Title

Geometric representations of evidence in models of decision-making

Permalink

https://escholarship.org/uc/item/01z86463

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors

Kvam, Peter D.
Favela, Luis H.

Publication Date

2016

Peer reviewed

Geometric representations of evidence in models of decision-making

Peter D. Kvam (kvampete@msu.edu)
Department of Psychology, Michigan State University
316 Physics Rd, East Lansing, MI 48824 USA

Abstract

Traditionally, models of the decision-making process have focused on the case where a decision-maker must choose between two alternatives. The most successful of these, sequential sampling models, have been extended from the binary case to account for choices and response times between multiple alternatives. In this paper, I present a geometric representation of diffusion and accumulator models of multiple-choice decisions, and show how these can be analyzed as Markov processes on lattices. I then introduce psychological relationships between choice alternatives and show how this impacts the sequential sampling process. I conclude with two examples showing how one can predict distributions of responses on a continuum as well as response times by incorporating psychological representations into a multi-dimensional random walk diffusion process.

Keywords: random walk; decision making; evidence accumulation; multi-alternative; continuous response

Introduction

Traditionally, models of the decision-making process have focused on the case where a decision maker has two options between which to choose. By far the most empirically successful accounts of this process are sequential sampling models, which are able to reproduce decision makers' observed choice proportions and distributions of response times with high fidelity.

Sequential sampling models can be broadly broken down into two categories: diffusion models and accumulator models. The former type of model posits that a person represents the information or evidence they have regarding the decision as a balance between the two choice alternatives. In inferential decisions, this balance can be formally represented as the log odds of one alternative (hypothesis) relative to another, requiring that the sum of the log odds for all hypotheses sum to zero (i.e. so that the probabilities of all hypotheses sum to one). A decision is triggered when the log odds for one hypothesis over the others exceeds a critical threshold value.

Accumulator models, by contrast, represent information or evidence for the alternatives as separate quantities [accumulators]. As a decision maker gathers or receives information, each accumulator is updated, and a decision is triggered once one of the accumulators reaches a critical threshold value. However, each of the accumulators is usually assumed to be independent such that incrementing one's value does not affect another's.

More recently, these models have been extended to account for decisions between multiple alternatives. In the case of models where evidence for alternatives is represented as multiple independent accumulators (one may think of these as 'pure' accumulator models), this is straightforward – one need only add an additional accumulator for each additional choice alternative. A decision is still triggered once any of these accumulators reach a critical threshold, so this addition can be done ad infinitum, or at least until a modeler runs out of computational resources.

For diffusion models, adding alternatives is only slightly more complicated. In a 3-alternative case, rather than decreasing the log odds of B by 1 every time the log odds of A increase by 1, one would have to decrease the log odds of B and C each by 1/2 when the log odds of A increase by 1. This ensures that the log odds across all hypotheses sums to zero / probabilities sum to one. As more alternatives are added, this decrement decreases so that incrementing evidence for a single alternative by 1 also decrements all other alternatives by $\frac{1}{n-1}$. The same decision rule applies – once the (log odds) evidence in favor of one alternative exceeds a critical threshold, a decision is triggered.

Although assuming independence of accumulators or at least a uniform distribution of negative evidence are convenient simplifying assumptions, they are perhaps unrealistic. A substantial body of literature has suggested that there are often strong and unbalanced interactions between different pairs of items in a set. For example, context effects arising from the inclusion of a third alternative – such as decoy, compromise, and similarity effects – substantially alter choices between an original set of two (see e.g. Trueblood et al., 2014). Similarly, in absolute identification tasks, adjacent categories (e.g. 50-60 and 60-70) interact more strongly than non-adjacent ones (50-60 and 80-90) (see Brown et al., 2008).

Models of decision making have been modified in a number of ways to account for these phenomena. For example, decision field theory (Busemeyer & Diederich, 2002) introduced an additional step in the decision process where pairs of items are contrasted against one another before computing accumulator values. The leaky competing accumulator model (Usher & McClelland, 2001, 2004) introduced competition and loss aversion to a similar effect, the multi-attribute linear ballistic accumulator model (Trueblood et al., 2014) includes pairwise comparisons as well as subjective attribute values, and the selective attention, mapping, and ballistic accumulation model (Brown et al., 2008) specifies and utilizes adjacency between categories to define the evidence accumulation process for separate accumulators. Similarly, models of confidence (see Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009) specify adjacency of judgment categories using ordinal states or accumulators.

