



A geometric framework for modeling dynamic decisions among arbitrarily many alternatives

Peter D. Kvam¹

The Ohio State University, Columbus, OH, USA
University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:

Received 27 November 2017
Received in revised form 1 March 2019
Accepted 4 March 2019
Available online 19 March 2019

Keywords:

Multi-alternative choice
Continuous response
Computational modeling
Random walk
Vector space
Optimal choice

ABSTRACT

Multiple-choice and continuous-response tasks pose a number of challenges for models of the decision process, from empirical challenges like context effects to computational demands imposed by choice sets with a large number of outcomes. This paper develops a general framework for constructing models of the cognitive processes underlying both inferential and preferential choice among any arbitrarily large number of alternatives. This geometric approach represents the alternatives in a choice set along with a decision maker's beliefs or preferences in a "decision space," simultaneously capturing the support for different alternatives along with the similarity relations between the alternatives in the choice set. Support for the alternatives (represented as vectors) shifts over time according to the dynamics of the belief / preference state (represented as a point) until a stopping rule is met (state crosses a hyperplane) and the corresponding selection is made. This paper presents stopping rules that guarantee optimality in multi-alternative inferential choice, minimizing response time for a desired level of accuracy, as well as methods for constructing the decision space, which can result in context effects when applied to preferential choice.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

From laboratory tasks like orientation judgments to real-world ones like price setting or spatial navigation, humans frequently make selections among large sets of potential responses. However, decision models have traditionally focused on the binary case where a decision-maker has only two or three alternatives available for selection, resulting in a relative theoretical and empirical underdevelopment in our understanding of how people deal with larger choice sets. Binary and trinary choice models have been extremely successful in accounting for choice proportions and response times in everything from simple perceptual tasks (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004) to more complex preferential and multi-attribute choice (Bogacz, Usher, Zhang, & McClelland, 2007; Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993), but extending them to modeling choices among larger sets with many alternatives comes with a host of new practical and theoretical challenges.

Perhaps the most critical challenge of moving toward modeling multi-alternative choices is that a modeler must account for

how alternatives in a choice set are related to one another. These relations come from similarity between choice options, which can arise from shared features or common neural activity resulting from overlapping activation patterns. These similarity relations are often incorporated into multialternative choice as additions to the mechanisms used in binary choice. For example, decision field theory (Roe, Busemeyer, & Townsend, 2001) computes a 'distance' between alternatives in a choice set based on their features, and applies a distance-dependent inhibition between alternatives that allows it to account for important phenomena in multialternative choice. Similarly, the leaky competing accumulator model (Usher & McClelland, 2001, 2004) explicitly includes lateral inhibition between the support for different alternatives, reflecting the interactive processes between alternatives in neural systems.

One drawback of these approaches is that they must use separate accumulators to store support for each alternative in the set, and then introduce dependencies between alternatives by building in pairwise interactions between the accumulators. The number of accumulators and binary comparisons needed to fully specify the decision process grows very large as the number of alternatives increases. In many cases, the number of parameters needed to specify all accumulation rates ($\sim n$ drift rates for n alternatives) and pairwise relations ($\sim n(n-1)/2$ comparisons for n alternatives) becomes unmanageable as the number of alternatives grows large, making them unsuited to

E-mail address: kvam.4@osu.edu.

¹ This research was supported by a fellowship from the National Science Foundation, USA, Grant No. DGE-1424871. The author would like to thank Tim Pleskac, Jerome Busemeyer, Philip Smith, and Jun Zhang for helpful comments on the ideas presented here.

predicting selections among many alternatives. This problem is exacerbated when considering responses on a continuum – such as indicating two-dimensional motion direction on a circle, or selecting a hue along a color wheel – where unique accumulators cannot reasonably be assigned to every possible response.

These types of tasks were recently addressed by the work of Smith (2016) (see also Ratcliff, 2018; Smith & Corbett, 2018), who proposed a multidimensional diffusion process with circular boundaries. This model makes very specific assumptions about how choice alternatives along the continuum are related to one another with alternatives on opposite sides of the circle providing evidence directly against one another when support for one is generated. For example, evidence favoring an alternative located at 45 degrees provides evidence against an alternative located at 225 degrees, and some evidence for alternatives located at 40 or 50 degrees. Although it does not provide all of the answers to modeling choice among many alternatives, adopting a similar spatial way of representing alternatives and their support can lead us to a more general theory of decision making among multiple alternatives.

1.1. A geometric theory of dynamic decision making

This paper explores a general theory for multi-alternative choice by establishing a geometric framework for constructing models of dynamic decision making. At its core, the geometric framework is a method for addressing three fundamental questions about the decision process:

- What alternatives does a decision-maker have and how are they related to one another?
- What information does the decision-maker consider and how does this change their beliefs/preferences relative to the alternatives?
- What conditions trigger a decision in favor of each of the available alternatives?

These questions bear a close resemblance to the search, stopping, and selection rules that are used to characterize heuristic decision strategies (Gigerenzer, Todd, & A.B.C. Research Group, 1999; Weitzman, 1979), but the first question elucidates an important factor that multi-alternative models in particular have to address. The decision process requires knowing or finding out what choice alternatives are available and how those alternatives are related to one another. The assortment of alternatives defines a context in which a decision takes place, resulting in the context effects relevant to multi-alternative choice.

The geometric framework proposes a direct answer to each of these questions while allowing room for domain-specific theory to inform the resulting models. In short, it states that (1) alternatives are represented as vectors in a multidimensional space, oriented in directions determined by their psychological representation; (2) beliefs or preferences are represented as a point in this space that shifts over time according to a sampling distribution that describes the information considered during the decision; and (3) an alternative is chosen when the belief/preference state meets a particular stopping condition, corresponding to a choice boundary formed by a hyperplane in the multidimensional space.

There is a great deal to unpack from this, so I begin by establishing what an optimal decision strategy prescribes. The first section of this paper establishes the basic elements of the framework by building an optimal model for decisions among n alternatives (where each piece of incoming evidence favors exactly 1 of the available options). The strategy this model describes minimizes the amount of time required to make a selection for a

desired level of accuracy. It can therefore be used as a benchmark optimal model against which we compare other models constructed using the geometric framework.

The second section of the paper examines cases where the assumptions of this model fail, primarily due to similarity relations between the available alternatives. In these cases, it is still possible to construct a model that optimizes decision time for a desired accuracy, but the rules for stopping the accumulation process and selecting an alternative must be conditioned on the similarity relations between options in the choice set. Examining how thresholds should be set to maintain accuracy allows us to make empirical response time predictions for different alternatives in these models, and I provide MATLAB code for deriving these predictions across different sets of alternatives and decision rules.

The third section of the paper extends the approach to modeling selections along a continuum, where the range of alternatives and the support for them must be represented as (nearly) continuous quantities. It examines the circular and hyperspherical diffusion model (Smith, 2016; Smith & Corbett, 2018) as a special case of the geometric framework, where each possible alternative in the space is available for selection and thresholds are held constant across alternatives. It then examines cases where constant thresholds might not be optimal and suggests avenues for correcting choice boundaries for the similarities between alternatives in continuous selection tasks.

The fourth section of the paper looks at assumptions and approaches that should be taken to construct geometric models of more complex multi-attribute and preferential choice. While the relations between alternatives in inferential choice can often be inferred from the physical properties of the stimuli or the common structure of neural systems, the subjective nature of preferential choice and the presence of correlated features make the decision space less straightforward to construct for these types of selections. In these cases, additional data such as binary confusion matrices or similarity judgments (used in conjunction with multidimensional scaling or singular value decomposition) or neural data (using multi-voxel pattern analysis to infer representational similarity) can be used to determine the relative spatial orientations of alternatives. It also illustrates how the geometric framework naturally results in context effects in these types of multi-alternative choices.

The paper concludes with some final comments about evidence dynamics and the possible applications of the geometric framework. The range of these applications is exceptionally wide, as the geometric framework is at its core a toolbox for developing models – a general theory with interchangeable parts rather than a domain-specific model.

2. Optimal geometric models

An obvious first question for models of multi-alternative choices is: how *should* people make decisions in these cases? In many choice scenarios, we might expect decision behavior to approximate optimality due to the strong selection pressures for decision success in our ancestral environments (Oaksford & Chater, 2007) (though see Kvam, Cesario, Eisthen, Schossau, & Hintze, 2015; Kvam & Hintze, 2018, regarding evolutionary claims to optimality in simple binary choice). Even when observed behavior does not appear to follow the normative or optimal pattern, establishing a normative or optimal model for n -alternative choice creates a benchmark for us to examine how and why behavior deviates from their predictions. This can allow us to better understand the computational constraints of the agents (human decision-makers) and how they result in departures from optimal performance (Anderson, 1990).

For these reasons, many existing models of binary and even trinary choice are based on an optimal decision policy that minimizes the amount of time it takes to make a decision for a particular desired level of accuracy. Most are derived from the sequential probability ratio test [SPRT] (Edwards, 1965; Wald & Wolfowitz, 1949), in which a decision-maker calculates and updates the odds of a pair or set of hypotheses, and continually checks these odds against an internally-set criterion θ . By stopping and making a decision when the odds exceed this threshold, the decision-maker can ensure a minimum level of expected accuracy, provided that the prior odds of the hypotheses were properly set.

In the binary version of the SPRT, information about the likelihood of the alternatives is represented as a balance between support for the two options: the evidence state is the quantified as the support for option A minus the support for option B. In an inferential decision, this difference $Ev(A) - Ev(B)$ is linearly related to the posterior log odds of one hypothesis relative to the other, $LO(A : B)$.² To ensure that the probabilities of the hypotheses sum to one in binary choice, it must be the case that the log odds always sum to zero $Ev(A) + Ev(B) = 0$ (this need not hold in multi-alternative choice, but the probabilities must still always sum to one). As the decision-maker gathers new information, this balance shifts toward A or toward B. A decision is made when the posterior log odds for one option relative to the other exceeds a criterion value θ corresponding to a desired level of accuracy. The minimum bound θ ensures that the decision-maker can be confident with at least $ce^\theta : 1$ odds that the hypothesis they have chosen is the correct one, where c is a scaling factor that corresponds to the baseline difficulty of discrimination between the two hypotheses (a measure of d' in signal detection theory).

When we visualize this evidence accumulation process as it unfolds over time, it generates a 1-dimensional random walk in log odds space, describing an optimal Bayesian belief updating procedure (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Palmer, Huk, & Shadlen, 2005). In terms of stochastic processes, it is described by a Markov chain where the position at time $t + 1$ depends on the position at time t and the incoming evidence, which determines the transitions between different levels of evidence for A relative to B (see Diederich & Busemeyer, 2003).

This process has a clear spatial representation, where the two alternatives are oriented in opposing directions such that gathering support for option A moves the state in one direction (up) while support for option B moves the state in the opposing direction (down, see Fig. 1A). The support for option A is the component of the evidence state in the direction of A, described by a vector v_A pointing upward. For example, if the evidence state is $s = +3$, the component of $[+3]$ along the positive direction vector $v_A = [+1]$ describes the evidence for A, and the component of $[+3]$ along the negative direction vector $v_B = [-1]$ describes the evidence for B. Since the evidence continuum is unidimensional the computations are elementary: the support for alternative A is $comp_{v_A}(+3) = 3$ and the support for B is $comp_{v_B}(+3) = -3$, resulting in a balance of $Ev(A) - Ev(B) = 6$, which is linearly related to the log odds of A:B by the scalar coefficient c .

Note that cases where the alternatives do not directly oppose one another can actually be mapped into this representation by adjusting the value of the scalar c (Bogacz et al., 2006), but this is only straightforward for binary choice. I return to this point later: uneven relations between alternatives are a fundamental part of multi-alternative choice, so they cannot often be handled in the same way.

² It should be noted that many diffusion models depart from the log odds framework (Link & Heath, 1975; Ratcliff, 1978). Models can still provide good descriptions of the decision making process without making explicit reference to the posterior odds of the various hypotheses.

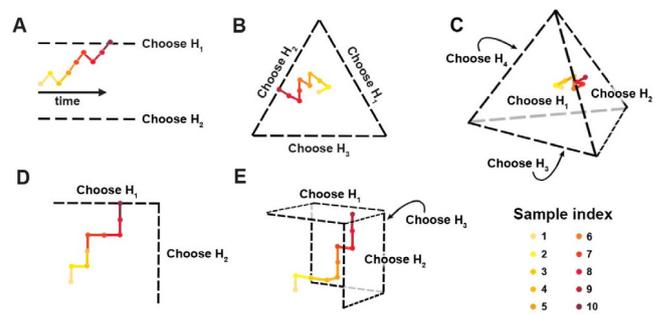


Fig. 1. Representation of a person's accumulated evidence and choice criteria for 2-, 3-, and 4-alternative relative evidence processes (A, B, C) and 2- and 3-alternative absolute evidence processes (D, E). An alternative is chosen when a person's represented evidence (yellow/red) crosses the corresponding edge for models A, B, and D, but the plane/face for models C and E. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Response time predictions from the random walk can be made by setting the threshold value θ as well as the probability of stepping up or down with each new piece of evidence (using parameter p for the probability of stepping up and $1 - p$ for stepping down) and attaching a distribution or fixed value to the time between sequential steps of the accumulation process (t). Typically this is done by assuming each step takes a fixed time, such as $t = 10$ ms. If we wish to preserve the Markov property when moving to continuous time, the time between steps should instead be exponentially distributed with rate parameter γ , such that the time between steps $t \sim Exponential(\lambda)$. The number of steps it takes to reach one of the boundaries $Ev(A) - Ev(B) \geq \theta$ (response favoring A) or $Ev(A) - Ev(B) \leq -\theta$ (response favoring B) can be put together with t to compute a distribution of response times. For example, taking 300 steps might result in a response time of 3s if t is fixed to 10 ms, or a distribution $RT \sim Gamma(300, \lambda)$ if t is pulled from a distribution $Exponential(\lambda)$.³

If evidence samples are arriving continuously rather than in discrete packets, we can take the limit as $t \rightarrow 0$ and the step size also approaches 0. This yields a Wiener diffusion process as described by Ratcliff (1978) and forms the basis for the diffusion decision model of binary choice (Ratcliff & Smith, 2004; Smith & Ratcliff, 2004). The response time predictions for binary choice models using random walks and diffusion processes are well-covered in the cognitive modeling literature (Diederich & Busemeyer, 2003; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016; Smith & Ratcliff, 2004), so I do not go into detail on them here. However, I do provide MATLAB code that can flexibly adjust to different numbers of alternatives (including binary choice) in order to derive the response time predictions for a normative model of n -alternative choice. To see how this is done, it is helpful to examine how the optimal model should change for selections among three alternatives and then build to the general case of n alternatives.

2.1. Simple optimal models

Moving to multi-alternative choice requires us to reconsider how support for one alternative interacts with other alternatives. This is simplest when all of the available choice alternatives are mutually exclusive, each new piece of evidence favors exactly one of the alternatives, and a decision-maker desires to minimize

³ The value of λ can also be fixed to an arbitrary value like 10 ms as long as the step size of the random walk is adjusted accordingly; it does not need to invoke an additional free parameter into the model.

response time for a specific level of accuracy (e.g. 80%) when selecting one of the options in the set. In such a case, evidence for one alternative is evidence against all other alternatives equally. Since the decision-maker only selects one option, the probability of choosing across all choice options (events) sums to one. These conditions lead to a multi-dimensional representation of evidence that unfolds within a set of response boundaries that form a simplex. However, this jump is not immediately obvious. I first cover the case of three alternatives and examine what must change from binary choice, then use this to build to the general case for n alternatives.

2.1.1. Three alternatives/triangle

Suppose a decision-maker has three mutually exclusive options – A, B, and C – and each new piece of evidence favors exactly one of the three. The probabilities of all alternatives still must sum to one, meaning that a +1% increase in support for A (in choice set {A,B,C}) must result in a –1% decrease in support for set {B, C}. Similarly, +1% support for B results in –1% support for set {A, C}, and +1% support for C results in –1 support for {A, B}. This necessarily means that +1% support for one alternative must decrease support for each the other two by 0.5%. The representation of this type of decision was laid out in early work by Audley and Pike (1965), although a different type of decision rule is used here.

Returning to the vector-based representations of alternatives, this implies that evidence moving the state toward A in direction v_A should move the state half that distance away from B, which is in direction v_B . For example moving 2 steps toward A should take the decision-maker 1 step away from B and 1 step away from C. This implies that the component of vector length 1 in direction v_A should have a component along v_B as well as v_C of $-\frac{1}{2}$. Put together, this means that the following relations hold:

$$\begin{aligned} \text{comp}_{v_B}(v_A) &= -\frac{1}{2} \\ \text{comp}_{v_C}(v_A) &= -\frac{1}{2} \\ \text{comp}_{v_B}(v_C) &= -\frac{1}{2} \end{aligned} \tag{1}$$

This necessarily implies that the angle between all pairs of vectors v_A , v_B , and v_C is exactly $\frac{2\pi}{3}$, or 120 degrees. This is because $\text{comp}_y(x) = \|x\| \cos(\phi_{xy}) = -\frac{1}{2}$, where ϕ_{xy} is the angle between the vectors, implies $\phi_{xy} = \cos^{-1}(-\frac{1}{2}) = \frac{2\pi}{3}$. The resulting arrangement of vectors for alternatives A, B, and C is shown in the left panel of Fig. 2. This results in a very simple geometric relationship between the evidence for different alternatives, which is at the core of the geometric framework. Given evidence E_A that favors option A, the amount of evidence it provides for any option B, E_B , depends on the angle ϕ_{AB} between v_A and v_B :

$$E_B|E_A = \|E_A\| \cos(\phi_{AB}). \tag{2}$$

The stopping conditions for this case will be similar to the binary choice case. Once the component of the evidence state along v_A , v_B , or v_C exceeds a criterion value θ , a choice is made in favor of the corresponding alternative. This guarantees that the net probability in favor of (e.g.) A relative to the set of all other alternatives is at least θ , guaranteeing a level of accuracy determined by $\theta - 1/n$ (assuming there is exactly one correct response).

