

MODELING DECISIONS AMONG MANY ALTERNATIVES

By

Peter Kvam

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Doctor of Philosophy

2017

ABSTRACT

MODELING DECISIONS AMONG MANY ALTERNATIVES

By

Peter Kvam

Many of the actions we take depend on being able to make selections among many alternatives or even along a continuum. However, our understanding of the decision processes underlying these selections is sparse, largely due to a traditional focus on developing models of binary decisions. Recent forays into modeling multi-alternative decisions have been forced to build in relations between representations of available alternatives. In this paper, I propose and test a general framework for modeling decisions between arbitrarily large numbers of alternatives that naturally incorporates psychological relationships between the representations of available alternatives. In the first study, I construct and evaluate the basic components of a model of this process by establishing benchmark empirical phenomena for decisions on a continuum. In the second study, I examine how the number of alternatives and the relations between them affect representations and the components of the decision process. Taken together, this paper establishes benchmark empirical results in a new choice domain (continuous selection), proposes and tests a new modeling framework that accounts for these phenomena, and brings together decision and representation models in order to develop an over-arching theory of how people make decisions among many alternatives.

ACKNOWLEDGEMENTS

The author would like to thank Timothy Pleskac, Susan Ravizza, Andrew Heathcote, Philip Smith, Taosheng Liu, Zach Hambrick, and David Johnson for helpful comments on the manuscript or the ideas contained herein. The author of the thesis was supported by a fellowship from the National Science Foundation (Grant No. 1424871) while pursuing this work.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1 INTRODUCTION	1
1.1 Benchmark phenomena	2
1.2 Relative evidence models	5
1.3 Absolute evidence models	8
1.4 Multiple alternatives	11
1.5 Geometric representations	13
1.5.1 Relative evidence	13
1.5.2 Absolute evidence	16
1.6 Random walks	17
1.7 Psychological spaces	20
1.7.1 Random walks in psychological space	21
1.8 Hick's Law	23
CHAPTER 2 GOALS AND PREDICTIONS	27
2.1 Study 1	28
2.2 Study 2	29
CHAPTER 3 STUDY 1 - EMPIRICAL HURDLES AND MODEL ELEMENTS	31
3.1 Methods	32
3.1.1 Task	33
3.1.2 Participants	35
3.1.3 Materials	35
3.1.4 Procedure	35
3.2 Results	36
3.2.1 Response times	37
3.2.2 Accuracy / response deviation	39
3.3 Modeling	41
3.3.1 Parameter-free predictions	44
3.3.1.1 Homogeneous response times	44
3.3.1.2 Von Mises responses	45
3.3.2 Free effects model	46
3.3.3 Simplified cue-inclusive model	48
3.4 Preliminary discussion	51
CHAPTER 4 STUDY 2 - EXPLORING RELATIONS BETWEEN ALTERNATIVES	53
4.1 Methods	54
4.1.1 Participants	54
4.1.2 Materials	54

4.1.3	Decision task	54
4.1.3.1	Practice trials	57
4.1.4	Similarity rating task	58
4.1.5	Procedure	59
4.2	Results	60
4.2.1	Number of alternatives	60
4.2.2	Conflation with difficulty	63
4.2.3	Multidimensional scaling	65
4.2.4	Summary of important effects	68
4.3	Modeling	69
4.3.1	Response boundaries	71
4.3.2	Discriminability	72
4.3.3	Changes in threshold	74
4.3.4	Model fitting and results	75
4.3.4.1	Posterior Predictions	79
4.4	Preliminary discussion	84
CHAPTER 5 GENERAL DISCUSSION		86
5.1	Limitations	87
5.2	Extensions	89
5.3	Conclusions	91
BIBLIOGRAPHY		93

LIST OF TABLES

Table 3.1	Mean estimates of coefficients for main effects and interactions of stimulus difficulty, speed manipulation, cue presence, and cue orientation on response times. The ranges containing the 95% Highest Density Interval (HDI) are also provided. Intervals excluding zero are starred.	38
Table 3.2	Mean estimates of coefficients for main effects and interactions of stimulus difficulty, speed manipulation, cue presence, and cue orientation on accuracy of responses. The ranges containing the 95% Highest Density Interval (HDI) are also provided. Intervals excluding zero are starred.	41
Table 3.3	Maximum likelihood estimates for the parameters of the restricted model.	49
Table 4.1	Observed results and the model components that can account for them. A check mark indicates that the model component can account for an effect, a blank means it cannot, and an X means it can produce the opposite effect.	75
Table 4.2	Bayesian information criterion (a function of the log likelihood with a penalty for the number of parameters; see Schwarz, 1978) for each model tested. The model with the lowest BIC is preferred and shown in bold.	77
Table 4.3	Maximum likelihood parameter estimates for the BIC-favored model of the data from Study 2.	78

LIST OF FIGURES

Figure 1.1	Diagram of three basic elements of the diffusion model: drift, diffusion, and threshold.	6
Figure 1.2	Diagram of basic elements of an accumulator model: starting point variability, multiple accumulators (with drifts), and threshold. As shown, drift can be represented as a function of sample arrival rate. This allows the accumulator model to also be described in terms of sampling rate and probability of sampling in favor of option A (or any other alternative).	10
Figure 1.3	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).	14
Figure 1.4	Lattice and state transition structure for a 3-alternative, relative evidence model of color identification.	19
Figure 1.5	Predicted response time proportional to the number of bits required to discriminate between alternatives. It is unclear what the response time prediction should be in terms of bit-based information for the continuous case, as Hick's law may not apply to this condition.	24
Figure 3.1	Time course of a trial of the orientation detection task.	34
Figure 3.2	Response times (left) and accuracy in terms of degrees deviation from the correct response (right) across difficulty levels and speed-accuracy manipulation. Note that higher values indicate less accurate responses. Error bars indicate pooled standard error across participants.	39
Figure 3.3	Mean response times (left) and absolute response deviation (right) by the orientation of the cue. Note that higher response deviations indicate less accurate responses. Error bars indicate pooled standard error across participants.	40
Figure 3.4	State representation, sampling, threshold, and evidence trajectories for the orientation matching task.	42
Figure 3.5	Drift magnitude (rate), threshold, and non-decision time estimates for individuals (colors) and the mean of the group (black) in the free effects model.	47
Figure 3.6	Effects of cue, speed-accuracy, and difficulty manipulations on drift magnitude, direction, and threshold for individuals (colors) and group mean effects (black).	50

Figure 4.1	Diagram of the decision task in Study 2. Depending on the condition, these will be comprised of 2, 3, 5, 8, or a continuous span of alternatives. Participants responded by moving their mouse across the arc corresponding to the desired response alternative.	56
Figure 4.2	Layout of the similarity rating task.	58
Figure 4.3	Relationship between the number of alternatives and mean response times in the colored dot mixtures task. Mean response times for the continuous condition (for comparison) are shown on the right. Each color corresponds to a different participant.	61
Figure 4.4	Relationship between the number of alternatives and accuracy in the colored dot mixtures task. Accuracy for the continuous condition (for comparison) is shown on the right. Each color corresponds to a different participant.	62
Figure 4.5	Relationship between the number of alternatives available and the discriminability of the target alternative (hue-based distance to the nearest distractor).	64
Figure 4.6	Mean accuracy and response time and accuracy for Participant 5 conditioned on the hue of the target color. The colored line is given by a kernel density estimator (bandwidth = 0.1) passed over the corresponding hue $\times$ accuracy or RT data.	66
Figure 4.7	Results of the MDS analysis of the similarity task included as part of Study 2 for each individual participant. The distance between any pair of hues is the approximate dissimilarity between the hues according to that participant.	67
Figure 4.8	Basic elements of the models used in Study 2. In addition to drift magnitude, threshold, and non-decision time, decision boundaries separated the different responses into correct and incorrect responses. These bounds were assumed to be optimal, so that they were halfway between the available alternatives in the discrete choice cases (2 and 3 alternatives shown at the bottom left and top right, respectively). In the continuous case, they were set to ± 7 degrees from the most prominent dot color, to reflect the criterion of ± 7 degrees tolerance for responses in the continuous condition.	70
Figure 4.9	Posterior predictions from the best-fitting models for each participant, overlaid on the actual response time and accuracy data.	80
Figure 4.10	Posterior model predictions (X) and observed mean response times (faded lines) for each participant (differentiated by color). Error bars indicate 1 unit of standard error in the data.	81
Figure 4.11	Posterior model predictions (X) and observed accuracy (faded lines) for each participant. Error bars indicate 1 unit of standard error in the data.	82

Figure 4.12 Posterior model predictions (dotted black line) of the relationship between target hue and response time or accuracy for each participant. Both data (colored curve) and model prediction curves are computed by passing a kernel density estimator over the hue $\times$ RT or hue $\times$ accuracy data (or prediction) from every trial. 83

CHAPTER 1

INTRODUCTION

Many of the tasks we accomplish in the laboratory or out in the world require us to make selections among many alternatives or along a continuum. Whether we are deciding how much time or money to invest, figuring out what direction to walk or drive, reproducing the orientation of a stimulus, coming up with the next word in a sentence, or even producing a musical note, there is a plethora of options available. However, our understanding of the decision processes underlying these sorts of selections is sparse. This is largely due to a traditional focus on developing theories and models of binary decisions, where a person only has two alternatives from which to choose. The goal of this thesis is to remedy this issue by developing and testing a general framework that can be used to model decisions among any number of alternatives.

In order to begin building a framework for modeling decisions among many alternatives, it is helpful to reflect on what approaches have been successful in explaining behavior on binary choice tasks. 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. The first section of this paper outlines the current state of sequential sampling models, looking at both diffusion and accumulator models of 2-choice decision-making. Each of these approaches to modeling binary decision behavior carries with it a set of assumptions about how choice alternatives relate to one another. I make these assumptions explicit, and examine the implications they carry for models of multiple-choice behavior. Carrying this further, I extend both diffusion and accumulator models to deal with multiple choice alternatives in a way that is consistent with their approach to modeling binary decisions. These two approaches can be related to one another by way of representing them geometrically – i.e. as the movement of an evidence state in a multidimensional decision space. I visit the different representations that allow both diffusion and accumulator models to be analyzed as random walk processes in these decision spaces. Finally, I provide a general framework for sequential sampling

models of multi-alternative as well as continuous-response decisions, formed by a random walk process unfolding in a multidimensional psychological space (e.g., a feature space). This provides the basis for constructing models of decisions among any number of alternatives.

Of course, the framework by itself is perhaps uninteresting without any evidence for its predictions. The second section of the paper is devoted to the first of two empirical studies, and is aimed at laying the empirical groundwork for a model of decisions among many alternatives. It sets benchmark empirical results using a simple perceptual task where participants must make a selection on a continuum, tests predictions of a model of this process, and applies the model to the resulting choice and response time data. In doing so, I establish empirical hurdles for models of this process, explore the elements (parameters) that allow the model to account for patterns of behavior on the task, and verify several independent predictions generated by the model.

The third section focuses on an second study aimed at understanding how the available choice alternatives interact with the decision process. It examines how the number of alternatives as well as their psychological representations affect decisions, looking at both accuracy and response times in a color-based task. It integrates a representation model based on multidimensional scaling with the decision model in order to explain and predict behavioral changes resulting from different combinations of available alternatives. In doing so, it illustrates the importance of considering the relationships between alternatives and provides a formal method for doing so.

The final section of the thesis reviews the findings and implications of the two studies, considers the limitations of the framework and present studies, suggests avenues for further research, and concludes with an examination of the contribution this work makes toward understanding the cognitive processes underlying decisions among many alternatives.

1.1 Benchmark phenomena

Before looking at the structure of sequential sampling models, it is worth first examining the behaviors that they seek to explain. To do so, it helps to consider a host of recurring empirical

phenomena that serve as benchmarks for model performance. The behavioral data in binary decision tasks is most frequently a) the response that a person makes to a stimulus, and b) the time it takes them to make this response. Much of the motivation for using sequential sampling models is their ability to predict both of these characteristics of decisions simultaneously.

In the case of inferential decisions where a decision-maker has two alternatives between which to choose, there is one correct and one incorrect response. The metric for choice performance in this case is accuracy, which is simply the proportion of responses a person makes that are correct. When the information that decision-makers receive is better, one can expect that their accuracy will be higher and/or that they will arrive at the correct decision more quickly. Conversely, when the quality of information they get is lower, decision-makers response more slowly or with lower accuracy. We refer to this straightforward relationship as the *difficulty effect*. It is well-supported by empirical data, and can often be directly estimated from the evidence a decision-maker is given (see Busemeyer & Townsend, 1993; Krajovich et al., 2012; Link & Heath, 1975; Palmer et al., 2005; Nosofsky & Palmeri, 1997; Ratcliff, 2014). This provides a first hurdle for decision models, indicating that a metric describing information quality may be important to account for choice behavior.

While accuracy and response time can often both improve with easier tasks (yielding faster response times and higher accuracy), they can trade off within a level of task difficulty. In these cases, accuracy suffers while response times get faster or response speed suffers while accuracy increases. This trade-off arises because higher accuracy often demands more information, whereas faster response times set a limit on how much information a decision maker can gather before responding. Decision-makers can improve their accuracy or response speed by adjusting how much information they gather, opting to require more (more accurate responses) or less (faster responses) information before making their decisions. This is referred to as the *speed-accuracy trade off*, where people can sacrifice the amount of information they gather to respond more quickly or vice versa (Bogacz et al., 2010; Heitz & Schall, 2012; Vickers & Packer, 1982; Wickelgren, 1977). This provides a second hurdle for decision models, and suggests the presence of an internal criterion

indicating the amount of information a person requires to make their decisions.

The final phenomenon I examine covers decision biases to respond in a particular way.¹ The alternative toward which a person is biased will experience faster response times, and the one against which they are biased will experience slower ones. When a person is biased toward the alternative which happens to be correct (congruent bias), we can expect that they will be more accurate, and less accurate when biased toward an alternative which is incorrect (incongruent bias). Bias may arise due to unbalanced rewards for hits versus correct rejections or unbalanced penalties for false alarms versus misses (Diederich & Busemeyer, 2006; Pleskac & Busemeyer, 2010), different base rates of stimuli (e.g. more targets than distractors J. M. Wolfe et al., 2007), or signals of true differences in prior plausibility of various hypotheses.

The pattern of results arising from a bias toward one alternative over another – faster and more accurate (congruent) or slower and less accurate (incongruent) responses toward the favored alternatives – is what I will refer to as a *decision bias effect*. This provides a final empirical hurdle, suggesting that decision models should have some mechanism allowing for a person to hold a pre-stimulus preference or belief.

Although these three phenomena certainly do not constitute an exhaustive list of choice and response time phenomena, they represent three of the most important empirical phenomena that have guided construction of binary choice models (Ratcliff & McKoon, 2008). They are particularly interesting because each effect corresponds to a different components of the decision-making process. Difficulty affects the rate of information accumulation in favor of the various alternatives, the speed-accuracy trade-off affects the internal criteria participants use to make their decisions, and the decision bias effect alters the information that a person considers in favor of a specific alternative. In the next section, I visit several common and successful sequential sampling models

¹In this paper, bias refers to a tendency to respond in favor of one set of alternatives over another by virtue of some information external to the target stimulus. It is not intended to refer to patterns of non-coherent behavior, as the term is used in the judgment and decision-making literature on heuristics and biases (Tversky & Kahneman, 1974; Gilovich et al., 2002). In fact, predecision biases in responding may actually be highly adaptive and correspond to an optimal prior for a given decision (Bogacz et al., 2006).

and examine how they account for each of these effects.

1.2 Relative evidence models

Sequential sampling models can be broadly broken down into two categories: relative evidence models and absolute evidence models (see Ratcliff & Smith, 2004, for a detailed breakdown of models). Though the conclusions that one draws from adopting one type rather than the other tend not to differ (Donkin et al., 2011), the assumptions that they make about how alternatives relate to one another is conceptually important when it comes to deriving a framework for multi-alternative decisions.

Relative evidence models like the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008) posit that a person represents the information or evidence they have regarding the decision as a balance between two choice alternatives. As they gather pieces of information, this balance shifts toward one alternative or the other. Once the balance shifts far enough in either direction, exceeding a criterion value specified by a particular threshold, a decision is triggered and a response entered. Their evidence state, which can be represented as a point, follows a particular path as the accumulation process unfolds. This allows us to display the choice process as an evidence trajectory, which moves along one dimension between favoring alternative A and favoring alternative B. The accumulation process describing this movement is shown in Figure 1.1.

One of the desirable properties of relative evidence models is that the balance of evidence between alternatives sums to zero. For inferential decisions, this means that the evidence balance between two alternatives [hypotheses] is linearly related to the log odds of one hypothesis relative to another.² This benefit derives from the diffusion model's roots in the sequential probability ratio test (Edwards, 1965; Laming, 1968; Wald & Wolfowitz, 1949), which treats decision-making as sequential tests of evidence against a criterion. The criterion corresponds to a desired internal

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

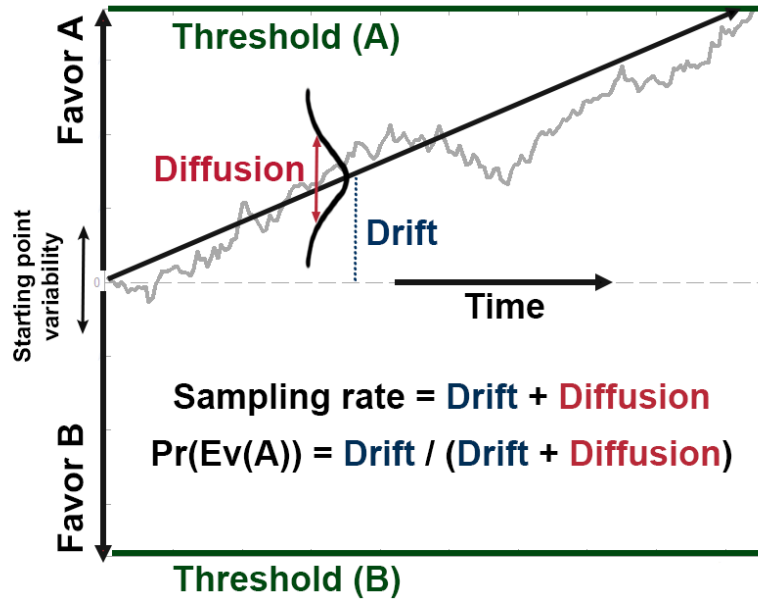


Figure 1.1 Diagram of three basic elements of the diffusion model: drift, diffusion, and threshold.

level of confidence set by the threshold parameter. For a particular desired level of confidence, the sample-by-sample evidence accumulation process constitutes a random walk in log odds space, describing an optimal Bayesian belief updating procedure (Bogacz et al., 2006; Palmer et al., 2005). As with the simple balance of evidence, a decision is triggered when the log odds for one hypothesis over the others exceeds the critical threshold value $\pm\theta$. The minimum bound θ ensures that the decision-maker can be confident with at least $(\gamma \times e^\theta) : 1$ odds that the hypothesis they have chosen is the correct one, where γ is a scaling factor that corresponds to the difficulty of discrimination between the two hypotheses (a measure of d' as in signal detection theory).

A very simple relative evidence model can often be characterized by a drift rate, diffusion, initial state, and threshold (as well as starting point variability, non-decision time, and drift rate variability; see Ratcliff & McKoon, 2008). The drift rate controls how quickly the evidence state moves toward one boundary or the other and diffusion controls the random noise, which moves the evidence state indiscriminately toward either alternative. The threshold sets the amount of information that a person requires to make a decision, and the initial state indicates where along the A-B continuum a person begins accumulating evidence. This initial state is frequently chosen

to reflect some pre-stimulus bias in favor of option A or option B, as it indicates what a person's state is before considering any information. In the case where they are unbiased, it is precisely in the middle of the thresholds corresponding to A and B (in the case of log odds representations, the initial state is 0). However, if a person receives some predecision cue or higher reward for one alternative over the other, they may favor that alternative in absence of any stimulus information. This favoritism is described by the starting point / initial state of the accumulation process.

The diffusion rate is often fixed at 1 in order to set the scale of the model – drift, diffusion, initial state, and threshold are otherwise confusable when permitted to vary freely within a condition. Therefore, the 'practical' number of parameters in this case is only 3, and it is frequently the case that diffusion will be ignored as a contributing factor the decisions and response times because it cannot be disentangled from the others.

Each of these components of the relative evidence accumulation process is shown in Figure 1.1. The three elements of drift, threshold, and initial state allow relative evidence models to account for the three choice phenomena I reviewed earlier. The difficulty effect corresponds to changes in drift rate – lower difficulty yields higher drift rates toward the correct alternative, giving both faster and more accurate responses. The speed-accuracy trade-off corresponds to shifts in the threshold – higher thresholds yield slower but more accurate responses, while lower thresholds yield faster but less accurate responses. Finally, the bias effect is achieved by moving the initial state around – information or reward favoring A will move it toward the A threshold, making responses in favor of A faster and more likely (but decreasing accuracy if A is not the correct alternatives). The reverse is true when information or rewards favor option B.

A final note regarding the diffusion rate is worth making. Rather than fixing it to 1, it is often more meaningful and more effective in terms of model fit to either set diffusion rates directly from the stimuli or constrain them to vary in a particular functional way across conditions (Donkin et al., 2009). In such a case, it is also possible to think of a diffusion process in terms of the overall sampling rate and the probability that each sample favors option A (versus option B) – in Kvam & Pleskac (2016), these are examined as independent factors called *weight* and *strength*, respectively.

This re-parametrization is helpful for developing a discrete Markov random walk formulation of the models (see Diederich & Busemeyer, 2003). Additionally, it enables more straightforward comparisons with other models of the decision process, including absolute evidence models and the more general continuous-response models I describe later.

1.3 Absolute evidence models

Absolute evidence models like accumulator models (Vickers & Lee, 1998; Smith & Van Zandt, 2000; Brown & Heathcote, 2005, 2008), by contrast, represent information or evidence for the alternatives as separate quantities [accumulators]. As a decision maker gathers or receives a piece of information, a single value describing the alternative that piece favors is incremented, while values for the other alternatives are unaffected. These values are frequently referred to as ‘accumulators,’ as each one stores the cumulative sum of information favoring the corresponding alternative. The accumulation process iterates for each new piece of information – as one comes in, the corresponding accumulator is updated, and a decision is triggered once one of the accumulators reaches a threshold value.

It is critical to note that each of the accumulators is usually assumed to be independent, such that only one accumulator’s value is updated with each new piece of information. This is what sets them apart from diffusion models – the total amount of information is retained and spread across the various accumulators, allowing each one to store the *absolute* amount of evidence for the alternative it matches.

Despite this major difference, there are a number of common mechanisms and analogous parameters that are shared between absolute and relative evidence models. Like relative models, absolute evidence models frequently use drift and threshold parameters in order to specify the evidence accumulation process. Each accumulator receives its own drift rate, though they can be set systematically using fewer than n parameters for n accumulators.³ The threshold works in a

³For example, it is typical to assume one drift rate for the accumulator corresponding to the

similar way to relative models – once a process crosses it, the relevant alternative is selected – but in cases of absolute models there is frequently only one threshold for all alternatives rather than one for each alternatives. In this case, it is which accumulator crosses the lone threshold rather than which threshold it crosses that determines choice outcomes.

Despite the similarities in terms of drift and threshold, some absolute evidence models depart further from relative models in terms of the source of response and response time variability. Ballistic accumulators like those of Brown & Heathcote (2008) assume no within-trial variability in evidence accumulation rate for an individual accumulator, so the diffusion component is removed in favor of other mechanisms that generate errors and response time variability. In these models, psychological or stimulus noise causes non-target accumulators to be incremented more often, sometimes allowing them to reach threshold before the target (correct) accumulator. However, this alone does not necessarily lead to errors; the drift rate for the target accumulator should still usually be the highest. In these cases, errors can be made when distractor / incorrect accumulators randomly start with higher values than the target / correct accumulators. This occurs due to variability in the initial state of the accumulator processes, which causes the value of each accumulator at the start of a decision to vary from trial to trial.

