as the precision is more than sufficient for our purposes. This could easily be extended to compute more precise distributions by a) adding more points to the multidimensional scaling and interpolation set, or b) using a more sophisticated interpolation algorithm such as polynomial or spline-based methods. Approximate Bayesian methods (Turner & Van Zandt, 2012; Turner & Sederberg, 2014; Turner & Van Zandt, 2018) that use model simulations to approximate the likelihood function are actually quite tractable in this case, as there are well-established algorithms for efficiently drawing symmetric normally distributed random variables that facilitate model simulation.

Readers interested in the exact technical details of this example, or interested in programming up a similar model themselves, can work their way through “Example_ArcCircle” code (which can be adjusted to form an arc or a full circle by changing the span of x) on the OSF site.

Example 2: Continuous choice among options arranged on a two-dimensional plane

Although circular sets of response options are perhaps the easiest to use, the GSR model can also be applied to multidimensional, non-circular sets of response alternatives. For example, the model could be used to explain the distribution of clicks on a computer screen (2 dimensions), locations for spatial foraging tasks (2-3 dimensions), or any number

of paradigms that require a decision maker to make a selection in continuous space. To illustrate how this can be done, we provide an example implementation of the model applied to a 2-dimensional response space. In these types of tasks, participants might respond somewhere on a plane, such as the screen of a computer or mobile device. The space of response alternatives is shown at the top-left of Figure 4, with two particular points in the space highlighted as A (yellow) and B (blue).

In this example, the two-dimensional plane spans $[0,1]$ in the x -direction and $[0,1]$ in the y -direction. Naturally, the actual span of response locations may vary and the σ_c parameter will have to be set relative to the scale. To generate the multidimensional scaling solution for this plane, we used a 9×9 grid of points, labeled 1-81. The multidimensional scaling solution was completed in 10 dimensions, which provided a reasonable approximation given the σ_c value of 0.05.

To compute the multidimensional scaling solution, the pairwise differences between all 81 points were computed and turned into pairwise similarities. These similarity relations were then used in the stress-based multidimensional scaling procedure to yield a 10-dimensional solution. In turn, the multidimensional scaling solution was used to construct the decision space in which evidence accumulation unfolded. As before, the evidence accumulation process used a single accumulator that started at $[0,0]$ and was pulled toward A and B in directions \mathbf{v}_A and \mathbf{v}_B , yielding a drift vector of \mathbf{v}_{AB} with random noise in all dimensions determined by the diffusion rate σ_s .

Once the distance from the origin exceeded the threshold θ , the process halted and the resulting state s was mapped onto the nearest response by interpolating its position in the grid from the nearest points of the multidimensional scaling representation. Rather than just an angle, this gives us both an x - and a y -coordinate corresponding to the response. Repeating this procedure many times results in a distribution of responses like the one shown on the bottom right of Figure 4 – as we might expect, there is a high concentration of responses in regions around A and B, and this concentration decreases as one moves away from either region.

Notably, the responses near A and B tend to be pulled somewhat from being centered directly on their target locations toward responses near the edges of the scale. This is because response alternatives that are grouped near the edge of the grid have fewer ‘competitor’ states: when evidence is accumulated in favor of a state near the edge (where there are nearby states on only 2-3 sides), the summed covariance with nearby states is lower than if evidence is accumulated in favor of a state near the middle of the scale (where it has neighboring responses on all 4 sides). As a result, evidence for options near the edges of the plane tends to accumulate more quickly than for those toward the middle. This allows the model to generate faster and more accurate responses when

the target is toward the edges of a scale. Naturally, this effect constitutes a testable prediction of the model, but work on absolute identification / judgment as well as confidence and other ratings have continually shown “bow effects” where responses are faster and more accurate toward the edges of a given scale (Luce et al., 1982; Pleskac & Busemeyer, 2010; Marley & Cook, 1984; Moran et al., 2015). Indeed, it is possible that the source of such effects may be attributable to the fact that choice alternatives toward the edge of a scale have fewer direct competitors than those toward the middle, allowing the SCDM and GSR model to provide a mechanistic explanation for bow and edge effects with no additional assumptions.

For the interested reader, we have supplied code for this example in an online repository as “Example_Grid”.

Example 3: Nonlinear similarity between choice options

One final example that illustrates the contribution of the GSR model over and above what models of continuous report have typically focused on is an example that emphasizes the importance of similarity in the representation. The experimental setup that gives rise to this type of example is one where participants are asked to make a perceptual magnitude judgment, and are presented with reference points that are designed to assist them in making these judgments. For instance, a participant might be presented with a transparent cup that is partly full of water and be asked to estimate what proportion of the glass is full – see Hollands & Dyre (2000), Figure 3(B).

In the simplest case, participants tend to overestimate how full the glass is when it is 0 to 0.5 full and underestimate how full the glass is when it is .5 to 1.0 full. However, this pattern is affected by the introduction of a reference point at the middle of the glass, drawn at the exact location where the glass is half full. In this case, participants tend to underestimate the fullness of the glass when it is slightly below .5 or slightly below 1.0 (between .25 and .5, and .75 and 1.0) and overestimate the fullness of the glass when it is slightly above 0 or .5 (between 0-.25, or .5-.75).

This example illustrates the importance of *reference points* or anchors (Marley & Cook, 1984), which are used by participants as contrasts to each stimulus they see. Perceptual discrimination is more precise near these reference points, suggesting that the representation of magnitudes near the reference points are easier to tell apart than magnitudes that are far away from the reference points. The pattern typically observed in experiments is shown in Figure 5, left panel: the objective amount of water in the glass (x , color coded to make it easier to read) is nonlinearly related to the subjectively experienced amount of water in the glass (y) due to the presence of three anchors: one at the bottom of the glass, one in the middle (line drawn at .5), and one at the top of the glass.