Since we assumed that each incoming piece of evidence favors exactly one alternative, we can represent this optimal accumulation process as a random walk on a hexagonal lattice, shown in Fig. 2. This allows for steps toward v_A at 60 degrees, v_B at 180 degrees, or v_C at 300 degrees with probabilities p , q , and $1 - p - q$ according to the stimulus presented. The 3-alternative case

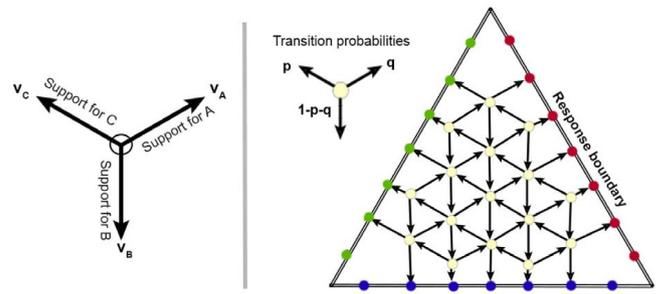


Fig. 2. Vectors describing direction of support for each alternative (left) and hex lattice for a 3-alternative random walk model with threshold $\theta = 2$ (right). Red, blue, and green nodes correspond to absorbing states for alternatives A, B, and C, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

only requires additional parameters specifying the probabilities of steps in the possible directions (2, giving p , q , and $1 - p - q$ as the step probabilities) and the time between steps in the chain λ .

In some cases, it can be useful to model this random walk as a Markov chain with a finite number of states. This allows for response time predictions to be analytically derived using a continuous-time, discrete-state model (see Diederich & Busemeyer, 2003).⁴ Critically, the model of Diederich and Busemeyer (2003) normalizes the sum of evidence for all alternatives to be zero, preserving the total probability as well as allowing for better model fits a similar transformation is used to improve model performance in Ratcliff and Starns (2013). Evidence that a person has at time t is specified by a vector $s(t)$, which gives the location of the evidence using a point within the lattice. In the lattice shown in Fig. 2, s would have 31 entries, designating the probability that a person’s beliefs correspond to each of the 31 possible belief states. A 31×31 transition matrix then specifies the probabilities of moving from one state to another, with 15 of these specifying absorbing states that generate a response. Code for this Markov chain model is provided at osf.io/75qv4/. However, as the threshold gets substantially higher (greater than $\theta = 2$ steps shown on the right of Fig. 2) a large number of states are required, and this approach to modeling quickly becomes impractical.

Whether using a Markov chain or simulating the random walk in continuous space, the stopping rule is the same. When the component of the state s along any of the vector describing the choice alternatives v_A , v_B , or v_C exceeds θ , a decision is triggered in favor of the corresponding alternative. For an optimal model with 3 alternatives and arbitrary choice criterion θ , the choice boundaries will form an equilateral triangle like the one shown above. If the process starts at the origin and moves in 2 dimensions, the system of equations specifying the interior of the triangle (the conditions under which evidence accumulation will continue) is

$$\begin{cases} y > -\theta \\ y < 2\theta + x\sqrt{3} \\ y < 2\theta - x\sqrt{3} \end{cases}$$

With the response boundaries established, we have the complete picture of the dynamics for the 3-alternative model. It should start at the origin unless some prior information is given

⁴ A discrete-time model can also be used as desired for situations where the number of steps rather than continuous RT distributions are useful. For example, predicting how many cards a person will draw from different decks or how many samples they will gather before choosing can make use of discrete-time Markov chains.

about the likelihood of the alternatives (in which case it should move to reflect the new prior probabilities of the alternatives). When the stimulus is presented, the state moves in direction v_A with likelihood p given by the proportion of samples that favor A, direction v_B with likelihood q given by the proportion of samples that favor B, and direction v_C with likelihood $1 - p - q$ given by the proportion of samples that favor C. A decision is made whenever the component of s along one of these vectors exceeds θ . The response time is given by the number of steps it took to cross the threshold, either multiplied by a fixed time value so that $RT = \Delta \cdot \lambda$ or pulled from an exponential distribution so that $RT \sim \text{Gamma}(\Delta, \lambda)$, where λ is the expected amount of time for each step and Δ is the number of steps it took.

Note that the main elements that had to change from the binary choice case are the positioning of the response alternatives (120 degrees from one another vs 180 degrees from one another) and the number of step probabilities (p, q , and $1 - p - q$ vs p and $1 - p$). These same properties will change when we move to the case of n alternatives.

2.1.2. N alternatives/simplex

The general ideas from the 3-alternative case allow us to scale the optimal model up to any arbitrarily large number of alternatives n . In the same way that the diffusion model (Ratcliff, 1978) implements the sequential probability ratio test (Wald & Wolfowitz, 1949), a multidimensional diffusion process on a triangle – or in the case below, a simplex – will implement the multihypothesis sequential probability ratio test [MSPRT] (Baum & Veeravalli, 1994; Bogacz, 2007, 2009) as the step size gets very small. The computations necessary for carrying out this type of inference process are thought to be carried out in the basal ganglia and cortex (Bogacz & Gurney, 2007), providing a link between the geometric model presented here and the neurobiology of these types of multialternative decisions (see also Niwa & Ditterich, 2008, which models neural accumulators using a triangular boundary approach)

So how can the optimal strategy be represented in a decision space for more than three alternatives, as in the triangular representation shown above? In order to maintain a sum of zero across evidence for all alternatives (total probability of one), information for alternative A must necessarily be evidence against the other $n - 1$ alternatives, such that for every alternative $i \neq j$,

$$\cos(\phi_{ij}) = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|} = -\frac{1}{n-1}. \tag{3}$$

Evidence for any individual alternative provides evidence against all others alternatives equally. The angle between any pair of vectors describing alternatives must therefore be $\phi_{ij} = \cos^{-1}(-\frac{1}{n-1})$. As before, a decision will be made when the criterion θ is reached along a direction corresponding to one of the alternatives, where θ is determined by the desired probability of making a correct answer.

Applying this decision rule to the case of four alternatives, the response boundaries corresponding to alternatives A, B, C, and D would each consist of a plane in a 3-dimensional space. Together they would form a tetrahedron (Fig. 1C), and evidence accumulation would unfold inside the figure bounded by these planes by sampling A (stepping in direction v_A) with probability p , B with probability q , C with probability r , and D with probability $1 - p - q - r$.

In order to accommodate n alternatives, this would naturally be extended to permit evidence accumulation in $(n - 1)$ dimensions. The evidence state would exist in the interior of a simplex (the general version of a triangle or tetrahedron in 4+ dimensions) formed by the choice boundaries. Each choice boundary would compose an $(n - 2)$ -dimensional facet of the simplex.

It will be more mathematically convenient to specify a standard simplex as the response boundaries than specifying each of the n -planes individually. A standard simplex uses n dimensions for n alternatives to specify an $(n - 1)$ -simplex in which evidence accumulation can unfold.

Taking this approach, the space of evidence accumulation will be the interior of the n -simplex denoted as Δ_n . The vertices of Δ_n have the following coordinate expressions:

$$[1, 0, 0, \dots, 0]$$

$$[0, 1, 0, \dots, 0]$$

...

$$[0, 0, \dots, 1]$$

The center of Δ_n is $C_{\Delta_n} = [1/n, 1/n, \dots, 1/n]$. The probability of an alternative k is given by the k th coordinate of the state. However, because all of the probabilities must sum to one, the entire n -dimensional space is not the space of evidence accumulation because the space is constrained by one dimension.

The facets of the simplex are formed by taking a subset $n - 1$ of these vertices, whose combination forms an $(n - 1)$ -simplex Δ_{n-1} . Each unique facet corresponds to a choice alternative – when the random walk exceeds the value θ for one of the alternatives, it crosses through a facet and a decision in favor of that alternative is triggered. Each subset of vertices gives an equation describing its composite facet: supposing each dimension is denoted x_i in $i = 1, \dots, n - 1$ (rather than x, y, z , etc.), the equation for the i th facet in n dimensions is given by

$$a_1x_1 + a_2x_2 + \dots + a_nx_n - a_ix_i = \theta \tag{4}$$

We can generate a system of equations by putting together all of the n facets and generating inequalities that point toward the interior of the simplex. This system of equation constitutes a list of checks that can be run at each step of the evidence accumulation process to examine whether a decision should be made (states violates an inequality) or continue (no inequalities are violated). The system consists of n inequalities:

$$\forall j, j = 1, \dots, n \left\{ -x_j + \sum_{i=1}^n x_i < \theta \right. \tag{5}$$

Essentially, one adds up the values along every dimension except for the j th dimension. So as long as the sum of the coordinates in every $(n - 1)$ -subset of the dimensions composing the simplex are less than θ , the process remains on the interior of the simplex. When one of these sums becomes larger than θ , a decision is triggered in favor of alternative j , where x_j is the omitted term from the equation for its facet.

2.2. Selecting or rejecting alternatives

In some decision situations, a decision-maker may not be looking to select from a set of alternatives, but rather make a binary choice on whether to select or reject each individual item in a set. This may apply to cases where they are interested in ‘pruning’ the set for bad alternatives, or assigning ranks to the alternatives. In these cases, two modifications are necessary. First, there must be two boundaries for each item: a ‘positive’ selection boundary v_{A+} and a ‘negative’ rejection boundary $v_{A-} = -v_{A+}$. Evidence favoring option A will move the state in direction v_{A+} , and evidence against option A will move it in the opposing direction v_{A-} . Once sufficient evidence in favor or against an option is gathered, that option will be chosen as the best in the set or rejected as the worst in the set and removed, respectively. In some cases decision makers may wish to have asymmetric rejection and acceptance thresholds (perhaps easily ruling items

out but being more careful when deciding which one is their favorite), θ_+ and θ_- respectively.

Second, the relative orientations of the items will need to change. We can think of each alternative in the set as its own miniature binary choice scenario taking place along a single dimension. Evidence for or against A will not necessarily be evidence for or against B, as they could both be selected or both rejected; they are not mutually exclusive in these types of choice tasks. As before, the selection or rejection thresholds will create a hyperplane orthogonal to the corresponding vector. An alternative is accepted if its coordinate along the corresponding dimension exceeds θ_+ , and rejected if its coordinate along the corresponding dimension dips below $-\theta_-$.

Accept i if: $x_i > \theta_+$

Reject i if: $x_i < -\theta_-$ (6)

This will create a geometric figure like the ones shown in Fig. 1D (for situations with two alternatives) or 1E (for situations with three alternatives). If all alternatives have the same positive and negative thresholds, the figure formed by their boundaries will be a hypercube, with n dimensions for n alternatives.

Negative choice boundaries provide a mechanism for modeling discrete choice tasks like ranking and best-worst decisions (Finn & Louviere, 1992; Hawkins, Marley, Heathcote, Flynn, Louviere, & Brown, 2014b; Marley & Louviere, 2005). When a positive boundary θ_{A+} is hit, alternative A is ranked as the best of the remaining set (so rank #1 if it is the first, #2 if it is the second boundary hit, and so on). When a negative boundary θ_{A-} in direction is hit, alternative A is ranked as the worst of the remaining set (rank # n if it is the first hit, # $n-1$ if it is the second, and so on). Evidence suggests that these two processes can happen coincidentally and seem to have similar properties (Hawkins, Marley, Heathcote, Flynn, Louviere, & Brown, 2014a), offering support for a geometric approach treating them as opposing ends of a single evidence accumulation space. Negative alternatives therefore provide a method for dynamically modeling these types of ranking paradigms, and indeed the best-worst linear ballistic accumulator model of Hawkins et al. (2014b) is possible to reconstruct (and potentially improve upon) using the geometric framework.

This completes the general solution to the question of how a decision maker should choose among n alternatives (if they wish to minimize response time for a desired level of accuracy) when the alternatives are mutually exclusive and each piece of incoming evidence favors exactly one of the possible alternatives, or when ranking or rejecting alternatives. When a single option is correct, setting the choice boundary θ guarantees that the probability of selecting the correct answer, no matter the number of options, is at least $\theta + 1/n$ when a choice is made. Unfortunately, not all decision situations are this simple; often, incoming evidence may favor a subset of the alternatives while disfavoring another subset. This creates violations of the assumptions of the normative model covered thus far. In the next section, we examine some of the ways these violations can occur and how a decision-maker can maintain the same level of accuracy when they do.

3. Similarity and the representation of alternatives

One of the key assumptions allowing us to construct the simplex model presented in the previous section is that each piece of evidence favors exactly one alternative in the choice set, but this assumption is frequently violated in real decisions. For example, suppose a decision-maker is determining which of a set of directions a dot stimulus is moving (as in the popular random dot motion/kinetogram task Ball & Sekuler, 1987; Julesz,

1971). Suppose motion direction is drawn randomly from a mean direction of 90 degrees (north), 70 degrees (north-northeast), or 270 degrees (south), with some noise around the mean in each case. The decision-maker's task is to determine which of these three mean directions is giving rise to the dot stimulus shown on a screen in front of them. Apparent motion toward 80 degrees might provide evidence for responses at both 70 and 90 degrees, but evidence against motion at 270 degrees. Even motion perfectly favoring one option (70 degrees) may provide some evidence for a second (90 degree motion) while disfavoring a third (270 degrees).

These situations have natural connections to the neural underpinnings of decision behavior as well. The activation functions for different alternatives in a choice set, such as 70- and 90-degree motion directions or orientations, may have overlapping response regions. In this case, both the set of neurons tracking 90-degree motion and the set tracking 70-degree motion might be activated in response to a stimulus with 80-degree motion. This makes similarity an unavoidable consideration in multi-alternative choice: due to the nature of neural representations in the brain, incoming evidence cannot be isolated to favor exactly one of the options in an available choice set while providing evidence against all others evenly.

In the geometric framework, this problem is solved quite naturally by placing the vectors describing choice alternatives at uneven orientations in the decision space, then adjusting thresholds to compensate for the interactions between them. Before delving into the details of how to do this, it is helpful to examine how such problems have been dealt with in other multialternative models.

3.1. Relation to previous approaches

In existing decision models, the support for each alternative in a set is typically represented as a distinct value called *accumulators* (LaBerge, 1962; Smith & Vickers, 1988; Vickers, 1979), and the similarity relations between different options are built into interactions between the accumulators for different alternatives. For example, decision field theory (Busemeyer & Diederich, 2002; Roe et al., 2001) includes an additional step in the decision process where pairs of items are contrasted against one another before computing accumulator values for each one. The leaky competing accumulator model (Usher & McClelland, 2001, 2004) includes inter-accumulator competition and loss aversion to a similar effect, including an explicit moment-to-moment interaction between accumulator values in the form of lateral inhibition. The multi-attribute linear ballistic accumulator model (Trueblood, Brown, & Heathcote, 2014) includes pairwise comparisons as well as subjective attribute values that result in interactions between alternatives in the set, and the selective attention, mapping, and ballistic accumulation model (Brown, Marley, Donkin, & Heathcote, 2008) specifies and utilizes adjacency between categories to define the evidence accumulation process for separate, correlated accumulators.

Models of confidence utilize similar approaches to handle the ordinal relations between different confidence levels — ordinal states or accumulators are fixed according to some adjacency to one another and often to some underlying distribution of stimulus strength (as in signal detection Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009). The RTCON2 model developed by Ratcliff and Starns (2013) explicitly normalizes the sum of evidence across accumulators, in effect introducing an inhibitory relationship between the values of different accumulators in the set as well.

Using vectors to represent alternatives and then orienting them in a “decision space” can yield the same predictions as correlated accumulators with fewer parameters, while maintaining

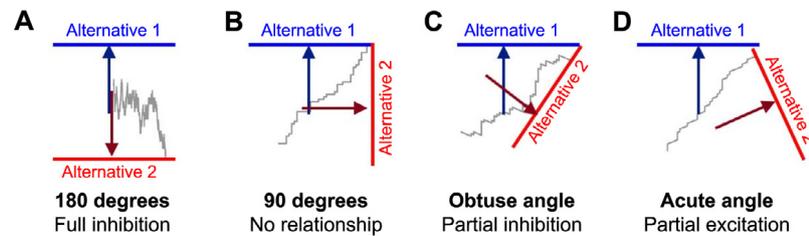


Fig. 3. Diagram of different orientations of response alternatives and the evidence accumulation processes they set involve.

the same connections to neural structures. Alternatives that are diametrically opposed and provide evidence against one another should be represented by vectors pointing in opposite directions, denoting inhibition, and ones that are extremely similar should be represented by vectors pointing in nearly the same direction, denoting coactivation between them. The cosine similarity relationship described in Eq. (2) is ideal for quantifying these relationships. The cosine function has been used to describe similarity in a multitude of vector space models (Bhatia, 2017; Harris, 1954; Nosofsky, 1997; Turk & Pentland, 1991) including those used in latent semantic analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Furnas, Deerwester, Dumais, Landauer, Harshman, Streeter, & Lochbaum, 1988; Landauer & Dumais, 1997) and quantum cognition (Busemeyer & Bruza, 2012; Pothos, Busemeyer, & Trueblood, 2013).