Returning to the case of non-ballistic, probabilistic accumulators, the drift rates of these accumulators can also be replaced by exponentially distributed inter-arrival times (see Figure 1.2). In the case where each new sample provides one ‘unit’ of evidence for the corresponding accumulator, the inter-arrival times simply describe the amount of time it takes between each sequential piece of arriving information. In a Poisson process – a common basis for accumulator models (Smith & Van Zandt, 2000; Smith & Vickers, 1988; Vickers & Lee, 1998) – the inter-arrival times for a particular accumulator will be exponentially distributed with a single parameter λ . On average, these function similarly to drift rates. However, due to the discrete nature of samples and stochastic nature of their arrival times, a single accumulator may behave somewhat differently than in the ballistic case.

target / true alternative and another for all of the distractor accumulators.

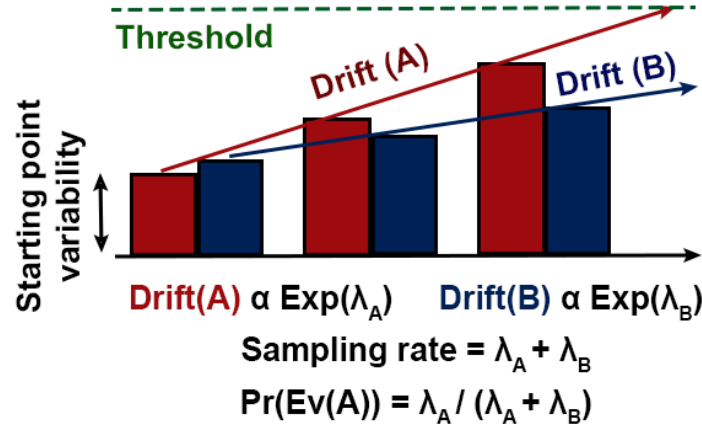


Figure 1.2 Diagram of basic elements of an accumulator model: starting point variability, multiple accumulators (with drifts), and threshold. As shown, drift can be represented as a function of sample arrival rate. This allows the accumulator model to also be described in terms of sampling rate and probability of sampling in favor of option A (or any other alternative).

As with the diffusion model, the parameters of a set of accumulators can be represented in terms of an overall evidence sampling rate and the probabilities of obtaining evidence in favor of option A and B (though the linear accumulator of Brown & Heathcote, 2008, has constant sampling rather than individual samples, so it can only be represented in terms of relative rates). Suppose that accumulators A and B are Poisson processes so that accumulator A is incremented with exponentially distributed time between each increment, $\text{Exp}(\lambda_1)$, and accumulator B is also incremented with exponentially distributed inter-arrival times $\text{Exp}(\lambda_2)$. They can be combined into a single Poisson process with exponentially distributed arrival times $\text{Exp}(\lambda_1 + \lambda_2)$, where the probability of incrementing accumulator A is $p_A = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ and the probability of incrementing accumulator B is $1 - p_A$ (see Ross, 2014, , Chapter 5). This transformation is also shown in Figure 1.2.

As with relative evidence models, accumulator models frequently fix a parameter in order to set the scale of the evidence accumulation process. The most common way of doing so seems to be to fix the sum of the accumulation rates across accumulators to 1 – in this case, the drift rates for individual accumulators can be viewed as the probability of that accumulator being incremented

next. Conveniently, this represents the evidence accumulation process in terms of a sampling rate (1) and probability distribution over possible sampling outcomes (accumulators). This makes these models easily comparable to the multi-dimensional random walks I present in the next section.

Perhaps the greatest benefit of these models is that their likelihood functions tend to be easily tractable and therefore the models are straightforward to fit to experimental data. As with the diffusion model, the difficulty effect is modeled as a shift in drift rate for the correct relative to the incorrect alternative. Correspondingly, the speed-accuracy trade-off is modeled as a change in the threshold parameter: lower values yield faster decisions and higher values more accurate ones. And finally, decision bias for one alternative is modeled as a change in the parameters corresponding to the initial state or as a shift in the relative values of the drift rates. However, because there are multiple initial states (one for each accumulator), the starting points and drift rates must be adjusted separately to account for a bias in favor of one alternative over another.

1.4 Multiple alternatives

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, 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 relative evidence models, adding alternatives is slightly more complicated. In a 3-alternative case with options A, B, and C, 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 $\frac{1}{2}$ when the log odds of A increase by 1. As more alternatives are added so that there are n outcomes, 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 (absolute evidence) or at least a uniform distribution of negative evidence (relative 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 (Huber et al., 1982), compromise (Simonson, 1989), and similarity effects (Tversky, 1972) – 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) (Brown et al., 2008). Confidence necessarily involves responses with at least ordinal relations to one another, so categories have to be structured with some adjacency properties as well (Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009).

Models of decision making have been modified in a number of ways to account for these phenomena. For example, decision field theory (Busemeyer & Townsend, 1993; 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 representations 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, for example.

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

1.5 Geometric representations

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).

1.5.1 Relative evidence

The basic uni-dimensional diffusion model used in recognition memory and other areas of psychology (Ratcliff, 1978) originally described the behavior of physical particles along a single dimension in space. Its natural geometric analogue is a random walk on a line (Figure 1.3A) (Link & Heath, 1975). 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 directly (at an orthogonal angle) 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)

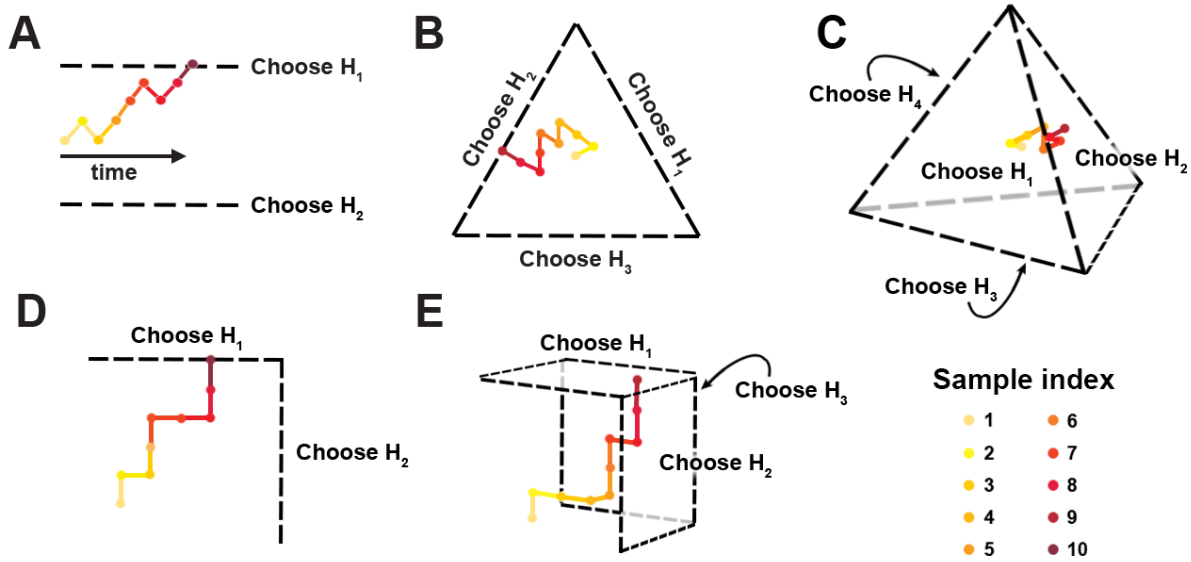


Figure 1.3 Representation of a person's accumulated evidence and choice criteria for 2-, 3-, and 4-alternative diffusion processes (A, B, C) and 2- and 3-alternative accumulator processes (D, E). An alternative is chosen when a person's represented evidence (yellow / red) crosses the corresponding edge (for models A, B, and D) or face (for models C and E).

a particular piece of evidence favoring hypothesis H_1 yields and the corresponding amount of evidence it provides for hypothesis H_2 :

$$Ev(H_2) \Leftarrow Ev(H_1) \cos(\angle D_1 D_2) \quad (1.1)$$

The cosine of this angle $\angle D_1 D_2$ can be quite naturally viewed as the similarity between the two choice alternatives – the cosine function has been used as a metric for similarity between vectors in a feature space, especially in latent semantic analysis (Wang et al., 2008; M. B. Wolfe & Goldman, 2003) as well as quantum models of similarity (Pothos et al., 2013). The evidence that information favoring H_1 provides for H_2 is therefore directly related to how similar these alternatives are.

In the two-alternative diffusion model, the two choice options are viewed as perfectly dissimilar – 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).⁴ This results in an evidence accumulation process that unfolds in a plane, contained within an equilateral triangle (Figure 1.3B). 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 a larger 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 $\{D_1, D_2, \dots, D_n\}$ satisfying the property $\cos(\angle D_i D_j) = -\frac{1}{n-1}$ for all $i \neq j$. In this case, evidence for any individual alternative provides evidence against all others alternatives equally. In the case of 4 alternatives, the boundaries corresponding to H_1, H_2, H_3 , and H_4 would each consist of a plane in a 3-dimensional space, together forming a tetrahedron (Figure 1.3C), 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 (the general version of a triangle or tetrahedron in 4+ dimensions), with the choice boundaries corresponding to each of its $(n - 2)$ -dimensional facets.

It is worth a note that the cosine relation in Equation 1.1 will preserve log odds in any n -dimensional space by virtue of every integral $\int \dots \int_0^{2\pi} \cos(\phi) d\phi = 0$. However, cases 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. If we suppose that the log odds of different hypotheses (given stored evidence) are the basis of probability and confidence judgments – an optimal strategy and one adopted by several cognitive and neural models of confidence judgment production (Bogacz et al., 2006; Edwards, 1965; Kepecs et al., 2008; Kiani & Shadlen, 2009; Meyniel et al., 2015; Pleskac & Busemeyer, 2010) – then this may lead to subadditivity or superadditivity of these judgments among sets of psychologically related alternatives.

⁴Note 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.

1.5.2 Absolute evidence

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. When there are 2 alternatives, H_1 and H_2 , evidence in favor of H_1 provides no information that changes a person's beliefs about H_2 . Using the relationship defined by equation 1.1, this means that the directions corresponding to each alternative must be orthogonal, $\cos(\angle D_1 D_2) = 0$. The choice boundaries can therefore be represented as two sides of a rectangle (Figure 1.3D), where the position of a person's state along one dimension (left / right) describes evidence for one alternative and the position along the other (up / down) describes evidence for the other.

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 some reference point (the initial state position, for example). 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 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. As in the two alternative case, log odds could be preserved by also allowing a theoretical alternative direction $-D_n$.

In the case of three alternatives, the evidence accumulation process would unfold on a figure bounded by three choice criteria constituting sides of a rectangular prism (Figure 1.3E). For n alternatives, the orthogonal choice criteria would compose a set of intersecting facets of an n -orthotope or in special cases an n -cube.

1.6 Random walks

An important benefit conferred by this geometric way of constructing response alternatives is that the relative and absolute evidence models can be constructed as different types of random walk models. Two important distinctions can be made between different random walk models of decision-making. The first is between continuous-time and discrete-time models. So far, I have covered primarily continuous-time models, which examine how evidence changes continuously across time. They do so by specifying distributions that describe how frequently evidence is updated with new samples or the rate at which evidence changes over time. However, discrete-time models, which break the evidence accumulation process into distinct units of time, are also useful. For example, they can be used to describe how evidence is updated sample by sample, as with drawing cards or receiving discrete chunks of information (Diederich & Busemeyer, 2003; Markant et al., 2015). The primary functional difference between the two approaches is that continuous-time random walks predict continuous response times (in minutes, seconds, etc.) as a function of distributional assumptions about how evidence arrives, while discrete-time walks predict discrete response times in terms of the number of steps it takes to arrive at a decision.

The second distinction between types of random walks is discrete space or continuous space random walks. Discrete-space walks describe how evidence changes over time using a set of individual evidence states. These are most frequently used when there is a meaningful set of non-overlapping cognitive states that a person could be in when gathering evidence. Formally, they are described by a mixed state that gives the probability of being in each particular state at a given time.⁵ For example, several models use separate evidence states that correspond to different levels of confidence, allowing for the distribution of evidence across states at any given time to give rise to a distribution of possible judgment responses (Busemeyer et al., 2006; Kvam

⁵Note that the probabilities in the mixed state reflect the modeler's – not the decision-maker's – uncertainty about their beliefs or preferences. The true location of the evidence is presumably fixed in the decision-maker's system (though see also Kvam et al., 2015), and it specifies the decision-maker's level of evidence.

et al., 2015; Pleskac & Busemeyer, 2010). Continuous-space random walks are more frequently used when such discrete levels of evidence are not necessary. They can also be conceptually simpler, as evidence states in continuous space frequently (barring the addition of unusual state- or time-dependent functions in the evidence) can be described using computationally convenient distributions like a normal distribution. The continuous and discrete space representations are closely related, however; a discrete-space model converges to a continuous-state model as the (unbounded) state space is more and more finely partitioned, i.e. as the number of states $s \rightarrow \infty$.

The geometric framework I have proposed lends it self to both discrete-time and continuous-time as well as discrete-space and continuous-space random walk representations, allowing it considerable flexibility in terms of the types of responses and decision processes it can be used to describe. For the models I have described so far, a discrete-space representation – useful for generating analytic distributions of responses and response times – can be gained 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-based stimulus into three categories: red, blue, or green. This corresponds to the triangular structure shown in Figure 1.3B. 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, corresponding to an unbiased 0/0/0 log odds distribution over the hypotheses, but as they view the stimulus they should sample pieces of evidence that favor green, red, or blue, causing them to step at angles $\frac{\pi}{6}$ radians / 30 degrees, $\frac{5\pi}{6}$ radians / 150 degrees, or $\frac{9\pi}{6}$ radians / 270 degrees 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

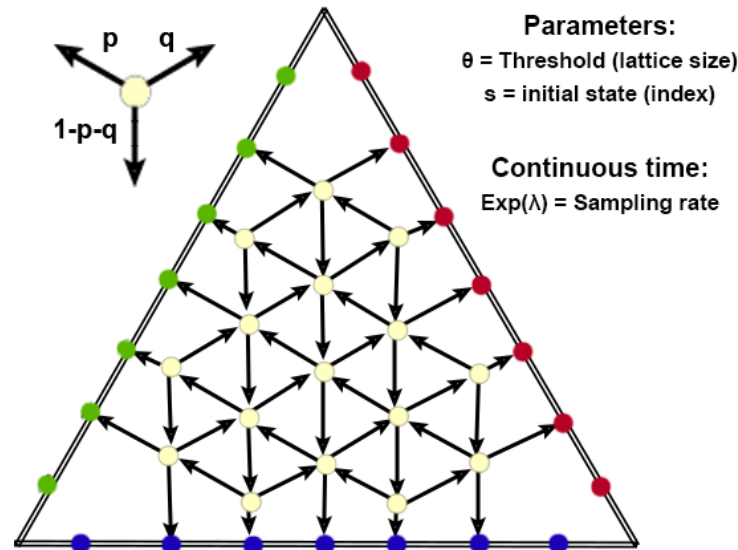


Figure 1.4 Lattice and state transition structure for a 3-alternative, relative evidence model of color identification.

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 continuous 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 – many of these are illustrated in Figure 1.3. The lattice shape will change based on the type of model and number of transition destinations, 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.

Note that a person's evidence representation will only be able to step toward available alternatives, not directly away from them as was the case in the two alternative diffusion model. In Figure 1.4, this is indicated by unidirectional transitions. Similarly, in accumulator models like the one shown in Figures 1.3D-E, evidence can only favor one of a set of orthogonal alternatives, meaning that there will never be evidence sampled against a particular one. At present, the accumulation

process specified by absolute evidence models will only consist of movement in half of the available directions (e.g. rightward or upward in the two alternative case, Figure 1.3D), though this assumption could be modified to generate interesting choice processes.

Each of these random walks requires some initial state. This can be a fixed point or it can be specified as a mixed state of a set of possible initial points. The mixed state, depending on how variable it is, is often used to produce error responses that are as fast or faster than correct responses. Bias in the initial state, corresponding to a tendency to select some alternatives over others in absence of any mitigating information, can be introduced by moving the initial state closer to one boundary or another.

However, this is not the only way to introduce bias into the system. It can instead be introduced as asymmetric transition probabilities, indicating that more information is gathered in favor of one alternative over the other. This bias is included in the drift rate on top of the information provided by the stimulus. Suppose that δ_s corresponds to the information from the stimulus, and δ_b corresponds to the biased information. Ordinarily, the drift rate would be δ_s when the stimulus favored option A and $-\delta_s$ when it favored option B. However, when drift bias favors option A, its drift rate will be $\delta_s + \delta_b$ when the stimulus favors option A but $-\delta_s + \delta_b$ when the stimulus favors option B.

1.7 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, as I covered in the section on multi-alternative models, 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, arranging them spatially by their features. The directions could also be constructed by using existing psychological theory or independent data like similarity judgments. I pursue each of these approaches in the studies that constitute the empirical component of this paper.

The second modification of the framework above is the assumption that a person’s representation of evidence is modified to directly reflect the information gathered from the stimulus. For example, if a person is trying to reproduce the orientation of a stimulus that can vary anywhere from 0 to 180 degrees, they must be able to sample and represent information that favors any direction between 0 and 180 degrees. 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.

1.7.1 Random walks in psychological space

Instead of the usual unidimensional random walk (or Markov chain) representation, 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 of 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.

Much as the discrete random walk converges to a continuous diffusion process as the step size becomes small, the sampling distribution approaches a diffusion process as the unit of distance for each step of the multidimensional random walk becomes small. This diffusion process can be described using a drift direction, which specifies what information is being accumulated most often, and drift magnitude, which specifies how rapidly this information is accumulated (signal to noise ratio). This diffusion process is described by Smith (2016), and it has a number of interesting properties. If we suppose that a person is using circular decision bounds and that decision-making starts in the center of this circle, the time it takes them to reach a decision (hitting the edge of the circle) is independent of where they hit the boundary. This is the continuous analogue of symmetric correct and incorrect response times that is found in diffusion models with no starting point or between-trial drift rate variability.

While the evidence at any given point in time can be given by its rectangular or Cartesian coordinates (x, y, z , etc.), it is often more convenient to represent it in terms of polar or spherical coordinates, where one coordinate describes its distance from the origin and the others describe its angle. In the case of the two dimensional model, this gives a radius r and single angle θ (note that this should not be confused with the threshold parameter discussed earlier, which describes the radius at which a decision is triggered). An important characteristic of the two-dimensional model is that the distribution of evidence at any given point in time should reflect a von Mises distribution over its θ coordinate. This is true during evidence accumulation and at the time of choice, meaning that the distribution of responses on a circle should also be described by a von Mises distribution.

Both of these constitute a priori predictions of the current 2-dimensional diffusion model: re-

sponse times should be uncorrelated with response error (accuracy) within a condition, and responses should be von Mises distributed when entered. As an important goal of this paper is to develop an empirically accurate model of decisions on a continuum, I test both of these predictions of the model on top of the three empirical phenomena describe earlier (difficulty effect, speed-accuracy trade-off, decision bias effect) in a continuous response task. Study 1 investigates these 5 predictions.

1.8 Hick's Law

One prominent phenomenon that relates closely to the models I have described here is called Hick's Law (Hick, 1952), which states that the time required to reach a decision increases as a function of the log of the number of alternatives. This relationship was originally explained in terms of the number of bits (Shannon information, Shannon & Weaver, 1949) it would require in order to distinguish between n alternatives. In theory, it should require 1 discriminating bit to distinguish between 2 alternatives, 2 bits to discriminate between 4, 3 bits between 8, and so on, yielding a log base 2 relation between the number of alternatives and the discrimination component of response time. This relationship is shown in Figure 1.5.

Alternatively, Usher et al. (2002) showed that the relationship might also be explained in terms of accuracy and the number of accumulators required to represent evidence for each of them – the threshold parameter in an accumulator model would have to be increased incrementally with the number of independent accumulators added in order to maintain the same desired level of accuracy. Therefore, if participants set their internal choice criteria so that they could maintain a constant level of accuracy, this would produce the log-linear relationship that Hick and others observed between the number of alternatives and response time.

Study 2 investigates Hick's law and how it relates to both discrete sets and a continuous span of alternatives. Of course, the infinite number of alternatives in the continuous case leads to the bizarre prediction that response times should be essentially infinite ($\log_2(\infty) = \infty$). However, peo-

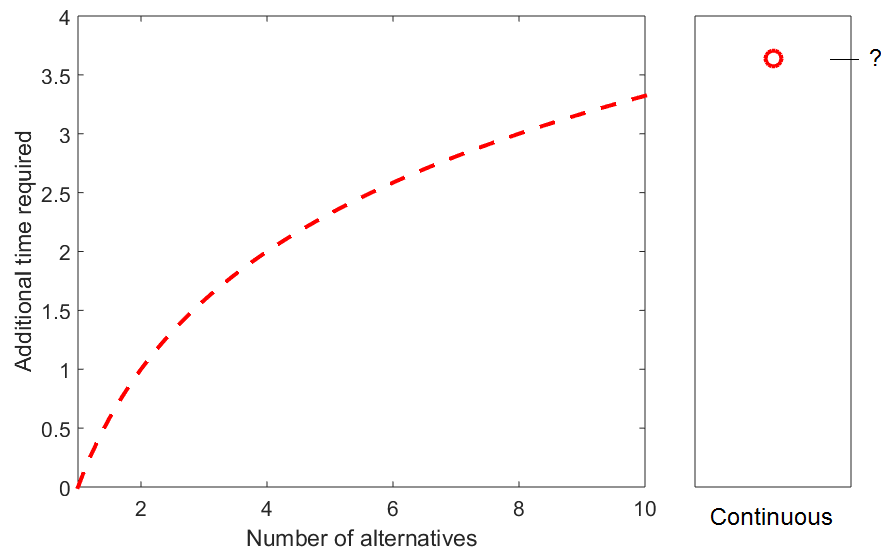


Figure 1.5 Predicted response time proportional to the number of bits required to discriminate between alternatives. It is unclear what the response time prediction should be in terms of bit-based information for the continuous case, as Hick's law may not apply to this condition.

ple could be using a smaller 'practical' number of alternatives. If so, it is unclear what number they may be using to approximate a continuous span of alternatives. As such, any finding in the continuous condition is informative, whether it takes a particularly long time or somehow falls within the range of discrete choice conditions.

The modeling component of Study 2 also investigates possible explanations for why response times changes as a function of the number of alternatives. As Usher et al. (2002) suggested, I investigate whether shifts in thresholds may be responsible. However, I also propose and test another potential explanation. It is likely that a larger number of alternatives can 'clutter' the decision space, such that alternatives tend to be more similar as more are introduced within the same range (e.g. 2 random values drawn between 0 and 1 will on average be further apart than 3, 4, or 5 random values). As such, pairs of alternatives within a larger set on the same range will on average be more difficult to distinguish from one another.

In practical terms, this may have two effects. The first is that the portion of the response space devoted to each alternative is smaller, and therefore each one is less likely to be selected

when selection depends on a random variable. For example, consider a situation where there are two alternatives – “greater than 50” and “less than 50” – as predictions for a random 0-100 draw. Compare this to the case where there are three alternatives, “less than 33” or “33-66” or “greater than 66.” In all cases, the probability of the first and last options in the 3-alternative case will be less than or equal to the first and second in the 2-alternative one. A similar concept may be at play in the case of multiple alternatives in other situations. However, this only explains differences in accuracy, which can only become differences in response time if participants make some trade-off in order to maintain accuracy at the expense of response speed.

