EXPLORING THE STRUCTURE OF COGNITIVE PROCESSES: DISCRETE AND CONTINUOUS THEORIES OF MEMORY AND PERCEPTION

A Thesis presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

APRIL R. SWAGMAN

Dr. Jeffrey N. Rouder, Thesis Supervisor

MAY 2013

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled:

EXPLORING THE STRUCTURE OF COGNITIVE PROCESSES:
DISCRETE AND CONTINUOUS THEORIES OF
MEMORY AND PERCEPTION

presented by April R. Swagman, a candidate for the degree of Master of Arts and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Jeffrey Rouder

Dr. Clint Davis-Stober

Dr. Chi-Ren Shyu

Dr. Jeffrey Johnson

For Stephanie, who is gone but never forgotten.

ACKNOWLEDGMENTS

I want to acknowledge the members of the Perception and Cognition Lab for helping make the work in this Thesis possible. I especially want to thank Dr. Jeff Rouder for continually setting me on the right path and Jory Province for helping me when I inevitably got lost.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Cognitive psychologists often debate whether cognitive processes such as recognition memory and visual perception are better described as continuous or discrete. Much of the literature is dominated by continuous latent-strength models such as signal detection theory. Here the author sought to test the effectiveness of discrete-state models at predicting performance on cognitive tasks.

Thirty-nine participants completed Experiment 1, which analyzed recognition memory and visual word identification performance in two-alternative forced choice tasks. Fifty participants completed Experment 2, which focused on word identification in both two-choice and one-choice designs. Data were analyzed via maximum likelihood estimation for comparable discrete-state and latent-strength models for all participants. The discrete-state models outperformed latent-strength models for the majority of participants in both tasks of both experiments. Evidence for discrete states was especially strong in two-choice word identification tasks. The results indicate that discrete-state models should not be ignored in cognitive processing, as they provide a good account of data in recognition memory and word identification tasks.

Chapter 1

Introduction

1.1 Background and Motivation

Scientists have studied the structure of human cognition and its many specialized components and systems for decades. One cognitive domain of particular interest is *recognition memory*. Recognition memory is our ability to recognize whether or not we have previously encountered a certain event. This type of memory can be vitally important, as it allows us to navigate the world and recognize potentially beneficial or harmful situations. Given its foundational importance, it is not surprising that recognition has been a topic of discussion since the inception of experimental psychology (see Allin, 1896).

Although research findings have been expansive, a debate still exists over the structure of the recognition process. Is recognition better characterized as discrete or continuous? Supporters of continuous theories suggest that cognitive phenomena have finely graded perceptual strengths that people have access to and can use to make decisions during recall. Discrete theories, in contrast, are characterized by discrete mental states which drive recall decisions. The probability a certain event is perceived or remembered depends only on which state one is in. Discrete processes are sometimes referred to as *all-or-none*. The discrete models I will focus on consist of two mental states: detection and guessing.

Assessing whether recognition is a discrete or continuous process is not a simple task. Human memory is not directly observable, and it is not clear how accurately behavioral data reflect the memory process. Although it is highly likely that the underlying latent cognitive process is continuous, it is not clear how much access people have to this continuous process. Stimuli can be perceived continuously but reach a threshold during retrieval that leads to discrete processing.

An example to clarify this concept is to think of the latent memory process as a meter or gauge. My brother used to have an old run-down car with a broken fuel gauge, meaning he never knew exactly how much gas was in his tank. The low-fuel light (indicating it is time to refuel), however, still worked. So he was only able to make a refueling decision based on *light on* or *light off*. The low-fuel indicator light is analogous to a discrete memory model: the information is either present, and the criteria is met, or it isn't. There is still a richness of continuous information that leads to the light turning on, but the driver has no access to it. An accurate fuel gauge could then represent a continuous model: the exact amount of fuel–or perceptual strength–at any given time is always known, and a decision is made based on this amount. In either case, an observer would only see the refueling decision the driver made, regardless of how he got there. Analogously, a researcher performing memory experiments only sees the decision a participant makes on a given trial.

1.2 Literature Review: What is Commonly Believed and Why

There are two current answers to the question of whether recognition is discrete or continuous. One answer comes from the older verbal-learning tradition in which perception is mediated by latent strength (Parks, 1966; Kintsch, 1967). This idea is captured by signal detection theory. Latent strength for an item is continuous and uni-dimensional, and strength increases in value as a result of a prior encounter. Accordingly, recognition is distinctly a continuous process, and the nature of latent strength may be accessed by behavioral data. A second answer comes from modern memory-systems theories in which memory is comprised of several possibly interrelated systems (Schacter & Tulving, 1994). Yonelinas (1994) in particular posits that memory consists of a mixture of a discrete process, called *recollection* and a continuous one, called *familiarity*. Accordingly, recognition memory is both continuous and discrete, and, moreover, behavioral results of recognition experiments reflect this mixture of processes. There is, however, one additional viewpoint not well represented in this debate: recognition is discrete.

Most researchers do not consider that cognitive processes might be discrete because of a particular pattern in the behavioral literature. I will first present the conventional confidence ratings paradigm and the corresponding receiver operating characteristic (ROC) curve analysis. Then I will show ubiquitous findings that are taken as evidence against discrete-state models. Finally, I show that these patterns rule out only a very constrained discrete-state model and not discrete-state models in general. This analysis leads to a startling fact: there is no strong evidence against the discrete-state recognition model, which, in turn, motivates this thesis.