What all of these models have in common is that they specify *psychological relationships* between alternatives in a choice set. Indeed, Trueblood et al. (2014) note explicitly that decisions are made between the psychological represen-

tations of alternatives rather than physical ones, and that this is the key component allowing the different models to account for context effects. These models each add components in order to avoid making the simplifying assumptions of independence and uniform negative evidence. Instead, they suggest that evidence for alternative A may also be evidence for alternative B and simultaneously be strong evidence against alternative C.

In the next section, I examine how the psychological relationships between alternatives can be thought of as geometric relations in a psychological space.

Geometry of evidence accumulation

In order to introduce a geometric representation of the decision process, I return first to the simple binary cases. Using the binary choice diffusion model, one can establish a rule to construct models relating accumulator and diffusion models, construct geometric models of evidence accumulation among multiple alternatives, and in turn derive a method for modeling evidence accumulation when the number of alternatives is very large or continuous (infinite).

Equal relative evidence models

The basic uni-dimensional diffusion model (Ratcliff & McKoon, 2008) originally described the behavior of particles along a single dimension in space. Its natural geometric analogue is a random walk on a line (Figure 1A). In inference tasks, this single dimension may correspond to the log odds of one hypothesis (H_1) relative to another (H_2).

Put simply, the closer a person's representation of evidence (or preference) is to a boundary corresponding to a choice alternative, the more they currently favor that alternative. Evidence that provides support for an alternative H_1 moves a person's state orthogonally toward the boundary corresponding to H_1 in direction D_1 and away from the boundary H_2, which is in direction D_2. This gives rise to a spatial relationship between the amount of evidence (Ev) a particular piece of evidence supporting one hypothesis H_1 provides for another hypothesis H_2:

$$Ev(H_2) \leftarrow Ev(H_1)cos(\angle D_1 D_2) \qquad (1)$$

In the two-alternative diffusion model, the evidence favoring alternative H_1 must move the state in the opposite direction from evidence favoring alternative H_2, so that $cos(\angle D_1 D_2) = -1$.

In order to obtain the case for three alternatives, where evidence for H_1 decreases the log odds of H_2 and H_3 equally, it must be the case that $cos(\angle D_1 D_2) = cos(\angle D_1 D_3) = -\frac{1}{2}$ (and of course it will also be the case with $cos(\angle D_2 D_3)$ to conserve total log odds).[1] This results in an evidence accumulation process that unfolds in a plane, contained within an

___
[1]Note that this approach offers an alternative solution to the relative-accumulator problem encountered by Nosofsky (1997), where there was more negative evidence than positive evidence added across accumulators if increments and decrements were restricted to values of one.

equilateral triangle (Figure 1B). A decision is triggered when a state crosses one of the sides of the triangle, each of which corresponds to choosing one of the alternatives (see Laming, 1968, for a similar proposal).

Extending this strategy to model decisions between any number of alternatives is relatively straightforward. In order to account for decisions where there are n alternatives, one must create a situation where there are n directions $\{D_1, D_2, ..., D_n\}$ satisfying the property $cos(\angle D_i D_j) = -\frac{1}{n-1}$, $i \neq j$, so evidence for any alternative provides evidence against all others equally. In the case of 4 alternatives, the boundaries corresponding to $H_1 - H_4$ would each be a plane in a 3-dimensional space, together forming a tetrahedron (Figure 1C), and evidence accumulation will unfold in a 3-dimensional space.

In order to accommodate n alternatives, this would naturally be extended to permit evidence accumulation in $(n-1)$ dimensions. The state would exist in the interior of a simplex, with the choice boundaries corresponding to each of its $(n-2)$-dimensional facets.

It is worth a note that the cosine relation in Equation 1 will preserve log odds in any n-dimensional space by virtue of every integral $\int \cdots \int_0^{2\pi} cos = 0$. However, later examples where decision bounds do not form regular figures show that log odds are preserved across the *theoretically possible* space of alternatives but not necessarily across all available ones.

Absolute evidence models

While it is often practically unnecessary to envision accumulator models in a geometric way, doing so illustrates the psychological assumptions that go into these models and allows them to be analyzed as a random walk. In the two alternative case, evidence in favor of H_1 provides no information regarding H_2. Using the relationship defined by equation 1, this means that the directions corresponding to each alternative must be orthogonal, $cos(\angle D_1 D_2) = 0$. The choice boundaries therefore form two sides of a rectangle (Figure 1D).