A second effect of response crowding is that there may be a shift in the discriminability of alternatives as a function of set size. This corresponds to changes in the ratio of sampling correct to incorrect information, a change in drift rate across conditions rather than thresholds. The major difference between these explanations is that a decrease in drift with more alternatives will result in both lower accuracy and slower response times, rather than maintaining accuracy at the expense of slower response times. Additionally, the precise nature of the relationship between number of alternatives and response times will not necessarily be log-linear. If the alternatives that are introduced are more similar to those already in the choice set, they will have a different effect on response times and accuracy than alternatives that are dissimilar to those already in the set. For example, adding *orange* to a set of color response alternatives $\{red, blue\}$ is likely to have a much more distracting effect when the target color is *red*. The discriminability between orange and red is much smaller than between orange and blue, so the rate of true information for *red* should be smaller when there is a red stimulus with orange distractor than the rate for *blue* when there is a blue stimulus with orange distractor.

Interestingly, this is perhaps a more accurate description of the relationship between the number of alternatives and response times than the strictly log-linear one. Longstreth et al. (1985) have suggested that, while the relation between the number of alternatives and response time appears to be positive, it is not necessarily of the form described by Hick (1952) and instead may vary depending on the assortment of available alternatives. By implementing the proposed elements

into a computational model in Study 2, I show a richer relationship between set size and response time that factors in the discriminability between alternatives and their interaction with internal choice criteria like thresholds.

CHAPTER 2

GOALS AND PREDICTIONS

So far, I have outlined a general framework for modeling decisions between any number of alternatives. When a modeler is interested in making assumptions of independence or equal interdependence between a finite number of alternatives, they can use the simplex or hypercube representations of alternatives shown in Figure 1.3. Doing so is similar to using a multi-alternative diffusion or accumulator model, but adds the additional option of using random walk variations when steps are among discrete states in a lattice. It is therefore useful for dealing with discrete and continuous models of decisions among uniformly-related or unrelated alternatives. However, psychological relations between alternatives must be introduced in order to account for a continuous span of alternatives (otherwise, we would have to work in an infinite-dimensional space) as well as to account for decisions between alternatives that share unequal similarity relations among them.

The primary goal of this thesis is to build and test models that account for decisions among alternatives that are related with one another in varying ways, and particularly to explore how people make responses on a continuum. To this end, I present two main studies. The first focuses on what components are important to include in such models by examining continuous analogues of common accuracy and response time decision phenomena. It investigates whether the model parameters change in sensible ways, and examines two additional a priori predictions that the model makes in order to gauge its appropriateness for modeling decisions on a continuum.

The second study examines how response times and accuracy change across varying numbers and arrangements of alternatives. It focuses on the case of color, where physically defined relations (hue) between alternatives may not map directly onto psychological relations between them. The modeling component of this study illustrates how these relations can be built into our modeling approach by combining a multidimensional scaling representation of colors with the decision model.

2.1 Study 1

Study 1 covers much of the groundwork in establishing whether the model framework is appropriate to explain and predict decisions on a continuum. Because the structure of the model I use for the continuous case is analogous to the 1-dimensional diffusion model, we should find similar accuracy and response time phenomena between the 2-alternative and the continuous case as well. I therefore test continuous analogues of the difficulty effect, speed-accuracy trade-off, and predecision cue bias effect.

The predictions of the first study are relatively straightforward. I expect each of the manipulations to have parallel effects to those that appear in the 2 alternative cases. Manipulations of the difficulty of the task should result in slower responses and larger errors for harder stimuli. In terms of model fits, manipulating difficulty should affect the estimates of the drift magnitude parameter ($|\mu|$) in the model.

Response times and accuracy reflecting the speed-accuracy trade-off should change in the same way as in binary decision studies as well. In the speed condition – relative to the accuracy condition – I expect to find faster response times and larger errors. This should be reflected in the model via changes in the threshold parameter (θ) estimates, which should be higher in the accuracy condition and lower in the speed condition.

In the case where there is a predecision cue designed to bias responses, responses should be faster and more accurate when the cue is congruent with the stimulus and slower and less accurate when the cue is less congruent with the stimulus. Ordinarily, this would produce changes in the initial state of evidence. However, due to computational limitations, it has not been possible to allow the initial state to vary. Instead, the bias can be modeled as if participants are integrating two sources of evidence – the cue and the stimulus. This should be reflected as shifts in two parameters: the drift direction (ϕ ; the cue should pull evidence sampling toward it and away from the stimulus mean when incongruent) and the drift magnitude (the cue should dilute evidence when it is incongruent with the stimulus, and enhance it when it is congruent). As the discrepancy

between the cue and the stimulus gets larger, I expect that the drift angle will deviate further from the stimulus mean and the drift magnitude will get smaller, i.e. the sampling distribution will be more diffuse.

Non-decision time estimates should not change too substantially across the various manipulations, as responses are entered in the same way within each task. Participants should be able to respond at relatively consistent speeds across conditions.

In addition to the manipulations derived from binary choice, the 2-dimensional diffusion model makes additional predictions regarding the distribution of responses and response times. It suggests that response time distributions should be unaffected by response deviation (within a condition), and that responses should be von Mises distributed. Study 1 tests each of these claims as well.

Ultimately, Study 1 should establish empirical phenomena that can be found in continuous response tasks, and establish what components should be included in a model of behavior on these tasks.

2.2 Study 2

The outcomes of the second study are somewhat less certain. If Hick's law holds, response times should shift as a function of the $\log_2$ of the number of alternatives when they are discrete, but it is unclear what will happen when there is a continuous span of alternatives. It is likely that accuracy will decrease with the number of alternatives (though this is not predicted by Hick's law), particularly because stimulus discriminability will change as more alternatives are introduced into the same space.

The threshold-based explanation of Hick's law (Usher et al., 2002) suggests that threshold estimates in the model should increase as the number of alternatives increases, and that this should be responsible for the response time (and accuracy) differences between conditions.

Participant's similarity ratings between pairs of alternatives should reflect how easy those alternatives are to discriminate. Therefore, we should find that alternatives they rate as less similar

– above and beyond the physical differences – should be faster and more accurate to discriminate. Those that they rate as more similar should be slower and less accurate to discriminate.

There are a number of components of the model, which are described in more detail in the section detailing Study 2. I expect that multi-dimensional scaling to obtain drift rates, shifts in category / choice boundaries based on the positions of alternatives, shifts in threshold for the number of alternatives, and thresholds that shift based on the (multi-dimensional scaling-based) discriminability of alternatives are all possible elements of the model. The modeling component of this study tests which components can be omitted without damaging the model's performance.

Ultimately, Study 2 seeks to examine how decisions among continuous and discrete numbers of alternatives are related, and explore how models can capture the psychological relations between alternatives (beyond physical relations) by integrating similarity and decision components together in a single model.

CHAPTER 3

STUDY 1 - EMPIRICAL HURDLES AND MODEL ELEMENTS

The purpose of Study 1 was to explore empirical phenomena relating to decisions on a continuum and their implications for modeling these decisions. Part of the reason that models of binary decision have been so well-developed is the presence of established empirical phenomena that relate the accuracy of responses to the speed at which they are entered. These phenomena impose strong constraints on formal theories, indicating what parameters and structures might be required in order to accommodate typical choice data. Ratcliff & McKoon (2008) focused on three particular findings that have shaped theory development: the speed-accuracy trade-off, where emphasis on response accuracy slows down responses while emphasis on response speed makes responses less accurate (Bogacz et al., 2010; Heitz & Schall, 2012; Wickelgren, 1977); stimulus difficulty, which can decrease accuracy as well as increase response times (Link & Heath, 1975); and predecision biases to make responses, which improve response speed and accuracy for favored responses and hurt them for disfavored ones (Diederich & Busemeyer, 2006; J. M. Wolfe et al., 2007).

Accuracy and response time are the basic metrics by which we evaluate decision models (Brown & Heathcote, 2008; Vickers & Lee, 1998), and yet no established accuracy and response time phenomena exist for continuous-response decisions; some previous work has looked at either accuracy or response times in continuous response tasks, but never both. For example, the accuracy of orientation reproductions has been used as a measure to distinguish between models of visual short-term memory (Van Den Berg et al., 2012). However, the interplay between accuracy and response times is nontrivial and may have serious consequences for studies that only consider one or the other. For example, response distributions that appear to be a mixture of a discrete set or continuum of von Mises distributions (which might be used to motivate a discrete or continuous-resource model of visual short term memory, respectively) could actually result from continuous or discrete combinations of difficulty, speed or accuracy emphasis, or decision biases.

Study 1 seeks to rectify the lack of empirical results in continuous choice by exploring contin-

uous analogues of three classic binary choice phenomena. It examines the effect of speed versus accuracy incentives, the effect of stimulus coherence (difficulty) on accuracy and response times, and the effect of predecision response cues on these same outcomes. If the decision process is similar between binary and continuous choice, we should find the presence of a speed-accuracy trade-off, higher response times and lower accuracy for more difficult stimuli, faster and more accurate responses for accurate predecision cues, and slower and less accurate responses for inaccurate predecision cues. I also examine two additional predictions made specifically by the particular model I use here: von Mises distributed responses and response times that are unrelated to error magnitude within a condition. In the following sections, I test these predictions using a simple perceptual task requiring responses on a continuum.

3.1 Methods

In the continuous case, accuracy is not binary. Rather, all responses can be seen as varying degrees of incorrect, as they can be closer or further from a true point answer. Therefore, 'accuracy' in these studies refers to how close a response is to the true answer. Frequently, I will refer to the *deviation* of a response from the true answer or a cue. Because the stimuli in Study 1 all fall on 0 to 180 degree orientations, any response deviations are bounded between 0 and 90 degrees.¹ Hence, perfect accuracy will result in average response deviations of 0, guessing responses will result in average response deviations of 45 degrees, and perfect inaccuracy (consistently giving the most wrong response) will result in average response deviations of 90 degrees.

As in studies of binary choice, a speed-accuracy trade-off was implemented by changing the incentive structure of the task to reward either response speed or response accuracy. Difficulty was manipulated by changing the coherence of the stimulus, and predecision bias was manipulated by

¹Note that being 120 degrees clockwise is the same as being 60 degrees counterclockwise from the true mean orientation. Because the stimuli fall only on a half-circle, responses at ± 180 degrees from one another are equivalent. The smallest distance to the target / true response is taken as the actual response deviation.

providing informative cues prior to showing the stimulus. Each of these was implemented in an orientation-based task, described below.

3.1.1 Task

The task is shown in Figure 3.1. This was primarily an orientation detection task, where participants saw a noisily jittering Gabor patch stimulus and had to select its mean orientation. The jittering Gabor stimulus was generated by initially choosing a random mean orientation from a uniform distribution on $[0,180)$ degrees. On each frame, a new Gabor patch would be drawn from a wrapped normal distribution centered on this mean orientation. The standard deviation of this wrapped normal depended on the difficulty manipulation. The standard deviations were 15, 30, or 45 degrees, corresponding to easy, medium, and hard difficulty (larger standard deviations result in noisier stimulus information). The difficulty was drawn randomly between these levels on each trial.

Participants' task was to produce the mean orientation of the stimulus by moving their mouse across the edge of a response circle (outer white circle). Their response could be made on the lower or upper half of the circle – those made at 30 and 210 degrees were equivalent. Participants received points in the task for both the speed and accuracy of their responses, contingent on whether a block of trials was a speed-emphasis block or accuracy-emphasis one. In the speed condition, they received 100 points for responding within 800 milliseconds of the stimulus onset and up to 100 points based on how close they were to the true mean orientation of the stimulus: each degree away decreased their reward by 1.1 points. In the accuracy condition, they received up to 200 points based solely on how close they were to the true mean orientation – in this case, each degree away decreased their reward by 2.2 points. Participants who accumulated a sufficient number of points during the task received additional research credit for their participation.

To begin a trial, participants clicked on a small white circle in the middle of the screen. Their cursor was then centered on the screen, which involved only minor adjustments as the white circle was already in the center of the screen. In the cued condition (top left of Figure 3.1), constituting

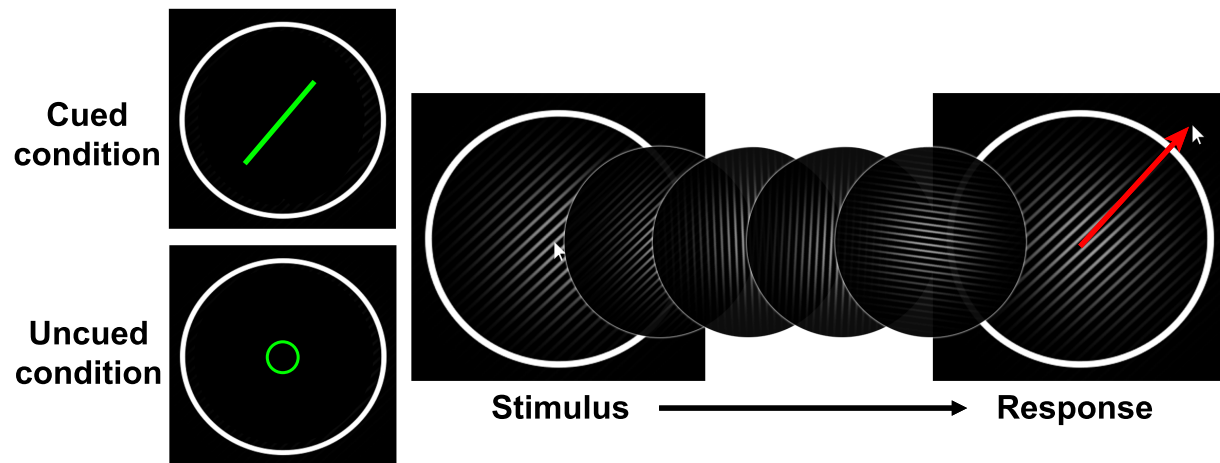


Figure 3.1 Time course of a trial of the orientation detection task.

half the trials, they then saw a green line on the screen. This line corresponded to the true mean orientation of the stimulus in 50% of cued trials, but was randomly set to 20, 50, or 70 degrees away from the true orientation on the remaining 50% of cued trials. This allowed the cue to be informative but not overwhelm stimulus information in terms of usefulness. For uncued trials, they simply saw an uninformative green circle appear (bottom left of Figure 3.1).

Either the green line cue or circle disappeared after 1 second, and then the jittering Gabor stimulus was presented. Participants could respond at any time by moving their mouse across the edge of the white response circle at their desired orientation. They were encouraged to have ballistic mouse movement by penalizing them based on the amount of time the mouse pointer spent between the center (circle that they clicked to start the trial) and the edge of the response circle. This was done both to control the non-decision time (time required for non-decision response processes like moving the mouse) across trials as well as avoid having participants externalize their beliefs before responding. The latter can create issues in terms of multiple sequential measurements – expressing one's beliefs in such a way may affect the distribution of subsequent responses.

3.1.2 Participants

A total of 12 Michigan State University students (8 female, 4 male) each completed 960 trials of the experiment for class credit. Participants were primarily 18-26 years old. One additional participant completed the task, but was removed from further analyses for having median response times well outside the normal range for participants (> 5000 milliseconds [ms], compared to 500-1500 ms for other participants). The full task, including introduction and practice trials, took approximately 1 to 1.5 hours to complete.

The sample size of 12 participants was selected so that the group-level posterior distributions of parameter estimates were both sufficiently constrained and suitably well-sampled. Because each participant provided a large number of data points and all parameters were estimated in a hierarchical Bayesian way, this sample was sufficient to obtain precise estimates on both individual and group-level parameters.

3.1.3 Materials

All stimuli were generated and responses recorded using MATLAB and Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007). Responses were recorded on the mouse.

3.1.4 Procedure

Upon entering the laboratory, participants were briefed on the rough content of the task and completed informed consent. They were then seated in a dark, windowless office to complete the task. They completed approximately 60 practice trials of the task (more trials were generated when responses were too inaccurate or slow), with immediate feedback on how quick and how accurate their responses were on each practice trial.

Upon finishing the practice trials, participants completed 960 trials of the main task as described above. These were organized into 12 blocks of 80 trials. The blocks were organized so that participants would see 3 blocks of accuracy with a predecision orientation cue, 3 blocks of

accuracy trials with no orientation cue, 3 blocks of speed with the orientation cue, and 3 blocks of speed trials with no predecision orientation cue. Difficulty manipulations took place within blocks. The order of the blocks of trials and the order of difficulty of trials within a block were both randomized.

Before each block, participants received instructions for the upcoming block's condition, indicating both how they would earn points (for the speed / accuracy manipulation) as well as the presence of the cue. They were also reminded before each trial whether they were in a speed or accuracy block with a small piece of text above the white pre-trial center circle, which read 'SPEED' or 'ACCURACY'.

Upon completing the experiment, participants were debriefed on the purpose of the study and told how much credit they would receive based on the number of points they accumulated during the task.

3.2 Results

The analyses focused on two main outcomes. The first was simple response times – the number of seconds it took for a person to enter their response. The second was their response deviations from the true orientation. For this metric, I took the mean orientation of the Gabor stimuli and compared it to participants' responses on the circle. Their response deviation was the number of degrees clockwise that their response fell relative to the true orientation. For example, if their response was 2 degrees counterclockwise from the actual mean orientation, their response deviation would be -2 degrees (all response deviations therefore fell between -90 and $+90$ degrees). On average, response deviations were approximately zero, indicating no clockwise or counterclockwise bias. In order to gauge accuracy, I used the *absolute* response deviation, which was simply the absolute value / magnitude of the deviation without regard to its direction.

Response times and response deviations depended on four main factors. The first was stimulus difficulty, operationally defined as the standard deviation of the orientation of the jittering Gabor

patches (15 / 30 / 45 degrees). The second was the speed or accuracy manipulation, coded in the models as 0 for accuracy and 1 for speed. Third was the presence of the cue; the default cue was assumed to be the orientation that matched the true stimulus orientation, so this factor describes the benefit conferred by an informative cue. The final factor was the orientation of the cue, which corresponded to the number of degrees that the cue deviated from the true stimulus orientation (i.e. 0 / 20 / 50 / 70 degrees). The inclusion of this final factor depended on the third factor, so that cue orientation was not considered when there was no cue.

Each of these factors was standardized before using them to predict responses or response times. All analyses used a hierarchical Bayesian linear model, where individual-level parameters (coefficients for the size of each standardized effect) were constrained by a group-level distribution. Here, I report only the group-level parameter estimates for each effect for the sake of brevity.

Unless otherwise specified, the Bayesian models use diffuse priors such that the data will have maximal influence over parameter estimates. In reporting, I provide the mean posterior parameter estimates as well as the range that includes the 95% most credible values from this posterior distribution of possible parameter values. This is referred to as the 95% highest density interval [95% HDI] (see Kruschke, 2014).

The data as well as the code for each of the JAGS models is available on the Open Science Framework at osf.io/ufe96.

3.2.1 Response times

To examine the effects of these manipulations on response times, I used a hierarchical Bayesian model with an intercept, a linear main effect of the difficulty of the stimulus, a binary main effect of whether each trial was a speed-emphasis trial, a binary main effect of whether a trial included a cue, a linear main effect for the orientation of the cue (which was paired with the binary indicator for cued trials so it did not come into play on non-cued trials) and all 2-, 3-, and 4-way interactions between these factors. However, note that cue orientation is not permitted to interact directly with difficulty or speed factors as it is contingent upon the presence of the cue. It can therefore only

appear as a higher order interaction.

Before analysis, response times were log transformed in order to make them approximately normally distributed. All predictors and outcomes were then standardized. The coefficients for each of the main effects and interactions were set hierarchically by participant, and the group level estimates are shown here.

Table 3.1 Mean estimates of coefficients for main effects and interactions of stimulus difficulty, speed manipulation, cue presence, and cue orientation on response times. The ranges containing the 95% Highest Density Interval (HDI) are also provided. Intervals excluding zero are starred.

Coefficient	Mean Estimate	95% HDI
Difficulty	0.06	[0.04, 0.08]*
Speed	-0.13	[-0.19, -0.06]*
Cue	-0.09	[-0.14, -0.03]*
Cue $\times$ orientation	0.04	[0.02, 0.05]*
Difficulty $\times$ speed	-0.02	[-0.03, 0.00]
Difficulty $\times$ cue	-0.02	[-0.04, 0.00]
Difficulty $\times$ speed $\times$ cue	0.00	[-0.02, 0.02]
Difficulty $\times$ cue $\times$ orientation	-0.00	[-0.02, 0.01]
Speed $\times$ cue	-0.02	[-0.06, 0.02]
Speed $\times$ cue $\times$ orientation	-0.01	[-0.01, 0.00]
Difficulty $\times$ speed $\times$ cue $\times$ orientation	0.01	[-0.01, 0.02]

The mean estimates and 95% Highest Density Intervals for each coefficient are shown in Table 3.1. As shown, the only effects with substantial contributions to differences in response times were the main effects, including the main effect of cue orientation that was contingent on the cue being present (and is hence presented as an interaction). Response times increased with difficulty as well as with less accurate cue orientations: the further the cue was from the true mean stimulus orientation, the longer participants took to respond. By contrast, response times decreased with speed instructions as well as with the presence of an informative cue.

Each of the interactions in the model included zero as a credible value, suggesting that the various manipulations did not interact strongly enough to produce substantial changes in mean response times.

This model confirms that difficulty, speed-accuracy, and cue / cue orientation manipulations had their anticipated effects. The raw data for difficulty and speed-accuracy conditions can be seen

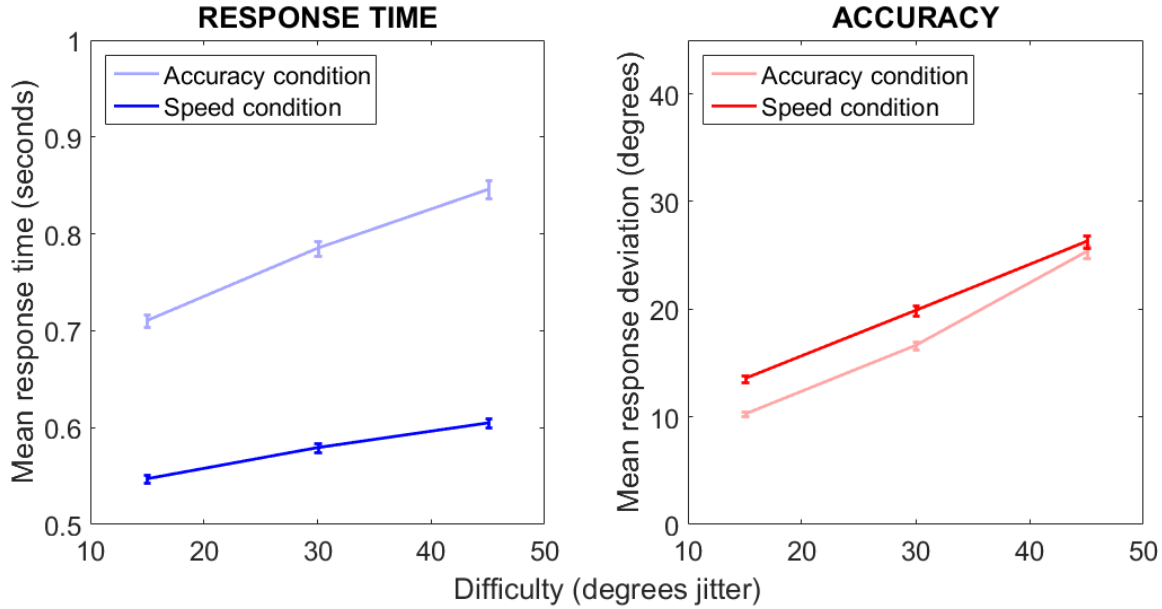


Figure 3.2 Response times (left) and accuracy in terms of degrees deviation from the correct response (right) across difficulty levels and speed-accuracy manipulation. Note that higher values indicate less accurate responses. Error bars indicate pooled standard error across participants.

in Figure 3.2 (left panel), and data for cue orientation conditions can be seen in Figure 3.3 (left panel).

3.2.2 Accuracy / response deviation

To predict the locations of participants' responses relative to the true mean stimulus orientation, I used a model that is highly similar to the one used for response times. However, rather than the mean location of responses – which is essentially zero, as they are generally centered on the correct response – each factor predicted the *variance* of the distribution of response deviations.² This gives us an estimate of how far away from the true answer we can expect responses to be, whether that deviation is in the clockwise or counterclockwise direction.