1.2.1 The Confidence Ratings Paradigm and ROC Plots

Many modern recognition studies evolved from the signal detection experiments of the 1950s. In a classic signal detection experiment, participants hear audio noise plus the presence or absence of a tone (signal) on each trial. The participant then has to decide whether or not the signal was present on that trial. Performance on the task could then be analyzed using hits and false alarms (Tanner & Birdsall, 1958). A hit occurs when a person successfully detects a signal, and a false alarm occurs when a person mistakenly reports signal on a noise trial.

In standard recognition memory experiments, participants study a list of items, and then, at test, they see an item and must decide if it was previously studied. The two possible responses are yes (or *old*) and no (or *new*). Performance can similarly be analyzed by looking at the rate each individual makes hits and false alarms. The hit rate is the proportion of times the participant correctly identifies a studied item as old. The false alarm rate is the proportion of times he or she incorrectly identifies a new object as old. Hit and false alarm rates both increase as the participant's tendency to respond *old* increases.

One problem with the use of yes/no paradigms is the limited amount of information a researcher can gather from these responses. Yes/no tasks produce only one set of hits and false alarms, making it difficult to reveal any details about the underlying cognitive structure. To mitigate this issue, a researcher can utilize confidence rating responses which produce several hit and false alarm rates for each participant, and thus a richer view of the underlying processes. Confidence ratings in a recognition task are fairly straightforward: on each test trial, the participant sees the test item and a confidence scale ranging from *Sure Old* to *Sure New*. She then makes a response based on how confident she is in her decision. Some experiments employ a continually-graded scale, while others give a limited set of options to choose from (e.g., *Sure Old*,

	Sure New	Likely New	Likely Old	Sure Old
Response Proportions				
Old Items	.02	.29	.38	.31
New Items	.31	.53	.14	.02
Cumulative Proportions				
Hit Rate	1	.98	.69	.31
False Alarm	1	.69	.16	.03

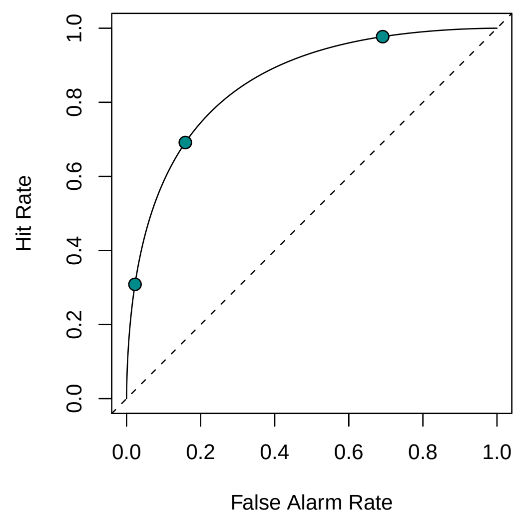

Figure 1.1: Construction of ROC plots. This curve was constructed using the data from the table.

Likely Old, *Guess Old*). Hit and false alarm rates are generated cumulatively across response options. The table in Figure 1.1 provides an example of response data with four confidence ratings. The numbers in the top two rows are response proportions (e.g. the participant responded *Sure Old* to old items 31% of the time). To calculate the hit rate, first look at responses to old items. Start with the response proportion for the most confident *Old* response and cumulatively sum across the other ratings. The same is done to calculate false alarm rates, except we sum across responses to new items. These rates will both necessarily sum to 1.0 because they are proportions.

Taken together, these multiple hit and false alarm rates give a rich picture of performance without loss of information. We can then use these rates to construct ROC curves, which are plots of hit rates as a function of false alarm rates. The plot in Figure 1.1 shows an example of an ROC plot with three points, drawn from the data in the table. The three points are generated cumulatively from the four confidence rating categories in this example. ROCs give a simple and straightforward visual representation of a participant's ability to discriminate old and new items across various conditions, and because of this, many researchers use ROC analysis to observe the details of underlying memory systems. The shape and curvature of ROCs are of particular interest as they may illuminate fundamental properties of the memory system. Continuous models and discrete-state models make different predictions about the shape of ROC curves; therefore, ROCs are often used as diagnostic tools for model comparison. I will next introduce a few basic models of recognition and then explain their ROC predictions.

1.2.2 The Rejection of High-Threshold Models

One of the earliest well-known discrete-state models was the Fechnerian high-threshold model. Blackwell (1963) formalized Fechner's view of a sensory threshold to explain performance in simple visual discrimination tasks similar to those of signal detection. The model has two parameters: a sensitivity parameter d and a guessing parameter g. In the model, when a signal is present, the participant either enters a detect state (with probability d), and says yes, or enters a guessing state (with probability $1-d$), and guesses either yes (with probability g) or no $(1-g)$. In a noise trial, there is no chance of detection, so the participant simply guesses yes or no with probability g and $1-g$, respectively. This model is illustrated in Figure 1.2 for signal and noise trials.

Another discrete model that was proposed to account for recognition memory was the *double-high threshold model*. This model's name comes from the idea that there exists an internal threshold for old items and another threshold for new items and only items that exceed these thresholds will be recognized as such. If neither of these thresholds is crossed, the participant simply makes a guess. This model differs from