However, the evidence state does not immediately have a clear log odds interpretation as it did in the diffusion models. One could potentially address this by assuming that there are two theoretical alternatives in directions $-D_1$ and $-D_2$ (if D_1 and D_2 are given by vectors) and anchor log odds to be zero at the initial state. This would allow computation of relative log odds of the hypotheses in the case a person wanted to make a relative judgment of the two alternatives (e.g. preference or confidence). However, doing so is not necessary for predicting choices and response times.

Extending accumulator models to three or more alternatives is relatively straightforward. One need only add additional, orthogonal dimensions to the evidence accumulation space, then set the choice criterion and new direction D_n for each new alternative as orthogonal to existing ones. In the case of three alternatives, this would yield a figure bounded by three choice criteria constituting sides of a cube (Figure 1E). For n alternatives, the orthogonal choice criteria would compose a set of intersecting facets of an n-cube.
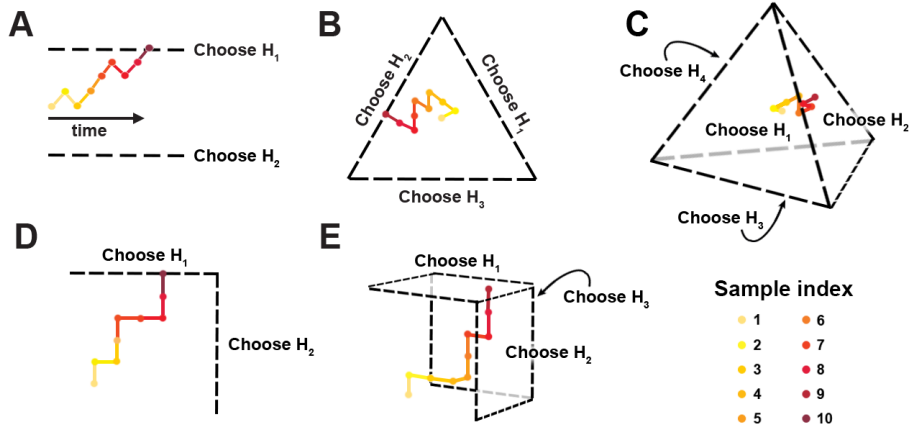
Figure 1: Representation of a person's accumulated evidence and choice criteria for 2-, 3-, and 4-alternative diffusion processes (A, B, C) and 2- and 3-alternative accumulator processes (D, E). An alternative is chosen when a person's represented evidence (yellow / red) crosses the corresponding edge (for models A, B, and D) or face (for models C and E).

Random walks

This geometric framework lends it self to both discrete-time and continuous-time as well as discrete-space and continuous-space random walk representations. A discrete-space structure can be given to the models I have described so far by imposing a lattice structure upon them.

For example, suppose we are interested in using an equal relative evidence diffusion model to describe how a person sorts a color stimulus into three categories: red, blue, or green. This corresponds to the triangular structure shown in Figure 1B. In order to produce a discrete random walk in this space, one can construct a triangular lattice bounded by the three choice criteria.

In this case, a person's internal representation of the stimulus in terms of the log odds of the three hypotheses (red, green, blue) corresponds to their position on the lattice. Initially, they might start out in the middle (unbiased / 0), but as they view the stimulus they should sample pieces of evidence that favor green, red, or blue, causing them to step at angles $\frac{\pi}{6}$, $\frac{5\pi}{6}$, or $\frac{9\pi}{6}$ radians on the lattice. The probability of taking a step in each direction is given by p (for green), q (for red), and $1 - p - q$ (for blue, to sum to one). The edges of the lattice defined by choice criteria consist of absorbing states; upon entering one of these states, the person halts the transition process and selects the corresponding alternative.

This state transition process can be represented as a Markov chain much as the 2-alternative diffusion model is (see Diederich & Busemeyer, 2003), with the caveat that each state has three rather than two transition destinations. It can be implemented as a continuous-time random walk by introducing the standard exponentially distributed transition time (requiring one additional sampling rate parameter), allowing prediction of choice probabilities as well as response times for each of the three possible choice alternatives.