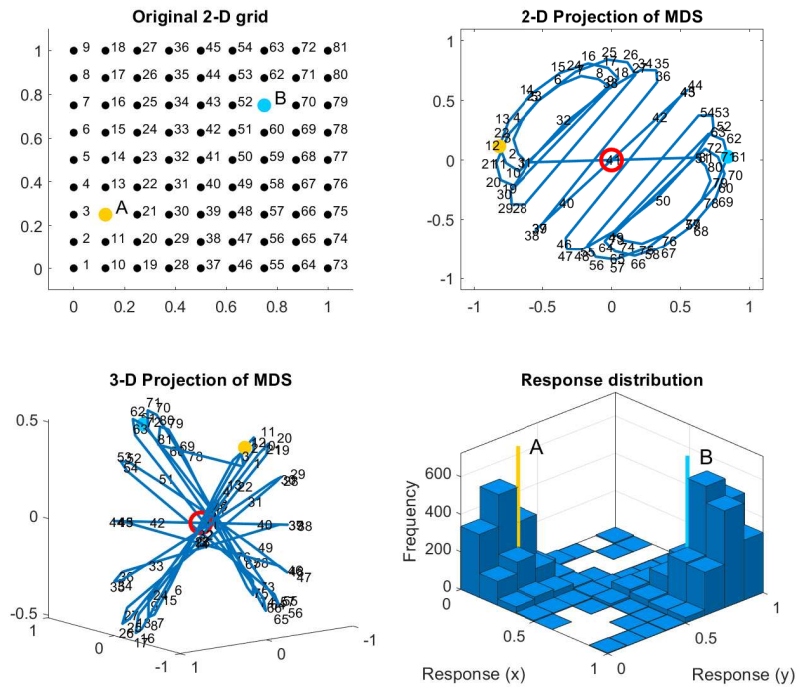


Figure 4

Diagram of the response space (top left), multidimensional scaling representations (top right / bottom left), and distribution of responses (bottom right) for an accumulation process driven by two main sources of information favoring A and favoring B (yellow and blue, respectively)

The GSR model provides a natural way to build this into our model of perceptual decision making by integrating the presence of the reference points into the multidimensional scaling procedure. Suppose that a participant has provided similarity ratings that allow us to construct a three dimensional representation of the stimuli. This representation is shown in the top-right panels of Figure 5: stimuli near the ends (0, 1) and middle (.5) of the scale are spaced further apart than stimuli near the quartiles (.25, .75) of the scale. This reflects the steeper parts of the subjective experience curve (left panels of Figure 5) near 0/.5/1 and flatter parts of the curve near .25/.75.

We can examine the predictions that the model makes for responses resulting from this representation of response options. The distributions of subjective magnitude estimates that we would expect for nine stimuli (with objective magnitudes of $x = 0, .125, .25, \dots, .875, 1$) are shown as colored distributions. If the true stimulus magnitudes were mapped directly onto subjective estimates, these distributions would be centered at the vertical gray lines depicted at the bottom right of Figure 5. However, responses that should be at .125 and .625 are shifted slightly upward and responses that should be at .375 and .875 are shifted slightly downward. This reflects exactly the underestimation and overestimation that we

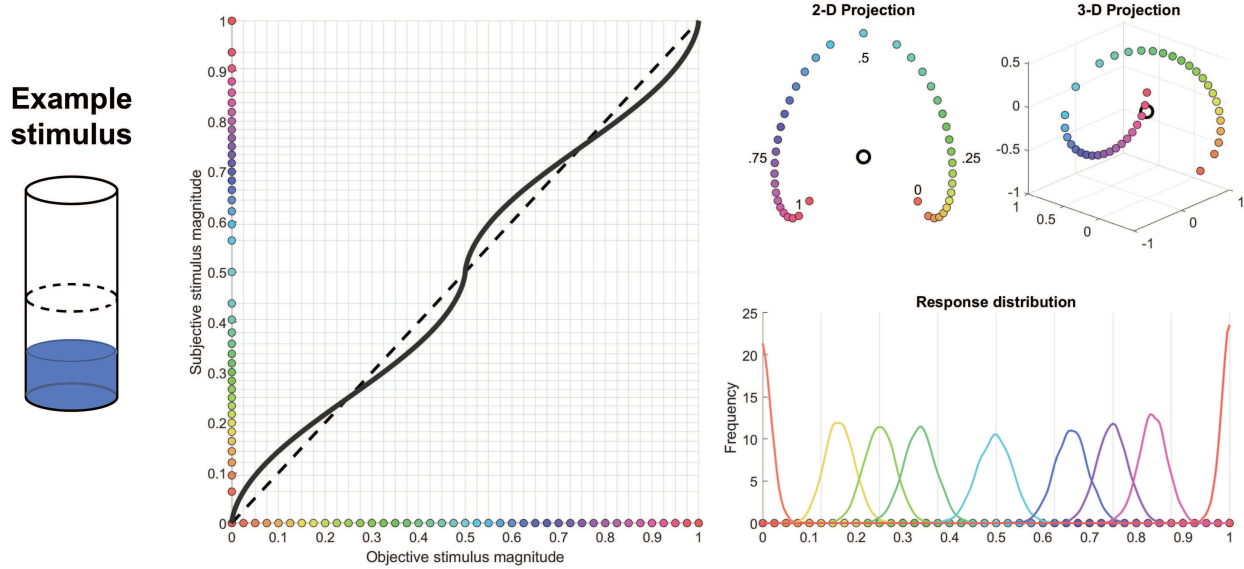
should expect from the subjective response curve (left of Figure 5).

Given that we used the subjective response curve to generate the representations of stimuli, this result should not be too surprising, but it illustrates an important aspect of the GSR model. Specifically, the GSR model accommodates representations of similarity that are generated from behavioral data like similarity ratings or confusion rates or that are known from a sophisticated theory of representation. The addition of subjective representations of response options, mapping onto the orientations of these options in the evidence accumulation space of the dynamic choice model, allows the GSR model to connect with models of representation in a way that the CDM and SCDM have not.

The code for this example is provided as “Example_NonLinearSimilarity” on the OSF site.

Discussion

The GSR model provides a natural and convenient framework for mediating between the CDM and SCDM by allowing the complexity of the representation to dictate the complexity of the response distribution. In our examples, we illustrated how it can overcome the challenge of multi-

**Figure 5**

Example of a GSR model that accounts for behavior on a task with nonlinear subjective representations of stimuli. When there are three reference points, the mapping from objective stimulus to subjective experience is “wavy” (left). This can be built into the multidimensional scaling procedure for the GSR model, yielding representations like the ones shown at the top-right. The resulting distribution of response produced by this model (bottom right) shows a corresponding overestimation of stimuli between 0-.25 and .5-.75, but underestimation of stimuli between .25-.5 and .75-1. GSR: Geometric Similarity Representation.

modality, extend to multidimensional responses, and accommodate complex nonlinear relationships between the objective features of a stimulus and its subjective representation. In the following sections, we elaborate on its relationship to the CDM and SCDM in terms of practical utility and model mimicry, outline procedures for fitting the model to empirical data, and then discuss potential connections to other levels of analysis such as its implementation in neural circuits.