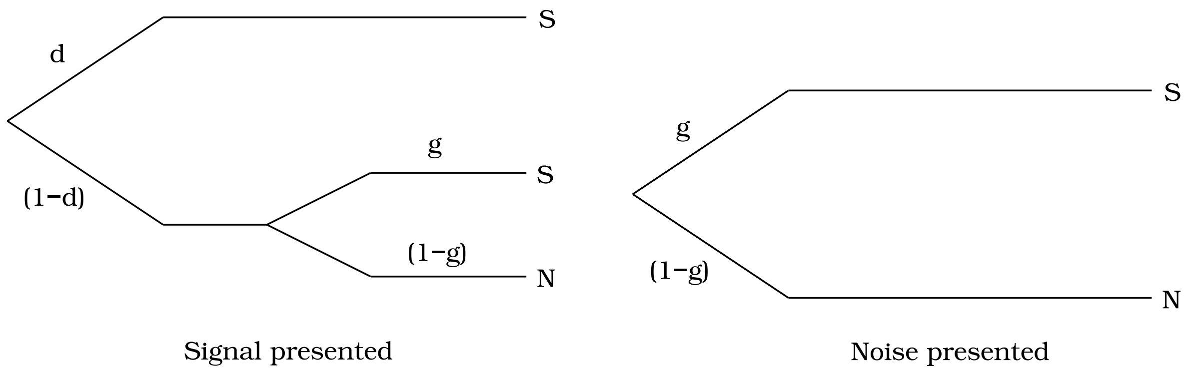

Figure 1.2: Fechnerian single high threshold model. S=Signal response, N=Noise response. The left panel shows the model for signal trials, and the right is noise trials.

the Fechnerian model in that participants have a chance of detecting noise or newness. These threshold models are discrete-state models, as they suggest an individual is in a state of either recognition or guessing. Single and double-high threshold models extend naturally from yes/no paradigms to confidence ratings, which I will expand on shortly.

Threshold models contrasted with continuous models, most notably signal detection theory (Green & Swets, 1966). The signal detection model consists of latent memory-strength distributions for old and new items, illustrated in Figure 1.3A. Without any loss of generality, the new-item distribution can be centered at zero with a variance of one. These are arbitrary choices which simply set a location and scale for the model. The old-item distribution is then shifted to the right by an amount dependent on how well a person can discriminate between old and new items. Each person sets a criterion between these strength distributions that will determine which response he or she will provide on a given trial. Figure 1.3B shows a signal detection model with confidence ratings. This model has multiple criteria values along the latent strength axis that determine which response a person will give. In addition to criteria values, the other central parameter in signal detection theory is d', or *d prime*,

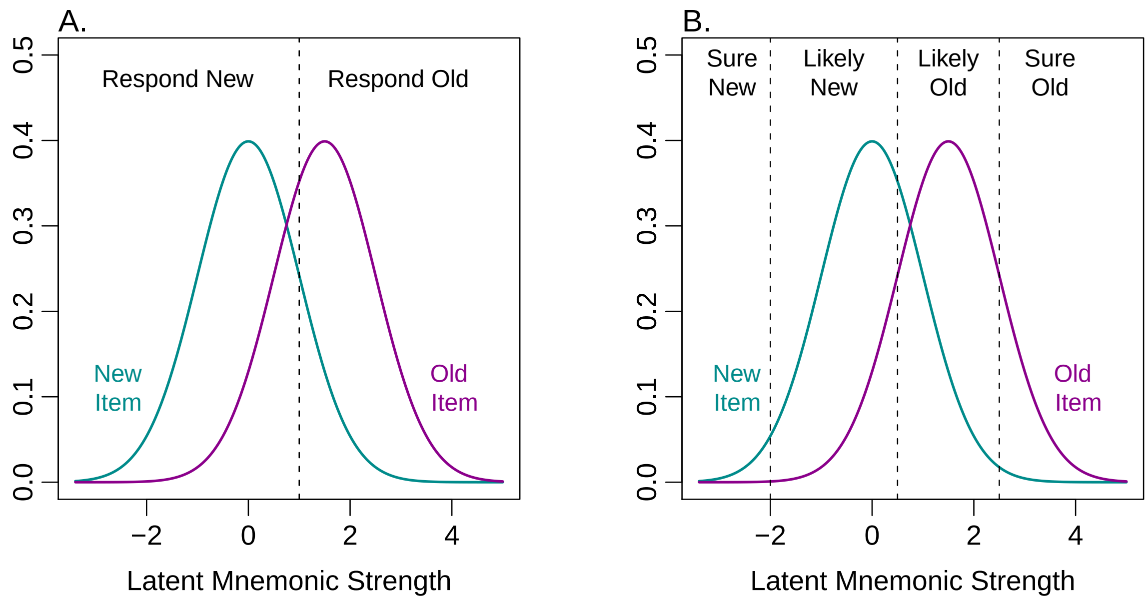

Figure 1.3: Signal detection theory, a continuous model of memory. The dotted lines represent criteria cutoffs. **A.** Model for old-new decisions. **B.** Extension for a confidence-ratings paradigm.

which is the distance between the means of the old and new item distributions. Hit and false alarm rates can be used to estimate these parameters for each participant. In both panels of Figure 1.3, the dotted lines represent criteria cutoff points, and d' is the distance between the maximum of the old and new distributions.

As I previously mentioned, these models make different predictions about the shape of ROC curves. Continuous models such as those of signal detection predict that ROCs will be curved. This curvature is a natural result of the assumptions underlying signal detection theory. See Figure 1.1 for an example of a curved ROC drawn from a signal detection model. Discrete models such as the Fechnerian and double-high threshold models, however, predict straight ROC lines. This again follows from the assumptions and constraints of the memory process within these models, which I will cover in more detail shortly. Researchers have been analyzing ROC data from memory studies for decades, and meta-analyses have revealed strong patterns:

ROCs formed from participant data are nearly always curved (Glanzer, Kim, Hilford, & Adams, 1999; Qin, Raye, Johnson & Mitchell, 2001). This curvature property in ROC data largely led to the abandonment of discrete-state models and their straight-line ROC predictions.

1.2.3 More General Discrete-State Models

Though discrete-state models were largely deserted by cognitive researchers, some felt the basis of the rejection was unstable. It seemed that only very constrained threshold models predicted linear ROCs. Broadbent (1966) noted that the double-high threshold model would only produce ROCs consisting of straight lines if we assumed a participant never used overlapping strategies. For example, in a signal detection study, a participant never says no when they may have heard something. In reality, humans are not always consistent and logical with their decisions. It is possible that a participant may occasionally doubt himself and guess randomly even when he does hear a tone.