As before, these outcomes were allowed to change as a function of main effects of difficulty

²Formally, they predicted the log of the variance of this distribution in order to allow the sum of all factors to take any value and to make the predicted variances log-normally distributed. A similar modeling approach predicting the variance of a drift-diffusion process can be found in Kvam & Pleskac (2016).

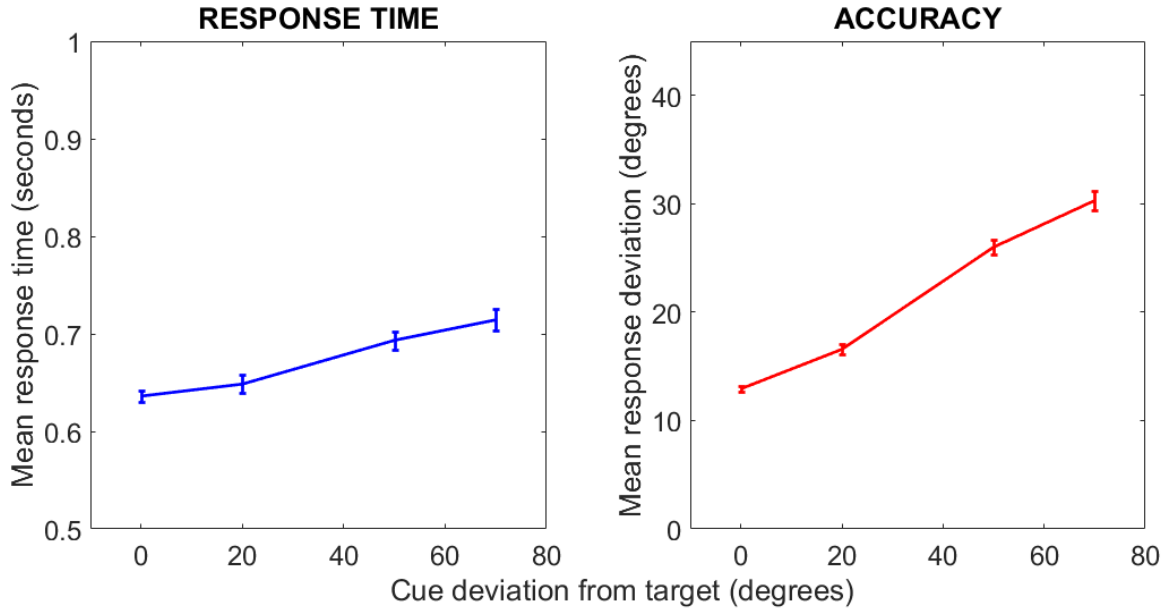


Figure 3.3 Mean response times (left) and absolute response deviation (right) by the orientation of the cue. Note that higher response deviations indicate less accurate responses. Error bars indicate pooled standard error across participants.

(linear), speed manipulation (binary), cue presence (binary), and cue orientation (conditional linear), as well as the interactions between all combinations of these factors. For simplicity, I discuss the results in terms of accuracy. This means that lower coefficients for effects on response deviation correspond to higher accuracy and higher coefficients correspond to lower accuracy.

Estimates for all coefficients are shown in Table 3.2. The main effects of manipulations were as expected, with difficulty, speed emphasis, and less accurate cues [orientation] all decreasing overall accuracy and the presence of accurate cues [cue] increasing accuracy.

However, unlike the response time results, the accuracy results were somewhat complicated by interactions between manipulations. The 2-way interaction between difficulty and speed emphasis as well as the 3-way interaction between difficulty / cue / orientation and 4-way interaction between difficulty / speed / cue / orientation, each of which increased accuracy, can be at least partially attributed to a floor effect. These manipulations put together would result in the lowest accuracy conditions – note the main effects of difficulty, speed, and cue orientation on accuracy – but average response deviations have a maximum value. Guesses will yield responses that are on average ± 45

Table 3.2 Mean estimates of coefficients for main effects and interactions of stimulus difficulty, speed manipulation, cue presence, and cue orientation on accuracy of responses. The ranges containing the 95% Highest Density Interval (HDI) are also provided. Intervals excluding zero are starred.

Coefficient	Mean Estimate	95% HDI
Difficulty	0.79	[0.70, 0.89]*
Speed	0.16	[0.07, 0.26]*
Cue	-0.31	[-0.49, -0.14]*
Cue $\times$ orientation	0.52	[0.40, 0.65]*
Difficulty $\times$ speed	-0.09	[-0.15, -0.04]*
Difficulty $\times$ cue	-0.24	[-0.46, -0.02]*
Difficulty $\times$ speed $\times$ cue	0.03	[-0.06, 0.13]
Difficulty $\times$ cue $\times$ orientation	-0.09	[-0.15, -0.03]*
Speed $\times$ cue	-0.05	[-0.21, 0.09]
Speed $\times$ cue $\times$ orientation	0.06	[0.00, 0.12]
Difficulty $\times$ speed $\times$ cue $\times$ orientation	-0.05	[-0.09, 0.00]

degrees away from the true orientation, so mean response deviations will generally not exceed these values in any conditions.

The interaction between difficulty and cue presence was the strongest effect of these and the only one to rival any of the main effects, but this interaction is simple to explain. When participants received an accurate cue, it could be used to make accurate inferences regardless of what was shown in the stimulus. Hence, the ill effects of difficulty were curtailed when a cue was present.

3.3 Modeling

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 180 degrees. For simplicity, assume that stimuli at orthogonal rotations (i.e. 45 / 135 degrees or horizontal / vertical) provide evidence against one another. Therefore, they can be arranged in a circle as shown in Figure 3.4.

The model assumes that participants complete the task by sampling orientation information piece by piece, and each piece of information they gather is assumed to be pulled from a von

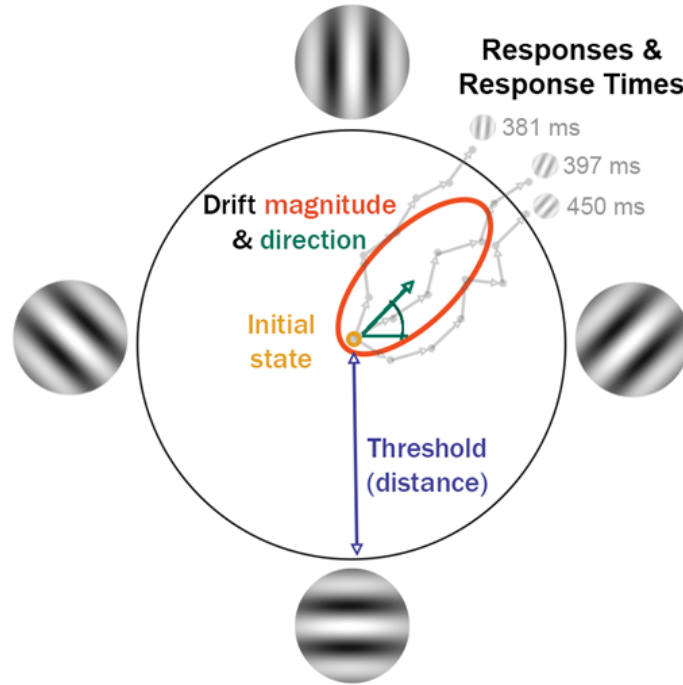


Figure 3.4 State representation, sampling, threshold, and evidence trajectories for the orientation matching task.

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.

With these assumptions, the model is functionally equivalent to the 2-dimensional random walk on a disc proposed by Smith (2016). This allows us to apply the analytic likelihood functions that Smith derived to the present task. The particular functional form of the model used here has 5 main parts: the drift magnitude $|\mu|$ (how quickly information in favor of the correct direction is sampled), the drift direction ρ (the mean orientation gathered from the stimulus or stimulus + cue), the threshold θ (how much information is needed to make a decisions), non-decision time ndt (the length of response times composed of processes other than the decision, such as perceiving the

³This 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 thesis.

stimulus and moving the mouse), and the initial position of the evidence. Unfortunately, due to the complexity of its likelihood function, the version of the model including variation in the initial state is intractable, so the initial state is set to the center of the response circle [0,0] and other parameters must compensate for its omission. The resulting likelihood function, covering the joint distribution of response times and responses $Pr(x,t)$, is given by Smith (2016) and shown below in Equation 3.1.

$$Pr(x,t) = \exp\left(\theta|\mu|(\cos(x)\cos(\rho) + \sin(x)\sin(\rho)) - \frac{|\mu|^2(t - ndt)}{2\sigma^2}\right) \cdot \frac{\theta^2}{\sigma^2} \sum_{k=1}^n \frac{j_{0,k}}{J_1(j_{0,k})} \exp\left(\frac{-j_{0,k}^2 \sigma^2 (t - ndt)}{2\theta^2}\right) \quad (3.1)$$

Because the parameters are non-identifiable when the noise parameter σ^2 is allowed to vary, it was set to 1. The function $J_1()$ is a first-order Bessel function of the first kind, and the elements $j_{0,k}$ are the zeroes of the zero-order Bessel function of the first kind. The series $\sum_{k=1}^n$ computes and evaluates the function for the first n zeroes. The true likelihood is given by an infinite series, $n = \infty$. However, for the practical purposes of this study, I found that $n = 151$ yields sufficiently precise approximations.

The remaining 4 parameters used in the likelihood – drift direction, magnitude, threshold, and non-decision time – vary from condition to condition based on the experimental manipulations. They must therefore be specified and estimated either freely or in some functional way across these conditions. Allowing each of these parameters to vary freely across conditions would yield 4 (parameters per condition) $\times$ 2 (speed / accuracy) $\times$ 3 (difficulty levels) $\times$ 8 (no cue + 7 cue deflections: $-70, -50, -20, 0, 20, 50$, and 70 degrees) = 192 free parameters. Instead of attempting to estimate them all simultaneously, I explore the manipulations in two stages in order to set the parameters progressively. First, I look at just the difficulty and speed-accuracy manipulations, fitting each condition independently. This allows me to explore what relationships each of these two manipulations has with drift, threshold, and non-decision time. In turn, these relations can be used to simplify parameter setting when we include cue manipulations in the model – for example, drift

magnitude appears to be close to a linear function of the standard deviation of the stimulus, so it can be set across conditions using just an intercept and slope parameter.

3.3.1 Parameter-free predictions

The model used here makes two predictions that can be tested without actually having to do any parameter estimation in the cognitive model. The first of these is the prediction that response times will not vary within a condition dependent on the magnitude of response deviations. This is analogous to the symmetric correct and error response time distributions predicted by 1-dimensional random walks. Introducing drift rate variability into the model will re-introduce a relationship between response time and deviation, so this test is designed to test whether such a component is necessary for the 2-dimensional model. I refer to the independence of response times from response deviations as “response time homogeneity” below.

The second prediction that can be tested without estimating the cognitive model’s parameters is that responses should be von Mises distributed when they arrive. This prediction could potentially be modified by introducing variability in the initial state or different sampling distributions to the evidence, but this may not be necessary. I refer to this property simply as “von Mises responses” below.

3.3.1.1 Homogeneous response times

I tested this assumption by looking at the correlation between response time and absolute response deviation. The group-level mean estimate (from a hierarchical Bayesian model) for the standardized correlation between response time and deviation, across all conditions and participants, was centered almost exactly at zero, $M(b_1) = 0.01$ (95% HDI = $[-0.04, 0.06]$). Within participants, there were occasional instances where response time was correlated positively with accuracy within the speed condition, and the HDI indicated that they could be greater than zero. These cases are likely due to these particular participants guessing in order to keep their response

times under 800 ms threshold required for speed-based rewards, and could be modeled as a contaminant distribution if one is particularly interested in the relationship between RT and absolute response deviation in each condition.

3.3.1.2 Von Mises responses

The second assumption made by the model, if there is a von Mises sampling distribution (or Brownian motion with drift) and no changes in the initial state of the process, is that all responses should be von Mises distributed when they hit the response threshold. Although testing this assumption doesn't require fitting the 2-dimensional diffusion model, it does require a descriptive model of the responses for each participant and condition.

To test this assumption, I estimated the maximum likelihood parameter values of a von Mises distribution for each condition and participant. This yields a central tendency or mean parameter μ and a concentration or variance parameter κ . I then used these maximum likelihood parameters to generate a random sample of 10,000 data points. In turn, I compared the randomly generated data to the actual data for each of the (12 participants $\times$ 3 difficulty levels $\times$ 2 speed / accuracy conditions $\times$ 8 cue conditions [7 cue orientations + 1 non-cue] =) 576 sets. The comparison was done using a Kuiper test (Louter & Koerts, 1970), the circular analogue of a Kolmogorov-Smirnov test. It tests the hypothesis that two independent samples were generated from the same underlying periodic distribution. In only 12 of the 576 sets (2%) was the hypothesis that the samples came from the same underlying distribution rejected at $\alpha = .05$ or lower.⁴ This is a reasonable rate to expect simply by chance, indicating that most sets of responses in the data can reasonably be considered von Mises distributed.

⁴At present there appears to be no Bayesian analogue to this test, so I ask the reader to forgive a mix of classical and Bayesian inferential statistics for the present circumstances.

3.3.2 Free effects model

The first cognitive model examines only non-cued conditions in order to establish relationships between the speed-accuracy and difficulty manipulations and the drift, threshold, and non-decision time parameters. There was no pre-stimulus information and no bias to respond clockwise or counterclockwise in the uncued conditions, so we can safely assume that participants were on average sampling orientation information based on what the stimulus indicated. Therefore, the drift direction was fixed for each trial at the true mean orientation of the stimulus.

Drift magnitude (the rate of sampling true information relative to random noise), threshold (the amount of required information), and non-decision time were permitted to vary freely across the 6 speed / accuracy and difficulty level conditions. Each of these parameters was estimated for each participant and condition (aside from cued ones) using maximum likelihood estimation. The resulting estimates for each participant, as well as the mean across these estimates, is shown in Figure 3.5

As expected, drift magnitude estimates decreased with increasing difficulty, indicating that less coherent stimuli were providing noisier information. Drift magnitude estimates were similar for speed and accuracy conditions as well, suggesting that participants were gathering information at roughly the same rate in both conditions. Importantly, drift magnitude decreases approximately linearly with stimulus difficulty – as a result, I assume a linear functional relationship between the two in the model predicting behavior the cued condition.

Threshold estimates were consistently higher in the accuracy relative to the speed condition, suggesting that participants were applying stricter criteria for their decisions to the evidence in the accuracy condition relative to the speed condition. This shift explains the longer response times and higher accuracy in the accuracy condition as shown in Figure 3.2. However, threshold estimates seemed to also decrease with increasing difficulty. While this was somewhat unexpected, similar results were found by Kvam & Pleskac (2016). The authors suggested that such an effect may be the result of increasing thresholds when high-quality information is available, suggesting

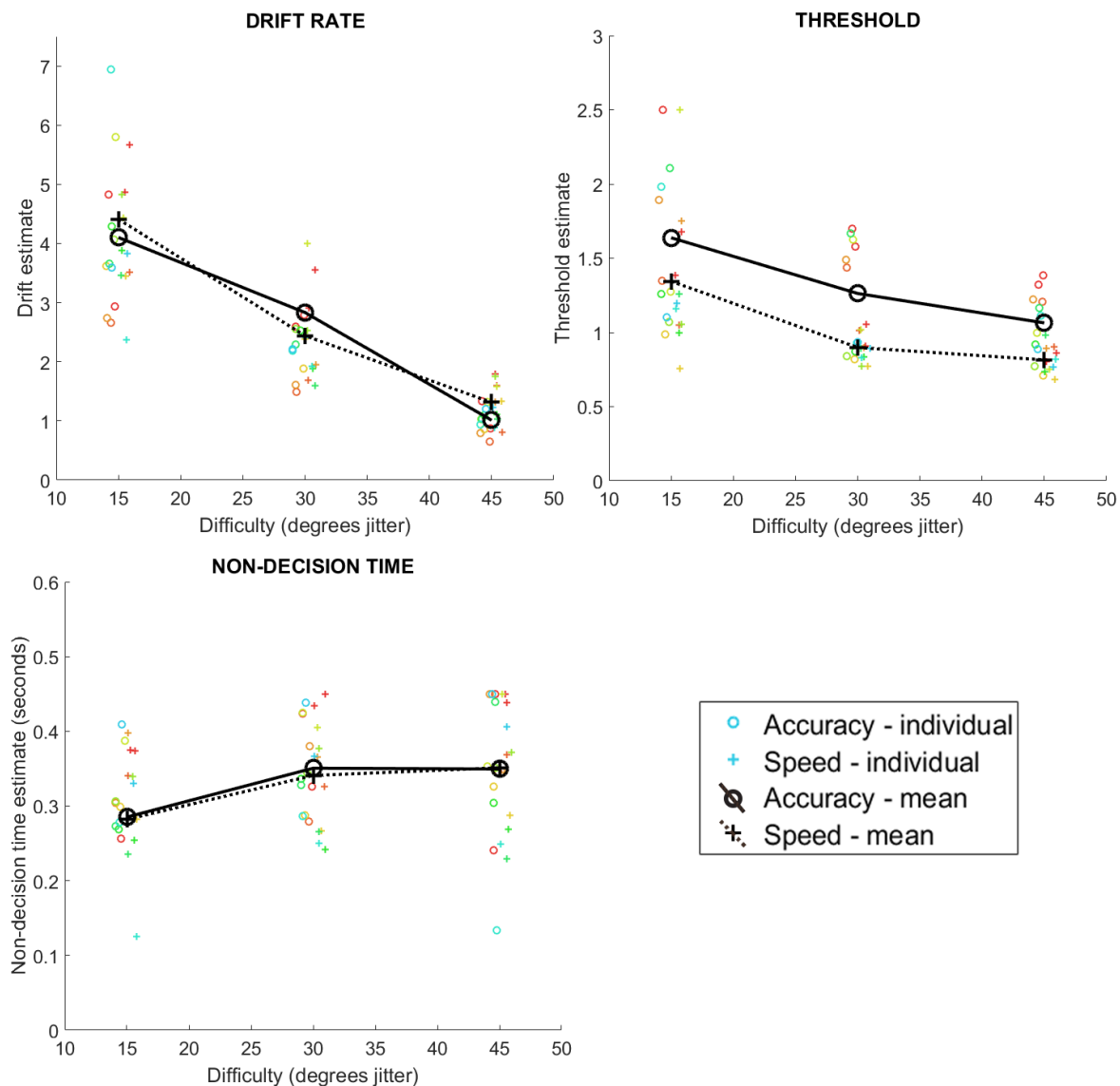


Figure 3.5 Drift magnitude (rate), threshold, and non-decision time estimates for individuals (colors) and the mean of the group (black) in the free effects model.

that participants are adjusting their choice criteria on-line. Alternatively, it could be the result of decision boundaries that collapse over time, indicating that participants are trading off accuracy for the amount of time it takes to finish a trial (Bowman et al., 2012; Drugowitsch et al., 2012; Ratcliff & Frank, 2012). Fortunately, fixed-boundary versus collapsing-boundary models predict highly similar distributions of response times (Voskuilen et al., 2016), suggesting that collapsing boundaries are unlikely to be a necessary component of the model at present. I therefore fix the

threshold of the model across difficulty levels in subsequent sections.

Finally, non-decision time appears not to change substantially between speed and accuracy conditions or across difficulty levels – any shifts are within approximately 40 milliseconds. I therefore fix non-decision time across conditions when modeling the effect of the cue manipulation.

3.3.3 Simplified cue-inclusive model

Given the results of the free-effects model and the uncued data, we can simplify the parameter space of the model. Instead of allowing them to be free across conditions, drift magnitude can set as a linear function of stimulus difficulty and threshold can vary between speed and accuracy manipulations. Non-decision time is fixed across conditions. Despite simplifying the model, even these assumptions don't buy enough simplicity unless we make some assumptions about how the parameters might change with the cue manipulations as well. Ordinarily, one might assume that initial state would be the parameter to change across cue manipulations, but the likelihood function including nonzero initial states is too complex to efficiently estimate the parameters of the model. Instead, the drift magnitude and drift direction are allowed to vary as linear functions of the cue angle. This shift in parameterization is akin to claiming that the cue is integrated as information along with the stimulus information as part of the decision process rather than as a fixed pre-stimulus bias. Its effects are not exactly the same as allowing initial state to vary, but this model is tractable and still interesting to examine.

Because there were 7 levels of cue orientations (-70 / -50 / -20 / 0 / 20 / 50 / 70), its effect on drift magnitude and drift direction were also linear functions. In total, this yielded 7 free parameters in the model. Drift magnitude had an intercept (m_0) and 2 slope parameters that set it as a linear function of the difficulty manipulation (m_1) and the cue orientation (m_2). Drift direction had an intercept (d_0) and a linear slope that depended on the cue orientation (d_1), the threshold shifted between speed (θ_s) and accuracy (θ_a) conditions, and non-decision time (ndt) was fixed to a single value across conditions.

As we might expect based on the results of the free effects model, estimates of the threshold

Table 3.3 Maximum likelihood estimates for the parameters of the restricted model.

Participant#	m_0	m_1	d_0	d_1	ndt	θ_s	θ_a
1	0.001	-0.057	0.139	0.004	0.360	0.947	1.305
2	0.001	-0.051	0.099	0.005	0.311	1.005	1.172
3	0.000	-0.036	-0.015	0.011	0.204	1.124	1.332
4	0.001	-0.038	0.024	0.005	0.344	0.874	1.425
5	0.010	-0.046	0.115	0.006	0.193	0.841	1.011
6	0.001	-0.048	0.011	0.004	0.204	1.115	1.203
7	0.001	-0.047	0.060	0.004	0.275	0.841	0.921
8	0.002	-0.040	0.198	0.002	0.127	0.961	1.203
9	0.001	-0.039	0.119	0.000	0.128	1.093	1.623
10	0.001	-0.044	0.099	0.010	0.240	0.848	1.657
11	0.000	-0.039	0.195	-0.004	0.124	0.963	1.198
12	0.001	-0.043	0.006	0.009	0.139	1.054	1.383

parameter were higher for the accuracy relative to the speed condition for every participant (see Table 3.3 and Figure 3.6, bottom right). Drift magnitude also still decreased with higher difficulty (Figure 3.6, top right), reflected in the negative slope of the linear function mapping difficulty onto drift magnitude.

The cue manipulation impacted both drift magnitude and direction. In the case of drift direction, nearly all participants saw the average information they sampled shift toward the cued direction, indicated by a positive slope (Figure 3.6, top left). Given that the cue was often informative, this is precisely what we should expect – it provided some information that was integrated with stimulus information, moving the drift away from center when the cue was not congruent with the stimulus.