The GSR also improves upon previous geometric models (Kvam, 2019a,b; Kvam & Busemeyer, 2020) by using representational models like multidimensional scaling to determine how response alternatives might be represented, rather than fixing their locations based on some fixed assumptions about the representational space. In past work, we had largely depended on the objective features of the stimulus translating directly into subjective representations, or a particular function that relates the two (e.g., a utility function; Kvam & Busemeyer, 2020). Here, we have shown that these assumptions can be relaxed and that the representation / decision space can be created based on pairwise similarity relationships among response options on the continuum. This addition should improve the models’ ability to capture subjective experiences and psychological representations that exist above and beyond the physical relationships between stimuli. As we discuss below, this brings the geometric framework closer to theories of representational similarity in

neural activation (Kriegeskorte et al., 2008; Love, 2015). It also provides the opportunity to build decision models that reflect semantic similarity between response options, as in vector space semantic models (Bhatia, 2017).

Comparing among the models

In this paper, we have shown that the GSR model can accommodate the similarity rule used in the SCDM into a generalized form of the CDM, where the similarity in representation forms the basis of mediation between the models. It can perfectly reproduce the behavior of the CDM and hypersphere models by definition because it is an extension of these approaches, but the degree to which the GSR model can fully account for the complex behavior of the SCDM has not been a focal point of our exposition. To examine this, we ran a model mimicry study where we generated data from the SCDM (for 100 simulated participants) and estimated the parameters of a five-dimensional GSR model. The details of this study are provided in Appendix B. To summarize, there was a high degree of correlation between the SCDM parameters used to generate the data and the GSR parameters estimated based on that data (with recovering the noise parameter σ_c creating slightly more difficulty than other parameters). Most importantly, the GSR model was able to accommodate the bimodal distribution of responses while simultaneously matching the distribution of response times in

the SCDM. This is not to say that the mimicry between the two models is perfect. There may be conditions we have not examined where the complexity of the SCDM is necessary and the GSR model is unable to accommodate the pattern of results, and we speculate below about what those might be.

Part of the reason the models are so similar is because the parameters of the two models are directly comparable: all of the models include identical parameters drift rates, drift direction(s), threshold, diffusion rates, and starting points as the main cognitive mechanisms underlying decision-making. The remaining parameter is the similarity kernel, which can be reproduced in the GSR model by turning it into spatial similarity relations between alternatives. The only limit on its ability to produce identical performance is the number of dimensions used to construct the decision space. If the GSR model uses as many dimensions as the SCDM has accumulators, the GSR model will yield similar performance to the SCDM in most cases. Mimicry is better when a convex kernel function is used for similarity, because it will satisfy the triangle inequality:

$$Ev(A, B) + Ev(B, C) \geq Ev(A, C) \quad (10)$$

While we can approximate sets of similarity relations generated from a squared exponential kernel or other kernels with concave components, the reason that they cannot be directly transformed into a distance metric (like cosine or inner product) is that they violate this inequality. Therefore, turning the similarities from a partially or fully concave kernel function to the vector space – via multidimensional scaling solution or singular value decomposition – will merely approximate all of the binary relations between alternatives in the SCDM. When we use a strictly convex kernel function, the similarities will not violate the triangle inequality and we can obtain an exact solution that perfectly reflects all of the pairwise relations between alternatives. In turn, this allows us to construct the GSR model such that it perfectly mimics the SCDM *or* the CDM.

Naturally, the GSR model will also be equivalent to the CDM whenever the decision space is constructed using two dimensions. This means that the GSR model is a general case of the CDM and can mimic the SCDM when a convex kernel function is used. The benefit of the more general GSR model is that it can be constructed on any arbitrary number of dimensions between two and m , where it will yield the multimodal distributions of responses that the CDM cannot and provide superior parsimony compared to the SCDM.

The models also differ in terms of their practical usability. In Appendix C, we examine several instantiations of the CDM, GSR, and SCDM. The CDM is reliably the fastest to fit when its analytic likelihoods can be used, but it cannot be used when there is starting point variability. In that case, the CDM can be simulated and performs about as well as the GSR. Of course, even the simulated CDM cannot han-

dle multimodality in response distributions (again, noting that this is for a single drift; a mixture of drifts / certain types of drift variability across trials can create multimodality Smith et al., 2020) and so the GSR appears to be the most efficient model for such cases. Across all of the simulations, spanning both complex and simple implementations of all the models, the GSR is consistently faster to simulate than the SCDM. This illustrates not only the practical utility of our approach but also its ability to condense the information from many correlated accumulators into a single state representation with similarities among response options.

Normalization and non-circular similarity

An important but perhaps not fundamental property of the SCDM is that the sum of evidence at each time step is constrained to be equal to zero, as with the model it was partially based upon, RCON2 (Ratcliff & Starns, 2013). This is also the case with the CDM, simply because options are arranged in a circle: the sum of support for one response option and its opposite (antipodal) alternative sums to zero (such as 0 and 180 degrees, or 80 and 260 degrees). Every point on the circle has an opposite, meaning the sum of evidence across all points on the circle (directions / alternatives) will always be zero. This will be the case with hyperspherical models (Smith & Corbett, 2018) as well, as each response option / point has a unique opposing point on the sphere (and formally, because $\int \dots \int_0^{2\pi} \cos(\phi) d\phi = 0$).

Typically, a multidimensional scaling solution will actually build this into the GSR model, with the sum across vectors corresponding to all alternatives equal to zero – for example, the solutions shown in Figures 3-4 will generate a decision space where evidence for 180 degrees is evidence against orientations at opposing orientations, such that the sum of evidence across all alternatives naturally sums to zero. Thus, normalization is commonly built into this construction of the alternatives. Despite this, normalization will not always be naturally built in through the multidimensional scaling procedure. This may not be an issue, as it is not clear that it is necessary or desirable for a model to have this property (Gluth et al., 2020). But in the case that it is useful, a modeler can simply build normalization into the translation between model similarities. This can be accomplished by computing the shift in evidence across all alternatives when information favoring each alternatives is gathered, taking into account that many of these shifts must be negative in order to conserve the sum.