The possibility of overlapping strategies is easily applied to recognition memory experiments as well. I will now present a simple memory paradigm and predictions from a couple discrete-state models. Before I go on, I want to introduce the concept of a two-interval (or two-alternative) design. Until now, we have only covered one-interval test designs; that is, at test, a participant sees one item and must decide old or new. A useful alternative is the two-interval design. At test, the participant sees two items—one was studied, and the other is new—and he or she decides which word was studied. Now, confidence ratings range from *Sure Left Word is Target* to *Sure Right Word is Target*. A two-interval design effectively resolves certain issues with the one-interval design, such as the meaningfulness of detecting noise and the ROC

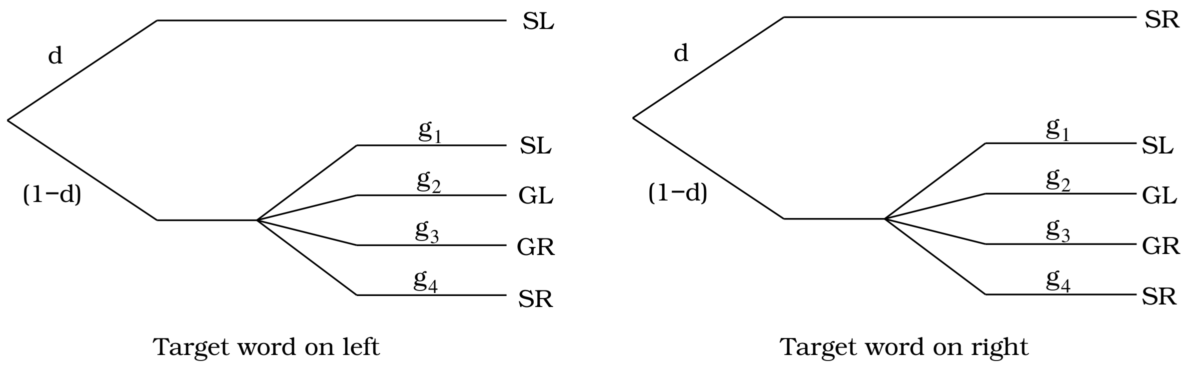

Figure 1.4: Double-high threshold model with certainty assumption. SR=Sure Right, SL=Sure Left, GR=Guess Right, GL=Guess Left

asymmetry that arises from old-new designs. Most importantly, a two-interval design allows the researcher to create a forced-guessing condition wherein the participant is presented with two new words and must decide which is old. I will go into more detail about this useful condition shortly.

Consider a two-interval recognition experiment, with two levels of confidence—*guess* and *sure*. First, the participant studies a list of items. Then, at test, the participant sees two items on a screen: a studied word, and a new word. Items are counterbalanced such that half of the time the old word is on the left, and half of the time it is on the right. On each test trial, the participant is asked to make a response from four choices: *Sure Left* (SL), *Guess Left* (GL), *Guess Right* (GR), *Sure Right* (SR). Under strict assumptions, the participant will always choose *Sure Left* when they detect that the studied word is on the left. We call this the *certainty assumption*. When they do not detect it, they enter a guessing state where they will choose one of the four responses depending on their guessing parameters. This model is illustrated in Figure 1.4. In the model, d is the probability of entering a detect state on a given trial. Naturally, $1-d$ is the probability of entering a guessing state, and g_1, g_2, g_3, and g_4 are the probabilities of endorsing each response given one is in the guessing state. Clearly, these guessing parameters must sum to one; i.e., $\sum_i g_i = 1,\ i = 1, \ldots, 4$.

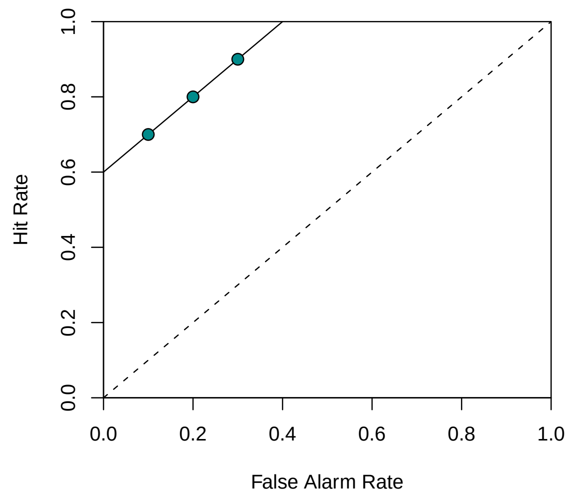

Figure 1.5: ROC curve drawn from a double threshold with certainty assumption model.

From this model, we can determine the probability a participant will choose any of the four responses. Let P_1, P_2, P_3, and P_4 denote the probability the *Sure Left*, *Guess Left*, *Guess Right*, and *Sure Right* response is selected, respectively, given that the target was presented on the left. Note that $\sum_i P_i = 1$. These probabilities are given as

$$P_1 = d + (1-d)g_1 \tag{1.1}$$

$$P_i = (1-d)g_i,\ i = 2, 3, 4. \tag{1.2}$$

Using these probabilities and simulating values for d and g_i, we can construct a hypothetical ROC curve. One ROC constructed from this model is shown in Figure 1.5. Upon inspection we confirm that a threshold model with a certainty assumption does indeed produce a straight ROC line.