This approach is shown in Fig. 3: full inhibition between alternatives resulting in zero-sum evidence is equivalent to the alternatives being oriented at 180 degrees relative to one another (Fig. 3A), which creates the representation of evidence observed in random walk and diffusion models (Link & Heath, 1975; Ratcliff, 1978; Stone, 1960).⁵ A representation of alternatives with evidence representations that do not interact, instead accumulating independently, is equivalent to a 90 degree orientation between alternatives (Fig. 3B), which is adopted by uncorrelated accumulator models of choice (Brown & Heathcote, 2005, 2008; Smith & Van Zandt, 2000). Partial inhibition or negative correlation between accumulators is generated by obtuse angles between the vectors for alternatives (Fig. 3C), and positive correlations between accumulators come from angles between alternatives that are less than 90 degrees (acute angles / Fig. 3D). These same relations can be obtained by deliberately building in correlations between options.

Note that these similarity relations can essentially be factored out when evidence accumulates in binary choice. An evidence state linearly related to the log odds of the two alternatives (assuming an inferential decision) can be computed by simply taking the difference between evidence for option A and option B. This creates an equivalence class of optimal models, as any degree of inhibition or excitation between them can be compensated for by simply scaling the difference in evidence between alternatives by a constant to compute the true posterior probability of A:B (Bogacz et al., 2006). In the next section, we will use a similar approach to control for the similarity relations among three or more alternatives with unequal similarity relations. Rather than a simple difference, we can subtract the support across all alternatives from the support for each individual alternative to compute the probabilities of each possible option and use this to set stopping and decision rules for each alternative.

⁵ Note that this refers to the relative orientation of the directions of the alternatives (the arrows in Fig. 3), not the choice boundaries.

3.2. Three or more alternatives

The first step in constructing a geometric model is to understand the relationship between alternatives so that similarity can be translated into geometric representations like those in Fig. 3. For simplicity, we can begin by directly using physical relationships between alternatives. Applying it to the previous example, suppose a decision-maker is determining 2-dimensional motion direction on a 360 degree circle. If the physical relations are preserved, 90-degree motion would be represented by the vector $v_{90} = [0, 1]$, 270-degree motion by the vector $v_{270} = [0, -1]$, and 70-degree motion as $v_{70} = [.34, .94]$. Using the cosine metric to quantify similarity on a $[-1, +1]$ scale, we find that $\phi_{70,90} = .94$ (similarity between 70 and 90), $\phi_{70,270} = -.94$ (similarity between 70 and 270), and $\phi_{90,270} = -1$ (similarity between 90 and 270).

The cosine metric is also used to quantify the amount of evidence allotted to each alternative with new information. New information ι is represented by vector v_ι . The direction of this vector describes which alternative it most favors (including those not in the set) and its length $\|v_\iota\|$ describes the amount of information it confers. In turn, the amount of support ι provides for option A is the scalar projection of v_ι onto v_A :

$$\text{proj}_A(\iota) = \|\iota\| \cos(\phi_{\iota A}) = \frac{v_\iota \cdot v_A}{\|v_A\|} \quad (7)$$

The evidence that a particular piece of information provides for each available alternative is therefore a function of the congruence of the alternative with that information. In turn, the distribution of support across alternatives will be determined by the congruence of each alternative with the new information. Naturally, this will result in correlations between support for different alternatives, unless they are all orthogonal as in Fig. 3B. The relative orientations of the alternative vectors embody our theory of how the alternatives are represented relative to one another, providing complete information about how evidence simultaneously affects support for each of the available choice options.

As evidence accumulates and the evidence state moves around, the evidence for each alternative shifts accordingly. As before, the evidence state s provides support for alternative j according to the component of the state along the vector describing alternative j , v_j :

$$\text{Support for } j: \text{comp}_{v_j}(s). \quad (8)$$

The rules for halting the evidence accumulation process will again take advantage of this geometric description of the support for each alternative. However, at this point there are diverging approaches that can be taken. The simpler *absolute* stopping rules halt evidence accumulation on the basis of the support for each alternative independently, and are frequently adopted by accumulator models of choice. However, these stopping rules are not optimal in the sense of minimizing response time for a desired level of accuracy. This kind of optimality requires *relative* stopping rules, which consider the degree of support for one alternative

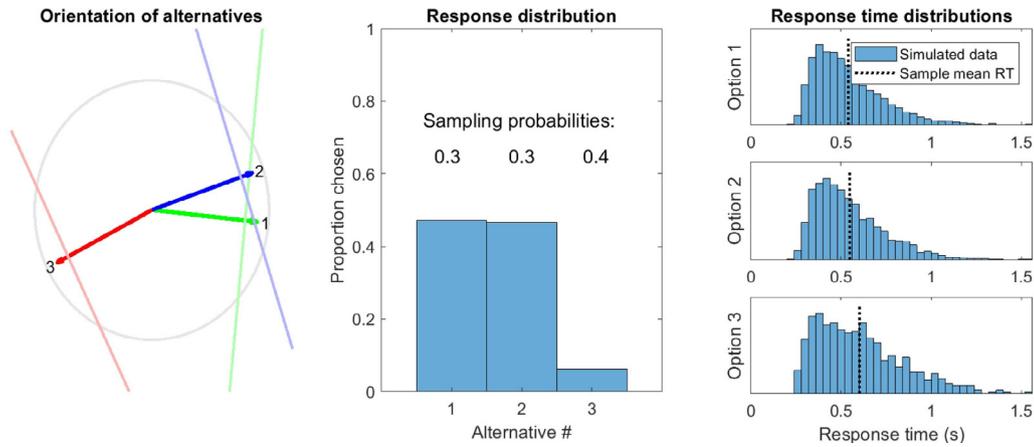


Fig. 4. Representation of three alternatives and their decision rules (colored vectors/lines) on top of the unit circle (gray) for an absolute choice rule where thresholds are the same for each alternative (θ_{abs}) (left). The middle panel shows the choice proportions for each alternative (probability estimate of choosing each option), and the right panels show the corresponding RT distribution for each alternative. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

over and above the support for all of the other alternatives in a choice set. Because optimality requires us to consider the probability of a target alternative being correct relative to all others in the set, it demands consideration of how support for the target alternative is related to support for the others.

3.2.1. Absolute stopping rule

Although it is not optimal in the sense of minimizing response time for a given accuracy, a simple stopping rule can be applied by comparing the evidence for each alternative j , given as $comp_{v_j}(s)$, to a criterion θ_{abs} . This is referred to as an *absolute stopping rule*: the decision process is terminated when evidence for any of the available choice options exceeds θ_{abs} . Such a rule is typical of accumulator models (Brown & Heathcote, 2008; Smith & Van Zandt, 2000; Smith & Vickers, 1988; Usher & McClelland, 2001), and is particularly convenient because it corresponds to a set of processes racing to a single finish line, such as individual neurons or neural populations firing until one reaches a movement initiation threshold (Schurger, Sitt, & Dehaene, 2012; Zandbelt, Purcell, Palmeri, Logan, & Schall, 2014).⁶

Geometrically, the boundaries for the choice rules are similar to the ones we saw in the simplex case. A hyperplane H_j is placed orthogonal to each vector v_j (such that $H_j \cdot v_j = 0$) at distance θ_{abs} from the origin. It will intersect v_j at point $(\theta_{abs} \cdot v_j)$, so the hyperplane that solves these conditions can be specified as a function of the coordinates (using a Cartesian system of coordinates in m dimensions described as x_1, x_2, \dots, x_m) of the alternative vector:

$$H_j : v_j \cdot [x_1, x_2, \dots, x_m] = \theta_{abs} \quad (9)$$

For example, an alternative A represented by $v_A = [\sqrt{.4}, \sqrt{.6}]$, $\theta_{abs} = 2$ would have response rule $H : x\sqrt{.4} + y\sqrt{.6} = 2$. When the line specified by this equation is crossed, i.e. $x\sqrt{.4} + y\sqrt{.6} \geq 2$ alternative A is chosen. More generally, response alternative j is chosen when the evidence state s meets the following condition:

$$\sum_{i=1}^n j_i s_i \geq \theta_{abs} \quad (10)$$

⁶ The orientation between alternatives will determine the correlation between separate accumulators. For those models that assume independence between evidence for different options where evidence for A does not affect evidence for B, this angle must be $\cos^{-1}(\phi_{AB}) = 0$, or $\phi_{AB} = 90$ degrees.

where j_i is the i th coordinate of v_j and s_i is the coordinates of the evidence state. Put together, there will be exactly m inequalities for m different response alternatives.

While this type of stopping rule is simple and easy to test, it is not strictly optimal as it does not result in the evidence accumulation process halting at a particular balance of odds between the alternatives. While it shares many properties with the relative or optimal stopping rules I describe next, the predictions do not line up exactly (see Figs. 4 and 5).

3.2.2. Relative/optimal stopping rule

In past work, evidence – particularly evidence representing the posterior odds of hypotheses – is commonly represented as a balance between support for two or more alternatives rather than separate accumulators. This is a central feature of diffusion models (Ratcliff, 1978; Ratcliff et al., 2016; Smith & Ratcliff, 2004) and other models based on diffusion processes or the sequential probability ratio test (Edwards, 1965; Link & Heath, 1975). This can be done easily when similarity relations are all equal, as with the simple optimal simplex models described in Section 2.1. However, a critical feature of alternatives with similarity relations is that the sum of the shifts in evidence across all alternatives j will often not automatically sum to zero. Fortunately, this does not mean that an optimal strategy minimizing response time for a desired level of accuracy cannot be constructed. The decision-maker simply has to control for how the probabilities across alternatives shift in the face of these similarity relations.

In multi-alternative choice, a balance of evidence is computed by taking the evidence for each alternative and normalizing the sum to zero by subtracting the mean shift in evidence across all alternatives from the support for each option (as in Ratcliff, 2018; Ratcliff & Starns, 2013). Computing the probability of each alternative in the geometric framework proposed here requires a similar approach: we have to factor out the support for the opposing options and compare it to the support for each individual option.

This modification means that each alternative will not gain support at the same rate. If an alternative has many similar ones around it, it will tend to accumulate more slowly because the mean shift across alternatives will be higher. Returning to our example with dot motion that could be centered at 70, 90, or 270 degrees, a +1 unit step toward 90-degrees ($v_{90} = [0, 1]$) will result in a shift of +.94 for the 70-degree alternative ($v_{70} \approx [.34, .94]$) and a -1 shift for the 270-degree alternative ($v_{270} =$

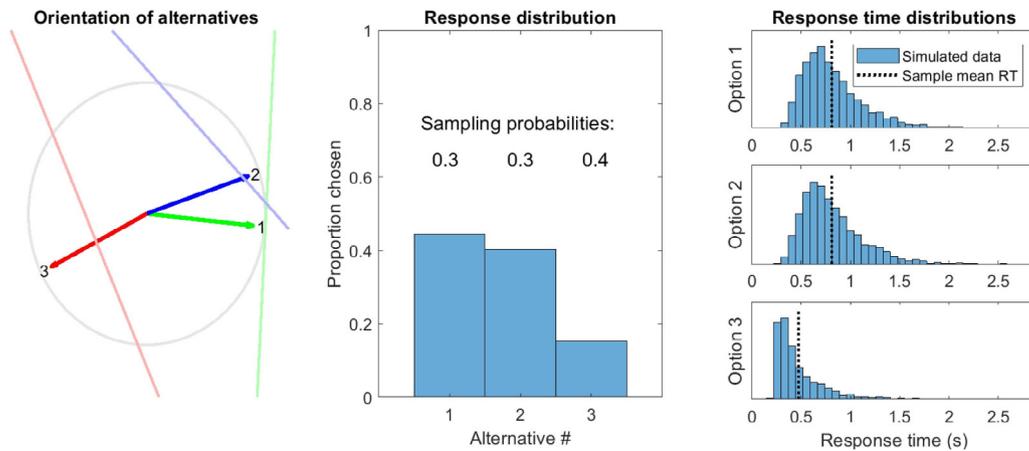


Fig. 5. Representation of three alternatives (different colored vectors) and their decision rules (colored lines) on top of the unit circle (gray) for a relative choice rule where thresholds (θ_{rel}) are adjusted to compensate for similarity relations among them (left). The middle panel shows the simulated choice proportions for each alternative, and the right panel shows the corresponding RT distribution for each alternative. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$[0, -1]$), for a net evidence shift of $(+1 + (-1) + 0.94 =) +.94$. Subtracting this from each of the alternatives equally results in a net increase in probability proportional to $+.69$ in favor of the 90-degree motion alternative. Conversely, suppose that a $+1$ unit step is taken toward the 270-degree alternative: this would result in -1 for the 90-degree alternative and $-.94$ for the 70-degree alternative, an average shift of $-.31$. This *increases* the probability shift in favor of the 270-degree alternative, resulting in a net $+1.31$ shift in favor of it rather than a single $+1$ unit shift.

The true support for each alternative – in terms of posterior probabilities – must be calculated to compensate for these net shifts. Because the probability shift differs depending on the similarity relations among alternatives, each option differs in the amount of support it will require to obtain the same odds of being correct if selected. Consequently, different thresholds for each alternative are required to ensure that the probabilities are matched across the decision rules for each alternative. For example, if the decision-maker desires an 80% chance of being correct, the thresholds for selecting 90-degree or 70-degree motion will have to be higher than the ones for selecting 270-degree motion.

To calculate a stopping rule which compensates for the similarity between alternatives, we first need to compute the conditions under which the probability for one alternative minus all other alternatives exceed the desired threshold. If the desired accuracy corresponds to a probability level specified by θ_{rel} , then the stopping condition for choosing option j (specified by vector v_j in the space) out of n alternatives given state s is given by:

$$comp_{v_j}(s) - \sum_{k=1}^n v_k \cdot s \geq \theta, k \neq j. \quad (11)$$

This stopping rule gives the optimal choice boundaries for any set of interrelated alternatives. As a natural consequence of Eq. (11), having a large number of similar alternatives in a choice set will force the alternatives to “share” evidence, diluting the gain in information provided by each new piece of evidence. In turn, the greater the number of similar choice alternatives a decision-maker has, the higher decision thresholds will need to be to make selections with a constant level of accuracy. As Usher, Olami, and McClelland (2002) note with regard to their model, adjusting the thresholds based on the alternatives present may provide an explanation for Hick’s law in multi-alternative choice: a greater number of alternatives will “crowd” the decision space, requiring a decision-maker to increase thresholds to maintain accuracy. In turn, the higher thresholds will result in slower response times as more alternatives are introduced.

3.3. MATLAB code for relative and absolute models

Given the two potential ways of implementing stopping rules for multi-alternative choice, a reasonable concern is what empirical properties each approach generates. To assist with exploring and visualizing the effects of different sets of choice options, MATLAB code that can generate the predictions of each kind of model accompanies this paper on the Open Science Framework at <https://osf.io/4ygvvs/>.

This code allows the user to specify a set of alternatives either by randomly generating n options in m dimensions or by putting in custom vectors for each alternative and a probability of sampling evidence in favor of each one at each time step. The user can also specify the desired threshold(s) and whether they prefer to have an absolute decision rule (θ_{abs}) or a relative one (θ_{rel}). The program will then generate a set of choice proportions covering each alternative in the set as well as a response time distribution for each one. Note that these predictions are generated via simulations (the number of trials is adjustable), as they do not have apparently straightforward analytic likelihoods.

For alternatives arranged in $m = 3$ dimensions, the program will visualize these vectors in three dimensions on a plot that can be pulled and rotated around. For those arranged in $m = 2$ dimensions, it will show not only the vectors for each alternative, but also a line illustrating the response boundary corresponding to each alternative. The user is free to also specify non-decision time, step size, step times (leading to different RT distributions), and the number of simulations to run to generate predictions, among other control variables. Although the program will generate response and response time simulated predictions for higher-dimensional decision spaces (of more than three dimensions), it will not illustrate them in the left panel.

Some examples of the plots generated by the program, using a two-dimensional arrangement with four response alternatives, are shown here. Fig. 4 illustrates an absolute decision rule, where each alternative has the same threshold θ_{abs} (Eq. (10)). As it shows in the left panel, the thresholds for the alternatives (labeled 1, 2, and 3) are orthogonal to the vector specifying their direction. In this case, all of the alternatives have the same right-skewed distributions of response times (similar to the symmetric error and correct response times predicted by random walk models; Stone, 1960). But because there are multiple alternatives in the same direction (1 and 2), these tend to be selected more often even though alternative 3 is the most frequently sampled.

Fig. 5 illustrates a relative decision rule, where an alternative is chosen only when the support for it is sufficiently greater than that of all other alternatives (Eq. (11)). Compared to the absolute stopping rule, there are two key differences in representation that can be seen in the left panel of the figure. First, the thresholds for different alternatives vary in height: thresholds for alternatives with no other neighboring similar alternatives (alternative 3) are substantially lower than those for alternatives with nearby/similar competitors (alternatives 1 and 2). Consequently, the lower threshold for alternative 3 results in it being picked more often (middle panel). It also creates a substantially faster distribution of response times for this alternative while maintaining the same level of accuracy (bottom-right panel).

A second thing to note is that the thresholds in the relative decision rule model are not quite orthogonal to the vectors specifying them, instead leaning slightly away from nearby alternatives to create a more tightly closed polygon. This allows for more complete coverage of the decision space, and as a result it generates faster response times for all of the alternatives while maintaining the same desired level of accuracy (given by θ_{rel}). Thus, while the absolute decision rule is generally simpler and likely easier to implement, there is a significant benefit in terms of accuracy and response times to using a relative decision rule.