Perhaps more interesting is the effect of the cue on drift magnitude, shown at the bottom left of Figure 3.6. A value of zero indicates that the drift was the same between the cued and non-cued conditions. In the cases where the cue indicated the correct answer, the drift rate was higher, reflecting the dual sources of true information. However, the drift magnitude was still higher in the case where the cue was 20 degrees off of the true stimulus orientation than in the uncued condition, indicating that the cue was helpful even when it was not precisely correct. The drift magnitude fell off as the cue became less informative, with 50 and 70 degree cue orientations hurting overall

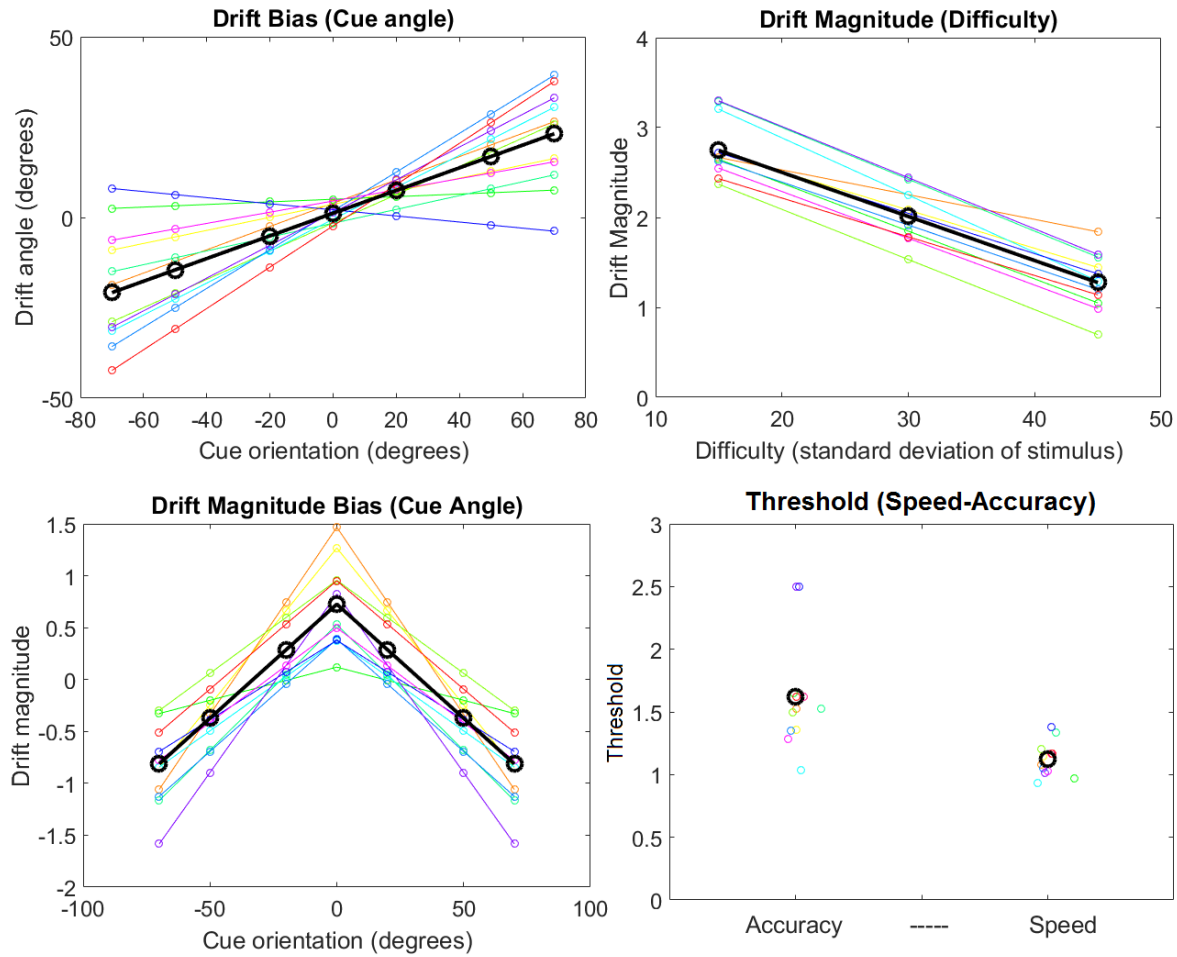


Figure 3.6 Effects of cue, speed-accuracy, and difficulty manipulations on drift magnitude, direction, and threshold for individuals (colors) and group mean effects (black).

performance.

One interesting characteristic of this particular relationship between cue orientation and drift magnitude is that it should be related to the tuning functions of orientation columns in the brain. Vertical columns will be most highly activated by vertical orientations of a cue, but they may also be activated somewhat by orientations of 5, 10, 15, or perhaps 20 degrees away from vertical (less and less as this discrepancy increases, of course). It may be the case that activity of the true stimulus direction may be facilitated by off-mean orientations simply because the tuning function of orientation columns overlaps with nearby (but not exactly matching) cue orientations.

3.4 Preliminary discussion

The results provide strong support for the presence for each of the three effects: speed or accuracy incentives produced a trade-off between response time and response accuracy, stimulus difficulty produced slower and less accurate responses, and predecision cues yielded faster and more accurate responses when they indicated the true orientation of the stimulus but slower and less accurate responses for cues that were less accurate.

Given the prevalence of analogous phenomena in binary choice, we might expect that a model with similar structure to diffusion or accumulator models would be well-suited to describing behavior in continuous-response tasks as well. This is reflected in how the parameters of the model change with manipulations. As in binary choice, speed-accuracy manipulations yield shifts in thresholds and difficulty manipulations yield changes in drift. Though it was unfortunately not possible to examine how initial states change with predecision cues, the drift magnitude and direction did, indicating that cues may have a similar effect as well.

In addition, two a priori predictions of the 2-dimensional diffusion model were validated as well. Across the various conditions and participants, error magnitudes were largely unrelated to response times, indicating that there is no need (yet) for modifications like drift rate or starting point variability in the model. Furthermore, responses on the task appeared to be approximately von Mises distributed, as the specified sampling distribution suggests.

Overall, this suggests that the recent proposals of Smith (2016) and Kvam (2016) provide a solid foundation for modeling how response times and responses on a continuum are distributed. Drift magnitude, direction, threshold, and non-decision time seem to be informative parameters for describing the decision process in such cases.

However, this does not quite give the full picture. An important feature of the geometric framework is that the relationships between response alternatives are built into the decision space. The relations between response options in this study are relatively trivial – orientation can be defined in terms of degrees and easily arranged on a circle. But what about more complex stimuli that do

not have an easy physical to psychological mapping? I examine this question in a second study, presented in the next chapter.

CHAPTER 4

STUDY 2 - EXPLORING RELATIONS BETWEEN ALTERNATIVES

The second study was aimed at showing the potential benefits of using the geometric framework for modeling decision-making among interrelated alternatives. It focused on how the available alternatives – rather than manipulations of the stimulus, cues, or rewards – affected the decision process. To investigate their contribution, I manipulated both the number of alternatives available to participants as well as the alternatives' (similarity) relations to one another.

This study also used perceptual stimuli, but investigated the potential sources and caveats of Hick's Law by manipulating the number of alternatives and their similarity to one another. In addition to looking at discrete numbers of alternatives, this study examined how these conditions related to one where participants had a continuum (a theoretically but not practically infinite) of alternatives from which to choose. Recall that Hick's law does not easily extrapolate to a continuous-response condition because such analog scales don't lend themselves easily to bit-based predictions; a major goal of this study was to uncover how it relates to discrete-choice conditions and how (if) Hick's law holds.

Beyond just response times, this study also looked at how accuracy changed with the number and assortment of alternatives. This gave 4 major outcomes to examine: 2 manipulations (number of alternatives, similarity between alternatives) $\times$ 2 outcomes (response times, accuracy). Each of the 4 patterns of results constituted a hurdle that a model of decisions among many alternatives should be able to clear. The modeling section focuses on what components of the framework – and what additions and modifications to it – allow it to account for variation in behavior on all 4 counts.

This section concludes with a review of the major empirical findings and model components suggested by Study 2. I focus in particular on how decision models and representation models can be used to inform one another, providing the groundwork for a method of constructing a decision space that relates the available choice alternatives.

4.1 Methods

Study 2 examined how participants' psychological representations of colors related to their physically defined hue. It consisted of two related tasks, which were run within participants in order to model each person's behavior individually. As such, there were few participants but a very large amount of data gathered from each one.

4.1.1 Participants

A total of 6 Michigan State University graduate students (4 female, 2 male) each completed 5 sessions of a decision task and 1 session of a similarity rating task, described below. All participants were 22-30 years old. They completed approximately 500-1000 practice trials of the decision task, 1400-2300 full trials of the decision task, and 435 trials of the similarity rating task. Each session took approximately 1 hour to complete. Participants were paid \$10 per session for participating, and informed of their average accuracy at the end of each session.

4.1.2 Materials

All study stimuli were generated and presented in MATLAB using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007). Analyses used the machine learning and circular statistics toolboxes (Berens, 2009). All responses were recorded from the mouse.

4.1.3 Decision task

The decision task used in Study 2 is shown in Figure 4.1. In this task, participants viewed a display of 78 dots scattered around a disc of approximately 10 visual degrees in diameter. The dots varied in hue such that no two colors had a hue that was within 0.04 units of one another (with hue ranging from 0.0 to 1.0, wrapping around such that $0.0 = 1.0$). The saturation of the dot colors was set to 1.0 and the value was set to 0.8. The display had a single dominant dot hue, of which there were

always exactly 18 dots. The dominant dot color was always a hue that exactly matched a hue in the available choice alternatives.

On top of the dominant dot color, there were 15 other dot hues present in the display, with 4 dots in each of these hues. In each block of trials, there were 8 hues that would appear on every trial whether they were the dominant color or not. These hues were partially determined by the available alternatives. For example, when there were 8 choice alternatives, each of their 8 hues would appear on every trial of the block. If there were 3 alternatives, those 3 hues plus 5 more fixed hues would appear on every trial, and so on. Within a block, the correct alternative was always one of these 8 fixed hues, even in the continuous condition (where the 8 block-fixed hues were generated randomly at the start of the block). The remaining 8 non-fixed hues present in the dot display were drawn randomly on every trial, again with the restriction that no pair of hues in the display be closer than 0.04 hue units apart. The non-fixed hues were never the target color, so they served strictly as distractors or noise in the stimulus. In total, this yielded the 78 (18 target + 7×4 non-target fixed + 8×4 non-target random hues) dots in the display.

Each block consisted of 30 trials of the decision task, plus practice (described below). Within a block, the set of alternatives was fixed. If there were 3 alternatives, the same 3 would be present on all 30 trials. A random alternative out of those available would be the dominant color on each trial. It is possible that this may have made decisions easier in the conditions where there were fewer alternatives, as this narrowed down the colors that a participant needed to attend, but it ensured that there was always a match between the true answer and the available ones.

The alternatives available to a participant were placed around the edges of a circle at approximately 20 visual degrees from the center, as shown in Figure 4.1. Each alternative took up a 14-degree arc along the edge of this circle in the discrete condition (top / left panels). In the continuous condition (bottom right panel), a hue circle was shown where every degree of the circle was a different hue, approximating a continuous gradient of hues.

As in Study 1, a participant began each trial by clicking within a small white circle in the middle of the screen, at which point their mouse cursor was centered. The alternatives for a block were

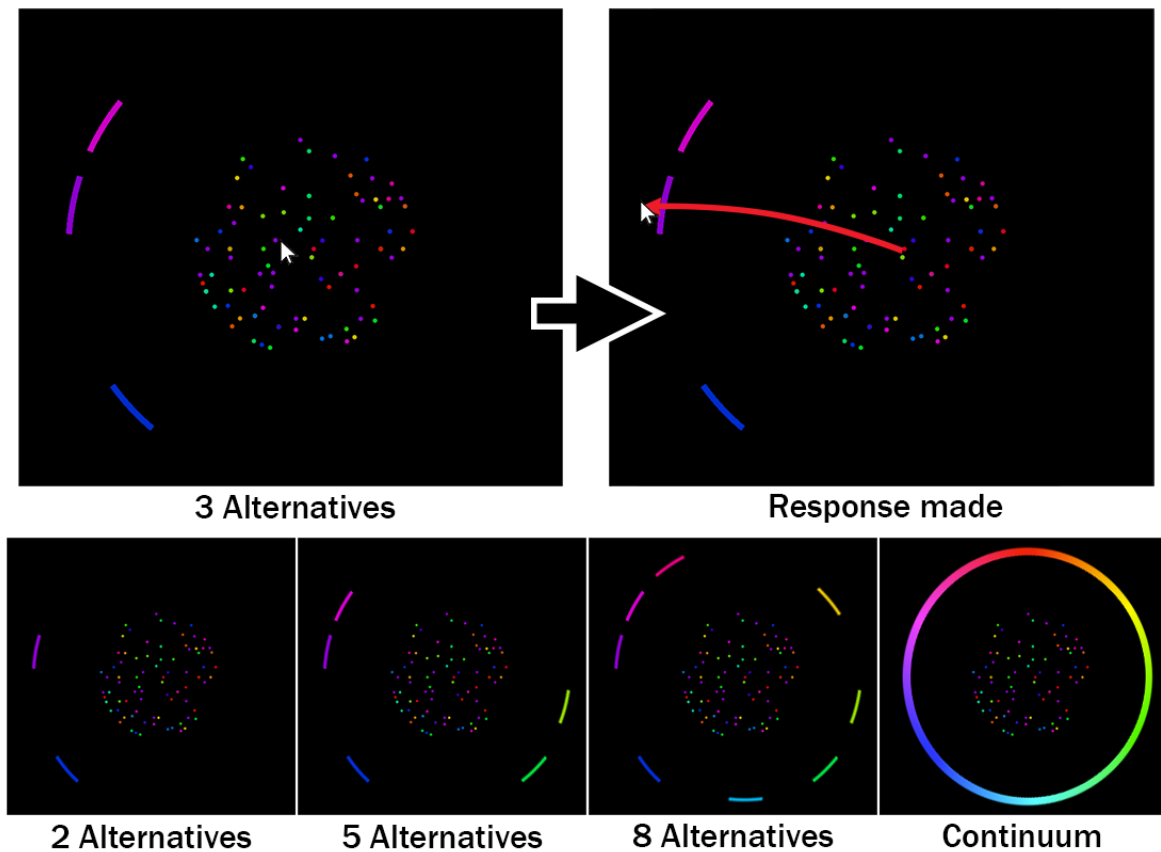


Figure 4.1 Diagram of the decision task in Study 2. Depending on the condition, these will be comprised of 2, 3, 5, 8, or a continuous span of alternatives. Participants responded by moving their mouse across the arc corresponding to the desired response alternative.

always on screen, but the stimulus did not appear until they clicked within the white circle. Once a trial had started, a participant entered their response by moving the mouse across the edge of the circle on which the alternatives were drawn. As soon as the mouse cursor crossed this boundary, their response time was recorded and their response was graded as correct or incorrect. In order to match the accuracy criterion across all conditions, responses were considered correct if they were within 7 degrees of the center of the location of the true dominant dot color. In the discrete case, this meant that responses were correct if they crossed the arc colored in the true dominant dot hue. In the continuous condition, it simply meant that responses that were within ± 7 degrees of the true dominant hue were considered correct. Participants were informed of this grading criterion prior

to beginning the study.

Within a session, the number of response alternatives in a block was evenly split between 2, 3, 5, 8, and a continuum of alternatives. The hue color wheel and the locations of alternatives were rotated randomly for each session, but kept constant within a session. This was done to ensure that response times were not unevenly distributed across hues due to their spatial locations on the screen.

At the end of each decision trial, participants would see feedback on whether or not their choice was correct by receiving 100 (correct) or 0 (incorrect) points for that trial. Similar to Study 1, ballistic mouse movement was encouraged by penalizing participants for straying between the dot display and available alternatives. This penalty was 1 point per every 20 ms after the mouse cursor was in between the dots and response alternatives for 300 ms.

4.1.3.1 Practice trials

In order to ensure that response times were affected as little as possible by practice effects, the physical locations of alternatives, and the time it took to make a ballistic movement to the edge of the circle, there were a large number of practice trials that preceded every block of decision trials. Each practice trial was similar to the decision trials, except that a single large, colored dot was shown rather than a noisy multicolored dot display. Therefore, instead of picking the dominant dot out of the display, participants simply had to match the color shown to the alternatives available. As in the decision task, their accuracy and response speed were gathered and stored.

The alternatives shown during the practice trials were the same as those in the succeeding decision trials. One of the goals of the practice was to make sure participants knew exactly where each of the alternatives was. Therefore, for each choice option present in the display of alternatives (2, 3, 5, or 8 for the discrete conditions), a participant saw at least 3 instances of the corresponding color appear in the center for them to match. In the continuous condition, they saw 10 or more random hues appear in sequence in the center. Anytime a participant made an incorrect assignment during the practice trials, an additional practice trial was added.

After each of the practice trials, participants would receive immediate feedback on the accuracy of their selections, including the hue they chose, its location on the screen (in terms of degrees around the circle), the correct hue, the correct hue's location on the screen (its center), and how far away in degrees their response was from the center of the correct hue.

4.1.4 Similarity rating task

Prior to the sessions with the decision task, participants completed a similarity rating session where they would see pairs of colors in the middle of the screen and a scale around the edges (see Figure 4.2). Their task was simple – they simply had to compare the two colors on the screen and assign a value from 0 (opposite) to 100 (identical) indicating how similar to one another they thought the colors were.

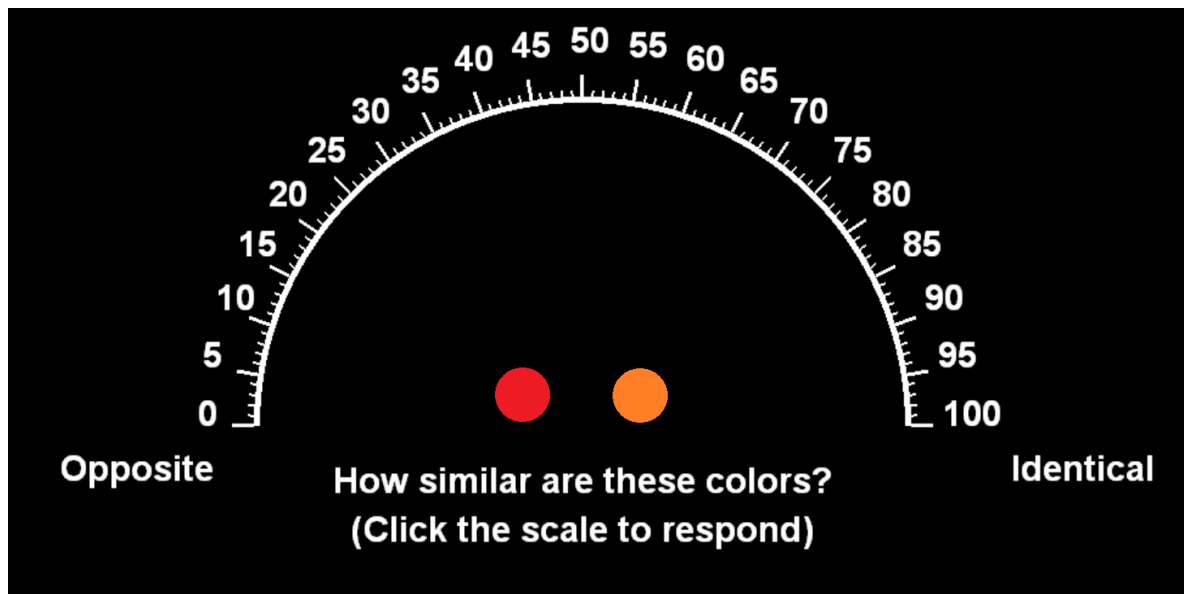


Figure 4.2 Layout of the similarity rating task.

The colors themselves were 30 hues equally spaced along the color wheel. Participants were presented with each possible combination of 2 non-identical colors exactly once, giving a rating for each pairwise comparison. These pairwise ratings were used to populate the upper diagonal of a similarity matrix, which was used to generate a multidimensional scaling [MDS] solution that

arranged the colors in 2 dimensions. This procedure is described in detail in the corresponding modeling section.

4.1.5 Procedure

After completing informed consent, participants were placed in a dark, windowless office and completed the similarity rating task for the duration of the first session. On later dates, they completed 5 sessions of the decision task.

The similarity rating task was self-paced, so that participants could take as long as they wanted to make exact similarity judgments and take breaks as they needed. They were encouraged to take a consistent amount of time on each trial and to make sure that their judgments were consistent (i.e. a rating of 30 should indicate that a pair of colors is more similar than 25, regardless of scale). Participants received \$10 for completing the similarity rating task and for each session of the decision task.

During the first session of the decision task, the experimenter would demonstrate how to perform the task in both discrete and continuous conditions, emphasizing that mouse movements should be consistent and ballistic – i.e., that participants should not move the mouse until they were ready to respond, but move it directly to the alternative when they were. In addition to their initial briefing and demonstration, participants completed an extra 30 practice trials at the outset of the first session that covered all numbers of alternatives they might see during the task. In the first session, they would then complete 10 or 15 blocks (dependent on time) of the decision task, including practice trials. In subsequent sessions, they would complete 15 or 20 blocks of the decision task, including practice trials.

Once all 6 sessions were completed, participants were debriefed on the purpose of the study and permitted to see some of the results of their performance, if desired.

4.2 Results

As in the previous study, the results are broken into descriptive and process / cognitive modeling sections. Both types of models examine accuracy as well as the amount of time it took participants to make a response. Recall that in the dot display task, a correct answer involves moving the mouse across the arc corresponding to the most prominent dot color in the discrete conditions or within 7 degrees of the most prominent dot color in the continuous condition. Any response not included in these ranges were counted as incorrect, including ones that did not correspond to any available alternative in the discrete condition (i.e., if their mouse crossed the circle on which alternatives were placed, but not in a location that had a color there). Response time was simply the length of time in seconds between when the stimulus appeared on the screen and when a participant's cursor crossed the response arc.

The initial descriptive and inferential statistics examine how different manipulations of the number of response alternatives, the combinations of alternatives, and the discriminability of these affected participants' response times and accuracy on the task. They utilize hierarchical Bayesian methods to estimate group-level relationships between these manipulations and outcomes.

4.2.1 Number of alternatives

A straightforward first question regarding the results is simply how response times and accuracy relate to manipulations of the number of response alternatives available. The relationships between them are shown in Figures 4.3 and 4.4.

Clearly, mean response times increase with the number of alternatives and mean accuracy decreases with the number of alternatives if we look simply at the discrete conditions. However, the precise relationship between the number of alternatives and mean response time does not necessarily follow a $\log_2$ relationship as Hick's law predicts. In fact, a model with a linear functional relationship between the number of alternatives and response times fit better than one which predicted response times as the $\log_2$ of the number of alternatives, $DIC(linear) = 22176 < DIC(\log_2) =$

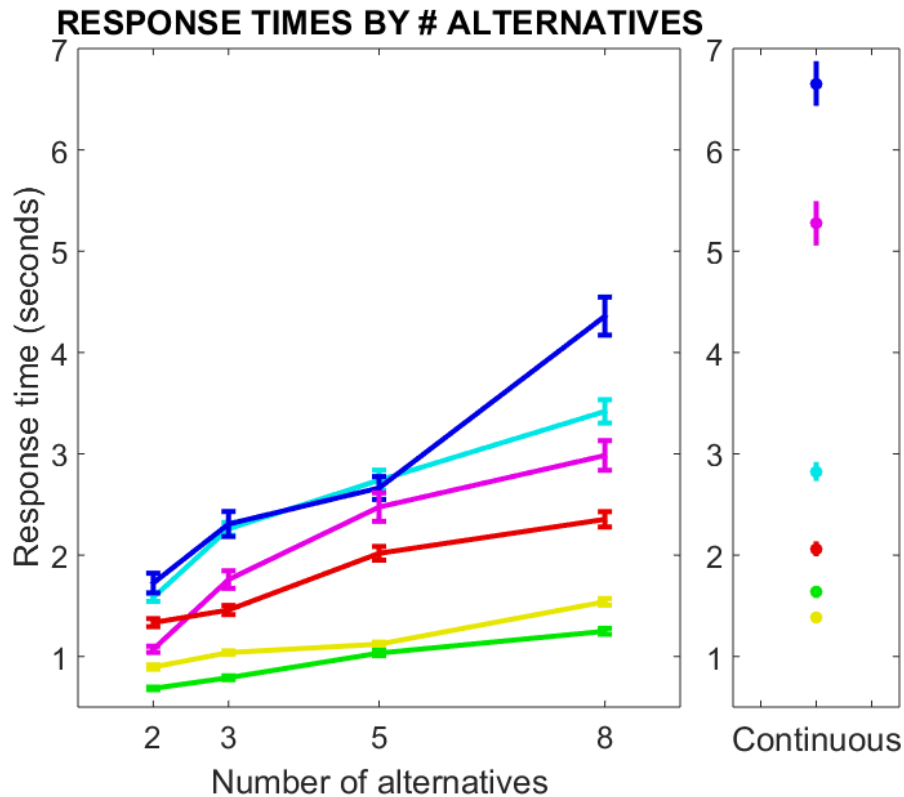


Figure 4.3 Relationship between the number of alternatives and mean response times in the colored dot mixtures task. Mean response times for the continuous condition (for comparison) are shown on the right. Each color corresponds to a different participant.