But what if a participant does not always follow this pattern? Anyone who has

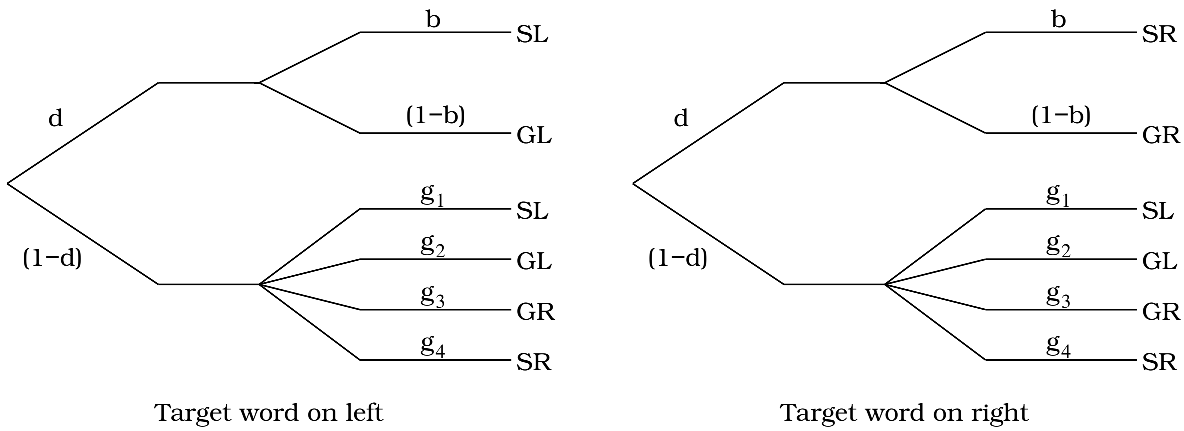

Figure 1.6: Double-high threshold model with no certainty assumption.

analyzed undergraduate participant data can attest to the unpredictability of an individual's test strategy. A simple tweak in the model can completely change the resulting ROC shape. Let's consider a similar model but with the introduction of b, a bias parameter within the detect state. When the participant detects the studied word is on the left, with probability b the participant picks SL, and with probability $1 - b$ she chooses GL. Responses from the guessing state remain the same as in the more constrained model. This model is illustrated in Figure 1.6.

Again we can easily compute the probability the participant will choose any of the four responses in this model, given the target was on the left:

$$P_1 = db + (1-d)g_1, \quad (1.3)$$

$$P_2 = d(1-b) + (1-d)g_2, \quad (1.4)$$

$$P_3 = d + (1-d)g_3, \quad (1.5)$$

$$P_4 = d + (1-d)g_4. \quad (1.6)$$

Using the above probabilities and simulating values for d, b, and g_i produces the ROC plot in Figure 1.7. Clearly with only four confidence ratings the plot does not

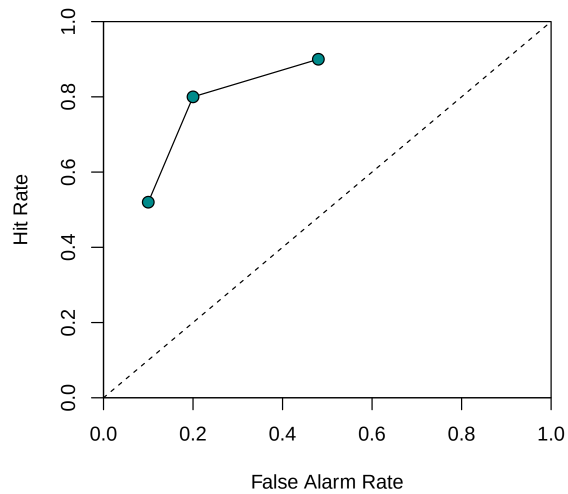

Figure 1.7: ROC curve drawn from a general threshold model with no certainty assumption.

look especially curved. But as we add more ratings to the paradigm, the plot will become more and more smoothly curved. The point is, a general threshold model has no problem producing non-linear ROCs. The variation in response strategies is what makes all the difference. Different response mappings from mental states to confidence ratings can result in straight or curved ROCs. Therefore, it is important to determine how participants are distributing their responses.

1.3 Testing Constraints of Discrete-State Models

As I have shown, discrete-state models can easily produce nonlinear ROC plots when the certainty assumption is relaxed. Without this assumption, discrete models predict ROC points that are connected with straight lines; therefore, these models can account for any single ROC curve and provide no constraint in a single condition. Most researchers were content to stop here, convinced discrete-state models were not worth further consideration. A few, however, realized that examining ROCs alone may not be informative (Lockhart & Murdock, 1970). Indeed the key to strong and testable constraints in discrete-state models lies not in single ROCs but in families of ROC curves. These constraints arise from a paramount characteristic of discrete-state models called *conditional independence*, which means that an individual's responses are determined only by which state one is in. Stimulus factors, such as the number of study repetitions, only affect the probability of entering the detect state. Regardless of how a person enters a state, responses will look the same once in it. In order to observe this constraint in data, the experimenter must implement multiple within-subject conditions. Ideally, the stimulus manipulation should produce clear differences in performance in each condition.

The conditional independence constraint can be observed in data multiple ways. One way focuses on the distribution of confidence rating responses across various conditions, which I will touch on shortly. To keep in line with the previous discussion, I will first focus on how this constraint is illustrated in ROC curves.

1.3.1 Discrete-State Model Constraints in ROC plots

We construct a realistic example in which targets are presented on the left or right across several stimulus conditions, such as repetition at study. Participants respond with a limited set of four confidence ratings. Let $i = 1, \ldots, 4$ index the confidence rating response (1=SL, 2=GL, 3=GR, 4=SR), $j = 1, 2$ index the side of the target (1=left, 2=right), and let $k = 1, \ldots, K$ index the stimulus condition.