Concerns in fitting real data

When fitting a model to data from a continuous response paradigms, a researcher is effectively trying to predict the joint distribution of responses and response times. The CDM has an analytic likelihood that spans these two dimensions,

and as we suggested earlier, the distribution of response times and hitting points in the GSR model will be identical to the CDM or hyperspherical extension (Smith, 2016; Smith & Corbett, 2018). There are a few instances where these likelihoods become more difficult to evaluate, such as when the number of dimensions in the models increases, where drift rate variability is introduced, where the start point is randomly distributed or systematically shifted (biased), or some combination of these factors. For the two-dimensional case, a characterization of drift rate variability is provided by Smith (2019), who derived the analytic distribution of response times and responses for a model with bivariate normally distributed drift rates. A bivariate normal may not be appropriate for all cases, and instead we may run into cases where the analytic distributions are very difficult or impossible to derive (e.g., the long-tailed distributions in Smith et al., 2020). In any case, drift rate variability results in a joint distribution of responses and response times where responses are fastest near the drift direction and slower as responses fall farther away from it. This is analogous to slow errors in binary choice (Ratcliff et al., 2016). Such a correlation between response time and response location has been found in some continuous response tasks (Smith et al., 2020) but not others (Kvam, 2019b; Ratcliff, 2018) suggesting that these types of tasks may inconsistently require drift variability to account for the relationship between responses and RTs.

Another case where the analytic likelihoods cause trouble is when the start point of the evidence accumulation process is not at the origin, indicating an initial bias toward some of the response options over others. In this case, it is useful to use other methods to obtain an approximate estimate of the likelihood of the data given the model. The most common is to simulate a large number of trials from the model of interest and evaluate the likelihood of the observed data in light of how frequently the same patterns appear in the simulations. One way to do this is to divide the observed and simulated data into quantiles (e.g., 10/30/50/70/90) and measure the degree of fit between the quantiles predicted by the model and those present in the data (Heathcote et al., 2002). However, this type of approach (and the region-based method used by Ratcliff, 2018) removes what is perhaps the most interesting property of this type of task – namely, that response and response times are both continuously distributed. Fitting RT quantiles loses information by fitting to summary statistics as opposed to complete likelihoods / distributions of response times (Turner & Sederberg, 2014).

Instead, we recommend an approach for model fitting based on kernel density estimation to turn the simulated data into a truly continuous, 2-dimensional distribution of responses and response times. This method has been effectively used to approximate the likelihoods of several types of simulation-based models (Palestro et al., 2018; Turner & Van Zandt, 2012; Turner & Sederberg, 2014;

Kvam & Busemeyer, 2020), is reasonably efficient especially with the addition of signal processing methods (Holmes, 2015; Lin et al., 2019), and can be easily adapted to a two-dimensional joint distribution like the one produced by the SCDM and GSR model. For these models, we can simulate a large number of trials from the model, use the kernel density method to generate an approximate likelihood, and then impute the likelihood of each combination of response and response time in the observed data set. This approach can similarly be applied to the CDM in order to build in bias in the starting points, for example (see Appendix C).

Connection to neural activity

As computational models are intended to explain behavior, a natural aspect of the model to consider is whether that model could be instantiated in the brain. The diffusion model and similar accumulation to boundary models have been connected to integrate-and-fire neural structures (Ratcliff et al., 2016; Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004), and even connected directly to neuroimaging data (Turner, 2015; Gold & Shadlen, 2007). Bidirectional constraints between behavioral and neural models improve both cognitive modeling and neuroscience (Forstmann et al., 2016, 2011) and place joint constraints on models of underlying cognitive processes (Love, 2015; Turner et al., 2019). The models we visit here are likely to be no exception: similarity is not only central to understanding behavioral data, but neural data as well. Comparable patterns of neural activation across different stimuli have been leveraged to understand “representational similarity” (Kriegeskorte et al., 2008). As the likely neural antecedent of behavioral similarity, neural representational similarity could be used to construct the multidimensional scaling solutions that we describe here.

On a more basic level, individual neurons at all levels of cortex have tuning functions that respond to more than just one stimulus (Ferster & Miller, 2000), including stimuli that are related in structure or function / meaning (Yee et al., 2010). The same can be said of the angles that the GSR model uses to represent similarity among response options – the subjective representations in the GSR model correspond to fundamental neural representations that drive every type of behavior. These overlapping activation functions, represented by the angles between response options (Kvam, 2019a), create both confusion and generalization between stimuli (Howard & Franklin, 2015) and permit decision makers to generalize to new contexts while retaining some specificity between neural activity and stimuli (Shepard, 1987).

Recent work delving into the multidimensional structure of similarity and neural representations (such as Bobadilla-Suarez et al., 2020; Hebart et al., 2020) promises to deepen the connections between neural representational similarity and the similarity representations that models like

the GSR use to model behavior. One promising candidate for constructing the similarity scaling solutions in the model is the multidimensional feature space representations that are thought to exist in the hippocampus (Aronov et al., 2017; Mok & Love, 2019). These hippocampal “grid” maps of object features could be used to evaluate the pairwise distances between choice alternatives, representing conceptual (Constantinescu et al., 2016) or perceptual similarity (Theves et al., 2019). The main conceptual challenges seem to be determining the scaling of different features as well as the type of distance metric (e.g., city block or Euclidean) that yields the most coherent translation between neural maps and behavioral data on continuous response tasks.

Conclusions

Overall, the geometric extension of the CDM provides a bridge between completeness and computational tractability in models of continuous report paradigms. It simplifies the SCDM by reducing the complex evidence accumulation dynamics of many correlated accumulators into a spatial relation between alternatives, represented in terms of directions and alignment between support for different responses. In translating the kernel-based similarity rule of the SCDM into the cosine-based similarity rule of the CDM (Equation 8), it also connects these two models together under a common geometric framework. Once this is done, the remaining properties of the model are all analogous, which is reflected by their common parameters: drift rate, drift direction(s), diffusion rate(s), thresholds, and non-decision time are all matched across the three models.