This approach and the MATLAB code can be used to predict choices among any discrete set of alternatives with known similarity relations; the relative decision rule describes the stopping conditions necessary to optimize response times for a desired level of accuracy. However, the sampling distributions and response distributions I have covered so far are mainly appropriate for small choice sets where a probability of sampling can be assigned individually to each alternative. In the next section, I extend the framework to continuous response sets and continuous sampling distributions, showing that the geometric framework can be flexibly applied to modeling both responses among both discrete and continuous sets of alternatives.

4. Continuous distributions of evidence and responses

A major benefit of the geometric approach is that it is convenient to move from small to large choice sets and even to a continuum of response options. This allows a modeler to scale up the approach to model behavior in situations where a decision maker has to specify a spatial location or direction, a color (e.g. hue), a price (\$), the magnitude of a stimulus, or any sort of continuous or near-continuous property like length or time. At the same time, it can still be used when there are a limited set of options to select. Representing alternatives as vectors permits a modeler to place more or fewer alternatives in the same multidimensional space, and specify their relations to one another by virtue of the angles between the vectors. Rather than having to set all of the pairwise correlations between accumulators for separate alternatives, a single evidence representation can be stored and related to each alternative by measuring its component along the vector for each alternative. This makes it more computationally tractable for large choice sets than decision strategies explicitly considering pairwise comparisons (Roe et al., 2001; Trueblood et al., 2014; Usher & McClelland, 2001), as all of the inter-alternative relationships are contained in their representation (which can often be set a priori) rather than added onto the sampling process itself.

The analytic likelihoods for most of the continuous models covered here are presented in work by Smith (2016), Smith and Corbett (2018). However, it is useful to build up to the continuous models to see how they are related to the discrete sets of alternatives. Both the representation of alternatives and the evidence dynamics of the models of Smith and Corbett follow as

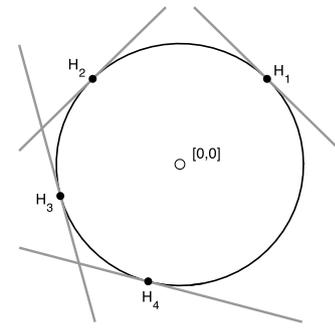


Fig. 6. The limit as the number of individual boundaries (H_{1-4}) equidistant from the origin approaches infinity creates a circular choice boundary in 2 dimensions.

asymptotic cases of the multialternative geometric models as the number of alternatives in a given space gets very large, resulting in a strong connection between the present geometric framework and their work, which I discuss later in this section. However, before discussing this work it is helpful to build some theoretical basis and intuition for how to generate a continuous span of responses in the first place.

4.1. Continuous spans of alternatives

In two dimensions, the vectors corresponding to response alternatives can point in any direction from 0 to 360 degrees. Naturally, the limiting case as the number of alternatives gets very large will be a continuous span of response alternatives along this entire range. If we assume that the same threshold height is used for all alternatives – as with the absolute stopping rules I described in the previous section – then the boundaries for all possible alternatives in this space create a circle, as shown in Fig. 6. This type of stopping rule is described by Smith (2016): the radius of the circle bounding evidence accumulation will be specified by the single parameter θ_{abs} , giving the total amount of evidence needed to stop and make a decision.

The lone inequality $\sum_{i=1}^n s_i^2 < \theta$, where s_i is the coordinate of state s along dimension i , is sufficient to divide states into decision and non-decision regions. The stopping rule therefore refers directly to the coordinates of the space in which evidence accumulation unfolds, rather than needing to make reference to the component of the state along individual vectors.

Higher-dimensional stopping rules can also be constructed. Formally, evidence accumulation unfolds with alternatives arranged as vectors specified on an n -sphere, where n is the number of dimensions needed. The n -sphere O^n with radius 1 is a figure composed of a set of points p_i that satisfy the condition

$$O^n = \{p \in \mathbb{R}^{n+1} : \|p_i\| = 1\}$$

Each of these points can correspond to a unique alternative vector, yielding an infinite number of potential choice options in a finite number of dimensions n . This is how the discrete representations of alternatives can be taken to its limit to represent a continuous span of alternatives.

A particularly desirable property of the n -sphere representation of alternatives is that the sum of the cosine similarity relation defined by Eq. (2) is zero-sum across all alternatives that could be represented in the n -dimensional space by virtue of every integral $\int \dots \int_0^{2\pi} \cos(\phi) d\phi = 0$. As with the optimal models, this connects the support among alternatives to a probability space representation, as the net probabilities (probability minus $1/\#$ of options) across all options must also sum to zero. As before, the probability of alternative A relative to alternative B given

an evidence state is computable as the relative values of the component of the state s along the vectors corresponding to their directions in the multidimensional space (v_A and v_B). The net probability of a span of alternatives will be given as the integral sum of the dot product between the state and each of the internal vectors between v_A and v_B :

$$\text{Net probability of } v_A \leq v_{correct} \leq v_B | s = \int_A^B s \cdot v_x dx \quad (12)$$

Sections of the space can therefore be compared by computing and comparing the integrals across different ranges. For example, a decision maker could use their evidence state to estimate the probability of a dot motion stimulus being oriented between 30 and 50 degrees versus between 110 and 120 degrees by evaluating the integral from Eq. (12) over these ranges. The upper and lower limits of the integral would naturally compensate for the larger range of the former, allowing the decision maker to come up with a posterior probability of 30–50 versus 110–120 that accounted for the base rate of each one occurring.

This gives us a basis for how to evaluate the likelihood of individual hypotheses in a continuous span, or subsets of the continuous span of alternatives, given a particular evidence state. It also provides an optimal mechanism for generating probability or confidence judgments, as the position of the state in the decision space gives the posterior probabilities of any hypothesis or span of hypotheses that can be constructed in the space. Next, it is important to describe how this state changes over time with new information.

4.2. Continuous states and dynamics

When there is only a discrete set of alternatives, describing how the state changes over time can be dealt with by assigning probabilities to each of the alternatives to specify the likelihood of stepping in the different corresponding directions. With a large number of alternatives, specifying all of the probabilities is cumbersome. Instead, to specify the dynamics of the state in a multidimensional space, one can simply specify a probability distribution over a continuous span of step directions. In this case, at each time point, a decision-maker takes a new piece of information drawn from some distribution of evidence based on the features or alternatives they are considering. This piece of information moves them distance $|\rho|$ (step size) in direction ϕ , a random variable specified by the sampling distribution. The direction and magnitude of this step changes the evidence with respect to the available alternatives. As before, the evidence it provides for alternative m , $\Delta s(m)$ is the angle between the step direction v_ϕ and the alternative vector v_m , multiplied by the magnitude of the step $|\rho|$.

$$(\Delta s(m)|\rho, \phi) = |\rho| v_m \cdot v_\phi \quad (13)$$

The angle of each step ϕ is the key random variable here. If there is a discrete number of step directions being sampled, then the distribution of samples over directions at each time step is specified by a categorical random variable X with parameters $p_1, p_2, \dots, p_n : \sum_{i=1}^n p_i = 1$ describing the probability of stepping in direction d_1, d_2, \dots, d_n at each transition, respectively. This may not always be practical, however; discrete-direction sampling process will require n parameters to describe a categorical distribution over step directions that unfolds in continuous time. This includes $n - 1$ probabilities (since the last probability p_n will necessarily be $1 - \sum_{i=1}^{n-1} p_i$) as well as the sampling rate parameter λ for the time between samples.

In many cases it will be simpler and more parsimonious to specify a distribution for the random variable ϕ over the entire

span of the n -sphere. This will be more convenient when evidence is sampled with noise, so that stimulus information (about dot motion, for example) favoring an orientation of 80 degrees might actually be perceived as 83 degrees. A continuous distribution of step directions in the model is also a more faithful representation when stimulus information truly varies along a continuum.

Some examples of continuous sampling distributions for a multidimensional random walk are given in Appendix B. The particular sampling distribution and dynamics used in a model will depend heavily on the stimulus displayed and the modeler's particular theory about how decision-makers sample information from the stimulus. For example, evidence accumulating in parallel or being sampled from multiple sources may require a bimodal distribution or multiple evidence states/accumulators (Kvam, 2019; Ratcliff, 2018). In other cases, deterministic dynamics may be more tractable and allow us to derive simple analytic solutions to the distributions of responses and response times as in the linear ballistic accumulator models (Brown & Heathcote, 2008). The most common form of state dynamics used in evidence accumulation are those of the continuous Brownian motion with drift – a Wiener diffusion process in a multidimensional space – covered next.

4.3. Diffusion on a hypersphere

One theoretically-motivated way to quantify how evidence shifts over time is to divide the sampling distribution into components describing signal (true stimulus information) and noise (components describing random stimulus or perceptual variation). Each step in the accumulation process sees the evidence state move a certain distance toward the 'best' response, with some (multivariate normal) distribution around the expected location. In this case, the dynamics can be quantified in terms of an average drift *direction* describing which alternative the average stimulus information favors, a drift *magnitude* describing how quickly 'signal' information favoring that alternative is gathered, and a *diffusion rate* describing the rate at which random noise is sampled. The accumulation trajectories resulting from these dynamics are shown in Fig. 7.

As the distance a person's state moves with each new piece of information (step size) gets very small and samples arrive more often ($\delta s \rightarrow 0$ and $t \rightarrow 0$), the random walk process described above approaches a continuous multidimensional diffusion process. Such diffusion processes are well covered in several references (see Itô, 1974; Stroock & Varadhan, 2007) and are useful for predicting the position of the multidimensional process in continuous time. These particular assumptions result in a multivariate normal distribution of evidence for any time $t > 0$ before the process crosses one of the response boundaries.

These evidence dynamics in combination with the n -sphere choice boundaries are a special case of the geometric framework that has been pursued in depth by Smith (2016) using two-dimensional circular boundaries and Smith and Corbett (2018) using four-dimensional hyperspherical choice boundaries. Smith and Corbett derive analytic likelihoods for the joint distribution of responses and response times for diffusion processes unfolding on the interior of a figure bounded by a circle or a hypersphere. The authors show that these evidence dynamics and choice rules result in a von Mises (in two dimensions) or von Mises–Fisher (in three or more dimensions) distribution of responses like those commonly observed in visual working memory experiments (Zhang & Luck, 2008), and a right-skewed distribution of response times typically found across domains of decision making tasks (Luce, 1986).

Critically, for the choice boundaries to form an n -sphere in the geometric framework, all response alternatives in the n dimensions must be available for selection, and they must all have

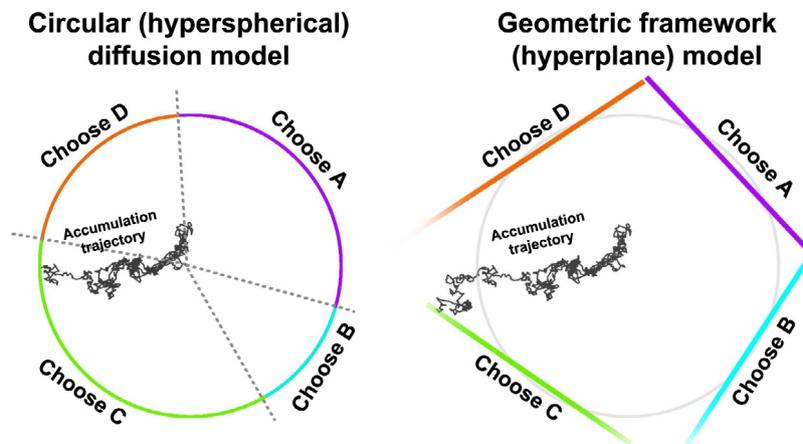


Fig. 7. Diagram of example choice boundaries and evidence accumulation (diffusion) process for the circular diffusion model (Smith, 2016, left) and the geometric framework (right) for a set of 4 alternatives.

the same threshold, θ_{abs} . This is equivalent to the absolute stopping/decision rule covered in Section 3.2.1. When the set of alternatives is not sufficient to approximate a continuum, its predictions no longer line up with those of the hyperspherical diffusion model. This becomes evident when we look at how the circular/hyperspherical and geometric framework approach discrete choice tasks: the stopping rules for selecting one alternative out of a small set are quite different.

In the hypersphere model, there are separate stopping and decision rules that trigger a decision and then sort evidence states into responses, respectively. Whether engaging in a continuous or discrete selection task, the decision maker's stopping rule is always the hypersphere (circle, in two dimensions), where evidence accumulation halts when the sum of the squared coordinates x_m exceed a specified threshold; evidence accumulates while $\sum_j x_j^2 < \theta_{abs}$. After this inequality is violated, the evidence accumulation process stops and states are mapped onto responses by dividing them into different categories as shown in Fig. 7, left panel. The decision rule or mapping stage is similar to the Thurstonian component of general recognition theory (Ashby & Townsend, 1986),⁷ where hyperplanes divide the multidimensional space of evidence states (points) into different regions corresponding to different responses.

Conversely, the geometric framework uses just one set of hyperplanes for both the stopping and decision rules. This is shown in Fig. 7, right panel. As soon as the process crosses a single boundary, the corresponding alternative is chosen, no matter if an absolute or relative stopping rule is used. Therefore, the primary difference between the two approaches to modeling selections among discrete sets of alternatives is related to the shape of the thresholds. Because the geometric framework uses hyperplanes for each choice boundary rather than a single hypersphere, the shape that the choice boundaries form in smaller sets of alternatives will often be a convex polytope rather than forming the smooth boundaries of a hypersphere.

This difference will be most notable in empirical predictions when the stimulus does not match perfectly with one of the available response alternatives, i.e., drift direction does not correspond exactly to the directions of any available alternatives. For example, the average drift of the evidence state based on incoming information could be between alternative C and D as with the trajectories shown in Fig. 7. The geometric framework

predicts that response times in these cases will be slightly longer than ones where the stimulus perfectly matches one of the alternatives in the choice set, because it will have a longer distance to travel on average before it crosses one of the response boundaries (right panel). By contrast, circular or hyperspherical boundaries imply that the distance to the response boundary from the origin is always the same, so response times should retain the same shape and mean regardless of the degree of match between the stimulus and available choice options (left panel). Since empirical work relating continuous to multi-alternative choice is sparse, there is currently no work that can speak directly to these diverging predictions. However, testing these contrasting hypotheses is certainly a promising direction for future research aimed at understanding the evidence representations underlying multi-alternative choice.

5. Attributes and preferences

Thus far, the decision situations and models considered have focused primarily on inferential decisions in perceptual choice problems. These choice scenarios are characterized by structured representations where the similarity relations between alternatives are straightforward and there is only one correct response (signal) among incorrect ones (noise). But this covers only a portion of the decisions people face in real life. When a decision maker has to choose between cars, foods, or other multi-attribute items, it will not always be immediately clear how to arrange the options in a spatial way. Furthermore, the relative weight or attention directed toward each feature of a choice alternative will differ from person to person or choice context to choice context. This occurs particularly often in preferential choice, where alternatives may be arranged differently based on a person's subjective beliefs about the similarity between items in a choice set. Individual differences in representation or evidence accumulation translate into differences in choice behavior, so they are critical to incorporate when constructing our geometric models.

The first hurdle when moving to preferential and multi-attribute choice is how people are representing the alternatives in front of them. Until now, we have taken it as a given what the relationships between alternatives are in constructing geometric models. In cases like motion direction, specifying the orientations of alternatives is fairly straightforward because there is a clear analogue between the directions of physical motion and the directions associated with choice alternatives: leftward motion can correspond to an alternative represented by a leftward pointing vector, upward motion can correspond to an alternative with

⁷ The primary difference is that the states resulting from the evidence accumulation process will lie along the edges of the figure that is used to specify the stopping conditions rather than being randomly distributed.

an upward pointing vector, and so on. But what about cases where the relations and orientations of alternatives are not so straightforward? The remaining problem is to take a given set of psychological or physical attributes or similarity relations and construct a decision space by arranging vectors of alternatives relative to one another. Like other vector space models, this will demand attention to the dimensionality of the representations and a method for deriving representations from similarity or attribute information.

5.1. Using features to construct the space

The simplest way to construct the vector space containing the alternatives is by using some objective feature space with a coordinate system where dimensions correspond to the feature values. In this case, the coordinates of the vectors describing choice alternatives directly describe their underlying features. Next, I show how to construct these representations directly from alternatives' values on each of their feature dimensions.

5.1.1. Unidimensional items

A simple place to start using features to construct the vector space representation of alternatives is with unidimensional options, which can be characterized by their value on only one feature. Naturally, two options will be more similar if they have a small difference in values on the feature, and less similar if they have diverging values of the feature. The case for binary features is quite simple; a feature value of 1 (feature present) can be represented as a positive vector $v_+ = [+1]$ and a feature value of 0 (feature absent) can be represented as an opposing negative vector $v_- = [-1]$. Alternatives defined by this single binary feature will either be perfectly the same ($v_+ \cdot v_+ = v_- \cdot v_- = 1$) or perfectly different ($v_+ \cdot v_- = -1$).

Unidimensional alternatives can also have a feature that varies along a continuum, varying from low to high on some quantitative scale. Since unidimensional alternatives with different values for a single feature all lie along a common continuum, the angles describing their similarity can also vary along a single scale. When unidimensional alternatives A and B have an angle ϕ_{AB} between them, the value of ϕ_{AB} corresponds to the difference along their single feature value. A value of $\phi_{AB} = \pi$ (180 degrees) will indicate alternatives with opposing feature values, such as northward motion versus southward motion, as in Fig. 3A, while a small value of ϕ_{AB} will indicate similar values on the feature, as in Fig. 3D.