The general discrete-state model may be given as

$$P_{ijk} = d_k b_{ij} + (1 - d_k) g_i. \tag{1.7}$$

For example, the probability of selecting *Sure Right* given the target was on the left in the second stimulus condition is $P_{412} = d_2 b_{41} + (1 - d_2) g_4$. The previous model shown in Figure 1.6 with a single bias parameter b may be implemented in (1.7) with the following restrictions:

$$b_{11} = b, \quad b_{21} = 1 - b, \quad b_{31} = 0, \quad b_{41} = 0, \tag{1.8}$$

and

$$b_{12} = 0, \quad b_{22} = 0, \quad b_{32} = 1 - b, \quad b_{42} = b. \tag{1.9}$$

Under these restrictions, P_{412} simplifies to $(1-d_2)g_4$. Going forward with our example, we will use four confidence ratings with three stimulus conditions termed *forced-guessing* ($k = 1$), *difficult* ($k = 2$), and *easy* ($k = 3$). In the easy condition, words are studied four times; in the difficult condition, words are studied once; and in the forced-guessing condition, words are not studied at all (but participants are still asked to make a response). As I mentioned before, the forced-guessing condition for memory experiments is only possible in a two-interval design, when two new words are presented and the participant is asked to choose which is old. This is a sneaky condition in that there is no correct answer, but it is exceptionally useful for modeling discrete-states. In these trials, participants have no chance of entering a detect state, so we can use these responses to accurately estimate their guessing distributions. We then have a clear picture of what each participant's guessing state looks like in terms of confidence ratings.

We can think of the three conditions in terms of d_k, the probability of detection. Since we now have multiple conditions, d_k takes on different values depending on the condition. In the forced-guessing condition, d_1 must equal zero, as the participant has no chance of detecting a studied word. The ROC curve for the forced-guessing

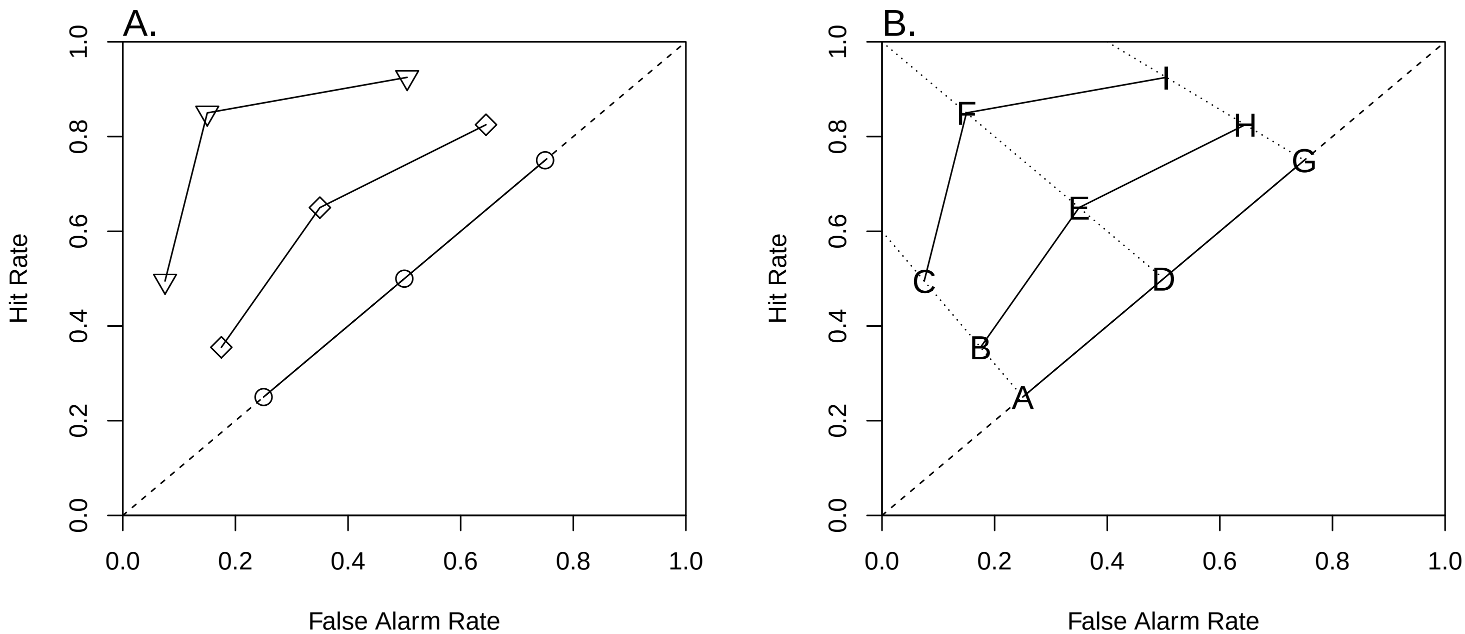

Figure 1.8: ROC curves for a two-interval design across three conditions. **A.** Triangles = easy ($k = 3$), diamonds = difficult ($k = 2$), circles = forced guessing ($k = 1$). **B.** Same ROC plot with points as letters for ease of explanation.

condition will be a straight line on the diagonal, indicating chance performance (see the circles in Figure 1.8A).[1] ROC curves are pulled off the diagonal as d becomes larger. Thus, in the difficult condition, the chance of detection may still be small, but is greater than zero. For this example, let's say $d_2 = 0.3$. In Figure 1.8A, the difficult condition ROC curve is represented by diamond points. Finally, d is largest in the easy condition. The chance of detection becomes high, and overall accuracy increases. ROC points from the easy condition are represented as triangles in Figure 1.8A, with $d_3 = 0.7$.