Although we do not provide fits to behavioral data here, there would be nothing surprising about the performance of the GSR model. Because it can act as a general case of both the CDM and the SCDM, in the worst case it will provide identical fits to those reported in previous work (Ratcliff, 2018; Kvam, 2019b). It also has the potential to provide more parsimonious accounts of behavior due to its ability to compress the covariance information in the SCDM into spatial relationships between responses, but the degree to which the dimensionality of the model can be interpreted as parsimony is certainly an open question. It could potentially be addressed through approaches analogous to probabilistic latent semantic analysis, where the ideal number of dimensions is identified through iterative model comparison when constructing a representation (Hofmann, 1999, 2001).

Of the models we have covered, only the GSR model and CDM provide a truly continuous joint distribution of responses and response times, because horse race multiple-accumulator models inherently produce a discrete distribution of responses when the decision rule is based on which accumulator wins the race. Whether the underlying representation should be continuous or discrete may never be definitively resolved, but certainly there are cases where

many more than 100 accumulators will be desirable. For instance, tasks where responses can be made in two or three dimensions – such as foraging and spatial navigation tasks (Kahana et al., 1999; Weisberg et al., 2014) or the 2-dimensional plane we showed in the example above – will require grids of m^2 or m^3 accumulators to obtain the same precision as single-dimensional tasks with m accumulators. In these cases, data compression via multidimensional scaling or singular value decomposition may be highly desirable.

Perhaps the most important frontier for the model is that of representations of similarity. Our third example, derived from real data on nonlinear subjective stimulus representations, illustrates the potential challenges associated with trying to predict continuous responses like magnitude estimates based only on the objective stimulus magnitudes. Integrating a subjective similarity structure into dynamic decision models like the GSR model, whether through multidimensional scaling or other dimension reduction techniques, will be paramount to accounting for empirical data in many realistic choice scenarios.

References

- Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647), 719–722.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622–628. doi: 10.1111/j.1467-9280.2007.01949.x
- Behboodiani, J. (1970). On the modes of a mixture of two normal distributions. *Technometrics*, 12(1), 131–139.
- Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, 124(1), 1.
- Blanchard, B. A., Riley, E., & Hannigan, J. H. (1987). Deficits on a spatial navigation task following prenatal exposure to ethanol. *Neurotoxicology and teratology*, 9(3), 253–258.
- Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., & Love, B. C. (2020). Measures of neural similarity. *Computational brain & behavior*, 3(4), 369–383.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. doi: 10.1016/j.cogpsych.2007.12.002
- Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, 115(2), 396–425. doi: 10.1037/0033-295X.115.2.396

- Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292), 1464–1468.
- Corbett, E. A., & Smith, P. L. (2020). A diffusion model analysis of target detection in near-threshold visual search. *Cognitive Psychology*, 120, 101289.
- Cox, T. F., & Cox, M. A. (1991). Multidimensional scaling on a sphere. *Communications in Statistics-Theory and Methods*, 20(9), 2943–2953.
- DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 968.
- Dowd, E. W., & Golomb, J. D. (2019). Object-feature binding survives dynamic shifts of spatial attention. *Psychological science*, 30(3), 343–361.
- Ekman, G. (1954). Dimensions of color vision. *The Journal of Psychology*, 38(2), 467–474.
- Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 23(1), 441–471.
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology*, 67.
- Forstmann, B. U., Wagenmakers, E.-J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? *Trends in cognitive sciences*, 15(6), 272–279.
- Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple Heuristics that Make Us Smart*. New York, NY: Oxford University Press.
- Gluth, S., Kern, N., Kortmann, M., & Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour*, 1–12.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, 30.
- Harvey, N. (1997). Confidence in judgment. *Trends in cognitive sciences*, 1(2), 78–82.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, 9(2), 394–401. doi: 10.3758/BF03196299
- Heathcote, A., Wagenmakers, E.-J., & Brown, S. D. (2014). The falsifiability of actual decision-making models. *Psychological Review*, 121(4), 676–678.
- Hebart, M., Zheng, C. Y., Pereira, F., & Baker, C. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgments. *Nature Human Behavior*, 4, 1173–1185.
- Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 289–296).
- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1-2), 177–196.
- Hollands, J., & Dyre, B. P. (2000). Bias in proportion judgments: the cyclical power model. *Psychological Review*, 107(3), 500–524.
- Holmes, W. R. (2015). A practical guide to the probability density approximation (pda) with improved implementation and error characterization. *Journal of Mathematical Psychology*, 68, 13–24.
- Holzmann, H., & Vollmer, S. (2008). A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU. *ASTA Advances in Statistical Analysis*, 92(1), 57–69.
- Howard, I. S., & Franklin, D. W. (2015). Neural tuning functions underlie both generalization and interference. *PloS one*, 10(6), e0131268.
- Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121(1), 1–32.
- Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M., & Madsen, J. R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature*, 399(6738), 781.
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2, 4.
- Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1–27.
- Kvam, P. D. (2019a). A geometric framework for modeling dynamic decisions among arbitrarily many alternatives. *Journal of Mathematical Psychology*, 91, 14–37.