Both the relative and absolute evidence models with any number of alternatives can be described in a similar way. The
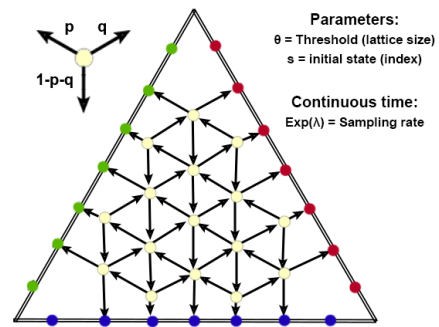

Figure 2: Lattice and state transition structure for a 3-alternative, relative evidence model of color identification.

lattice shape will change and the number of transition destinations at each step will grow along with the number of alternatives, but the sampling rate and thresholds operate in a similar way.[2]

Note that a person's representation will only be able to step toward available alternatives (not directly away from them) – in Figure 2, this is indicated by unidirectional transitions. Similarly, in accumulator models like the one shown in Figure 1D, the representation will only be able to step to the right or upward. It is perhaps worth noting that the transition probabilities and sampling rate in such a case can be obtained from the accumulation rates of independent Poisson accumulators (as in Smith & Van Zandt, 2000), making it possible to swap between the separate race and single geometric random walk process representations (see, e.g., Ross, 2014).

Such a random walk still requires an initial state. Bias can be introduced by moving the initial state closer to one boundary or another, and starting point variability (required for pre-

---

[2] Indeed, a fixed sampling rate but a larger number of alternatives will yield slower response times, which may give rise to Hick's Law (Hick, 1952)

dicting fast errors) can be introduced by specifying a mixed state across starting positions.

Psychological spaces

Thus far I have focused mainly on the geometric structure of models where evidence for one alternative has no net effect on the evidence balance between other alternatives. However, this is often an unrealistic assumption. Returning to color categorization, suppose that a participant must match a stimulus to one of 4 categories: red, yellow, green, or blue. One might expect that a stimulus emitting light peaking at a wavelength 610 nm (orange) would provide evidence in favor of both "red" and "yellow" responses, but provide evidence against a "blue" or a "green" response.

In such a case, it makes little sense to treat red, yellow, blue, and green categories as independent or equally related alternatives. Instead, they must be related to one another by constructing a psychological space describing the cognitive representations of the stimulus and choice alternatives.

Doing so requires two generalizations of the framework described in the previous section. First, the directions corresponding to alternatives are permitted to vary. There are several ways to do so. For example, they could be released to vary as free parameters – in 2 dimensions, this could simply be the angle relative to a reference direction, though this would require more parameters when moving to 3 or more dimensions. Alternatively, the directions could be set a priori by the modeler. This could be done by using the physical characteristics of the stimulus, by using existing similarity judgments, or by using existing psychological theory. I give examples of these approaches in the next section.

Second, a person's representation of evidence is modified to reflect the type of information they gather from the stimulus relative to the type of responses they can make. For example, if a person is trying to reproduce the orientation of a stimulus that can vary anywhere from 0 to π radians, they must be able to sample and represent information that favors any direction between 0 and π radians. This will often mean that discrete-state Markov chain representations of the evidence accumulation process are no longer possible, except as approximations or in the rare case that the choice alternatives are arranged in a psychological space so that their orientations allow for a convenient lattice to be superposed upon them.

Random walks in psychological space

Instead of the discrete-state Markov chain, evidence accumulation in a psychological space is enabled by utilizing a multidimensional random walk. In this framework, a person's cognitive state representing evidence they have gathered from a stimulus is described by a point in a multidimensional (e.g. feature-based) space. As a person integrates a new piece information, this state representation is updated by drawing a random variable ϕ that describes a direction in the space and moving one unit of distance in that direction.

The distribution of ϕ is determined by the stimulus and the psychological space in which it is represented. The arrival

time of each piece of evidence is again described by an exponential distribution, $Exp(\lambda)$. As before, once the state representation crosses one of the boundaries corresponding to a choice alternative, that alternative is chosen, yielding a choice and response time.

Examples

It is perhaps helpful to visit some examples of how to construct a psychological space and sampling distribution. For each example, we must consider several important factors: what type of response is being elicited (are they discrete / continuous, time-dependent?); how these responses relate to one another in terms of their psychological representation (are they similar, what physical dimensions overlap, how are they processed at a sensory level?); and how a person gathers and represents information about the stimulus relative to the response options (how much evidence does percept X provide for alternative Y?). I do so for two cases: an orientation estimation task where responses can fall anywhere along a continuum and a color identification task where participants must match a color stimulus to a set of categories.