In order for the angle between alternatives to vary along a continuum to create a scale for a feature, we must have least 2 dimensions to represent a feature. For simplicity, we can fix one end of the feature scale, such as setting the lowest value of the feature range to $v_{low} = [1, 0]$. Where the high end of the scale is located will depend on the type and range of the feature. For example, a feature like monetary value with a range of \$0–20 might vary from 0 ($[1, 0]$) to $\pi/2$ ($[0, 1]$), whereas a range of \$–100000 to +1000000 might see the ends of the scale point in opposite directions, varying from ($[1, 0]$) to π ($[-1, 0]$).⁸ The exact range of the scale will vary based on a modeler's theory about how high and low values of the feature are represented relative to one another (e.g., are they truly opposites or just

ends of a range?); it can also be estimated by including a free parameter such as $\phi_{max-min}$.

For features that vary along a range, each independent feature requires two dimensions to allow the angle ϕ to vary with the value of the feature. The upper and lower bounds of the range of angles in these two dimensions serve as high and low 'anchor' values for a feature (a concept established in absolute identification, see Marley & Cook, 1984), and the value of an alternative along that feature corresponds to its angle relative to the high and low feature value vectors. Formally, the angle ϕ describes a choice alternative's value x along a feature γ , which varies from γ_{min} (low anchor) to γ_{max} (high anchor), as

$$\phi = \phi_{max-min} \cdot \frac{x - \gamma_{min}}{\gamma_{max} - \gamma_{min}} \quad (14)$$

A feature value with an angle of ϕ that is closer to $\phi_{max-min}$ indicates it is nearer the maximum feature value γ_{max} , and closer to 0 indicates it is closer to the minimum feature value γ_{min} .

Such an approach is well-suited to modeling rating judgments in particular. Suppose a decision-maker must rate their confidence in a decision from 0 to 100, for example. The span of alternatives on the scale ranges from $\gamma_{min} = 0$ to $\gamma_{max} = 100$, and the possible responses will form an arc between v_0 and v_{100} . For example, supposing that $\phi_{max-min} = \pi/2$, a rating of 70 will fall at $\phi = \frac{\pi}{2} \cdot \frac{70}{100}$, represented by vector $v_{70} = [\cos(\phi), \sin(\phi)] = [.454, .891]$. This allows for continuous rating scales – where all vectors v_{0-100} are represented in the space – as well as discrete scales where only some subset are represented, such as $v_0, v_{25}, v_{50}, v_{75}$, and v_{100} . The same approach could be applied to construct a set of alternatives in any sort of rating or estimation task, including Likert scales, probability or frequency estimation, numerosity, or even price judgments like willingness to pay or willingness to accept.

5.1.2. Multidimensional items

Moving to multidimensional stimuli requires a modeler to consider how the features are related to one another. When all features are represented as independent and separable, our model should use orthogonal dimensions to describe each one. For binary features, each feature adds only a single dimension quantifying the presence versus absence of the feature. Each continuous feature requires two dimensions for the angle to vary corresponding to an item's value, as shown above in the unidimensional case. Therefore, the number of dimensions required to represent m (separable, independent) binary features and n (separable, independent) continuous features is simply $2n + m$.

The representation of a choice alternative is then set by turning the angles across all features into a set of coordinates. Each angle gives two coordinates describing its values along the low feature value and high feature value axes. If an alternative only has one feature γ and an angle describing its value on that feature ϕ , its coordinates are $v_A = [\cos(\phi), \sin(\phi)]$. This row vector serves as the description of the alternative in the decision space.

Adding independent features to this representation is simple. For an alternative A described by a set of n independent, continuous features $\gamma_1, \dots, \gamma_n$ and feature values x_1, \dots, x_n , we can calculate the angle with respect to feature vectors using Eq. (14). This gives us a set of feature angles ϕ_1, \dots, ϕ_n , and the vector describing the alternative A, v_A , will have coordinates given as

$$v_A = [\cos(\phi_1), \sin(\phi_1), \cos(\phi_2), \sin(\phi_2), \dots, \cos(\phi_n), \sin(\phi_n)]. \quad (15)$$

Any binary features can be added as coordinates of either +1 or –1, adding a single dimension per binary feature. Once constructed, the vector for each alternative can be normalized to a vector v'_A of length 1 by setting $v'_A = v_A / \|v_A\|$.

⁸ Work by Dodds, Rae, and Brown (2012) has actually suggested that the ends of some scales may come back around and become more similar to one another, creating a sort of 'ring' shape to the similarity between alternatives. This is easily handled in the geometric framework by allowing angles $\phi_{AB} > \pi$, which results in cosines to "bounce back" from –1 by coming back around the opposite side of the circle for items on opposing ends of a scale (forming a sort of horseshoe shape to the representation of the span of a feature).

5.2. State dynamics with features

These representations also pave the way for different evidence accumulation dynamics. As a person considers a particular feature, they move toward values of the feature that they prefer. For example, if a person is considering the quality and cost of a car as its two features, they might move toward low-cost and high-quality cars. Supposing that $\phi_{max--min} = \pi/2$ for all features, low cost will be in direction $v_{cost,low} = [1, 0, 0, 0]$, high cost will be in direction $v_{cost,high} = [0, 1, 0, 0]$, low quality will be in direction $v_{quality,low} = [0, 0, 1, 0]$, and high quality will be in direction $v_{quality,high} = [0, 0, 0, 1]$. When considering cost, the person will move toward $v_{cost,low}$ and when considering quality, they will move toward $v_{quality,high}$.

As in multiattribute decision field theory, we might assign a probability or *weight* to each feature, which gives the likelihood of considering feature m at any given point in time. If we suppose that the decision-maker considers cost half the time and quality half the time, we might find a probability vector of $p = .5 \cdot v_{cost,low} + .5 \cdot v_{quality,high} = [.5, 0, 0, .5]$. This can then be represented either as a random walk with discrete steps or as a continuous diffusion process; for now we just examine the latter. The motion of the preference state s over time (ds/dt) will be driven by the sampling rate λ , the weight vector p , and any momentary noise in the accumulation process (σ_t):

$$\frac{ds}{dt} = \lambda(p + \sigma_t) \quad (16)$$

On average, the support generated for a particular alternative A per unit time (dA/dt) will be a function of how closely alternative A matches this preference vector, again considering the sampling rate (λ) and any noise in the evidence accumulation process (σ_t):

$$\frac{d}{dt}Ev(A) = \lambda(p + \sigma_t) \cdot v_A \quad (17)$$

Naturally, support will accumulate fastest in favor of alternatives that possess features that are desirable to the decision maker, and the state will be most likely to reach the choice boundaries corresponding to alternatives that possess desirable features. Functionally, this is essentially the same as the diffusion process described by Smith (2016); the drift is simply driven by weighted preference for different features as in decision field theory (Busemeyer & Townsend, 1993; Roe et al., 2001).

5.3. Estimating relations among alternatives

So far, I have only covered the simplest case where features are fully separable and represented as independent quantities. But again this does not fully reflect the complexity of real-world decisions. In many cases, different features will be correlated; for example, decision makers in a variety of choice environments perceive and experience a trade-off between the potential payoff of a choice and the chance of receiving that payoff (Pleskac & Hertwig, 2014). These two features are represented as negatively related, which should correspond to a difference in representation of feature vectors such that the features are non-orthogonal. Furthermore, different features may carry different weight in computing the similarity of two items. In such cases, it is more difficult to derive the relation between features and similarity, as this requires another layer of modeling to impute similarity in representation from objective features.

For complex multi-attribute items, an alternative approach is to construct a vector space based on estimating the psychological representations of alternatives rather than using the objective features they possess. This requires additional information about

the individual differences between decision makers and their representations of the choice problem. For example, one could use similarity rating data (Shepard, 1962), infer similarity from neural data with approaches like multivariate pattern analysis (Kamitani & Tong, 2005; Kriegeskorte, Goebel, & Bandettini, 2006; O'Toole, Jiang, Abdi, Pénard, Dunlop, & Parent, 2007), or use binary confusion data to generate a similarity matrix relating each alternative to another. Once this is accomplished, the estimates of inter-alternative similarity can either be fed directly into calculating the angles between alternatives, or used in concert with multidimensional scaling to construct a distance-based representation of these alternatives in a feature space.⁹ The distance between items in the multidimensional scaling representation can in turn be mapped onto angles ϕ_{ij} between all pairs of alternatives $i \neq j$. The form of the function mapping similarity derived from behavioral, neural, or MDS analyses to angles is a potentially interesting topic for empirical work, but it must at least be monotonic with its output bounded between 0 and π .

When performed for individuals, this approach would yield predictions about decision patterns and response times that vary as a function of individual differences. For example, we might expect red (R) and green (G) hues in a red-green colorblind participant to appear similar to one another (small ϕ_{RG}), whereas participants with normal vision tend to rate them as very different (large ϕ_{RG}). We would therefore expect high confusion between red and green colors when a participant is asked to assign a target color-based stimulus to one category or the other (red or green), and longer response times when trying to discriminate between them (if relative thresholds are used) for the colorblind participants.

Another way to implement these relations in a geometric model is to use singular-value decomposition [SVD] to turn the similarities derived from ratings or multivariate pattern analysis of neural data into low-dimensional vector space representations of the alternatives (Deerwester et al., 1990; Furnas et al., 1988; Landauer & Dumais, 1997). The resulting vector space of alternatives (used in place of the usual concepts in latent semantic analysis) gives us the angle ϕ_{ij} for all $i \neq j$ among choice alternatives. Such an approach has been pursued with some success by Bhatia (2017): singular value decomposition based on latent semantic analysis was able to construct a vector space representation of alternatives that was predictive of the choices people made. Supplementing this vector space model with a dynamic evidence accumulation process forms a more complete picture of how people make decisions given the associations they have previously learned.

5.3.1. Negative features and rejection boundaries

If each feature of an item is represented as a balance between high and low feature values where $\phi_{max--min} = \frac{\pi}{2}$, alternatives with objective feature values can be represented in the first orthant of a Cartesian coordinate space (positive values along all dimensions). However, this is not necessary. In some decision situations, the best representation of available alternatives may make use of the full manifold on which alternatives can be situated. Negative values along a dimension correspond to features that have coherent opposites: opposing directions (north-south, up-down, left-right, etc.), gains versus losses, or opposing color pairs (yellow-blue, red-green) would be represented in opposing directions v and $-v$.

The hue color wheel is a good example of this. Colors can be arranged such that colors with opposing hues are situated across

⁹ As is typically done, multidimensional scaling representations with varying numbers of dimensions can be subjected to a model comparison to determine how many dimensions is optimal for a given set of alternatives

from one another, resulting in a circular arrangement of hues (see Hurvich & Jameson, 1957; Shepard, 1962). Although the psychological representation of these colors is slightly different than the hue metric and colors can be represented in different coordinate systems that better reflect perception (Tkalcic & Tasic, 2003), it seems that blue / yellow and red/green are often situated across from one another, providing natural opposing choice alternatives in color hue-based selection tasks. Clear examples also exist in 2- or 3-dimensional motion and orientation tasks: rightward opposes leftward, forward opposes backward, and upward opposes downward.

This provides opportunities for rejecting alternatives as in Section 2.2: if an alternative A is too low on a certain feature, the state may move away from it and cross the boundary θ_- in direction $-v_A$. At this point, alternative A is eliminated from the choice set and the space of alternatives is simplified as in elimination by aspects (Tversky, 1972). This can reduce the dimensionality of the representation of remaining alternatives when the remaining options vary on fewer dimensions (for example, if the remaining alternatives all match on cost when an expensive option is removed).

5.4. Context effects

In ideal circumstances, multidimensional scaling or singular value decomposition this can provide an immense simplification of the choice problem for both decision-makers and modelers. For example, in the case where two features are perfectly correlated, alternatives can be represented in just two dimensions rather than four, as high values of one feature will correspond perfectly to high (or low, if the correlation is perfectly negative) values of the other features. The relationships between alternatives will therefore depend on the items in the choice set — if there are four features but only three alternatives, the decision space can be reduced into a three-dimensional representation. Critically, the resulting three-dimensional representation will depend on the particular alternatives available and their features. For example, if all the alternatives share a feature (e.g., the same cost), the two feature dimensions can be collapsed into one dimension because that feature is non-discriminating for that choice scenario.

As a result, this approach can yield context effects based on the features of items included in the choice set. For example, the work of Tversky (1977) showed that the distribution of selections between pairs of items in a set depended on the third alternative (“Which country is most similar to Austria?” when the options were Sweden/Norway / Hungary versus Sweden/Poland / Hungary), resulting in a violation of independence from irrelevant alternatives [IIA] assumption made by many axiomatic theories of decision making. Tversky postulated that this was because the relevant features associated with Norway — as opposed to those associated with Poland — resulted in different attributes being involved in constructing representations of the other countries in the set. This same idea is reflected in the geometric framework: as the alternatives in the set change, the space of features most relevant to the available options also changes by virtue of how the space is collapsed into lower dimensional representations. In turn, this changes the weight of each feature based on the available choice alternatives in a set. This effect therefore provides a structural explanation for violations of IIA and diagnosticity in these paradigms. A proof of this is provided next.

Introducing additional alternatives or removing alternatives may change a person’s anchors for feature values as well — for example, introducing a very high cost/low quality alternative may make the other alternatives seem lower in cost and higher in quality or even change the attributes that the decision-maker is attending. This range effect results in context dependence

between alternatives in choice sets across many domains like risky choice (Birnbbaum, 2008), single attribute (Braida & Durlach, 1972; Lockhead, 2004; Marley & Cook, 1984) and multiattribute judgments (Mellers & Cooke, 1994; Trueblood, Brown, Heathcote, & Busemeyer, 2013), and especially consumer choice and pricing (Cunha Jr & Shulman, 2010; Hutchinson, 1983; Janiszewski & Lichtenstein, 1999). Range and other context effects are therefore an important feature for a model of preferential choice to handle. In Appendix B, I provide a simple geometric model that produces similarity, decoy, and compromise effects to illustrate how mechanisms within the geometric framework can be leveraged to produce these interesting phenomena. This material is somewhat tangential to the main point, however, which is that context effects are naturally produced by the principles I described above. In the next section, I provide a proof that singular value decomposition results in context effects in the geometric framework.

Proof of context effects. We want to prove that the relative support generated for alternatives A and B differs between when they are presented alongside alternative C versus presented alongside alternative D, due to the differences in singular value decomposition resulting from sets A, B, C and A, B, D. Recall that in the geometric framework, the weight vector \mathbf{p} describes the weight attached to each feature dimension, giving the average support generated for alternatives that have that feature. The average amount of support per unit time generated for each alternative based on this weight vector is given in Eq. (17): the average support generated for A is given by $\mathbf{p} \cdot v_A$, the support generated for B is $\mathbf{p} \cdot v_B$, and so on. Suppose that \mathbf{p} is the attention weight vector representing the average motion of the state in four dimensions, \mathbf{p}'_{ABC} is its value when collapsed into three dimensions using singular value decomposition of A, B, C, and \mathbf{p}'_{ABD} is its value when collapsed into three dimensions using singular value decomposition of A, B, D. The goal of the proof is to show

$$\frac{v'_A \cdot p'_{ABC}}{v'_B \cdot p'_{ABC}} \neq \frac{v'_A \cdot p'_{ABD}}{v'_B \cdot p'_{ABD}} \quad (18)$$

where v'_A is the new vector for alternative A following decomposition and v'_B is the new vector for alternative B following decomposition.

To perform singular value decomposition on a set of four-dimensional vectors v_A, v_B, v_C (which describe alternatives A, B, and C), we first create a 3×4 matrix where each vector is a row,

$$M_{ABC} = \begin{bmatrix} v_A \\ v_B \\ v_C \end{bmatrix} \quad (19)$$

Because there are more dimensions/columns (4) than vectors/rows (3), there exists a linear span of the set of vectors, referred to as $S_{ABC} = \text{Span}(M_{ABC})$. The span S_{ABC} is sufficient to describe the set of alternatives and preserve all similarity relations between them, because it contains all of the linearly independent vectors needed to describe the set of alternatives. Singular value decomposition is a method for computing a set of vectors in fewer dimensions (in this case, 3 dimensions) that preserve all of the internal relations between the original set (vectors in 4 dimensions) by using this span.

The singular value decomposition of M_{ABC} is given as $M_{ABC} = U_C \Sigma_C V_C^{-1}$ (simplifying our notation to C rather than ABC, since A and B are always in the choice set). The $^{-1}$ indicates the inverse of a matrix. The rows of the matrix U_C represent a linearly independent set of vectors that covers the range of alternatives in the set. If all alternatives are linearly independent (i.e., if none of them can be thought of as linear combinations of one another,

such as northeast motion is a combination of northward and eastward motion), each one will be represented as a row of U_C .