- Kvam, P. D. (2019b). Modeling accuracy and response times in decisions among many alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 301–318.
- Kvam, P. D., & Busemeyer, J. R. (2018). Dynamic and distributional properties of prices. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 653–658).
- Kvam, P. D., & Busemeyer, J. R. (2020). A distributional and dynamic theory of pricing and preference. *Psychological Review*.
- Lin, Y.-S., Heathcote, A., & Holmes, W. R. (2019). Parallel probability density approximation. *Behavior Research Methods*.
- Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Topics in cognitive science*, 7(2), 230–242.
- Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Attention, Perception, & Psychophysics*, 32(5), 397–408.
- Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology*, 37(2), 136–151. doi: 10.1111/j.2044-8317.1984.tb00797.x
- Mellers, B. A., Chang, S.-j., Birnbaum, M. H., & Ordonez, L. D. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 347–361.
- Mok, R. M., & Love, B. C. (2019). A non-spatial account of place and grid cells based on clustering models of concept learning. *Nature communications*, 10(1), 1–9.
- Montello, D. R., & Pick Jr, H. L. (1993). Integrating knowledge of vertically aligned large-scale spaces. *Environment and Behavior*, 25(3), 457–484.
- Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: {N}ovel data and a computational account. *Cognitive Psychology*, 78, 99–147. doi: 10.1016/j.cogpsych.2015.01.002
- Naveh-Benjamin, M. (1987). Coding of spatial location information: An automatic process? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 595.
- Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. Springer.
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864. doi: 10.1037/A0019737
- Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, 12(13), 21–21.
- Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review*, 125(6), 888–935.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. doi: 10.1162/neco.2008.12-06-420
- Ratcliff, R., & McKoon, G. (2020). Decision making in numeracy tasks with spatially continuous scales. *Cognitive Psychology*, 116, 101259.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4), 260–281.
- Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, 120(3), 697–719. doi: 10.1037/a0033152
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27(3), 219–246. doi: 10.1007/BF02289621
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317–1323.
- Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123(4), 425–451. doi: 10.1037/rev0000023
- Smith, P. L. (2019). Linking the diffusion model and general recognition theory: Circular diffusion with bivariate-normally distributed drift rates. *Journal of Mathematical Psychology*, 91, 145–158.
- Smith, P. L., & Corbett, E. A. (2018). Speeded multielement decision-making as diffusion in a hypersphere: Theory and application to double-target detection. *Psychonomic bulletin & review*, 1–36.

- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in neurosciences*, 27(3), 161–168.
- Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dhafarov (2014). *Psychological Review*, 121(4), 679–688.
- Smith, P. L., Saber, S., Corbett, E. A., & Lilburn, S. D. (2020). Modeling continuous outcome color decisions with the circular diffusion model: Metric and categorical properties. *Psychological Review*.
- Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, 53(2), 293–315. doi: 10.1348/000711000159349
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, 32(2), 135–168. doi: 10.1016/0022-2496(88)90043-0
- Snizek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational behavior and human decision processes*, 43(1), 1–28.
- Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional feature space. *Current Biology*, 29(7), 1226–1231.
- Turner, B. M. (2015). Constraining cognitive abstractions through Bayesian modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (p. 199–220). New York: Springer.
- Turner, B. M., Forstmann, B. U., Steyvers, M., et al. (2019). *Joint models of neural and behavioral data*. Springer.
- Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic bulletin & review*, 21(2), 227–250.
- Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate bayesian computation. *Journal of Mathematical Psychology*, 56(2), 69–85.
- Turner, B. M., & Van Zandt, T. (2018). Approximating bayesian inference through model simulation. *Trends in Cognitive Science*, 22, 826–840.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592. doi: 10.1037/0033-295X.108.3.550
- van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological review*, 121(1), 124.
- Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 669.
- Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.
- Yee, E., Drucker, D. M., & Thompson-Schill, S. L. (2010). fmri-adaptation evidence of overlapping neural representations for objects related in function or manipulation. *NeuroImage*, 50(2), 753–763.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233.

Appendix A: Number of modes in multimodal distributions

Although there SCDM can in principle produce multimodal data, whether or not a mixture of two distributions of evidence will result in multimodal response distributions depends on how similar the representations of the two options are relative to the width of the similarity kernel σ_c . If there are two sources of information, and there is a Gaussian similarity relation between them, then information for the two alternatives, summed together, will not be multimodal unless the width of σ_c is below a certain value. Exactly how small it has to be can be derived from a test for unimodality of the mixture of two normal random variables, such as Holzmam & Vollmer (2008). Formally, if the location of one alternative is at m_1 and the location of the other is at m_2 , then the aggregate distribution of evidence resulting from combining evidence evenly from both alternatives will be multimodal if and only if

$$|m_1 - m_2| > 2\sigma_c \quad (11)$$

In the case that there is a probabilistic mixture of the two sources of evidence – such as the case where m_1 is the ‘true’ globally maximal signal and m_2 is a ‘distractor’ local maximum – the distribution of evidence will not be multimodal unless the following condition is met (Behboodan, 1970):

$$|m_1 - m_2| > 2\sigma_c \sqrt{1 + \frac{|\log(p) - \log(1-p)|}{2}} \quad (12)$$

As a result, a set of alternatives spanning the range $[a, b]$ will be able to have at most $\min(N_{\text{modes}}, n)$ unique modes in the response distributions, where n is the number of accumulators and N_{modes} is computed as

$$N_{\text{modes}} = f\left(\frac{a-b}{2\sigma_c \sqrt{1 + \frac{|\log(p) - \log(1-p)|}{2}}}\right), \quad (13)$$

Here, $f(x) = \lceil x \rceil$ is the “ceiling” function (rounding up to the nearest integer). Thus, the number of unique modes that can exist in the data using a Gaussian similarity rule to relate the accumulators is $\lceil \frac{a-b}{2\sigma_c} \rceil$. Conversely, the key to multiple modes in the GSR model is that there can be no alternative on the hypersphere O^d that is located nearer to the geodesic between the two vectors corresponding to the location of the modes than either of the constituent alternatives.

The maximum number of unique modes that can be represented in d dimensions using a cosine similarity rule – for a set of alternatives arranged on a hypersphere – is equal to $2(d-1)$. This is because a complete cycle of $y = \cos(x)$ ($x \in [0, 2\pi)$) will cross every value of $y \in [-1, 1]$ exactly twice, but there cannot be one full cycle without at least two dimensions. Each additional dimension confers one additional cycle in cosine similarity, such that there is one cycle in

two dimensions, two cycles in three dimensions, three cycles in four, and so on. Therefore, the number of dimensions necessary to represent an evidence accumulation process with Gaussian similarity kernel width σ_c is