24290. Of course, some participants (see Figure 4.3) still seem to have negative curvature in how their response times increase with the number of alternatives.

Perhaps more problematic for Hick's law is the pattern of response times that appears in the continuous color wheel condition. For half of the participants, mean response times in the continuous condition are contained within the range of those produced by discrete numbers of alternatives (usually between 5 to 8). In both the linear and log-linear cases, one would predict much longer response times in the continuous than in any of the discrete-alternative conditions simply because there are many more options available. In order to fit the results of the continuous case within the bounds of Hick's law, we would be forced to make the claim that there was some "practical" number of alternatives that participants were using in the continuous case, and that this practical number was frequently fewer than 8 (often about 7 alternatives).

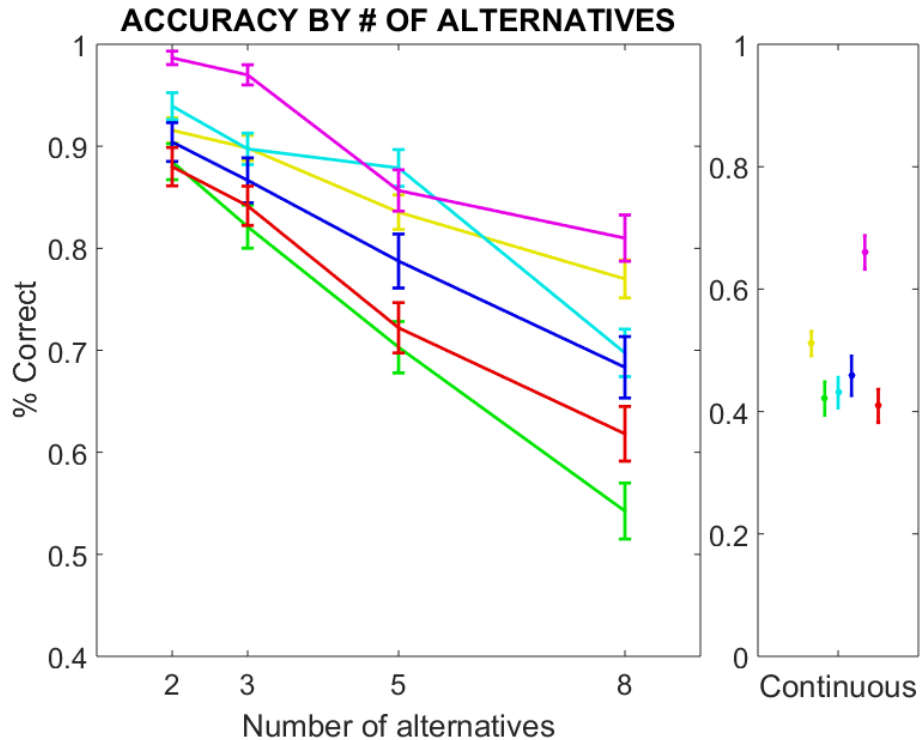


Figure 4.4 Relationship between the number of alternatives and accuracy in the colored dot mixtures task. Accuracy for the continuous condition (for comparison) is shown on the right. Each color corresponds to a different participant.

The superior performance achieved by some participants in the continuous case relative to the 8-alternative one is particularly curious. It suggests that the strategy participants implement in the continuous condition is actually more effective than the one they use for discrete numbers of alternatives. Indeed, ignoring the available alternatives and instead applying the continuous-response strategy for the 8-alternative condition would yield performance that is at least as good as in the continuous condition. It may be the case that there is a mapping component (from evidence onto available alternatives) that is introduced in the 8-alternative condition that slows down participants' responses.

Alternatively, participants could be setting different thresholds for evidence across different conditions. This would suggest that participants trade speed for accuracy between the 8-alternative condition and the continuous one, meaning that those participants with shorter response times in the continuous condition would have lower accuracy, and those with longer response times would have

higher accuracy (as in Study 1). There is some suggestion that this may be a plausible explanation, as those participants whose response times in the continuous condition were shorter than in the 8-alternative condition showed lower accuracy in their responses in the continuous condition than the 8-alternative one (see yellow and cyan participants in Figures 4.3 and 4.4). I investigate this possibility further in the Modeling section.

4.2.2 Conflation with difficulty

One issue with examining differences between the number of alternatives across conditions is that the number of alternatives in these (and most other) tasks is conflated with how easy it is to distinguish between the alternatives. Anytime there is a fixed range from which alternatives can be drawn – such as within the hue wheel, audible sounds, visible colors, the size of the screen, or with any artificial bounds on stimuli dimensions – a greater number of alternatives will result in more ‘crowding’ in local areas of the space of alternatives. To illustrate this point, consider the case where we draw uniform random variables on $[0,1]$. When we draw 2 random variables, each one will be about 0.33 units away from one another. With 3 random draws they would be on average about 0.25 units away from one another, with 4 draws we would expect them to be 0.20 units away, and so on. As the number of random draws (alternatives added) increases, the space is divided into smaller and smaller parts, with each draw being closer to its neighbors on average.

The implication for analyses based on the number of alternatives is quite important. If there is a greater number of alternatives, we can expect that any given target alternative will have more distractor alternatives in close proximity than when there are fewer alternatives. It is therefore more difficult to discriminate targets from distractors when there are more alternatives, meaning that accuracy should be lower and response times higher when a greater number of alternatives are available (as I showed with the difficulty effect in Study 1).

The difference in discriminability between conditions for this study is shown in Figure 4.5. As we might expect, the distance between the target hue and its nearest distractor decreases with the number of alternatives. Note that because each available hue alternative in the discrete condition

spanned 7.2 degrees in either direction, no pair of alternatives was closer than 0.04 units of color hue distance (14.4 degrees). This yields the value for the continuous condition as well – the nearest distractor was taken to be the nearest hue that would be counted as ‘incorrect.’

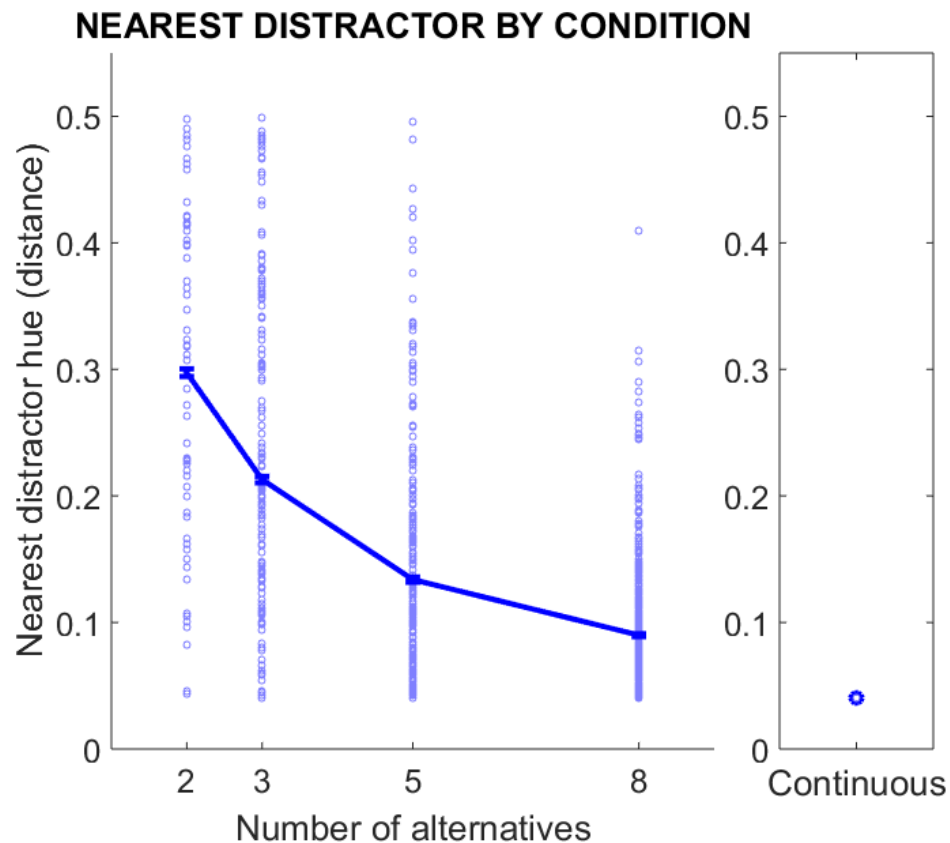


Figure 4.5 Relationship between the number of alternatives available and the discriminability of the target alternative (hue-based distance to the nearest distractor).

Ultimately, there is a strong case to be made that the discriminability of alternatives is largely responsible for the relationship between the number of alternatives and decision makers’ accuracy and response times. This study was not designed specifically to investigate this issue, but it certainly hints that this is a line of investigation worth pursuing. If increasing the number of alternatives while maintaining their discriminability results in consistent response times and accuracy, then the number of alternatives is not actually an important independent factor in all decisions. Instead, Hick’s law may arise from ‘downstream’ factors like the difficulty of tasks featuring different numbers of alternatives.

Despite this, there still seems to be a general negative trend in accuracy and positive trend in response times as a function of the number of response alternatives available. As I suggested before, this could potentially be due to varying decision thresholds that people set across different conditions – a possibility which I investigate in the modeling section.

4.2.3 Multidimensional scaling

The hue of colors as shown on the color wheel is a case where the physical distances (in hue units) do not map directly onto the psychological representations of the stimuli. For example, the distance between two shades of green may be the same as the distance between a green and a yellow, but the latter will almost certainly be more distinguishable. This suggests that there is an additional component that ought to be added to the model in order to map the physical relations between alternatives onto psychological ones. But first, it is worth visiting the effects that physical-psychological differences introduce into the choice data.

The mean accuracy and response times for Participant 5 as a function of the most prominent dot hue in the display (target) are shown in Figure 4.6. Note that both shift massively depending on what the most prevalent dot color was in the display, and in reverse to one another – when accuracy decreases, response time increases and vice versa. This suggests that decisions for this participant tended to be more difficult when the most prominent dot color was (e.g.) green, and easier when it was (e.g.) purple.

The difference in difficulty between these conditions suggests that some hues were easier to discriminate from their neighbors than others were. Green colors were hard to tell apart from nearby hues, and purple ones were easy to tell apart from nearby hues. Notably, the troughs in accuracy (and corresponding peaks in response times) tend to correspond to the locations of two of the three main colors of cone receptors in the retina: green and blue. In other participants, similar troughs tended to appear around red as well (see Figure 4.12). This is likely because colors are hard to discriminate at the peak ranges of these receptors, as the difference in activation (and therefore ease of discrimination) between two colors depends on the slope of the activation function for red,

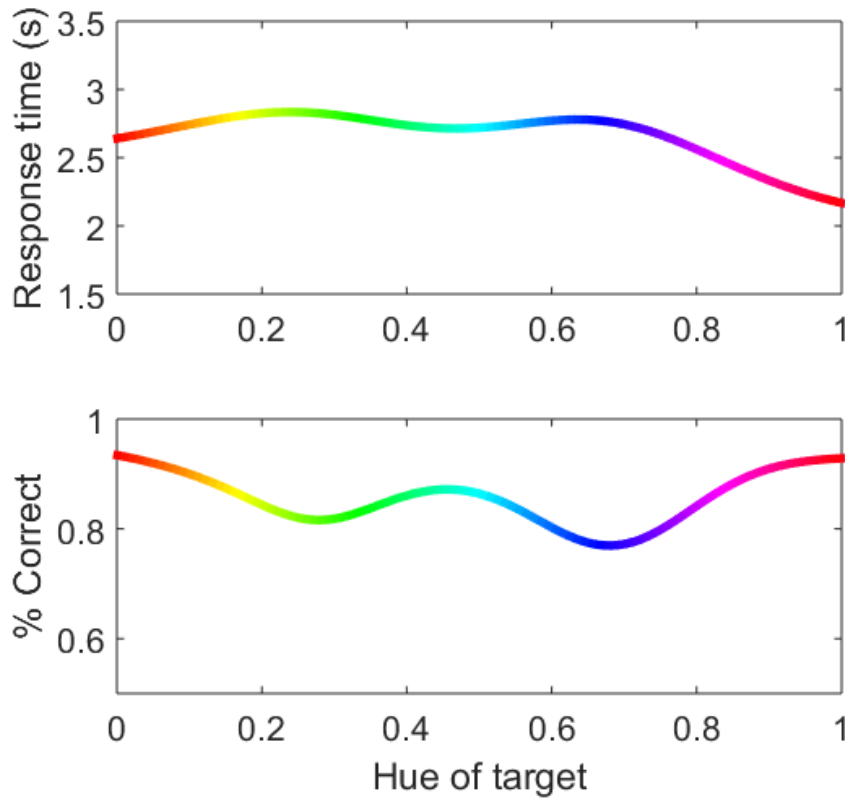


Figure 4.6 Mean accuracy and response time and accuracy for Participant 5 conditioned on the hue of the target color. The colored line is given by a kernel density estimator (bandwidth = 0.1) passed over the corresponding hue $\times$ accuracy or RT data.

green, and blue detectors. The slope is smallest at the peak activation levels, and therefore the distance between two hues will yield the smallest difference in activation near the center of these three main receptors.

These differences in discriminability from the decision task are reflected in the results of the similarity rating task as well, giving independent confirmation of (and capacity to estimate) differences in difficulty across target hues. For each participant, the similarity data was used to construct the upper diagonal of a 30×30 similarity matrix. These values were scaled to between 0 and 1 (from 0 to 100) and then fed into a stress-based 2-dimensional multidimensional scaling [MDS] process.¹

¹Formally, this minimized Kruskal's normalized stress criterion, a process which is implemented in MATLAB using mdscale.

The results of the MDS procedure are shown in Figure 4.7. The distance between any pair of points in these diagrams corresponds approximately to the rated dissimilarity between them. Recall that the hues were originally sampled uniformly across the color wheel, so clustering of hues indicates that people rated clustered hues as more similar than non-clustered ones.

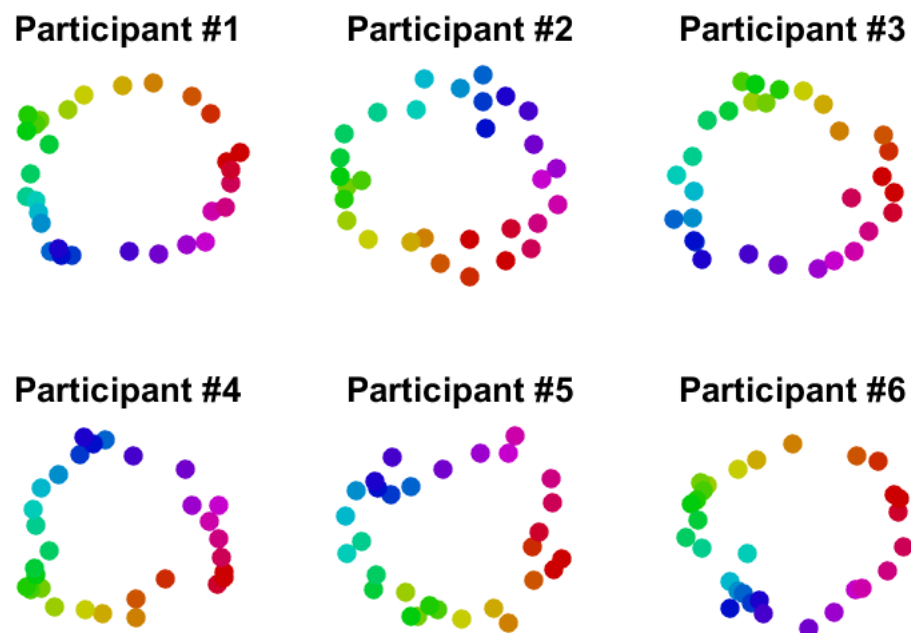


Figure 4.7 Results of the MDS analysis of the similarity task included as part of Study 2 for each individual participant. The distance between any pair of hues is the approximate dissimilarity between the hues according to that participant.

The most notable clusters of hues occur toward the middle of the green, red, and blue ranges. This indicates that participants felt that these hues were more similar to one another than to other adjacent hues, and reflected their inability to distinguish between these hues relative to 'boundary' colors between the three major hues. These differences in discriminability may be related to language as well as the major cone tuning frequencies (Heider, 1972; Witthoft et al., 2003), but the influence of both of these contributions should be reflected in both the multidimensional scaling output as well as decision data.

The MDS results were used to inform the decision model by using it to estimate the difficulty of each dot set of response options. The metric distance between a pair of colors in the MDS space was taken as an independent source to predict the discriminability of a pair of colors. The locations of hues, including both target colors and distractors, were imputed using cubic spline interpolation (De Boor, 1978) based on the 30 example hues that were included in the similarity rating task.² This yielded a location for each hue used in the study, and allowed us to compute the distance between the target hue and distractor hues in the MDS-generated space. This distance was taken as the inverse discriminability of a target from its distractors. In turn, the discriminability was used as a predictor of the drift magnitude and threshold in the cognitive model below.

4.2.4 Summary of important effects

Across the second study, there were a few main results that stood out as particularly important for a model of the decision process to capture. The first of these was the positive relationship between the number of response alternatives and mean response times (Figure 4.3). For all participants, regardless of the particular functional form, there seemed to be a monotonically increasing relationship between number of alternatives in the discrete conditions and mean response times. Whether this can be attributed to the difficulty of the task is an open question, but the pattern of results is clear.

The second thing to note is that there is a decreasing relationship between accuracy and the number of response alternatives (Figure 4.4). Like mean response times, this could be at least partially attributed to conflation of the number of alternatives with task difficulty, but a model should be able to account for this pattern as well.

Third, the heterogeneity of response times as a function of the target hue (Figure 4.6, bottom panel) is a particularly striking feature of the data. This may be due to shifts in drift rate (because

²Note that this included all colors used in the study – since hue values were continuous, the colors shown in Figure 4.7 were never perfectly replicated in the experiment. The spline interpolation method is executed within MATLAB's spline function.

of the discriminability of alternatives) or threshold (also based on discriminability) in the model.

Fourth, the heterogeneity of accuracy as a function of the target hue (Figure 4.6, top panel) should also be explainable by the model. However, it cannot be explained simultaneously with the heterogeneity in response times by having thresholds change across target hues (or discriminability) – an increase in threshold would see accuracy increase and response times increase, but accuracy appears to increase when response times decreases. This suggests that drift rates are likely to be involved in explaining both effects rather than threshold, but shifts in threshold could be at play as well.

One final explanation for the accuracy-related effects could be that participants are using different decision boundaries across decisions. This is discussed in the response boundaries section below.

4.3 Modeling

The basic elements of the 2-dimensional diffusion model used in Study 1 were also used in the models of Study 2. Its interpretation is essentially the same: at each point in time, a participant gathers information from the stimulus display that indicates the presence of a particular dot color. As they view the stimulus, they gather many of these samples and integrate them over time. Once the participant has enough information (the distance from the evidence state to the center / unbiased initial state is large enough), a decision is triggered in favor of the dot color the participant favors at that point in time. This gives rise to a response which depends on the match between the available alternatives and the participant's current beliefs about the most prominent color in the display. Finally, the time at which the decision process crosses the boundary defined by the threshold, in addition to the time taken for non-decision components of making a response, gives rise to a participant's overall response time for the trial.

The model again predicts both responses and response times. As in Study 1, the decision process can be defined chiefly by the drift magnitude, initial state, threshold, and non-decision time.

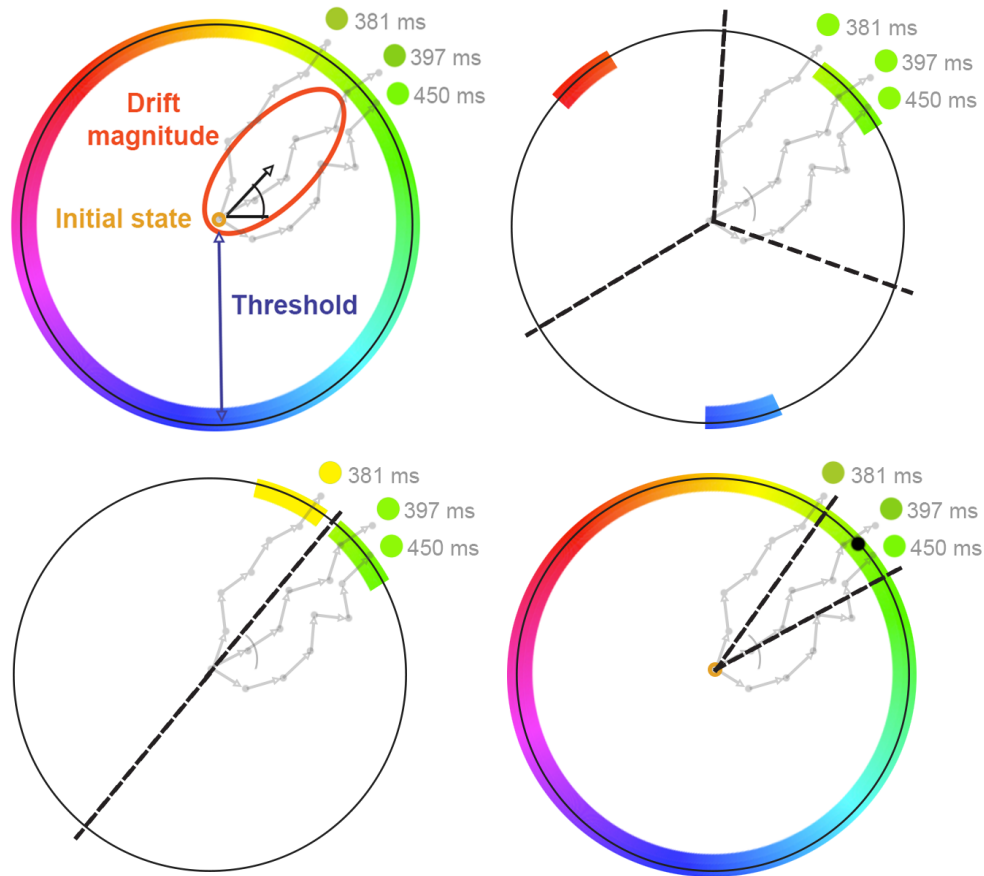


Figure 4.8 Basic elements of the models used in Study 2. In addition to drift magnitude, threshold, and non-decision time, decision boundaries separated the different responses into correct and incorrect responses. These bounds were assumed to be optimal, so that they were halfway between the available alternatives in the discrete choice cases (2 and 3 alternatives shown at the bottom left and top right, respectively). In the continuous case, they were set to ± 7 degrees from the most prominent dot color, to reflect the criterion of ± 7 degrees tolerance for responses in the continuous condition.

As before, the initial state was set to $[0,0]$. This was partly done in order to ensure computational tractability, but since there were no pre-stimulus cues in this study there was also no reason to assume a biased initial state. For this study, the drift direction was assumed to be centered on the true dominant dot color present in the stimulus display.³ A diagram of this model is shown in

³This assumption could be made because there was no pre-stimulus cue as to the most prominent color, though a fair argument could be made for allowing it to vary across conditions. Notably, it could be adjusted in the discrete-choice case to reflect the fact that the available alternatives actually gave some indication of what the most prominent dot color could be.

Figure 4.8.

4.3.1 Response boundaries

Because Study 2 classified responses as either correct or incorrect, the model also includes a mechanism for transforming the location at which the evidence trajectory crosses the threshold into a correct / incorrect diagnosis. For example, when a participant has to choose between green and yellow responses (bottom left of Figure 4.8), they may hit the threshold with evidence that favors an ‘orange’ response. This must be mapped onto one of the available alternatives, so the participant must apply some criterion in order to map it onto either green or yellow. To do so, the model assumes that responses are given based on the alternative that is most similar to the represented evidence at the time it crosses the decision threshold. In essence, decision boundaries are placed evenly between the available response alternatives on all discrete-choice trials. For continuous trials, these response boundaries are not necessary, but responses can be classified as correct or incorrect based on whether or not they are within 7 degrees of the true dominant dot color. Therefore, decision boundaries can be placed at ± 7 degrees from the true dominant color in order to separate responses into correct and incorrect.