Once the relation between M_{ABC} and U_C is established, we can take any vector in the decision space \mathfrak{N}^4 and map it into the new decision space \mathfrak{N}^3 using the same transformations. For example, we can take the vector \mathbf{p} specifying the weight assigned to each dimension in the four-dimensional space and examine the resulting weight \mathbf{p}'_{ABC} of each dimension in the three-dimensional space.

$$\mathbf{p}'_{ABC} = \mathbf{p}V_C \Sigma_{C,right}^{-1} \tag{20}$$

Here, V is the inverse of V^{-1} , and $\Sigma_{C,right}^{-1}$ is the right inverse of Σ , $\Sigma \Sigma_{C,right}^{-1} = I_3$. Once this is done, we have a new vector \mathbf{p}_{ABC} that specifies the evidence accumulation in \mathfrak{N}^3 under the singular value decomposition of M_{ABC} . We can easily calculate the rate of accumulation for A, B, and C by multiplying $\mathbf{p}'_{ABC}U_C^{-1}$ (the inverse of U_C), provided A, B, and C were linearly independent in the first place. If they are not linearly independent, a vector \mathbf{v}_C that was 'lost' in the decomposition can be calculated as $\mathbf{v}'_C = \mathbf{v}V_C \Sigma_{C,right}^{-1}$, then take the dot product $\mathbf{p}'_{ABC} \cdot \mathbf{v}_C$ to calculate its resulting support. The rows of the resulting vector $\mathbf{s}_{ABC} = \mathbf{p}'_{ABC}U_C^{-1}$ will give the new rate of evidence accumulation in favor of A, B, and C in the lower-dimensional space.

As with set A, B, C, one can perform a similar singular value decomposition for set A, B, D. As we before, this will take a matrix M_{ABD} and decompose into $M_{ABD} = U_D \Sigma_D V_D^{-1}$. The resulting weight vector can be calculated as $\mathbf{p}'_{ABD} = \mathbf{p}V_D \Sigma_{D,right}^{-1}$, and the support generated for each alternative is given as $\mathbf{p}'_{ABD}U_D^{-1}$.

Putting the two decompositions together, we can compare the support generated for A and B as the first two entries of $\mathbf{p}'_{ABC}U_C^{-1}$ and $\mathbf{p}'_{ABD}U_D^{-1}$. The support generated for A, B, and C is $\mathbf{s}_{ABC} = \mathbf{p}V_C \Sigma_{C,right}^{-1}U_C^{-1}$, while the support generated for A, B, and D is $\mathbf{s}_{ABD} = \mathbf{p}V_D \Sigma_{D,right}^{-1}U_D^{-1}$. To calculate the relative support for A and B, we can either divide the first entry of the resulting matrix by its second entry, or calculate them individually and compare them. Substituting the result into Eq. (18), we find

$$\frac{(\mathbf{p}V_C \Sigma_{C,right}^{-1}) \cdot (\mathbf{v}_A V_C \Sigma_{C,right}^{-1})}{(\mathbf{p}V_C \Sigma_{C,right}^{-1}) \cdot (\mathbf{v}_B V_C \Sigma_{C,right}^{-1})} \tag{21}$$

which will not be equal to

$$\frac{(\mathbf{p}V_D \Sigma_{D,right}^{-1}) \cdot (\mathbf{v}_A V_D \Sigma_{D,right}^{-1})}{(\mathbf{p}V_D \Sigma_{D,right}^{-1}) \cdot (\mathbf{v}_B V_D \Sigma_{D,right}^{-1})} \tag{22}$$

The matrices V_D and V_C are just rotations in four dimensions that preserve all internal dot products, so they do not directly affect the balance of support for A and B. The key is how they interact with $\Sigma_{C,right}^{-1}$ and $\Sigma_{D,right}^{-1}$, which are both projectors that reduce the state from four dimensions to three dimensions. In singular value decomposition, the rotation V_C or V_D moves the vectors describing alternatives (\mathbf{v}_A , \mathbf{v}_B , \mathbf{v}_C , and \mathbf{v}_D) such that their last coordinate will become zero, making their overall length unaffected by the projection operators. However, this is not true for the effect on \mathbf{p} because it is not considered in computing the decomposition so its last coordinate will not be zero when rotated. When $\Sigma_{C,right}^{-1}$ is applied to $\mathbf{p}V_C$, it will have a different effect than when $\Sigma_{D,right}^{-1}$ is applied to $\mathbf{p}V_D$ because a different coordinate will be affected by the projection. As a result, \mathbf{p} will wind up generating different support for A and B dependent on whether $\Sigma_{C,right}^{-1}$ (in case ABC) or $\Sigma_{D,right}^{-1}$ (in case ABD) is applied.

To illustrate this point, it is helpful to provide an example. Let us suppose that $\mathbf{v}_A = [\sqrt{8}, \sqrt{2}, 0, 0]$, $\mathbf{v}_B = [\sqrt{2}, \sqrt{4}, \sqrt{2}, \sqrt{2}]$, $\mathbf{v}_C = [0, 0, 1, 0]$, and $\mathbf{v}_D = [\sqrt{5}, \sqrt{5}, 0, 0]$. Additionally, for simplicity, let us assign equal weight initially to all dimensions,

$\mathbf{p} = [.5, .5, .5, .5]$ (note that $\|\mathbf{p}\| = 1$, and the lengths of all alternative vectors are also 1). Before decomposition. Our initial matrices are

$$M_{ABC} = \begin{bmatrix} \mathbf{v}_A \\ \mathbf{v}_B \\ \mathbf{v}_C \end{bmatrix}$$

$$M_{ABD} = \begin{bmatrix} \mathbf{v}_A \\ \mathbf{v}_B \\ \mathbf{v}_D \end{bmatrix}$$

Applying singular value decomposition, we obtain matrices for U_C , Σ_C , and V_C .

$$M_{ABC} = \begin{bmatrix} \mathbf{v}_A \\ \mathbf{v}_B \\ \mathbf{v}_C \end{bmatrix} = U_C \Sigma_C V_C^{-1}$$

$$= \begin{bmatrix} -.59 & .55 & .59 \\ -.71 & 0 & -.71 \\ -.39 & -.84 & .39 \end{bmatrix} \begin{bmatrix} 1.35 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & .43 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} -.63 & .49 & .50 & .35 \\ -.53 & .25 & -.42 & -.69 \\ -.52 & -.84 & .17 & 0 \\ -.23 & 0 & -.74 & .63 \end{bmatrix}$$

Calculating the weights for the features in three dimensions as $\mathbf{p}'_{ABC} = \mathbf{p}V_C \Sigma_{C,right}^{-1}$, we find $\mathbf{p}'_{ABC} = [-.71, -.05, -.58]$.

Conversely, the decomposition for set A, B, D is

$$M_{ABD} = \begin{bmatrix} \mathbf{v}_A \\ \mathbf{v}_B \\ \mathbf{v}_D \end{bmatrix} = U_D \Sigma_D V_D^{-1}$$

$$= \begin{bmatrix} -.58 & .49 & -.65 \\ -.54 & -.83 & -.14 \\ -.61 & .27 & .75 \end{bmatrix} \begin{bmatrix} 1.61 & 0 & 0 & 0 \\ 0 & .60 & 0 & 0 \\ 0 & 0 & .21 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} -.74 & .42 & -.52 & 0 \\ -.64 & -.20 & .74 & 0 \\ -.15 & -.62 & -.30 & -.71 \\ -.15 & -.62 & -.30 & .71 \end{bmatrix}$$

The resulting weight vector \mathbf{p}'_{ABD} is $[-.52, -.86, -.89]$. With both decompositions computed, the average evidence for A, B, and C under decomposition A, B, C can be computed as $\mathbf{s}_{ABC} = \mathbf{p}'_{ABC}U_C^{-1} = [0.05, 0.92, 0.09]$. The average support for A, B, and D under decomposition A, B, D is given as $\mathbf{s}_{ABD} = \mathbf{p}'_{ABD}U_D^{-1} = [0.46, 1.12, -.059]$. To complete the comparison, we compute the ratio of support for A and B under ABC as $.05/.92 = .0543$ and the ratio under ABD as $.46/1.12 = .4107$. The difference in this ratio gives us the context effect from the decomposition. This is obviously an extreme example, but hopefully it illustrates how substantial the effect of decomposition can be.

Interestingly, a side effect of this transformation is that the resulting \mathbf{p}' will also have a different length depending on the alternatives used to compute the decomposition. Thus, not only will the relative support for A and B depend on whether they are decomposed using ABC or ABD, but the total support for all alternatives $\|\mathbf{p}'_{ABC}\|$ and $\|\mathbf{p}'_{ABD}\|$ will also depend on the choice set. It is an open question whether \mathbf{p}'_{ABC} and \mathbf{p}'_{ABD} should be normalized; if not, then response times should actually change as a function of the third alternative in a choice set. The inclusion of C versus D could make the decision overall slower or faster because it changes the rate of information sampling.

In simple terms, the result of the decomposition of the decision space is that it re-weights different features according to the alternatives in the choice set. This alters the sampling vector \mathbf{p} that describe the emphasis a person puts on the different dimensions of the original higher-dimensional space, resulting in a revised sampling vector \mathbf{p}' for different sets of alternatives. As

a result, the support generated for each alternative depends on the other alternatives in the set. This can result in violations of independence of irrelevant alternatives and diagnosticity, just as re-weighting features in the attribute weighting model of Tversky (1977) produces these effects.

6. Discussion

As shown, the geometric framework provides methods for modeling choice behavior in both inferential and preferential choice, offering tools for constructing models where objectively optimal behavior might be expected as well as scenarios where choice is entirely subjective. It connects binary and multialternative choice to asymptotic cases where responses can fall along a continuum, providing a natural explanation for Hick's law when relative stopping rules are implemented. The geometric framework also emphasizes the importance of representing the relationships between alternatives as part of a decision model, and presents a simple way for integrating these relationships into our models of the decision process. In preferential choice, the framework provides a way to inform models of the decision process with representational models using methods like singular value decomposition, which in turn provides a potential mechanism explaining context effects.

As part of this approach, the geometric framework also allows for decision models to be related to neural data and processes. Like the leaky competing accumulator model (Usher & McClelland, 2001), it includes inhibitory and excitatory connections between different alternatives (although these do not result in the same nonlinear dynamics because the inhibition and excitation relations are passive rather than active; see Fig. 3) and accumulation-to-threshold as two of the primary mechanisms underlying decision dynamics. It even prescribes decision rules for making optimal decisions in light of these interactions between alternatives. Using multivariate pattern analysis of neural data may even allow us to construct the decision space a priori, using the similarity in neural representations of alternatives to generate similarity-based representations of alternatives for making behavioral predictions.

Despite this array of applications, there are a number of outstanding issues in the framework. These include theoretical issues like where the geometric approach can and cannot be used (including conditions for falsifiability) as well as practical ones like how computationally tractable the representations and dynamics will be.

6.1. Theoretical considerations

The geometric framework allows for modeling of responses, judgments, and response times among any number of alternatives with varying assortments of similarity relations and underlying feature structures. It is intentionally quite flexible in order to allow it to be applied across choice domains, but constructing a model using the geometric framework is often quite straightforward and principled. A modeler expresses the assumptions going into their model by specifying the multidimensional space of alternatives, the sampling distribution, and the system(s) of equations that map states onto responses. Working within these assumptions, there are a few ways to theoretically rule out its mechanisms as plausible explanations of behavior.

The main theoretical limitations concern where the geometric approach will be able to be applied. The psychological theory built into the modeling architecture we have described here relies heavily on the specification or calculation of angles between vectors (describing alternatives) and distances along those vectors. It therefore requires a Riemannian metric on a manifold to describe

the relations between preference or evidence states and the alternatives that can be chosen. This makes it more convenient to apply to alternatives that have attributes measured in Euclidean metrics (e.g. quantitative values along a dimension) or a Hilbert space.

Interestingly, the geometry of neural representations has been considered in terms of simplices Reimann, Nolte, Scolamiero, Turner, Perin, Chindemi, Dlotko, Levi, Hess, and Markram (2017) like the ones constructed in the initial optimal models, and there are a number of avenues to draw connections between the neural data and the geometric structure ascribed to behavioral data. However, it may not be the case that the geometry of neural representations is restricted to manifolds in the same way. Finding neural structures that suggest alternatives are not represented on a Riemannian manifold could potentially require an extensive revision of the geometric framework.

Unless cases are found where a Riemannian manifold is ruled out, the geometric framework can in principle be used for constructing representations of choice alternatives and evidence. Discrimination therefore happens at the level of psychological theory that is built into the geometric model: how alternatives are arranged, what sampling distributions are used, and what response rules are applied. The specification of these components constitutes testable hypotheses about the decision process, and model comparison can be done by contrasting different implementations of the framework. The types of models that can be implemented are partly restricted by computational considerations, which leads to a number of practical issues in the geometric framework.

6.2. Practical considerations

One of the more noticeable consequences of the optimal models presented in the first section is that the number of dimensions required to represent the alternatives increases linearly with the number of alternatives. This can make it impractical to derive predictions when the number of alternatives gets large. One way to get around this problem has already been covered: restricting the model to lower-dimensional representations by using multidimensional scaling or singular value decomposition will help reduce the dimensions needed to represent the alternatives. This will be most powerful when a modeler has a strong theory about the number of dimensions needed to represent the alternatives in their choice set, and of course model comparison will help arbitrate between representations or decision models with varying numbers of dimensions.

6.2.1. Tractable likelihoods

Perhaps the greatest hurdle to utilizing some of the geometric models is the lack of closed-form solutions to the likelihood functions giving the choice probabilities and response time distributions as a function of the model parameters. Some particular combinations of sampling distributions and response rules will have analytic likelihoods, such as the von Mises or von Mises–Fisher distributions resulting from diffusion inside circular or spherical choice bounds (Smith, 2016; Smith & Corbett, 2018). These likelihoods are extremely convenient when they can be used, but shifts in the start point or irregularly-shaped (non-hyperspherical) response boundaries can result in situations where they cannot be used.

Instead, many of the models I presented here will require simulation methods for generating approximate likelihood functions. These work by taking a set of parameters, generating a set of simulated trials (e.g., 10,000 simulations), and estimating the joint distribution of responses and response times from the simulated sample using methods like kernel density estimation

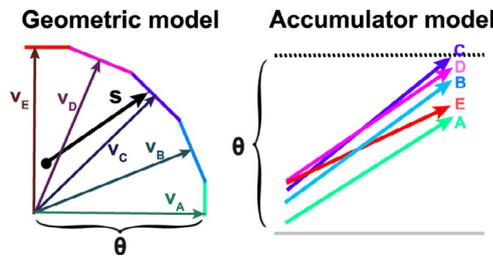


Fig. 8. Equivalent geometric and multiple-LBA representations of a decision. Note that the rates and starting points in the LBA will be correlated according to the relative orientations of the alternative vectors ($d_{1=5}$).

(for Bayesian methods along these lines, see [Holmes, 2015](#); [Turner & Sederberg, 2012](#)). This approximate likelihood method will often be necessary when response boundaries form irregular figures, the starting point of evidence accumulation is not centered at the origin, and where multiple alternatives with asymmetric similarity relations are involved (due to the unusual shapes formed by the response boundaries).

6.2.2. Deterministic dynamics

Using models that have deterministic evidence evolution (such as linear ballistic accumulators [Brown & Heathcote, 2008](#)) will alleviate the problems associated with simulated likelihoods immensely, as a random variable does not need to be drawn for each step of the accumulation process as in multinomial or diffusion-based stochastic dynamics. The description thus far of the sampling process has been characterized by independent samples of evidence drawn with noise at each time step of the random walk or diffusion process.

Removing the within-trial noise associated with the random walk can potentially make geometric models much easier to work with, resulting in deterministic accumulation like that of the linear ballistic accumulator model ([Brown & Heathcote, 2008](#); [Rouder, Province, Morey, Gomez, & Heathcote, 2015](#)). In this model, the trajectory of evidence accumulation is a ray directed at an angle drawn from a probability distribution, making it computationally convenient to derive where the trajectory hits choice boundaries. Using the geometric framework, we can simply find the points of intersection between the ray trajectories generated by a multidimensional ballistic accumulation process and the choice boundaries that correspond to the set of alternatives. Specifying the multidimensional ballistic accumulators is relatively straightforward; it requires a distribution of starting points in n dimensions, a direction of accumulation (e.g., a mean direction plus variability), and a rate of accumulation or distribution of rates along the various dimensions of accumulation

For example, one might suggest a rate of accumulation $r_1 \sim \text{Normal}(\mu_1, \sigma_1)$ along feature 1 in direction v_1 , and a rate of accumulation $r_2 \sim \text{Normal}(\mu_2, \sigma_2)$ along feature 2 in orthogonal direction v_2 . The starting points vary along the same dimensions, so that there is a starting point along feature 1 $s_1 \sim \text{Uniform}(0, a_1)$ and a starting point along feature 2 $s_2 \sim \text{Uniform}(0, a_2)$. The angle of accumulation between v_1 and v_2 would be $\phi = \tan^{-1}(r_2/r_1)$ and the overall rate of accumulation would be $r_{\text{total}} = \sqrt{r_1^2 + r_2^2}$. The point at which this accumulation process crosses the response boundaries (hyperplanes orthogonal to v_A, v_B, \dots in an absolute stopping rule model, or ones defined by the difference between alternatives for relative stopping rule) gives the response, and the time at which it crosses the first boundary yields a response time.

This accumulation process, for a set of alternatives v_{A-E} , is shown in [Fig. 8](#). Interestingly, the multidimensional geometric

representation of a linear ballistic accumulation process is equivalent to a multiple accumulator model with correlated rates and starting points. For alternatives A, B, \dots in directions v_A, v_B, \dots , the starting point of accumulator n is determined by the component of the starting point $[s_1, s_2]$ along vector v_n . The rate of accumulation in favor of alternative n is the component of the drift vector $[r_1, r_2]$ along v_n . This allows us to derive a starting point and drift rate distribution for a set of alternatives based on the starting point and rates of the multidimensional, geometric model. These rates and starting points will naturally be correlated according to the relationships between alternatives and features (dimensions), as the starting points and drift rates of each accumulator are all derived from the same underlying parameters r_1, r_2, s_1 , and s_2 .