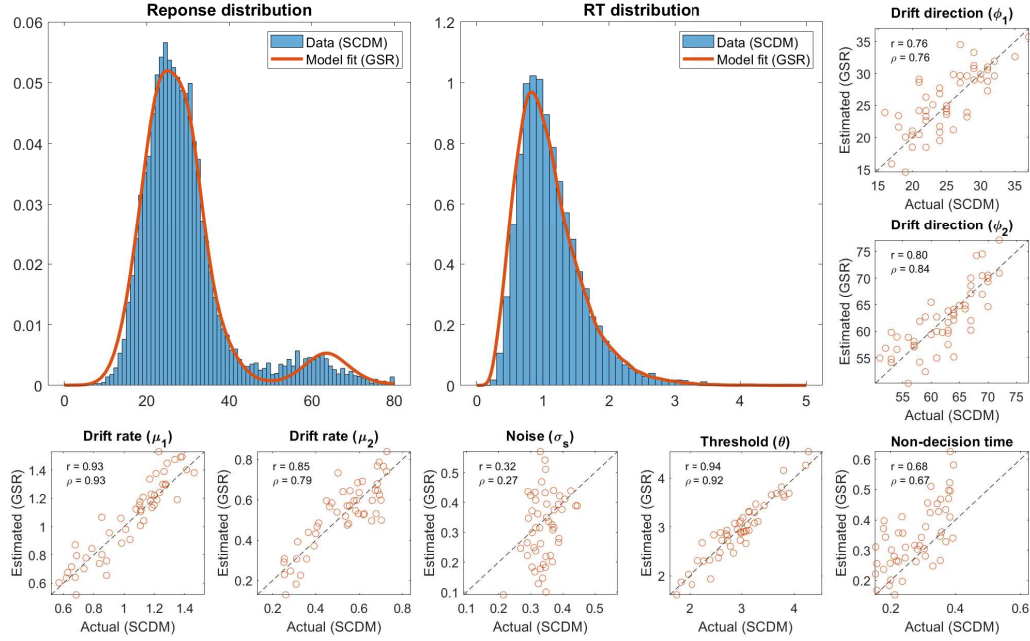
$$d_c = \left\lceil \frac{1}{2} \left\lceil \frac{a-b}{2\sigma_c} \right\rceil + 1 \right\rceil \quad (14)$$

Hence, d_c is the maximum number of dimensions needed for the multidimensional scaling algorithm to produce a representation with sufficient flexibility to allow multi-modality in the data. Therefore, we know that we will need no more than d_c dimensions in the GSR model to account for data with $d_c - 1$ modes, providing an upper bound on the required model complexity for the GSR model that keeps it from being as complex as the SCDM.

Appendix B: Model mimicry study

In this article, we showed that the GSR model can mimic the most important properties of the SCDM, including multimodality. However, it is not necessarily clear that both models will give a researcher the same information. On the surface, the drift rate and direction, noise, threshold, and non-decision time all have the same interpretation as one another, in addition to reflecting the one-dimensional case of the diffusion model (Smith & Corbett, 2018). To test whether the GSR model was capturing the same information as the SCDM, we performed a model mimicry analysis. For this analysis, we simulated a large volume of data from the SCDM for multiple “participants” all with different values of the important parameters of the SCDM. To make a realistic and demanding situation from SCDM, we focused on a case where responses could fall along a line (i.e., were not wrapped around a circle) and where there was a stimulus with two modes – a “central” location and a “secondary” location. This setup can be compared to the bright / dark dot patches used by Ratcliff (2018), Experiment 3.

For the simulations, we generated 50 artificial participants who each completed 200 trials of a task with a bimodal stimulus. To generate the simulated data from each participant, we used 80 accumulators in the SCDM. The locations of the true stimuli modes (drift direction), drift rates for the two main stimuli, diffusion rate / noise, threshold, and non-decision times were randomly varied across participants. Drift locations for the primary stimulus locations were drawn according to a normal random variable with mean of 25 and standard deviation of 5, whereas locations for the secondary stimuli were drawn from a normal distribution with mean at 62 and the same standard deviation. The drift rates for these stimuli were drawn from uniform distributions from .5 to 1.5 and .25 to .75, respectively, such that the secondary stimulus could sometimes be stronger than the primary one (which should test the model’s flexibility to accommodate many patterns of stimuli). Threshold and diffusion rate / noise were

**Figure 6**

Fits of the GSR model to data generated from the SCDM. Large plots show the best fit predictions generated by the GSR model (orange lines) overlaid on a histogram of data generated from the SCDM (histogram) for response (left) and response time (right) distributions. Smaller plots show the relationship between SCDM parameters used to generate the data (x-axis) and the estimated parameters of the GSR model based on maximum likelihood fits to the data (y-axis). GSR: Geometric Similarity Representation. SCDM: Spatially Continuous Diffusion Model.

drawn from normal distributions with mean 3 and standard deviation .5 and mean .35 and standard deviation .05, respectively, truncated so they could not go below zero. Non decision time was drawn from a uniform distribution $U(.15, .4)$ to reflect values of the nondecision time parameter that are typically observed in similar perceptual decision making experiments.

Once the data were generated from the SCDM, we fit the GSR model to the joint distribution of responses and response times. The GSR model had five dimensions, and used 5000 simulated trials for each participant to approximate the likelihood surface (predicted joint distribution of responses and response times) for each combination of parameters. To estimate the parameters of the GSR model, we used maximum likelihood estimation (`fminsearch` in MATLAB) with 100 different start points. Start points were generated by drawing a multivariate normal random variable around drift directions of 20 and 60, drift rates of 1 and .5, threshold of 3, noise of .3, and non-decision time of .2. Once the parameters were estimated, we compared the maximum likelihood estimates from the GSR model against the generative parameters of the SCDM, and compared the predicted distribution of responses and response times (from the GSR model) against

the ones generated by the SCDM (i.e., the data that were being fit).

Figure 6 presents the results. The top left panel shows the observed distributions of responses (blue histogram) and predictions from the GSR model from the best-fitting model parameters (orange), whereas the top middle panel shows the observed distributions of response times (also shown in blue histogram) and predictions from the GSR model (also shown in orange). Generally, the GSR model had no trouble recovering the pattern of data generated by the SCDM, as we might expect from our examples in the main text.

As for the parameters themselves, most parameters showed a close relationship between the generative parameters from the SCDM and the estimated parameters from the GSR model. Each parameter is shown around the edges of Figure 6, labeled to indicate which parameter is shown. For reference, we include both the linear relationship between generative and estimated parameters (r) and rank correlation between them (ρ). In all cases, these relationships were substantial and positive. The only cases where they dipped below .6 was in the noise parameter (σ_s), where the GSR model tended to over-predict the variability in this parameter relative to the true SCDM parameters. This is likely due to the

relative modest effect of σ_c in this dataset: there were not that many ‘pure’ mistakes, and so slightly smaller or larger values of the diffusion rate did not create too substantial of differences in the observed data. As a result, the GSR model did not have much variability in the data to meaningfully inform its estimates of the noise.