This setup allows for optimal discrimination based on the information available to a participant; however, it is an assumption that could be amended. For example, if one category were rewarded more than another, the choice criteria may be adjusted asymmetrically. If green were rewarded more than blue, for example, blue-green evidence could be classified as green rather than blue. Such changes to the response boundaries could be modeled by setting the boundaries as free parameters or adjusting them as some function of the experimental manipulations applied to move them around. However, for the purposes of the present study this is not necessary and so the boundaries are assumed to be unbiased.

Note that the response boundaries themselves actually may account for differences in accuracy between different numbers and sets of response alternatives. Compare the top right against the bottom left panels of Figure 4.8 – in the former case, there are 3 relatively distinct response alter-

natives (blue, green, red) that are easy to tell apart and so all 3 responses shown in the diagram are classified as green / correct. However, in the case where the alternatives are more difficult to tell apart (bottom left, green versus yellow), one of these responses is classified as yellow simply because the response boundaries must be placed more closely in order to distinguish between them.

This same property may be at least partially responsible for the differences in accuracy between numbers of response alternatives shown in Figure 4.4. Because the alternatives are naturally grouped more closely together when there is a larger number of them, we can expect that the span of the decision boundaries will on average be narrower in these cases (compare the boundaries on the top right against the bottom right in Figure 4.8, for example). This will result in fewer evidence representations being classified correctly, and correspondingly lower accuracy.

Therefore, the incorporation of choice boundaries may actually be able to account for both important accuracy effects observed in the empirical data. Decision boundaries will be narrower when alternatives are grouped closely together (lower discriminability), resulting in lower accuracy when there are more alternatives or more similar alternatives available. However, they do not account for the variation in response times, neither as a function of discriminability nor as a function of the number of alternatives. There must therefore be another component of the decision process that leads to variation in response times in the task.

4.3.2 Discriminability

The major advantage conferred by the multidimensional scaling solution is that we can estimate the discriminability of any pair or set of choice alternatives presented in the study. The most straightforward use of this information is to feed it into the drift rate of the decision process – when the set of alternatives is easier to discriminate, the drift magnitude should be higher. Because the multidimensional scaling solution was computed independently, this can be done by adding just one additional free parameter that sets the drift rate as a (linear) function of the distance between the target hue and the nearest non-target / distractor response present in the display.

The reason for using the distance to the nearest distractor and not the discriminability of all

available alternatives is twofold: first, it is necessary in order to consider the continuous condition. Because there is an infinite number of alternatives in the continuous condition, it makes little sense to try to consider the similarity of all of them to the target hue. Instead, the nearest distractor can be used to estimate the difficulty of the decision, which will be smaller when a particular hue is difficult to discriminate from the hues around it and larger when it is easier. This allows for a consistent metric across all conditions to be used to set the drift magnitude. Second, the drift magnitude only really needs one additional reference point to be set. If we think about the magnitude as the inverse of the width of the sampling distribution (its variance), it can be assessed by examining how quickly the distribution falls off from its center. This can be done by taking two points with a fixed relative density (e.g. 3:1) and identifying where they are on the circle – for example, comparing the center of the sampling distribution to the location of another point with $\frac{1}{3}$ the density. The nearest distractor can be seen as a point with $\frac{1}{n}$ times the density of the mean, and therefore its location can be used as an index that gives the approximate variance of the sampling distribution. The location of the nearest distractor is therefore sufficient to provide an approximate estimate of the drift magnitude. A simpler way to put this is that the sampling distribution requires 2 free parameters, a mean and a variance (concentration). The mean is set at the target, so we can set the sampling distribution by referring to the location of the distractor to set the variance / concentration as the second parameter.

The greatest advantage of incorporating multidimensional scale distance into the drift magnitude in the model is that it can theoretically account for all four major phenomena observed in the empirical data. It accounts for variation in response time and accuracy as a function of target hue because the nearest distractor to these hues will tend to be further (higher discriminability, faster & more accurate responses) or closer (lower discriminability, slower & less accurate responses) depending on the target. In addition, because hues tend to naturally be grouped closer together when there are more alternatives, it predicts faster and more accurate decisions when there are fewer alternatives from which to choose.

4.3.3 Changes in threshold

Although changes in drift magnitude (as a function of multidimensional scaling distance) are in principle sufficient to account for variation in accuracy and response times across targets and conditions, it may well be the case that participants systematically adjust their own internal choice criteria across trials as well. This would reflect a desire to gather more or less information dependent on some characteristics of a particular decision.

Perhaps the most likely instance where participants could change their thresholds is when the number of alternatives changes. Usher et al. (2002) suggest that this is the mechanism that actually produces the changes in response times across numbers of alternatives. Given the evidence they provide, it seems reasonable to at least test this assumption by allowing thresholds to change as a function of the number of alternatives available to a participant on a trial. Therefore, one of the potential model modifications that I tested was to allow the threshold to vary as a linear function of the number of alternatives. Because it is unclear how the continuous condition relates in terms of the discrete number of alternatives it corresponds to, the threshold was permitted to vary independently in the continuous condition. This resulted in three additional free parameters: one intercept parameter for the threshold in the discrete conditions, one parameter that set the slope as a function of the number of alternatives, and one that set the threshold for the continuous condition.

However, this is not the only manipulation that may have changed the thresholds that participants applied. The discriminability of pairs of alternatives within the choice set could also cause them to apply stricter or more lenient choice criteria. The threshold was therefore permitted to vary as a linear function of the discriminability of the target hue and its nearest distractor, just as the drift magnitude varied. This introduced one free parameter to the model, a slope that set the threshold as a function of discriminability.

Either of the modifications to thresholds could account for some changes in response time and accuracy across conditions. However, they cannot account for both simultaneously. Because response time increased while accuracy decreased – both as the number of alternatives varied as

well as the multidimensional scaling distance between them – changes in threshold cannot account for all 4 effects. This is because a shift in threshold can create either increase both response time and accuracy or decrease both response time and accuracy. Alone, it does not predict discordant changes in the two outcomes.

4.3.4 Model fitting and results

In all, there were four possible additions to the basic model which included a base threshold, non-decision time, and drift magnitude. The effects of each of these modifications is shown in Table 4.1. The first of these, response boundaries, is necessary to include because it separates correct from incorrect responses. There is no straightforward alternative model which does not incorporate a similar mechanism. However, adding this element to the basic framework may by itself account for variation in accuracy across the conditions of the task. Because it was required in order to generate accuracy predictions and because it added no free parameters to the decision model, no model which excluded this modification was tested.

Table 4.1 Observed results and the model components that can account for them. A check mark indicates that the model component can account for an effect, a blank means it cannot, and an X means it can produce the opposite effect.

Empirical effect	Model component			
	Response boundaries	MDS distance $\mapsto$ drift	MDS distance $\mapsto$ threshold	# alternatives $\mapsto$ threshold
Accuracy $\times$ # alternatives (Figure 4.4)	✓	✓	✓	X
RT $\times$ # alternatives (Figure 4.3)		✓	X	✓
Accuracy $\times$ target hue (Figure 4.6, top)	✓	✓	✓	
RT $\times$ target hue (Figure 4.6, bottom)		✓	X	

The second addition was to set the drift magnitude as a function of the MDS distance between the target hue and its nearest distractor. This addition can in theory account for all four important

effects, and it is the only modification considered that can create differences in response times as a function of target hue (without adversely creating the opposite of the observed effect in accuracy).

The final two additions are not absolutely necessary in order to create the observed accuracy and response time phenomena. These are modifications that allowed thresholds to vary across conditions, either as a function of the number of alternatives or as a function of the discriminability of the target.

In order to gauge whether the inclusion of each of these elements outweighed the associated increase in complexity, I compared 3 models excluding each individual element against the full model. The first alternative model allowed only a single drift magnitude across conditions, ignoring the discriminability between alternatives as a factor impacting drift. The second alternative model allowed thresholds to vary only as a function of the number of alternatives, and the third alternative model allowed thresholds to vary only as a function of the discriminability of alternatives (generated from the MDS solution). Each model included all other elements not specifically omitted from the full model, so that each alternative tested the hypothesis that a single modification could be omitted without reducing model performance.

Each of the four models was fit using maximum likelihood estimation. The likelihood function incorporated both the accuracy and response time of each decision participants made. The predicted accuracy of a response was given by integrating the projected distribution of responses over the range of values that would be classified as the true dominant color given optimally placed decision bounds (see Figure 4.8). Note that predicted responses will be von Mises distributed as a function of the drift magnitude ($|\mu|$) and threshold (θ), and the location of the counterclockwise / lower decision bound (L) and clockwise / upper decision bound (U) will vary by trial. Therefore, the likelihood of a correct response $Pr(C)$ will vary as a function of the condition, choice alternatives, and target color.

$$Pr(C) = \int_L^U \frac{\exp\left(\frac{\theta|\mu|}{\sigma^2} \cos(\phi)\right)}{2\pi I_0\left(\frac{\theta|\mu|}{\sigma^2}\right)} d\phi \quad (4.1)$$

In this case, ϕ is the angle at which a response can hit the threshold, L is a negative number indicating how far away the nearest counterclockwise decision boundary is, and U is a positive number indicating how far away the nearest clockwise decision boundary is. The function $I_0()$ is a modified Bessel function of the first kind, of order 0. As the distribution of responses is assumed to be centered on the true dominant dot color, the center of the von Mises distribution will be 0, giving the above equation.

Computing the joint distribution of responses and response times requires that we multiply the conditional response time distributions by the probability of being correct. The conditional distribution of response times is given by conditioning Equation 3.1 on the response angle x . Notably, the predicted response time distributions are simply scaled copies of one another, as their relative frequency does not depend on the response angle.

These were used to derive log likelihoods for every model and participant, which were in turn used to compute a Bayesian information criterion [BIC] (Kass & Raftery, 1995; Schwarz, 1978). The BIC includes a penalty for the number of parameters in the model, favoring the more parsimonious model with fewer free parameters when their likelihoods are similar. Therefore, if one of the model elements (free parameters) conferred insufficient benefits, the BIC should favor the model which omits that feature. The results of the model fits are shown in Table 4.2.

Table 4.2 Bayesian information criterion (a function of the log likelihood with a penalty for the number of parameters; see Schwarz, 1978) for each model tested. The model with the lowest BIC is preferred and shown in bold.

Participant #	Model			
	Full	Fixed drift magnitude	Threshold invariant to # alternatives	Threshold invariant to discriminability
1	497.39	509.48	737.48	3153.5
2	2845.9	2866.5	3502.5	3072.5
3	8430.6	8553.1	9769.1	9449.1
4	11364	11477	14037	11591
5	28851	31502	34482	34816
6	7449.9	8288.5	7728.5	7524.5

For each participant, the BIC favored the full model including drift that varied by MDS dis-

tance, thresholds that varied by MDS distance, and thresholds that varied by the number of alternatives. The modeling results suggest that the discriminability and number of alternatives both play important roles in the decision process. However, given that both of them are parts of the best-fitting model, the effect of the number of alternatives cannot be isolated to just shifts in threshold or just changes in the discriminability of alternatives. Instead, the results paint a somewhat richer picture where the number of alternatives affects multiple facets of the decision process, including both the information participants consider as well as the choice criteria they apply.

Since the full model was favored for all participants, the same parameters are present in each of the best-fitting models. The maximum likelihood parameter estimates are given in Table 4.3, for drift magnitude intercept μ_0 , drift slope as a function of MDS distance μ_s , non-decision time ndt , threshold in the continuous condition θ_c , threshold intercept for the discrete alternative condition θ_d , threshold slope as a function of the number of alternatives (for the discrete conditions) θ_{sa} , and the threshold slope as a function of the MDS distance between target and nearest distractor θ_{sb} .

Table 4.3 Maximum likelihood parameter estimates for the BIC-favored model of the data from Study 2.

Participant	μ_0	μ_s	ndt	θ_c	θ_d	θ_{sa}	θ_{sb}
1	3.19	8.74	0.16	3.15	2.97	0.32	0.33
2	1.36	9.91	0.39	2.60	0.99	0.12	0.35
3	1.26	6.42	0.37	3.87	3.35	0.33	2.26
4	2.27	-2.30	0.48	12.88	0.24	0.91	-0.21
5	1.35	16.55	0.46	7.79	3.68	0.44	-1.35
6	1.95	2.06	0.44	2.56	1.58	0.38	-1.31

For nearly all participants, drift magnitude increased with a larger MDS distance between the target hue and its nearest distractor (i.e., discriminability; $\mu_s > 0$). Thresholds also consistently increased with the number of alternatives ($\theta_{sa} > 0$), as suggested by Usher et al. (2002). Curiously, there was substantial variation in the effect of discriminability on thresholds (θ_{sb}). Non-decision times (ndt) were largely within normal ranges expected for a rating task of this sort. Interestingly, the threshold estimates for the continuous condition (θ_c) fell all along the spectrum of thresholds for the discrete conditions (compare θ_c to $\theta_d + \theta_{sa} \times \#$ alternatives). This suggests that there may

be substantial disagreement between participants as to how much information they should gather in the continuous condition relative to the discrete ones – the thresholds in the continuous condition were equivalent to the thresholds for anywhere from 1 to 14 alternatives in the discrete conditions.

The parameter estimates themselves are somewhat interesting, but how does the resulting model perform? In theory, the model elements included should allow it to pick up the variation in behavior across numbers of alternatives, the continuous condition, and different discriminabilities of alternatives.

4.3.4.1 Posterior Predictions

For each participant, I took the best-fitting model (the full model) and derived its predicted response time and accuracy data. Because the parameters varied based on what the available alternatives were, how many of them were available, and which one was the dominant color, I generated a response time distribution and predicted accuracy for every trial. These were used to generate aggregate distributions of response times and accuracy as shown in Figure 4.9. With the slight exception of Participant #4, the resulting models provided a good account of the aggregate response time and accuracy data.

Because the model generates a joint prediction for each trial, we can also examine its predictions for response times and accuracy across target hues as well as across numbers of alternatives. Predicted (Xs) and actual (faded lines with error bars) mean response times by the number of available alternatives are shown in Figure 4.10. The models recover these mean response times quite well, but tend to slightly underestimate the amount of time it takes to respond in the continuous condition. The most likely explanation for this error is that participants are taking longer to actually enter their responses in the continuous condition – the differences are reflected mainly in a mean shift in response times but not reflected in accuracy. This suggests that there may be an additional component of non-decision time in the continuous condition that is not present in the discrete-alternative conditions. If this is the case, then it is likely that different experimental designs will produce varying differences between continuous and discrete choice conditions.

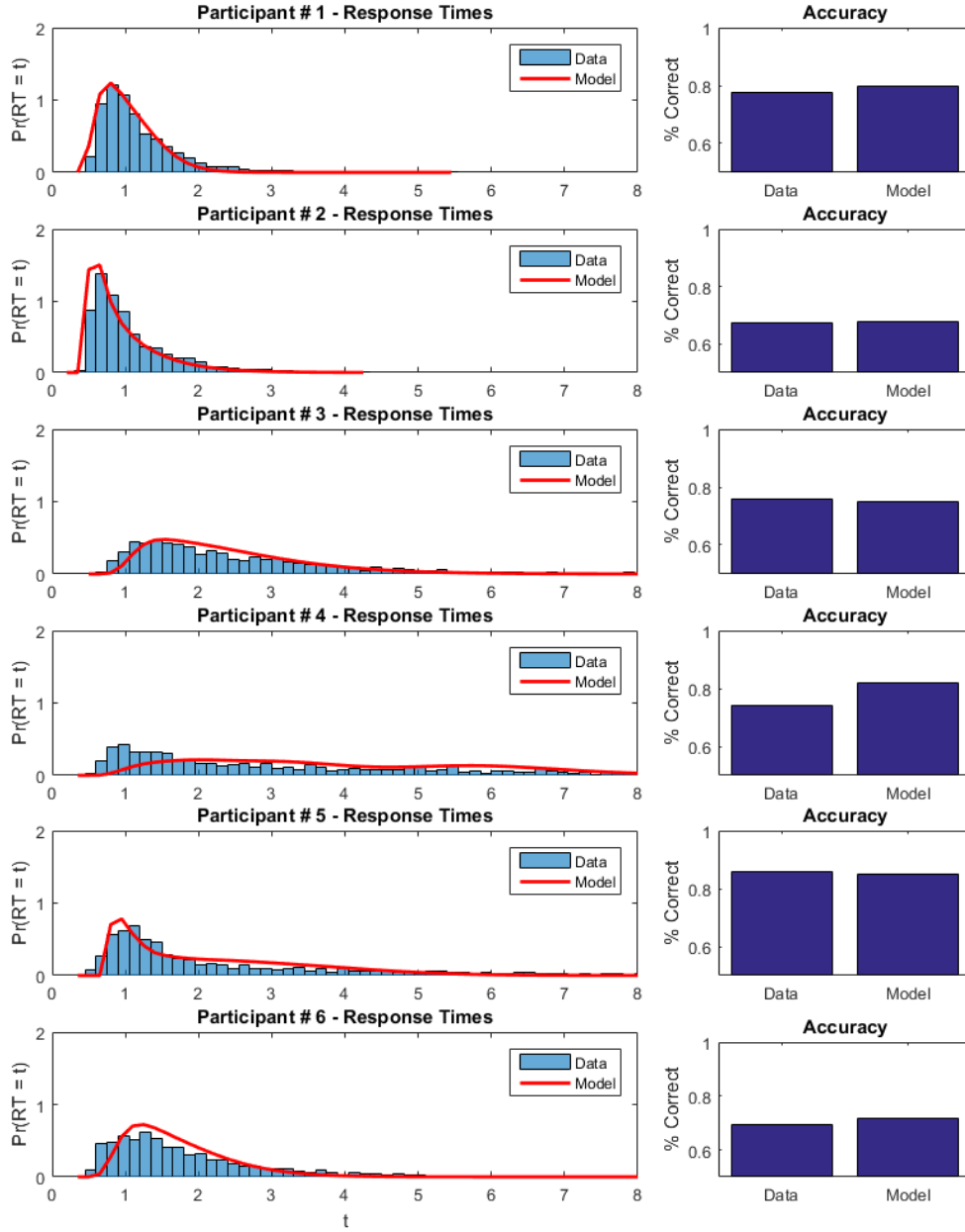


Figure 4.9 Posterior predictions from the best-fitting models for each participant, overlaid on the actual response time and accuracy data.

Note that the models do not necessarily predict exactly linear nor exactly log-linear relationships between the number of alternatives and mean response times. As the data and model fits reflect, response time is also heavily influenced by the discriminability of alternatives. This factor substantially changes the drift magnitude, and to a lesser extent the threshold as well (see Table

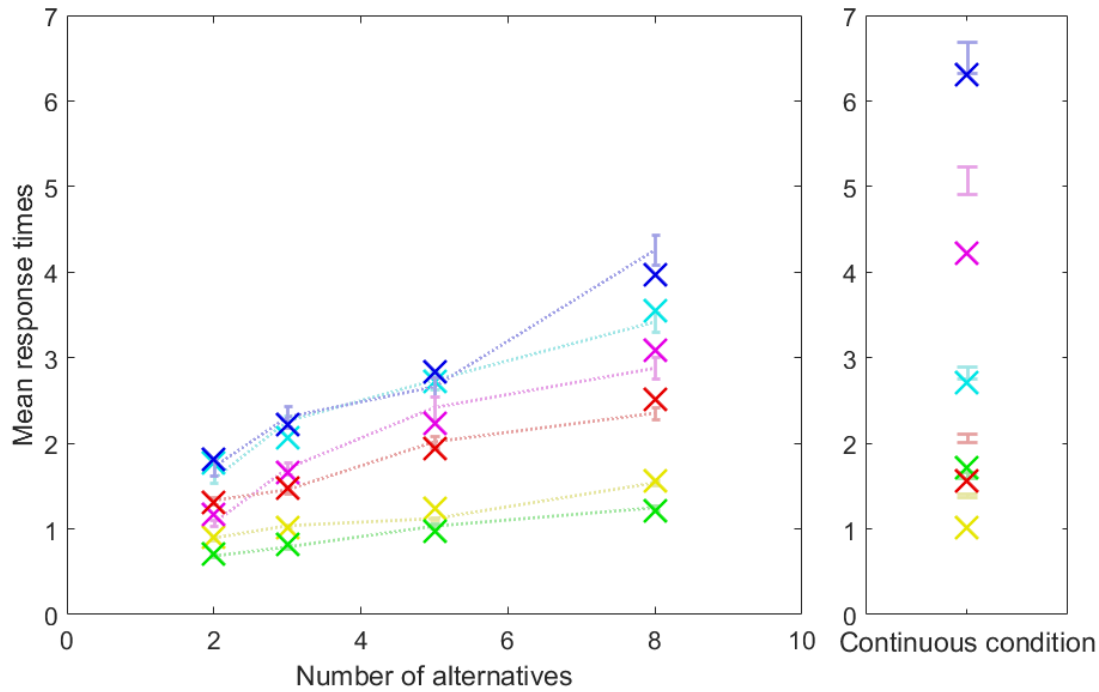


Figure 4.10 Posterior model predictions (X) and observed mean response times (faded lines) for each participant (differentiated by color). Error bars indicate 1 unit of standard error in the data.

4.3). As a result, the predicted and observed relationships between the number of alternatives and mean response times are much richer than that which is suggested by Hick's law.

Similar model predictions for accuracy as a function of the number of alternatives can also be generated. As with response time, the model prediction (Xs) and observed accuracy (faded lines with error bars) are shown. As with the response time distributions, there is substantial misfit for the continuous condition data of Participant #4 (blue). It seems likely that this participant had substantial trouble with the continuous condition, to the point where each response was both very slow and very inaccurate. Despite this, the model predicts well the accuracy data for the discrete-alternative conditions.

Although thresholds did seem to change across the number of alternatives (indicated by the model fits with and without this factor), this was apparently insufficient to maintain constant accuracy across the discrete-choice conditions. As a result, both mean response time and accuracy shifted as a function of the manipulations of the number of alternatives. It is possible that if partic-

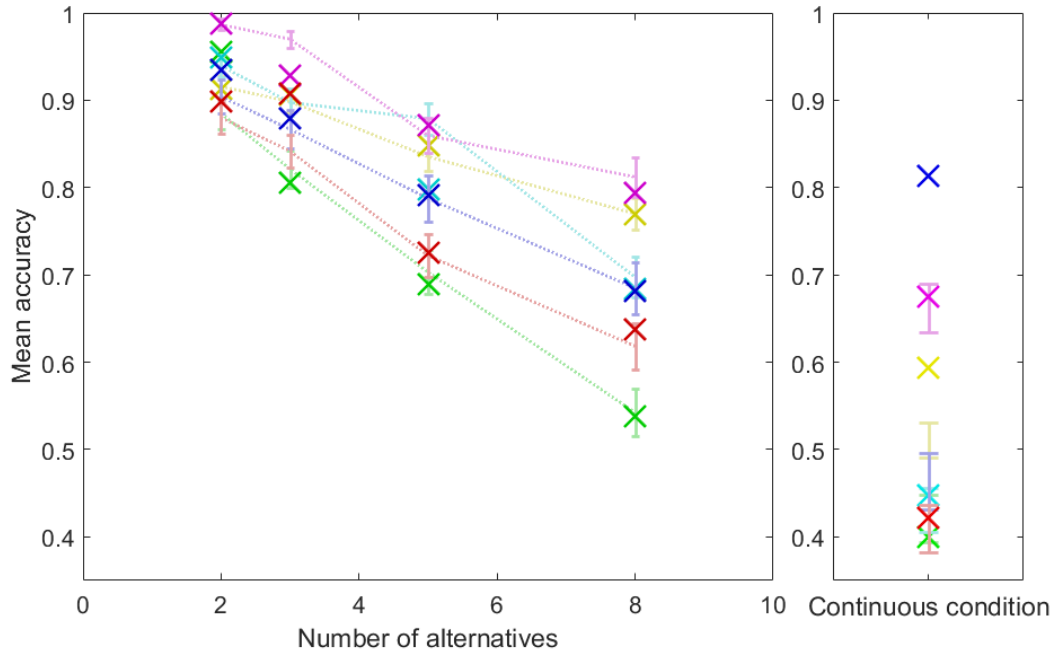


Figure 4.11 Posterior model predictions (X) and observed accuracy (faded lines) for each participant. Error bars indicate 1 unit of standard error in the data.

ipants were able to maintain accuracy, their response times would be log-linearly distributed, but this is highly unlikely. The extra processing time needed to maintain accuracy in conditions with a large number of alternatives would lengthen response times in these conditions – i.e., we could expect that striving to maintain accuracy would only increase the marginal response time added by each alternative. As a result, mean response time would probably be a convex function of the number of alternatives – not a concave one as Hick’s law predicts.