Of course, a similar representation of evidence could be derived for each alternative in a stochastic accumulator model as well — the rate of accumulation (and noise) for each stochastic accumulator is given by the differential equation shown in [Eq. \(17\)](#). In either case, the deterministic dynamics will simplify the model relative to stochastic dynamics when it comes to simulation-based likelihoods, and may prove to be a more practical approach to implementing geometric models in the long run.

6.3. Conclusions

The geometric framework developed here has a multitude of potential applications, and provides a principled way of building evidence accumulation models for small or large sets of alternatives. In principle, it can be used to explain and predict selections among any set of choice alternatives that can be represented on a Riemannian manifold. In most cases, the primary hurdle is constructing the decision space that relates the alternatives to one another. Research in some domains has already mapped this structure — for example, pitches and tones one can select are often modeled as toroidal or helical shapes ([Lerdahl, 2004](#); [Shepard, 1982](#)). We might reasonably expect that a decision space where someone has to select (or produce) a pitch or tone should reflect these psychological representations, both in terms of what errors they make and how quickly they are able to reach their decisions. This is only the tip of the iceberg — any domain in which the psychological relations between alternatives can be empirically estimated is a domain in which the modeling framework can in principle be applied to predict decisions among them.

Like its predecessors, such as diffusion models, the geometric framework also provides a link between decision behavior and underlying neurobiology ([O'Connell, Shadlen, Wong-Lin, & Kelly, 2018](#)). The cosine rule that quantifies the association between two alternatives ([Eq. \(2\)](#)) naturally relates to shared activation or inhibition between feature dimensions or alternatives. And as illustrated above, it can implement an optimal decision procedure that is thought to be carried out via computations in the cortex and basal ganglia ([Bogacz, 2007](#); [Bogacz & Gurney, 2007](#)). The multidimensional state additionally allows us to represent how widespread activation across populations of neurons, which are not inherently associated with a single particular choice outcome, gives rise to degrees of support for a set of choice options through its component-based associations ([Ma, Beck, Latham, & Pouget, 2006](#); [Niwa & Ditterich, 2008](#)).

My hope is that the material presented here serves as a useful tool for constructing models of the decision process across domains, including binary and multiple choice, ordinal ranking judgments, best-worst selections, categorization, ratings like Likert scales, and numerical judgments like confidence, estimation, and price. The geometric framework serves as a general case that

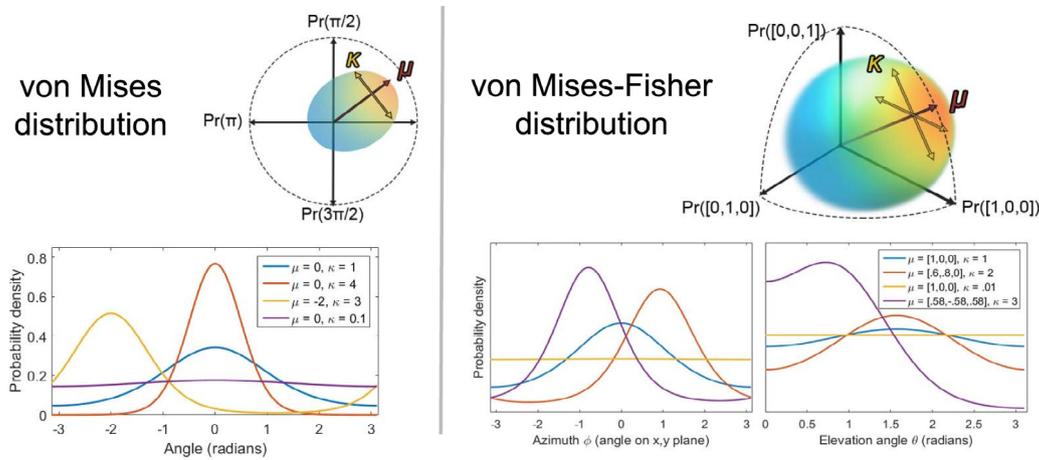


Fig. 9. Parameters and shapes of von Mises and von Mises–Fisher distributions. The von Mises–Fisher distribution can be specified in arbitrarily many dimensions, but its 3-dimensional structure is shown for reference.

brings multiple models together in the same umbrella, connects these models to neurobiology, and provides a method for constructing new models of the decision process that are sensitive to the alternatives contained in a choice set.

Appendix A. Continuous distributions for sampling

Although they result in convenient likelihoods in many cases, the diffusion dynamics described in Eq. (17) may not perfectly match the perceptual information provided by every stimulus. When there is a substantial mismatch or unusual properties to a stimulus (such as bimodal distributions or correlated dimensions), it may be more prudent to construct a sampling distribution that reflects the properties of a stimulus. In this case, a modeler must define evidence dynamics in terms of a probability distribution over step directions (what is sampled and how it relates to orientations in the decision space) and the rate at which samples arrive. The sampling distribution has to cover a potentially high-dimensional space, as evidence accumulation must unfold in the same number of dimensions as the decision space.

For a two-dimensional continuous sampling distribution, one convenient candidate for the distribution of ϕ (step direction) is the von Mises distribution $\phi \sim VM(\mu, \kappa)$, which is similar to a normal distribution wrapped around a circle (Mardia, 2014; von Mises, 1918). The distribution of ϕ over possible directions x given its parameters κ and μ is given by the probability density function:

$$Pr(\phi = x | \mu, \kappa) = \frac{\exp(\kappa \cos(x - \mu))}{2\pi I_0(\kappa)} \tag{23}$$

The von Mises distribution has two parameters: μ , which specifies the direction of the central tendency of the samples; and κ , which specifies the precision of the samples (the inverse of the variance). Higher κ indicates more tightly concentrated samples, and lower κ indicates ones that are more spread out over the circle. In the case of alternative or attribute wise sampling, these can be interpreted as the average attribute or alternative a person favors (specified by direction μ) and the noise or variation around it ($1/\kappa$). As κ decreases toward zero, the von Mises distribution approaches a circular uniform distribution, where steps in all directions are equally likely. The von Mises distribution can be simulated efficiently, so drawing many random variables from this distribution to predict the dispersion of responses on a circle is not too computationally demanding (Best & Fisher, 1979).

When the representations of alternatives uses a higher number of dimensions, the sampling distribution will have to change appropriately. For example, a stimulus providing evidence in 3 or more dimensions might be better characterized by a von Mises–Fisher distribution $\phi \sim VMF(\mu, \kappa, \nu)$, which specifies the dispersion of vectors on an n -sphere (Banerjee, Dhillon, Ghosh, & Sra, 2005; Fisher, 1953; Watson & Williams, 1956).

$$Pr(\phi = x | \mu, \kappa, \nu) = \frac{\kappa^{\nu-1}}{(2\pi)^\nu I_{\nu-1}(\kappa)} e^{\kappa \mu \cdot x} \tag{24}$$

Similar to the von Mises distribution, this sampling distribution has a central tendency μ – described by an $(n - 1)$ -dimensional vector using spherical coordinates – that specifies the modal direction of the samples. It also has a similar concentration parameter κ that specifies the precision of the samples; higher values of κ mean it will move more consistently in directly μ , lower values mean it will move around more randomly. The von Mises–Fisher distribution also adds another parameter ν that simply allows it to adapt to the number of dimensions: its value is $\nu = \frac{n}{2}$ where n is the number of dimensions in the representation/decision space.

This allows the von Mises–Fisher distribution to be used to specify a sampling distribution over any n -sphere, making it convenient for modeling decisions among many alternatives that have many features or dimensions underlying their representations. It might be used to describe an object moving in three dimensions or a choice option with many attributes. For example, a decision among many different cars might use dimensions for price, longevity, handling, acceleration, physical appearance, color, and so on that contribute to high-dimensional representations.

The probability density functions for the von Mises and von Mises–Fisher distributions are given in Eqs. (23) and (24) respectively, and depicted in Fig. 9. The function $I_m(\cdot)$ is a modified Bessel function of the first kind, order m , which frequently appear in circular statistical distributions that need to wrap where 0 and 2π meet.

Fortunately, the sampling rate gets no more complex as the number of dimensions involves increases. The arrival time of each piece of evidence/step in the accumulation process can still be described by a fixed increment t or an exponential distribution $t \sim Exponential(\lambda)$, just as it is in other continuous-time models. Not all of the parameters of the distribution will be simultaneously meaningful, though – a lower sampling rate with higher precision and a lower threshold may give the exact same results as one with a higher sampling rate, lower precision, and

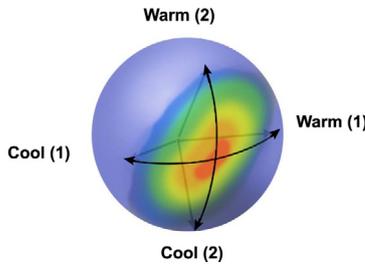


Fig. 10. Asymmetric probability distribution on a sphere. Red indicates the most likely directions, with yellow, green, and blue indicating progressively less likely directions for each sample to move the evidence state. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

higher threshold (and matched mean direction). Therefore, the parameters have to be either constrained across conditions or fixed within a condition in order to set their scale, just as with diffusion and accumulator models.

A.1. Asymmetric distributions

One drawback of using the von Mises–Fisher distribution described in Eq. (24) is that it has the same level of variability along all dimensions; it cannot have greater noise in the representation of one feature relative to another. This means that the state will be symmetrically distributed around the mean accumulation direction when it reaches a specified distance from the starting point (i.e. with spherical response boundaries).

This assumption is often reasonable, but may not make sense for all decision situations. When selecting ingredients to put into a meal or drink, a person may prefer to have items that match in temperature – for example, they may have a meal made of all cold ingredients or all warm ingredients, but not mix warm with cool. In a three-dimensional space where azimuth angle corresponds to the temperature of one ingredient and horizon corresponds to the temperature of another, the preferred combination will be a diagonally oriented distribution like the one shown in Fig. 10.

With only a von Mises–Fisher distribution for sampling, a modeler cannot generate this pattern of responses because the samples would be forced to be symmetric. Instead, some way of implementing a distribution that looks like a multivariate normal rather than a symmetric normal wrapped around a sphere. There is a generalized version of the von Mises–Fisher distribution that does this called the Fisher–Bingham distribution (Mardia, 2014), but a more frequently used (and mathematically convenient) distribution is the Kent distribution (Kent, 1982). This distribution describes a probability density function over three-dimensional vectors x on the 3-sphere using five parameters γ_{1-3} , κ , and β .

$$Pr(\phi = x | \gamma_1, \gamma_2, \gamma_3, \kappa) = \frac{1}{c(\kappa, \beta)} \cdot \exp\left(\kappa \gamma_1 \cdot x + \beta [(\gamma_2 \cdot x)^2 - (\gamma_3 \cdot x)^2]\right) \quad (25)$$

The term $c(\kappa, \beta)$ is a normalization constant:

$$c(\kappa, \beta) = 2\pi \sum_{j=0}^{\infty} \frac{\Gamma(j + .5)}{\Gamma(j + 1)} \beta^{2j} \left(\frac{2}{\kappa}\right)^{2j+.5} I_{2j+.5}(\kappa) \quad (26)$$

As before, Γ is the gamma function and $I_{2j+.5}()$ is the modified Bessel function of the first kind, order $(2j + .5)$. The parameters γ are three orthogonal unit vectors. The first, γ_1 , specifies the mean direction of accumulation (mean orientation); the second, γ_2 , specifies the major axis along which responses are dispersed

(i.e., the direction along which the distribution will be stretched); and the last, γ_3 specifies the minor axis along which responses will be normally dispersed. As in the vMF and vM distributions, the parameter κ controls the general precision (inverse variance) of the distribution, while β controls how elliptical/asymmetric the noise is on the major relative to the minor axis.

The Kent distribution can be extended to higher dimensions as well. For each additional dimension n , there will need to be an additional vector γ_n specifying another axis along which the distribution can be stretched, and an additional parameter β that specifies how widely the distribution is stretched along that axis. In this way, it functions very similarly to a multi-variate normal distribution. The full probability density for the random direction variable ϕ pulled from an n -dimensional Kent distribution is

$$Pr(\phi = x | \gamma_1, \gamma_2, \gamma_3, \beta, \kappa) = \frac{1}{c(\kappa, \beta)} \times \exp\left(\kappa \gamma_1 \cdot x - \left(\sum_{i=2}^n \beta_i (\gamma_i \cdot x)^2 + \sum_{j=2}^{n-1} \beta_j (\gamma_j \cdot x)^2\right)\right) \quad (27)$$

The normalizing constant $c(\kappa, \beta)$ acts as a multiplier that simply scales the probability density according to the β and κ parameters. The β parameters are subject to the restrictions $\frac{\kappa}{2} \geq |\beta| \geq 0$ and $\sum_{j=2}^n \beta_j = 0$.

Applied to our example of ingredient matching, we might expect that the mean accumulation direction is toward $\gamma_1 = [1, 0, 0]$, indicating no strong preferences for warm or cool ingredients. The major axis would be something like $\gamma_2 = [0, \sqrt{.5}, \sqrt{.5}]$, creating a line from cool, cool (low values of y and z) to warm, warm (high values of y and z). This would be the axis along which the distribution would be wider dependent on the value of β , and thinner along the remaining minor axis $\gamma_3 = [0, -\sqrt{.5}, \sqrt{.5}]$. Then the parameter κ would control how widely the selections were dispersed across directions.

The von Mises, von Mises–Fisher, and Kent distributions represent only a handful of potential distributions over step angles. Multinomial or other multidimensional distributions may prove useful in modeling stimuli with unusual evidence properties. In some cases, it may also be reasonable to have multiple accumulation processes running at the same time, which seems to be important for predicting multimodal response distributions (Kvam, 2019; Ratcliff, 2018).

Appendix B. A simple model of context effects

At this point it is critical to note that context effects are permitted in the framework in principle by virtue of its ability to violate IIA and diagnosticity, as shown in the proof in the main text. This will yield context effects merely by the particular decomposition of the feature space for a set of alternatives, which generates different weights on different features depending on the choice set. The toy model of context effects I describe in this section is mainly to demonstrate how a simple geometric model can generate several of the most common context effects. It is not intended to be construed as a new competitor to existing models; rather, it is likely that equivalence relations or simple modifications of the geometric framework will generate models that behave like the multiattribute linear ballistic accumulator (MLBA Trueblood et al., 2014), 2N-ary choice model (Wollschläger & Diederich, 2012, 2017), multialternative decision field theory (Roe et al., 2001), the leaky competing accumulator model (Usher & McClelland, 2001), and other models that generate context effects. The geometric framework would therefore be a general case of several implementations that can create the effects in principle, and particular theories instantiated as special cases of the geometric framework can be compared in terms of the

particular mechanism they posit for each effect or combination of effects.

The toy model presented here assumes that each alternative has two features that can vary continuously. First, we relate the response alternatives to one another. Each alternative possesses a value along each of the feature dimensions, and the initial choice alternatives' values are negatively correlated (i.e. one is high on dimension 1 and low on dimension 2, the other is low on dimension 1 and high on dimension 2). Alternative A is therefore represented by a vector that has a high x value (dimension 1) and low y value (dimension 2) as $v_A = [1, 0]$, and alternative B is initially represented by a vector with a low x value and high y value as $v_B = [0, 1]$.

Next, we specify the sampling distribution. In this example, assume that a decision-maker values an item that is high on dimension 1 and high on dimension 2 equally.¹⁰ When considering dimension 1, this valuation moves their preference state rightward on average, and when attending to dimension 2 it moves their preference state upward on average. The net movement of this state – putting together rightward movement, upward movement, and noise – will be toward the upper-right. We therefore model it as a diffusion process with mean drift in direction $\mu = \frac{\pi}{4}$ radians (45 degrees). The precision of the sampling distribution is set to $\kappa = 1$ to set the scale of the process.

Finally, the response boundaries are formed by specifying a line tangent to each of the alternative vectors at height $\theta = 2$ (again, arbitrarily chosen). This will result in the equation $x \geq 2$ as a rule for selecting alternative A, and $y \geq 2$ as the rule for selecting alternative B. Put together, this will result in roughly 50% of responses favoring A and 50% favoring B, as shown in the top panels of Fig. 11.

B.0.1. Similarity

So how do we get context effects by introducing a third alternative? The three effects are shown in Fig. 11. The similarity effect is perhaps the most straightforward, so it provides a good starting point. Suppose that we introduce a third alternative C with attributes similar to alternative B. The typical finding is that doing so reduces the proportion of responses favoring B relative to A, tipping response proportions away from 50/50 (Sjöberg, 1977; Tversky, 1972) toward alternative A and resulting in a violation of the property of independence from irrelevant alternatives.

The addition of a third alternative C requires adding an additional vector v_C and its corresponding orthogonal boundary. For an alternative similar to A, this response boundary might be placed at $v_C = [\sqrt{.1}, \sqrt{.9}]$. Given that $\theta = 2$, the point at which this vector intersects its orthogonal boundary will be $2 \cdot v_C = [1.90, 0.63]$, and therefore the response boundary will be roughly $y = -0.33x + 2.57$. This response boundary will intersect the ones for A and B, as shown with the yellow line in the middle-left panel of Fig. 11. However, its intersection with $y = 2$ (the boundary for B) will be at $x = 1.71$ while its intersection with $x = 2$ (the boundary for A) will be at $y = 1.90$.