Comparing single ROC curves across participants or experiments may not be very meaningful. But comparing ROCs within participants across conditions can be very

[1]I discuss all predictions in the large-sample limit. Clearly, for any observed data, the ROC points will cluster around the diagonal but not fall exactly on it.

informative. Examining ROC families is where we can see the concept of conditional independence manifested.

Statement of Constraints

The two constraints in ROC plots from discrete-state models can be stated as follows:

1. **Points From A Particular Confidence Rating Category** The segment connecting all ROC points from the same confidence rating across different conditions will be a straight line.

2. **Intercepts** Under the restrictions in (1.8) and (1.9), these connecting lines will have the following intercepts: for the *Sure Left*, *Guess Left*, and *Guess Right* categories, the lines pass through points $(0, b)$, $(0, 1)$, and $(1-b, 1)$, respectively.

Figure 1.8B illustrates these constraints. Points A, B, and C are from the *Sure Left* response for the forced-guess, hard, and easy conditions, respectively. According to the first constraint, the slope between points A and B and points B and C must be equal. According to the second constraint, the y-axis intercept is b when the restrictions in (1.8) and (1.9) are assumed. Points D, E, and F are ROC points from the *Guess Left* response; the slopes between points D and E and points E and F are equal. These points do not rely on b; the slope is instead constrained to pass through the point $(0, 1)$, the upper-left corner of an ROC plot. Finally, the same equal-slope constraint applies to points G and H and H and I (for the *Guess Right* response), and a line through these points will cross the point $(1 - b, 1)$. A key property of this constraint is that none of the slopes rely on d, the probability of detection. Once a person enters a given state, the distribution of responses is invariant across conditions because of conditional independence. A proof of this invariance for the case when (1.8) and (1.9) hold is provided.

Proof of ROC Constraint

Let $s(AB)$ denote the slope of the line from points A to B in Figure 1.8B, and let the other slopes be notated equivalently. The invariance of interest is given by the following three equalities: $s(AB) = s(BC)$, $s(DE) = s(EF)$, and $s(GH) = s(HI)$. Let $x(A)$ and $y(A)$ denote the x-axis and y-axis values for point A in Figure 1.8B, respectively, and let the x-axis and y-axis values for the other points be notated equivalently.

We start with the first equality, $s(AB) = s(BC)$. Expressions for the x-axis and y-axis values of the three points in terms of the discrete-state model parameters are

$$\begin{aligned}
x(A) &= P_{121} = g_1 \\
y(A) &= P_{111} = g_1 \\
x(B) &= P_{122} = (1 - d_2)g_1 \\
y(B) &= P_{112} = d_2 b + (1 - d_2)g_1 \\
x(C) &= P_{123} = (1 - d_3)g_1 \\
y(C) &= P_{113} = d_3 b + (1 - d_3)g_1
\end{aligned}$$

The slopes are

$$\begin{aligned}
s(AB) &= \frac{y(B) - y(A)}{x(B) - x(A)} = \frac{g_1 - b}{g_1} \\
s(BC) &= \frac{y(C) - y(B)}{x(C) - x(B)} = \frac{g_1 - b}{g_1}
\end{aligned}$$

Clearly, the slopes are equal and do not depend on d. Moreover, the slope equation ensures a line through these points will intersect the y-axis at the value of b.

In the next equality of interest, $s(DE) = s(EF)$, the expressions for x-axis and y-axis values are

$$
\begin{aligned}
x(D) &= P_{121} + P_{221} = g_1 + g_2 \\
y(D) &= P_{111} + P_{211} = g_1 + g_2 \\
x(E) &= P_{122} + P_{222} = (1 - d_2)g_1 + (1 - d_2)g_2 \\
y(E) &= P_{112} + P_{212} = d_2 b + (1 - d_2)g_1 + d_2(1 - b) + (1 - d_2)g_2 \\
x(F) &= P_{123} + P_{223} = (1 - d_3)g_1 + (1 - d_3)g_2 \\
y(F) &= P_{113} + P_{213} = d_3 b + (1 - d_3)g_1 + d_3(1 - b) + (1 - d_3)g_2
\end{aligned}
$$

The slopes are

$$
\begin{aligned}
s(DE) &= \frac{y(E) - y(D)}{x(E) - x(D)} = \frac{g_1 + g_2 - 1}{g_1 + g_2} \\
s(EF) &= \frac{y(F) - y(E)}{x(F) - x(E)} = \frac{g_1 + g_2 - 1}{g_1 + g_2}
\end{aligned}
$$

As predicted, these slopes are equal and do not depend on the probability of detection parameter. Also, a line through points D, E, and F is constrained to cross the point $(0, 1)$.

Finally, we can prove the last slope equality, $s(GH) = s(HI)$, using the following expressions for the x-axis and y-axis values of points G, H, and I:

$$
\begin{aligned}
x(G) &= P_{121} + P_{221} + P_{321} = g_1 + g_2 + g_3 \\
y(G) &= P_{111} + P_{211} + P_{311} = g_1 + g_2 + g_3 \\
x(H) &= P_{122} + P_{222} + P_{322} = (1-d_2)g_1 + (1-d_2)g_2 + d_2(1-b) + (1-d_2)g_3 \\
y(H) &= P_{112} + P_{212} + P_{312} = d_2 b + (1-d_2)g_1 + d_2(1-b) + (1-d_2)g_2 + (1-d_2)g_3 \\
x(I) &= P_{123} + P_{223} + P_{323} = (1-d_3)g_1 + (1-d_3)g_2 + d_3(1-b) + (1-d_3)g_3 \\
y(I) &= P_{113} + P_{213} + P_{313} = d_3 b + (1-d_3)g_1 + d_3(1-b) + (1-d_3)g_2 + (1-d_3)g_3
\end{aligned}
$$

The slopes are

$$
\begin{aligned}
s(GH) &= \frac{y(H) - y(G)}{x(H) - x(G)} = \frac{g_1 + g_2 + g_3 - 1}{g_1 + g_2 + g_3 + b - 1} \\
s(HI) &= \frac{y(I) - y(H)}{x(I) - x(H)} = \frac{g_1 + g_2 + g_3 - 1}{g_1 + g_2 + g_3 + b - 1}
\end{aligned}
$$

Again, we confirm that the slopes are equal and depend only on the bias and guessing parameters. This slope equation guarantees that a line through the points will cross $(1-b, 1)$.