Finally, the variation in response times and accuracy across the discriminability of alternatives is shown in Figure 4.12. While the posterior model predictions do not explain this variation as neatly, they did pick up on some important characteristics of the data. Participant #5 provides a good example – the slower and less accurate responses when green or blue was the target are predicted by virtue of these colors being closer to their neighbors in the MDS space (Figure 4.7) and therefore harder to discriminate. As with some of the other posterior predictions, Participant #4 was a bit of an outlier; while the model predictions followed the general trend of accuracy and response times for other participants, it predicted almost the reverse of the empirical trend for

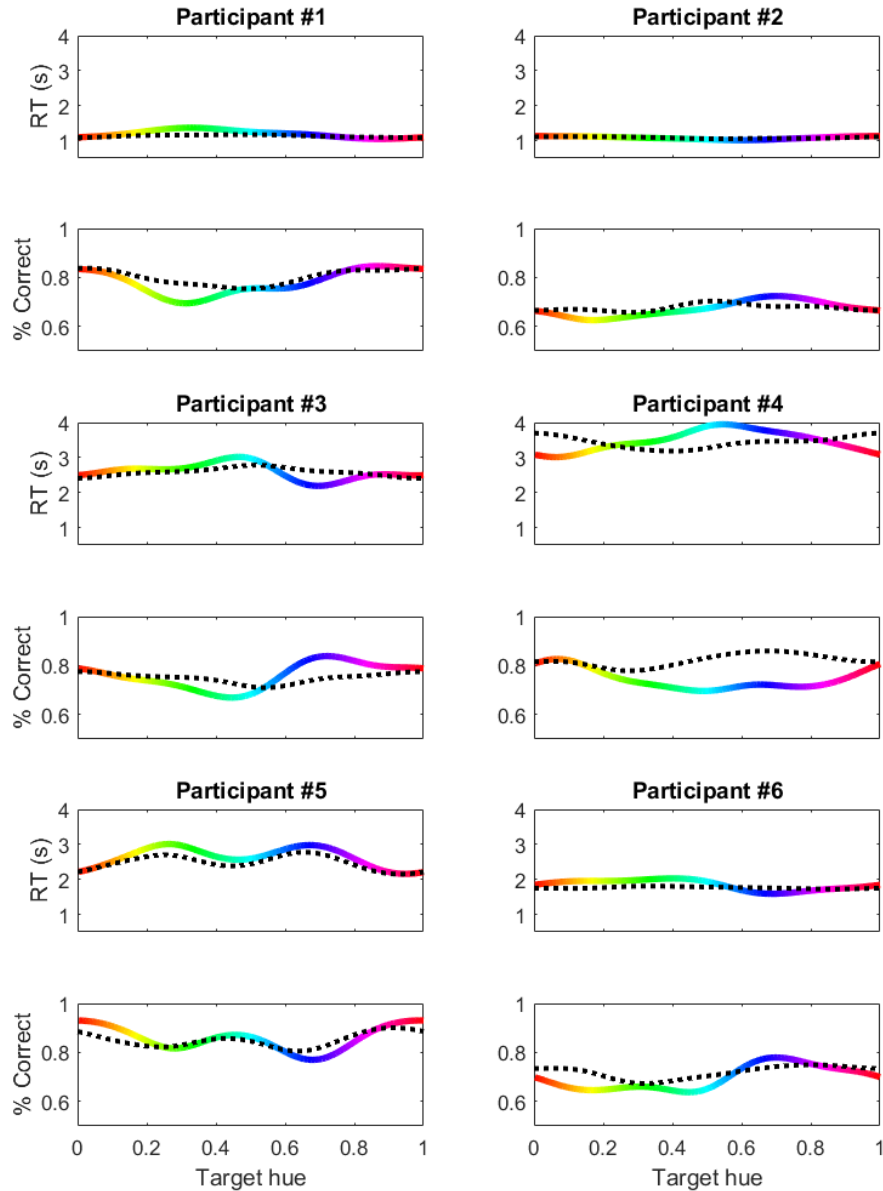


Figure 4.12 Posterior model predictions (dotted black line) of the relationship between target hue and response time or accuracy for each participant. Both data (colored curve) and model prediction curves are computed by passing a kernel density estimator over the hue $\times$ RT or hue $\times$ accuracy data (or prediction) from every trial.

Participant #4.

Overall, the full model seems to account quite well for the patterns of mean response times and accuracy across different numbers of alternatives, and captures some of the important trends in how accuracy and RT respond to differences in target hue.

4.4 Preliminary discussion

Study 2 provided an examination of how the number and assortment of alternatives affects the decision process. I found that accuracy decreases and mean response times increase as the number of alternatives a person has to choose between gets larger. The functional relationship, especially taking the continuous condition into consideration, does not seem to strictly follow Hick's law. Instead, the patterns of response time and accuracy that result from manipulations of the number of alternatives arise from several contributing factors. The discriminability of the target from distractors in the stimulus played a large role, as did shifts in choice criteria resulting directly from the number of alternatives available. The locations of choice boundaries that divided responses into the available alternatives also contributed to shifts in accuracy across varying numbers of alternatives.

Although the patterns of response times and accuracy are fairly straightforward for discrete choice conditions, the continuous-response condition did not fit neatly in with these results. Some participants were able to maintain accuracy and response times in these conditions as if they were dealing with only 2-8 alternatives; others had great difficulty and suffered both in terms of accuracy and response speed. It seems likely that different participants apply different strategies in order to cope with a continuum of responses, meeting with variable levels of success. Despite this, responses in the continuous case did seem to be impacted by the discriminability of the target hue just as they were in the discrete alternative cases. This suggests that the characteristics of the information participants use seem to be the same – which we might expect given that the stimuli were nearly identical – but that different decision rules are applied depending on whether the task has discrete options or a continuum. There are hints that the way participants enter their responses (i.e. non-decision components of the task) may differ in the continuous condition as well.

The addition of a multidimensional scaling component based on similarity rating data particularly helped to understand the factors giving rise to variation in behavior. By using this data to construct a multidimensional scaling representation of the colors for each participant, I was able

to predict the discriminability of a pair of colors independent of the decision data. And as the model fits suggest, incorporating this multidimensional scaling representation into the decision model helped to account for both response time and accuracy outcomes. This study therefore provides and favorably tests a method of integrating representation and decision models into the same framework, and emphasizes the importance of considering the relationships between alternatives when accounting for behavior.

CHAPTER 5

GENERAL DISCUSSION

The framework proposed in the introduction serves as the focal point of this paper, but the empirical investigations are relevant sans model as well. The speed-accuracy trade-off, difficulty effect, and predecision cues all have analogues in a continuous-response task, providing some of the first phenomena that relate error magnitudes to response times on a continuum. The von Mises distributed responses and (non-)relationship between RT and error magnitude provide additional dimensions to the empirical results. Furthermore, the effects of the number and similarity of alternatives in both discrete and continuous choice conditions provide hurdles for any theory of how people make selections on these tasks.

The responses and response times alone tell us that stimulus characteristics, cues, rewards, and assortment of choice options all have dramatic impacts on decision processes. The added value of the model is to describe precisely and concretely how the outcomes change as a function of these manipulations. In turn, it has many potential uses. One of these is that it can serve as an embodiment of our understanding of decision processes in choices among many alternatives. We can understand the effects of our manipulations in terms of shifts in parameter values, allowing the framework to serve as a measurement model. The parameters are themselves psychologically meaningful: drift magnitude tells us how quickly a person can gather valid information information, threshold describes the strictness with which they make decisions, drift direction gives us a lens onto biases and inaccuracy in information they use, and non-decision time lets us discriminate the components of action and perception beyond decision processes. Estimates of the relationships between alternatives are also particularly important: we can uncover how people are representing their available choice options, and in turn project what errors are likely and what decisions will be most difficult. In turn, we can devise interventions based on these parameter estimates, focusing on speeding up decisions, alleviating bias, improving discrimination between sets of alternatives, or encouraging people to set better criteria for making their decisions.

The integration of the decision model with representation components is particularly notable in that it brings together multiple independent processes in order to make better predictions. Eventually, our model of the whole brain and its computational processes will have to incorporate multiple layers of cognitive models – from perception to representation, belief and preference updating, decision, action, feedback, and strategy revision. Using multidimensional scaling to construct decision spaces puts at least two pieces of this puzzle together, allowing us to explain phenomena in behavior that results from the interaction of multiple cognitive processes.

My hope is that the framework and data proposed here have opened up new questions as well as provided hints as to how we represent and process information. At worst, it at least provides a first pass at attempting to model how people make selections among many alternatives – a task which decision modeling has stalled on for perhaps too long (Townsend, 2008).

5.1 Limitations

Of course, the present studies explore only a small portion of potential tasks and models that could be used in research on continuous or multi-choice decisions. Perhaps most notably, the experiments presented here are both perceptual, inferential choice tasks where there is a true correct answer. Given the differences between perceptual and preferential choice even with the same stimuli (Zeigenfuse et al., 2014; Dutilh & Rieskamp, 2015), it seems likely that the decision processes in selections among many alternatives are likely to change based on the goals of the participant as well. Fortunately, these changes may be possible to explain through changes in parameters rather than complete shifts in model structure.

More complex multi-attribute decisions are also likely to require some modifications to the structure of the models presented here. Fortunately, the architecture presented in the introduction is well-suited to such tasks, allowing the space of available alternatives to reflect a multi-dimensional feature space in which they might exist. In these cases, the alternatives may have to be constructed based on their raw physical characteristics, though multidimensional scaling may help to construct

the decision space in such cases as well.

As the number of feature dimensions increases, however, so does the complexity of the decision space and corresponding decision model. When the alternatives can be arranged so that they create a hypersphere (for example, one alternative taking up each octant of a sphere), there are analytical solutions to the likelihood function that can be derived by substituting the von Mises distribution of evidence with a von Mises-Fischer distribution. Given the tractability of even the 2-dimensional case, these are likely to be quite complex and difficult to use practically. In many other cases where the alternatives do not form a neat closed hypersphere, response time and response distributions may have to be derived by repeatedly simulating the decision process.

The particular functions attached to the sampling distribution (drift), threshold, and non-decision times may also not be the best candidates in all situations. When a person switches slowly between two attributes of a stimulus, for example, the sampling distribution may need to be time-dependent, as in the binary decision models of Diederich & Busemeyer (2006). Similarly, as Study 2 suggests, non-decision components may vary between tasks (even discrete versus continuous selection tasks) in their mean or functional form, rather than being set as fixed or uniform distributions (Verdonck & Tuerlinckx, 2016). Given the apparent shifts in thresholds as a function of stimulus manipulations in Study 1 (Figure 3.5), thresholds may need to vary across time. These assumptions lead to different distributions of response times and decisions, and further work should explore exactly how sensitive model predictions are to these particular distributional assumptions.

Finally, recent work on the sequential effects between and within trials have suggested that decisions and judgments made in sequence – whether about different stimuli or the same one – affect one another (Brown et al., 2008; Kvam et al., 2015). Trial-to-trial effects may have a different underlying source than within-trial effects, but response time and accuracy distributions (as well as confidence) for sequential judgments have been well-described by decision models constructed using a quantum probability framework (Fuss & Navarro, 2013). It is possible that a quantum random walk in multiple dimensions may provide a more complete of multi-choice or continuous-response decisions as well.

5.2 Extensions

Most of the limitations I have discussed so far actually offer opportunities for future work as well. Notably, sequential choice effects, collapsing response boundaries, quantum random walks, and preferential decisions all offer directions for further investigation. However, there are a number of other applications of the model that do more than explore binary choice phenomena in multi-choice or continuous-response cases.

One particularly interesting extension is to spatial tasks, where different locations correspond to different rewards or risks. Deciding where to look for rewards draws strong parallels to foraging, where agents have to make spatial selections when gathering food. These different locations almost certainly correspond to variations in predation risks, environmental hazards, and potential payoffs (food), where foragers have to make compromises in order to survive and reproduce. Our understanding of risky decision processes in humans and other animals can be well-informed by examining what conditions affect the utility of different outcomes, produce risk avoidance or risk seeking behavior, impact exploration or exploitation orientation, or yield varying representations of rare events. It seems likely that experience in a spatial domain produces different representations and foraging (decision) behavior than descriptions or maps of the territory, and these continuous representations should be richer than those that give rise to the description-experience gap in binary gambles (Hertwig & Erev, 2009).

Another immediate extension of the modeling framework is to begin predicting confidence or likelihood ratings that people give. There are two directions to pursue in this case. In the simple case where there are two alternatives but a continuum of judgments to be made between them, the model framework can offer a method of predicting the distributions of confidence judgments as well as the speed with which these judgments are made. It is likely that the model framework can be used to reproduce many of the important effects observed in judgment distributions and judgment response times (Pleskac & Busemeyer, 2010; Moran et al., 2015), but as of yet this is an untested claim.

Another direction that could be pursued is to examine how judgments are made between many alternatives, or even on continuous spans (e.g. “How likely is it that the population of city X falls between 170,000 and 200,000?”). In principle, the location of the multidimensional random walk process (when it unfolds on a hypersphere) corresponds to the log odds across all available hypotheses or alternatives in that space. There is therefore an ‘true’ / optimal probability or confidence rating that one should give based on the available information or evidence. Moreover, the optimal case imposes strong constraints on the additivity of judgments – because the log odds always sum to zero, probability judgments made across all alternatives should sum to one. It is possible that evidence is re-mapped onto confidence ratings (Ratcliff & Starns, 2009; Pleskac & Busemeyer, 2010), in which case this relationship could be broken. In either case, the model framework provides a basis from which we can study judgments of many alternatives on top of just choices among many alternatives. Understanding how these judgments are apportioned may also help us understand how people store and revise evidence, allowing for appropriate revisions of the model structure in terms of how information is represented.

Behavioral data is not the only province for decision models like the one proposed here, either. More recently, sequential sampling models have been linked to neural data (Hanes & Schall, 1996; Schurger et al., 2012; Shadlen & Newsome, 2001; Ratcliff & McKoon, 2008) and model-based analyses of joint neural and behavioral data are becoming easier and more effective (Turner et al., 2013). The multidimensional walk proposed here could certainly be used for these purposes as well. For example, activation of orientation columns could be mapped onto the sampling distribution from the model in Study 1. The parameter estimates in Figure 3.6 (bottom left) gives an idea of how varying cue orientations might facilitate or inhibit decisions in favor of related orientations – this is almost certainly related to the tuning curves of orientation detectors in the brain. Each of the model parameters should describe some facets of both behavior and neural activity; they motivate exploration of what behaviors and regions are related as well as what parameters are most appropriate for describing both the activity and products of the brain.

In principle, the modeling framework can be extended to explain and predict any selection

among alternatives that are psychologically related. 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 be applied to predict decisions among them.

5.3 Conclusions

This thesis has covered a broad range of considerations that go into decisions among many alternatives, but the most important points are relatively simple. First, the decision process that goes into multiple choice and continuous response decisions reflects a sequential sampling evidence accumulation process. As we saw in the first study, the empirical phenomena and model fits all reflect a decision architecture that resembles (but is a general case of) that used to predict binary decisions and response times.

Second, the combinations of available options have substantial effects on the decision process. As we saw in the second study, the number and the similarity between alternatives can slow down choices and make them more error-prone. However, the architecture of the decision model can be informed by independent data on the relationships between the psychological representations of the available alternatives. This integration allows it to account for much of the variation in decision patterns across combinations of alternatives.

Finally, the model framework proposed here opens up a wide range of new behaviors to predict and new questions to address. It lets us move toward predicting and explaining behavior on tasks like investment, time management, confidence and likelihood judgments, price setting, spatial

navigation, and even things like pitch production or foraging. Empirical studies of each of these domains can inform our understanding of decision processes, and the framework I have established in this thesis provides the formal quantitative theory to embody and extend our knowledge of behavior on decisions among many alternatives.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Berens, P. (2009). Circstat: A Matlab toolbox for circular statistics. *Journal of Statistical Software*, 31(10), 1–21.
- 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. doi: 10.1037/0033-295X.113.4.700
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in neurosciences*, 33(1), 10–16. doi: 10.1016/j.tins.2009.09.002
- Bowman, N. E., Kording, K. P., & Gottfried, J. A. (2012). Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron*, 75(5), 916–927.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Brown, S. D., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological review*, 112(1), 117–128. doi: 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. doi: 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. doi: 10.1037/0033-295X.115.2.396
- Busmeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, 43(3), 345–370.
- Busmeyer, 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. doi: 10.1037/0033-295X.100.3.432
- Busmeyer, J. R., Wang, Z., & Townsend, J. T. (2006). Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, 50(3), 220–241. doi: 10.1016/j.jmp.2006.01.003
- De Boor, C. (1978). *A practical guide to splines* (Vol. 27). Springer-Verlag New York.
- Diederich, A., & Busmeyer, 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. doi: 10.1016/S0022-2496(03)00003-8
- Diederich, A., & Busmeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, 68(2), 194–207. doi: 10.3758/BF03193669

- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes? *Psychonomic bulletin & review*, 18(1), 61–69. doi: 10.3758/s13423-010-0022-4
- Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, 16(6), 1129–1135.
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, 32(11), 3612–3628.
- Dutilh, G., & Rieskamp, J. (2015). Comparing perceptual and preferential decision making. *Psychonomic bulletin & review*, 23, 1–15.
- 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.
- Fuss, I. G., & Navarro, D. J. (2013). Open parallel cooperative and competitive decision processes: A potential provenance for quantum probability decision models. *Topics in Cognitive Science*, 5(4), 818–843. doi: 10.1111/tops.12045
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, 274(5286), 427–430. doi: 10.1126/science.274.5286.427
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of experimental psychology*, 93(1), 10–20. doi: 10.1037/h0032606
- Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed-accuracy tradeoff. *Neuron*, 76(3), 616–628. doi: 10.1016/j.neuron.2012.08.030
- Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in cognitive sciences*, 13(12), 517–523. doi: 10.1016/j.tics.2009.09.004
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1), 11–26. doi: 10.1080/17470215208416600
- 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.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430), 773–795. doi: 10.1080/01621459.1995.10476572
- Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210), 227–231. doi: 10.1038/nature07200

- Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928), 759–764. doi: 10.1126/science.1169405
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, 36(14), 1.
- Krajovich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, 3. doi: 10.3389/fpsyg.2012.00193
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Kvam, P. D. (2016). Geometric representations of evidence in models of decision-making. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1098–1103).
- Kvam, P. D., & Pleskac, T. J. (2016). Strength and weight: The determinants of choice and confidence. *Cognition*, 152, 170–380. doi: 10.1016/j.cognition.2016.04.008
- Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences*, 112(34), 10645–10650. doi: 10.1073/pnas.1500688112
- Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Academic Press.
- 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. doi: 10.1007/BF02291481
- Longstreth, L. E., El-Zahhar, N., & Alcorn, M. B. (1985). Exceptions to hick's law: Explorations with a response duration measure. *Journal of Experimental Psychology: General*, 114(4), 417–434. doi: 10.1037/0096-3445.114.4.417
- Louter, A., & Koerts, J. (1970). On the kuiper test for normality with mean and variance unknown. *Statistica Neerlandica*, 24(2), 83–87. doi: 10.1111/j.1467-9574.1970.tb00110.x
- Markant, D., Pleskac, T. J., Diederich, A., Pachur, T., & Hertwig, R. (2015). Modeling choice and search in decisions from experience: A sequential sampling approach. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1512–1517). Cognitive Science Society, Austin, TX.
- Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, 88(1), 78–92. doi: 10.1016/j.neuron.2015.09.039
- Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: novel data and a computational account. *Cognitive psychology*, 78, 99–147. doi: 10.1111/j.1467-9574.1970.tb00110.x

- 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.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104(2), 266–300. doi: <http://dx.doi.org/10.1037/0033-295X.104.2.266>
- 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. doi: 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. doi: 10.1037/A0019737
- Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review*, 120(3), 679–696. doi: 10.1037/a0033142
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. doi: 10.1037/0033-295X.85.2.59
- Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 870–888. doi: 10.1037/a0034954
- Ratcliff, R., & Frank, M. J. (2012). Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Computation*, 24(5), 1186–1229. doi: 10.1162/NECO_a00270
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. doi: 10.1162/neco.2008.12-06-420
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, 111(2), 333–367. doi: 10.1037/0033-295X.111.2.333
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116(1), 59–83. doi: 10.1037/a0014086
- Ross, S. M. (2014). *Introduction to probability models*. Academic press.
- 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. doi: 10.1073/pnas.1210467109
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. doi: 10.1214/aos/1176344136
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area *lip*) of the rhesus monkey. *Journal of neurophysiology*, 86(4), 1916–1936. doi: none

- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.
- Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological review*, 89(4), 305. doi: 10.1037/0033-295X.89.4.305
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research*, 16(2), 158–174. doi: 10.1086/209205
- Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123(4), 425–451. doi: 10.1037/rev0000023
- Smith, P. L., & Van Zandt, T. (2000). Time-dependent poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, 53(2), 293–315. doi: 10.1348/000711000159349
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, 32(2), 135–168. doi: 10.1016/0022-2496(88)90043-0
- Townsend, J. T. (2008). Mathematical psychology: Prospects for the 21st century: A guest editorial. *Journal of Mathematical Psychology*, 52(5), 269–280. doi: 10.1016/j.jmp.2008.05.001
- 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. doi: 10.1037/a0036137
- Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, 72, 193–206. doi: 10.1016/j.neuroimage.2013.01.048
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281. doi: 10.1037/h0032955
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. doi: 10.1126/science.185.4157.1124
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3), 550–592. doi: 10.1037/0033-295X.108.3.550
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, 111(3), 757–769. doi: 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. doi: 10.1006/jmps.2002.1420

- Van Den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109(22), 8780–8785. doi: 10.1073/pnas.1117465109
- Verdonck, S., & Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review*, 123(2), 208. doi: 10.1037/rev0000019
- Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences*, 2(3), 169–194. doi: 10.1023/A:1022371901259
- Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, 50(2), 179–197. doi: 10.1016/0001-6918(82)90006-3
- Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, 73, 59–79. doi: 10.1016/j.jmp.2016.04.008
- Wald, A., & Wolfowitz, J. (1949). Bayes solutions of sequential decision problems. *Proceedings of the National Academy of Sciences*, 35(2), 99.
- Wang, H.-C., Chang, C.-Y., & Li, T.-Y. (2008). Assessing creative problem-solving with automated text grading. *Computers & Education*, 51(4), 1450–1466. doi: 10.1016/j.compedu.2008.01.006
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, 41(1), 67–85. doi: 10.1016/0001-6918(77)90012-9
- Witthoft, N., Winawer, J., Wu, L., Frank, M., Wade, A., & Boroditsky, L. (2003). Effects of language on color discriminability. In *Proceedings of the 25th annual meeting of the cognitive science society* (pp. 1247–1252).
- Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, 136(4), 623–638. doi: 10.1037/0096-3445.136.4.623
- Wolfe, M. B., & Goldman, S. R. (2003). Use of latent semantic analysis for predicting psychological phenomena: Two issues and proposed solutions. *Behavior Research Methods, Instruments, & Computers*, 35(1), 22–31. doi: 10.3758/BF03195494
- Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition*, 131(2), 181–194. doi: 10.1016/j.cognition.2013.12.012