Overall, the parameters of the GCDM and GSR model showed relatively close agreement. A GSR model with more dimensions is likely to show even better performance by yielding a representation of response options that has higher fidelity. Meanwhile, it is possible that we could modify the SCDM to make even more difficult patterns of data to accommodate. However, this exercise illustrates that the GSR model does exactly as we described in the main text, which is to condense the information in the correlated accumulators of the SCDM into representations of similar response options. Ultimately, the GSR model performs well in accommodating the data from the SCDM, while decreasing the dimensionality of the evidence representation down to 1/16th of its original complexity.

Appendix C: Implementational study

Considering simulation and analytic implementations (where available) of the three approaches leaves us with four potentially interesting models with different capabilities: (1) the analytic version of the circular diffusion model (CDM), (2) a simulated variant of the circular diffusion model, (3) the new geometric similarity representation (GSR) approach introduced in this paper, and (4) the spatially continuous diffusion model (SCDM). To examine the practical implications of choosing to use each of these models, we apply each model in several different scenarios and examine the relative amount of time it takes to evaluate the likelihood function of each model. The speed with which the likelihood can be evaluated constrains a modeler’s ability to test new models, add new mechanisms, and compare within and across different models. It is therefore not only a concern of efficient use of the modeler’s time, but one of theoretical innovation as well: if a model will take several weeks to fit, the modeler can only test so many different alternative accounts of the data.

To thoroughly examine this aspect of the models, we implemented 2-4 versions of each model, depending on what conditions they qualitatively could implement. For all models, we set the baseline threshold to 5, drift rate to 0.05, diffusion rate to 0.20, drift direction was π , and non-decision time to 250 ms. These were chosen to create a realistic likelihood function where responses were distributed across approximately half the scale and response times were concentrated between 500-3500 ms, roughly reflecting the properties of past experimental results. For the simulated models, the time between steps of the process was 10 ms. This time can be used directly as the time for each step or, if the

modeler wishes to preserve the Markov property of the diffusion process, the time between steps can be exponentially distributed as $\text{Exp}(0.01)$. The latter approach also allows for a continuous distribution of response times, to maintain consistency with models that predict continuous distributions of responses.

The results of four different tests are shown in Table 1. The first of these tests was a “high resolution” test of the models, where great demands were placed on precision for each of the models as would be the case in realistic modeling scenarios. For the analytic CDM, this meant computing 150 terms of the zeros of the Bessel function in Equation 2, which provides a smooth likelihood and low occurrence of unusual behavior on the leading edge of the response time distribution (when fewer terms are used, there can be erratically high probability densities around the leading edge of the RT distributions). For the simulated models, this meant that they used 5000 simulations to generate the likelihood surface. The high resolution GSR model used 10 dimensions, allowing for at least 11 modes in the data to be distinguishable. The high resolution SCDM used 100 accumulators, allowing for a distribution of responses that could nearly approximate a continuum.

To make the simulation times somewhere in an achievable range for a modeler with modest computing resources, we implemented the models on a laptop computer with a 4-core, 2.4 GHz intel Core i7 processor in MATLAB 2017b. This is not meant to be a comprehensive study of how well the models can be fit on a variety of the most advanced supercomputing resources with massively parallelized processes, but to give the reader an idea of the *relative* efficiency of the different modeling approaches. The analytic CDM naturally provided the best performance in terms of computing time, while the simulated CDM and GSR model took about 2 seconds per likelihood computation – with a total time of just over half an hour. The SCDM took almost two orders of magnitude longer than either of the other simulated models, taking several days for just a thousand samples in a single condition, which would be effectively prohibitive for many modeling goals.

The second test evaluated very coarse approximations of each model, allowing them to reduce the computational demands at the expense of making precise predictions. This test was designed to evaluate what the lower bound on reasonable computing time for each model would be. In this test, there were only 40 terms computed for the zeros of the Bessel function for the analytic CDM, 100 simulations to generate the likelihood surface of the simulated models, 3 dimensions for the GSR model, and 20 accumulators for the SCDM. Here the results show greater parity – the SCDM only takes twice as long as the other simulated models, reflecting our suggestion above that the number of accumulators and their covariance is the main culprit in terms of long

Table 1

Time taken to complete 1000 likelihood computations for a (simulated) data set of 2000 trials. High resolution models tested precise predictions generated using many simulations, low resolution models tested very rough predictions using few simulations, start bias models tested conditions where the initial point was not [0,0], and multimodal models tested conditions where the stimulus had multiple favored response directions.

Model	High resolution	Low resolution	Start bias	multimodal
CDM (analytic)	4.76 s	0.42 s		
CDM (simulated)	2,230.19 s	214.49 s	1,075.34 s	
GSR	2,868.83 s	246.68 s	1,160.03 s	1,430.28 s
SCDM	168,141.68 s	531.12 s	5,152.75 s	6,062.06 s

computation times.

The third and fourth tests were designed to illustrate the capabilities of each of the models as well as the length of time it might take to fit them to data. The third test examined how starting point shifts impacted the performance of the models. In this test, we put intermediate demands on computations such that the simulated models used 1000 simulations per likelihood computation, the GSR model used 5 dimensions, and the SCDM used 50 accumulators. The major finding of note is that the analytic CDM cannot be used in cases where pre-stimulus information or asymmetric rewards induce a bias in the start point of the model (although there are suggestions that this is not how bias due to predecision information works in continuous paradigms; see Kvam, 2019b). This leaves the simulated CDM and GSR model as the fastest working models, within 100 seconds of one another, while the SCDM takes approximately 5 times as long as either one.

Finally, the last test was designed with multimodality in mind, and used multiple drift directions to drive the evidence accumulation process (such as a stimulus with two main competing spatial locations). Because it uses a cosine similarity rule, the analytic and simulated CDM could not accomplish a reasonable account of the multimodal response distribution. The GSR model, however, worked well in this case and was able to out-perform the computation time of the SCDM by a factor of 4.