This means that the C boundary will intersect the B boundary earlier along its length than it will intersect the A boundary, cutting short the response rule for B more than for A. Consequently, alternative C will 'steal' responses from B more often than it will from A. The intuition here is quite clear: when there are two similar items, they tend to split responses between them, while the lone dissimilar item has no competitors that are high on its same attributes and therefore experiences no splitting.

¹⁰ It is worth a quick note that context effects also occur in inferential decisions (Trueblood et al., 2013). The geometric account does not rely on preference-specific mechanisms like loss aversion (Usher & McClelland, 2004) and therefore it is consistent with the observation that these effects occur in both preferential and inferential choice paradigms.

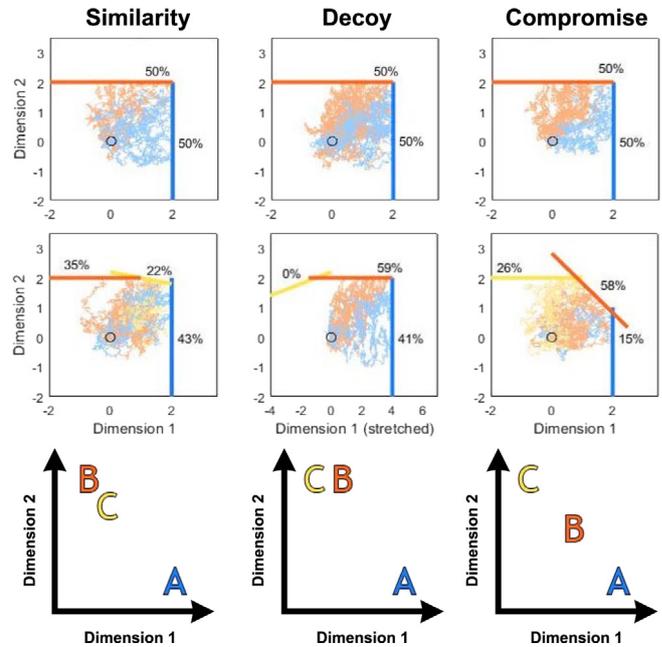


Fig. 11. Summary of geometric explanation for context effects, examining how choice proportions change from binary choice (top row) to ternary choice (middle row). The displayed choice proportions are based on 50,000 simulated trials from the model per plot. Example trajectories (20 in each plot) are color coded based on which choice outcome they triggered, and the values of the alternatives on the two feature dimensions are shown via their location in the bottom panels.

B.0.2. Decoy/attraction

Another observed effect from introducing a third alternative is referred to as the phantom decoy or attraction effect (Huber, Payne, & Puto, 1982; Ratneshwar, Shocker, & Stewart, 1987). In this scenario, the addition of a third option that is dominated by one of the initial alternatives (but not by the other) results in an increased frequency of choices in favor of the alternative that dominates it. When we have alternatives with two features, it should be inferior or equal to one alternative (B) on both features, and superior to the other alternative (A) on one feature.

There are multiple ways of generating the decoy effect in the geometric framework – notably, an alternative that breaks the $r = -1$ correlation between features means that they should actually be represented in three or four dimensions. But for simplicity, let us focus on an attention-based explanation. The diagram for such a setup is shown in the middle panels of Fig. 11: alternative C is added such that it is inferior to alternative B on dimension 1 and equal/inferior on dimension 2. Alternative C is still higher than alternative A on dimension 2. We begin again with alternative A at $v_A = [1, 0]$ and B at $v_B = [0, 1]$, with response rules $x = 2$ and $y = 2$, respectively. The alternative C is added in direction $v_C = [-1/3, 1]$, with response rule given by the orthogonal line $y = x/3 + 2.57$.

The explanation for this effect posited by the model is an attentional one: essentially, adding alternative C crowds the representation of alternatives by occupying space near alternative B, and the participants pay particular attention to details that can distinguish the two. Its effect is that the vectors v_A , v_B , and v_C all have their dimension 1 (x-)values multiplied by a constant. For demonstrative purposes, let us assume this constant is simply 2. This creates the situation shown in the center panel of Fig. 11: the response rule for A is shifted to $x = 4$, while the response rule for B is unaffected (as B has a 0-value along dimension 2).

The sampling distribution should reflect the re-scaled decision space, and so its new direction is $\mu = \text{atan}(4/2) = 0.46$ radians,

or about 27 degrees. This reflects a balanced emphasis on dimension 1 and dimension 2, so as not to claim that the addition of another alternative changes a person's underlying desires to obtain a product that is high on both dimension 1 and dimension 2. Because the threshold is higher for alternative A than alternative B, the noise to threshold ratio is higher for B than it is for A. As a result, the stochastic nature of the evidence accumulation process will cause it to more frequently cross the response boundary for B than the boundary for A even though equal emphasis is placed on the two feature dimensions. This creates the choice proportions shown in the center panel of Fig. 11, which favor alternative B over alternative A when alternative C is added.

Wedell (1991) actually ruled out attention-based accounts based on range and frequency explanations for similarity effects, favoring one where the property of dominating another alternative (C) specifically increases support for a choice option (B) over its non-dominated competitor (A). Such a model could also be implemented in the geometric framework, but this adjustment makes it so that the features of the available alternatives cannot be perfectly (negatively) correlated, meaning the three alternatives would need to be represented in three or more dimensions rather than two. Because this is hard to show visually, it is not displayed here, but such a representation is perfectly reasonable to implement.

B.0.3. Compromise

In our simple example, compromise is essentially a product of a range effect that is introduced by adding an additional alternative. Supposing that alternative A has feature values [2,0] and alternative B has feature values [1,1], they can initially be positioned at $v_A = [1, 0]$ (maximum value on feature 1, minimum value on feature 2) and $v_B = [0, 1]$ (minimum value on feature 1, maximum value on feature 2). As before, a decision-maker desiring high values of feature 1 and feature 2 equally will drift in direction $v_\mu = [1, 1]$, resulting in equal choice proportions between A and B.

Adding alternative C with feature values [0,2] changes the minimum and maximum value along dimension 2 and the minimum value along dimension 1, resulting in a re-positioning of alternative B in the decision space. Consequently, the new vectors for alternatives are $v_A = [1, 0]$, $v_B = [\sqrt{.5}, \sqrt{.5}]$, and $v_C = [0, 1]$. Since the mean accumulation direction is $v_\mu = [1, 1]$, the most frequently chosen alternative will then be alternative B.

B.0.4. Correlations between effects

The astute reader may have noticed that adding alternative C in the decoy paradigm (bottom-middle panel of Fig. 11) should also make alternatives A and B appear more similar on dimension 1. A fair argument could be made that v_B should be therefore be shifted rightward and the angle ϕ_{AB} should shrink. This would make the orientation of the response boundaries look somewhat like those in the compromise effect (middle-right panel). Participants sensitive to the compromise manipulation should therefore be sensitive to the decoy manipulation as well. This prediction is validated in empirical data on context effects, as attraction and compromise effects tend to co-occur (Berkowitsch, Scheibehenne, & Rieskamp, 2014; Trueblood, Brown, & Heathcote, 2015).

Conversely, a tendency to stretch the decision space when a similar alternative is added near one of the alternatives (as in the similarity effect manipulation, right panels of Fig. 11) would reduce the similarity effect. To compensate for the similarity between B and C, a participant might stretch the decision space along dimension 1 to move the alternative vector v_B leftward and make it distinctive relative to v_C . This would mean that the intersection between the response rule for B and the response rule for C would be further away, and C would 'steal' fewer responses

from B. As a result, we might expect that participants prone to re-scaling the decision space to create decoy/compromise effects would show attenuated similarity effects. This precisely what happens in empirical data, where decoy and compromise effects are negatively correlated with the similarity effect (Trueblood et al., 2015).

Put together, the mechanisms of the geometric framework allow for all three types of context effects with minimal tertiary assumptions. While the model presented here is merely a toy model to illustrate the effects, it seems likely that it could be developed into a fully fledged theory of how context affects choice. With an eye on enabling this development, the code for these simulations is provided at osf.io/75qv4/.

References

- Anderson, J. R. (1990). *Cognitive psychology and its implications*. WH Freeman/Times Books/Henry Holt & Co.
- Ashby, F., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2), 154–179.
- Audley, R., & Pike, A. (1965). Some alternative stochastic models of choice 1. *British Journal of Mathematical and Statistical Psychology*, 18(2), 207–225.
- Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*.
- Banerjee, A., Dhillon, I. S., Ghosh, J., & Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research (JMLR)*, 6(Sep), 1345–1382.
- Baum, C. W., & Veeravalli, V. V. (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6).
- Berkowitsch, N. A., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, 143(3), 1331–1348.
- Best, D., & Fisher, N. I. (1979). Efficient simulation of the von Mises distribution. *Applied Statistics*, 152–157.
- Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, 124(1), 1–20.
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, 115(2), 463–501.
- Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, 11(3), 118–125.
- Bogacz, R. (2009). Optimal decision-making theories. In *Handbook of reward and decision making* (pp. 373–397). Elsevier.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700–765. <http://dx.doi.org/10.1037/0033-295X.113.4.700>.
- Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, 19(2), 442–477.
- Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, 362(1485), 1655–1670.
- Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *The Journal of the Acoustical Society of America*, 51(2B), 483–502.
- Brown, S. D., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, 112(1), 117–128. <http://dx.doi.org/10.1037/0033-295X.112.1.117>.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. <http://dx.doi.org/10.1016/j.cogpsych.2007.12.002>.
- Brown, S. D., Marley, A., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, 115(2), 396–425. <http://dx.doi.org/10.1037/0033-295X.115.2.396>.
- Bussemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. New York, NY: Cambridge University Press.
- Bussemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, 43(3), 345–370. [http://dx.doi.org/10.1016/S0165-4896\(02\)00016-1](http://dx.doi.org/10.1016/S0165-4896(02)00016-1).
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459. <http://dx.doi.org/10.1037/0033-295X.100.3.432>.
- Cunha Jr, M., & Shulman, J. D. (2010). Assimilation and contrast in price evaluations. *Journal of Consumer Research*, 37(5), 822–835.

- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.
- Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, 47(3), 304–322. [http://dx.doi.org/10.1016/S0022-2496\(03\)00003-8](http://dx.doi.org/10.1016/S0022-2496(03)00003-8).
- Dodds, P., Rae, B., & Brown, S. (2012). Perhaps unidimensional is not unidimensional. *Cognitive Science*, 36(8), 1542–1555.
- Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2(2), 312–329. [http://dx.doi.org/10.1016/0022-2496\(65\)90007-6](http://dx.doi.org/10.1016/0022-2496(65)90007-6).
- Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing*, 12–25.
- Fisher, R. (1953). Dispersion on a sphere. In *Proceedings of the royal society of London A: Mathematical, physical and engineering sciences: Vol. 217* (pp. 295–305). The Royal Society.
- Furnas, G. W., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., & Lochbaum, K. E. (1988). Information retrieval using a singular value decomposition model of latent semantic structure. In *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 465–480). ACM.
- Gigerenzer, G., Todd, P. M., & A. B. C. Research Group (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162.
- Hawkins, G. E., Marley, A., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014a). The best of times and the worst of times are interchangeable. *Decision*, 1(3), 192–214.
- Hawkins, G. E., Marley, A., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014b). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science*, 38(4), 701–735.
- 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.
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1), 90–98. <http://dx.doi.org/10.1086/208899>.
- Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, 64(6p1), 384–404.
- Hutchinson, J. (1983). On the locus of range effects in judgment and choice. *Advances in Consumer Research*, 10(1), 305–308.
- Itô, K. (1974). *Diffusion processes*. Wiley Online Library.
- Janiszewski, C., & Lichtenstein, D. R. (1999). A range theory account of price perception. *Journal of Consumer Research*, 25(4), 353–368.
- Julesz, B. (1971). *Foundations of cyclopean perception*. University of Chicago Press.
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5), 679–685.
- Kent, J. T. (1982). The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 71–80.
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103(10), 3863–3868.
- Kvam, P. D. (2019). Modeling accuracy, response time, and bias in continuous orientation judgments. *Experimental Psychology: Human Perception and Performance*, 45(3), 301–318.
- Kvam, P. D., Cesario, J., Eisthen, H., Schossau, J., & Hintze, A. (2015). Computational evolution of decision-making strategies. In *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 1225–1230).
- Kvam, P. D., & Hintze, A. (2018). Rewards, risks, and reaching the right strategy: evolutionary paths from heuristics to optimal decisions. *Evolutionary Behavioral Sciences*, 12(3), 177–190.
- LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, 27(4), 375–396.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211.
- Lerdahl, F. (2004). *Tonal pitch space*. Oxford University Press.
- Link, S., & Heath, R. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40(1), 77–105. <http://dx.doi.org/10.1007/BF02291481>.
- Lockhead, G. R. (2004). Absolute judgments are relative: A reinterpretation of some psychophysical ideas. *Review of General Psychology*, 8(4), 265–272.
- Luce, R. D. (1986). *Response times: Vol. 8*. Oxford University Press.
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432–1438.
- Mardia, K. V. (2014). *Statistics of directional data*. Academic press.
- Marley, A., & Cook, V. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology*, 37(2), 136–151.
- Marley, A. A., & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology*, 49(6), 464–480.
- Mellers, B. A., & Cooke, A. D. (1994). Trade-offs depend on attribute range. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1055.
- von Mises, R. (1918). Über die "Ganzzahligkeit" der Atomgewichte und verwandte Fragen. *Physikalische Zeitschrift*, 19, 490–500.
- Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *Journal of Neuroscience*, 28(17), 4435–4445.
- Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. In A. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 347–365). Mahwah, NJ: Lawrence Erlbaum.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. *Trends in Neurosciences*.
- O'Toole, A. J., Jiang, F., Abdi, H., Pénard, N., Dunlop, J. P., & Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of Cognitive Neuroscience*, 19(11), 1735–1752.
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5(5), 1. <http://dx.doi.org/10.1167/5.5.1>.
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864–901. <http://dx.doi.org/10.1037/A0019737>.
- Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, 143(5), 2000–2019. <http://dx.doi.org/10.1037/xge0000013>.
- Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review*, 120(3), 679–696. <http://dx.doi.org/10.1037/a0033142>.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <http://dx.doi.org/10.1037/0033-295X.85.2.59>.
- Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review*, 125(6), 888–935. <http://dx.doi.org/10.1037/rev0000117>.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. <http://dx.doi.org/10.1162/neco.2008.12-06-420>.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333–367. <http://dx.doi.org/10.1037/0033-295X.111.2.333>.
- 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. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116(1), 59–83. <http://dx.doi.org/10.1037/a0014086>.
- 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. <http://dx.doi.org/10.1037/a0033152>.
- Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research*, 13(4), 520–533.
- Reimann, M. W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., Dlotko, P., Levi, R., Hess, K., & Markram, H. (2017). Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, 11, 48.
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108(2), 370–392. <http://dx.doi.org/10.1037/0033-295X.108.2.370>.
- Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, 80(2), 491–513.
- Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109(42), E2904–E2913.
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27(3), 219–246. <http://dx.doi.org/10.1007/BF02289621>.
- Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305–333. <http://dx.doi.org/10.1037/0033-295X.89.4.305>.
- Sjöberg, L. (1977). Choice frequency and similarity. *Scandinavian Journal of Psychology*, 18(1), 103–115.
- Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123(4), 425–451. <http://dx.doi.org/10.1037/rev0000023>.

- 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*, <http://dx.doi.org/10.3758/s13423-018-1491-0>.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168.
- 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. <http://dx.doi.org/10.1348/000711000159349>.
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168. [http://dx.doi.org/10.1016/0022-2496\(88\)90043-0](http://dx.doi.org/10.1016/0022-2496(88)90043-0).
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260.
- Stroock, D. W., & Varadhan, S. (2007). *Multidimensional diffusion processes*. Springer.
- Tkalcic, M., & Tasic, J. F. (2003). *Colour spaces: perceptual, historical and applicational background: Vol. 1*. IEEE.
- Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*(2), 179–205. <http://dx.doi.org/10.1037/a0036137>.
- Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, *122*(4), 848–853.
- Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, *24*(6), 901–908.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.
- Turner, B. M., & Sederberg, P. B. (2012). Approximate Bayesian computation with differential evolution. *Journal of Mathematical Psychology*, *56*(5), 375–385.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. <http://dx.doi.org/10.1037/h0032955>.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. <http://dx.doi.org/10.1037/0033-295X.84.4.327>.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. <http://dx.doi.org/10.1037/0033-295X.108.3.550>.
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769. <http://dx.doi.org/10.1037/0033-295X.111.3.757>.
- Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology*, *46*(6), 704–715. <http://dx.doi.org/10.1006/jmps.2002.1420>.
- Vickers, D. (1979). *Decision processes in visual perception*. Academic Press.
- Wald, A., & Wolfowitz, J. (1949). Bayes solutions of sequential decision problems. *Proceedings of the National Academy of Sciences of the United States of America*, *35*(2), 99–102.
- Watson, G. S., & Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, *43*(3/4), 344–352.
- Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 767–778.
- Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, *64*(1), 641–654.
- Wollschläger, L. M., & Diederich, A. (2012). The 2N-ary choice tree model for N-alternative preferential choice. *Frontiers in Psychology*, *3*.
- Wollschläger, L. M., & Diederich, A. (2017). A computational model for constructing preferences for multiple choice options. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 1351–1356). Austin, TX: Cognitive Science Society.
- Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences*, *111*(7), 2848–2853.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.