Orientation task

Perceptual tasks often lend themselves to straightforward psychological representations. A common stimulus used to examine how people (as well as other animals) perceive the orientation of objects in the world is the Gabor patch (Figure 3) (see, e.g., Wilson et al., 1983). They may be used in tasks where participants must distinguish one from a set of stimuli, decide between left- or right-leaning orientation, or be combined with color or motion. We examine the simple case where a participant in a task must reproduce the orientation of a Gabor patch stimulus.
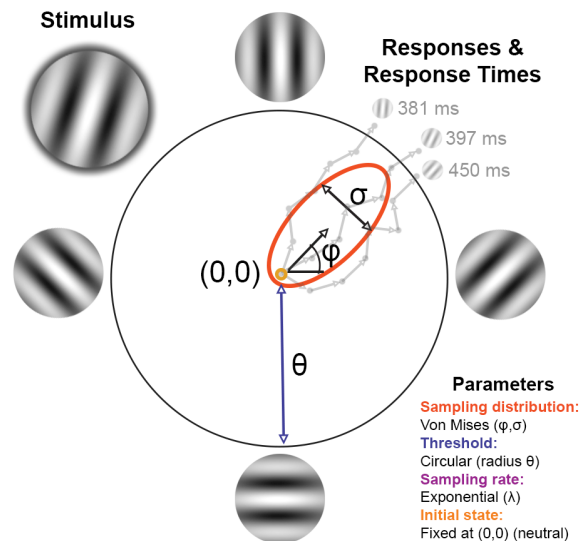

Figure 3: State representation, sampling, threshold, and evidence trajectories for the orientation matching task.

Because these stimuli are symmetric across horizontal and vertical axes, meaning they have no top and bottom, the orientations of these stimuli and the possible responses to them vary from 0 to π radians. For simplicity, assume that stimuli at orthogonal rotations (i.e. $\frac{\pi}{2}$ / $\frac{3\pi}{2}$ or horizontal / vertical) provide evidence against one another. Therefore, they can be arranged in a circle as shown in Figure 3.

In order to complete the task, a participant must gather pieces of information from the stimulus or from memory (if the stimulus has been removed and masked) to construct a cognitive representation that they match to the possible response options. Each piece of information they gather is assumed to be pulled from a von Mises distribution.[3] They sample information until a criterion level of certainty is met, given by the circular threshold, and the point at which the walk crosses the threshold gives the orientation response.

The model still requires additional assumptions regarding the initial representation of the stimulus (before information is gathered) and specification of the sampling rate. For the former, assume a neutral or unbiased initial state, indicating that a participant believes that all orientations are equally likely a priori. Of course, this assumption could be modified if the generating stimuli were not evenly distributed across orientations or if a person was believed to be biased toward (e.g.) a "vertical" response.

The mean of the von Mises sampling distribution (ϕ) can be set directly from the stimulus, given that the true orientation is the most likely to be sampled. Then the error, threshold, and sampling rate are free to vary and can be estimated. In the simple case of diffusion in 2 dimensions with a von Mises sampling distribution and circular decision bounds, first passage times can be derived analytically (see Smith, 2016, for a thorough discussion of this matter). However, in more complex cases with different assumptions regarding non-decision time, functional forms of starting point variability, decision bounds, or trial-to-trial parameter variability, the model may need to be simulated in order to compute the likelihood and/or require approximate fitting criteria such as joint quantiles (as in Heathcote et al., 2002).

Color identification task

I return now to the color identification task, where a participant must classify a fixed-luminosity stimulus as red, green, yellow, or blue. Each category corresponds to a direction D_R, D_G, D_Y, or D_B.

The particular direction corresponding to each color may be fixed or free to vary. One of the benefits of taking the approach presented here is that each direction can in fact be fit by imbuing it with a free parameter or fixed based on a particular theory the modeler has in mind. This allows for model comparison between 2-dimensional representations

based purely on hue (Figure 4A), Hering's opponent-process representation (Figure 4B), a multi-dimensional scaled space representation (as in Shepard, 1962)(Figure 4C), or a situation where each of the directions is given by an additional parameter of the model.

Each of these theories makes very different predictions about how stimuli and responses should interact. For example, a stimulus showing light at 430 nm (violet) should result in very long response times in hue or multidimensional scaling space, but not in an opponent-process theory space. Similarly, the multidimensional scaling space predicts far more green-yellow and green-blue confusions than do hue or opponent process constructions. Because of this, the pattern of responses from a decision task can actually inform and distinguish between the theories used to construct the space, or even develop better theories when parameters giving D_R, D_G, D_Y, and D_B are estimated.

As with the orientation task, a neutral initial state, von Mises sampling distribution (or the von Mises-Fisher in a higher-dimensional representation), and exponential arrival times for new samples are reasonable assumptions for how a person's cognitive representation of the stimulus changes over time. These can of course be modified if the base rates of different colors or stimulus information change.

Discussion

One may wonder what the point of modeling different tasks in this way is. The answer is many-fold. One potential use is as a measurement model – the parameters of the model are psychologically meaningful, with sampling rate describing how quickly a person can gather new information, threshold describing the strictness with which they make decisions, and the amount of error in drift or starting point location in the system (e.g., in a von Mises distribution or mixed initial state) allows us to potentially diagnose sources of bias and inaccuracy as with the uni-dimensional diffusion model (Ratcliff & McKoon, 2008; Pleskac & Busemeyer, 2010).

Second, modeling decisions among many alternatives in this way allows for testable predictions and model comparison. For example, without starting point variability, the models presented in the examples predict that response times should be longer for larger errors. This means that one would expect an incidental "red" response to a blue stimulus or a left-leaning response to a right-leaning stimulus to take longer than more similar responses like "green" or vertical-leaning (respectively).

Furthermore, psychological representations of both stimuli and potential responses can be informed by established models of how people represent objects and concepts. Similarity judgments or categorization data can be used to construct a multidimensional feature space that relates available choice alternatives, in turn making a priori predictions about response and response time distributions about decisions among them. Similarly, neural data about activity during decision-making can be connected to the psychological

---

[3]This is similar to a normal distribution on a circle. Each piece of evidence could be sampled from momentary activation across orientation columns in the visual cortex. This opens up the question of whether activation across the columns mimics a von Mises distribution, which is interesting but beyond the scope of the paper.
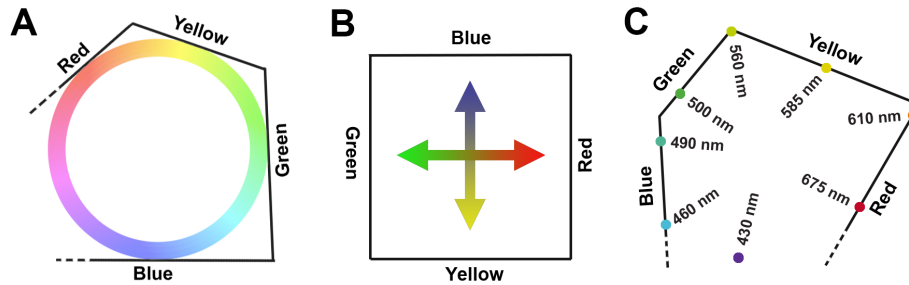
Figure 4: Accumulation space representation for 4-category decision based on hue (A), visual opponent-process theory (B), and Shepard (1962)'s multidimensional scaling construction of color (C).

representations of the stimuli and response alternatives.

When put together, this approach may also provide natural and parsimonious explanations of existing psychological phenomena, such as context effects (Trueblood et al., 2014), Hick's law (Hick, 1952), selection of prices and certainty equivalents (Busemeyer & Diederich, 2002), confidence and confidence response times (Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009), and bow effects and lateral interactions in absolute identification (Brown et al., 2008).

In short, I hope that the approaches and models presented here are able to inform our understanding of the decision process. At best, they should open up new questions as well as bring together different perspectives and sources of data on how to represent and process information. At worst, they should at least provide a first approach toward modeling how responses and response times arise when decisions are made among many alternatives or along a continuum.

Acknowledgments

References

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.

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*(3), 345–370.

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.

Heathcote, A., Brown, S., & Mewhort, D. J. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic bulletin & review*, *9*(2), 394–401.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Laming, D. R. J. (1968). *Information theory of choice-reaction times.* Academic Press.

Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. *Choice, decision, and measurement: Essays in honor of R. Duncan Luce*, 347–365.

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.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*(1), 59-83.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. ii. *Psychometrika*, *27*(3), 219–246.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*.

Smith, P. L., & Van Zandt, T. (2000). Time-dependent poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, *53*(2), 293–315.

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.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, *108*(3), 550–592.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review*, *111*(3), 757–769.

Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision research*, *23*(9), 873–882.