Examples of ROC Constraint

Up to this point, I have drawn all ROC plots with the set parameter values $g1 = g2 = g3 = 0.25$ and $b = 0.6$ for the sake of clear, discernible graphs. The slope constraint I just proved is easy to visualize in Figure 1.8 because the points all line up nicely and are evenly spaced apart. Obviously this is not always the case in real data, but the constraint will still hold no matter what the values are for g_i, b, and d_k. I illustrate this fact in Figure 1.9, which displays three ROC plots with widely varying parameter

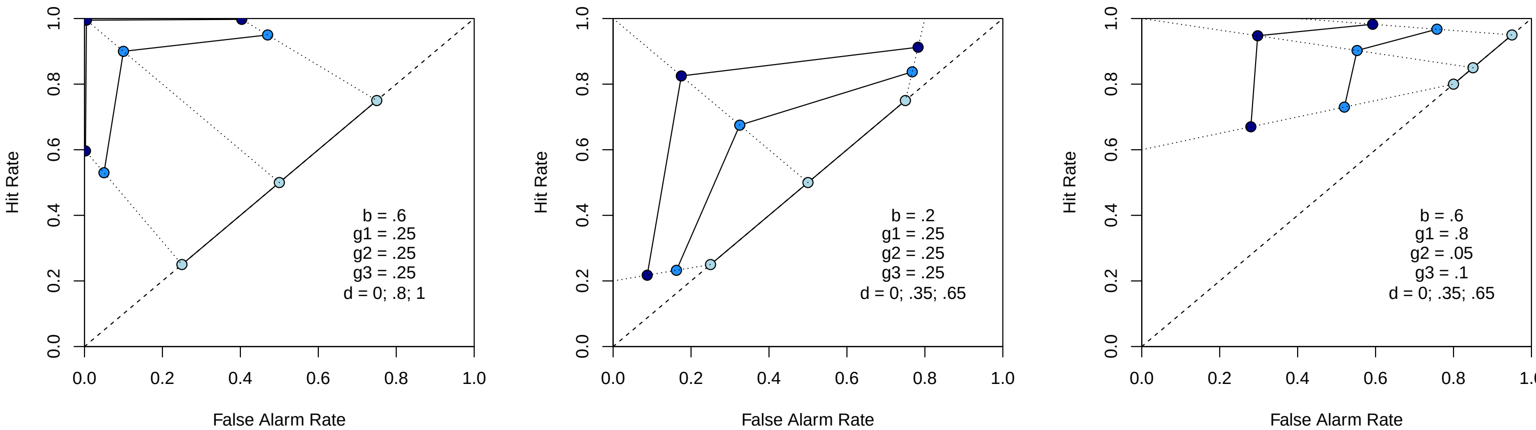

Figure 1.9: Three ROC plots drawn from the same discrete-state model with varying parameter values. The dotted lines connecting points show that the slope and intercept constraints always hold.

values. These plots also demonstrate what changing different parameter values will do to a resulting ROC curve.

Again, the point to stress here is that the conditional independence constraint in discrete-state models can only be seen in families of ROCs; therefore, it is necessary to test participants across multiple conditions.

1.3.2 Discrete-State Model Constraint in Confidence Ratings

Though the conditional independence constraint in ROC families is robust, it is difficult to see on the individual level because of a limited number of trials and the amount of noise in individual data. There is, however, another way to visualize this constraint in participant data: histograms of individual confidence rating responses. Because of the nature of discrete states, it is necessary that a participant's responses across confidence ratings reflect only a mixture of guessing and detect states. For example, in a forced-guessing condition, a participant distributes responses in a way that reflects his guessing state, taking into account his own biases and strategies. In the easiest condition, when accuracy is high, responses mostly reflect a participant's detect state. The key to conditional independence lies in intermediate conditions. Discrete-state theory predicts responses in these conditions will be clearly distributed as a mixture of the two states. It is easiest to see this constraint graphically. Predictions from a discrete model for various stimuli conditions are displayed in Figure 1.10.

Figure 1.10A shows hypothetical response distributions for one participant's detect and guessing states, given the target was always on the right. These distributions were drawn from a paradigm that utilizes a continuous scale of response ratings. The guessing distribution is centered around the middle of the confidence scale, with a large variance. The detect distribution is skewed toward a high-confidence *Sure Right* response, with a tighter variance. The solid line in Figure 1.10B shows the participant's response distribution in the forced-guessing condition. As you can see, responses in this condition are a pure reflection of the guess-state distribution. The solid line in Figure 1.10C shows the distribution of responses in a difficult stimulus condition. It is a mixture of predominantly the guessing distribution with some detect state input. Finally, the distribution in Figure 1.10D is from the easiest stimulus condition. Here, the mass is mostly in the detect state, with some still in guessing. This mixture of guessing and detect states in response distributions is a direct prediction of discrete-state models.

A latent-strength model, in contrast, would predict a single, normal distribution of responses that simply shifts toward higher confidence ratings as the task becomes easier. The latent-strength model predictions for responses are the same as the discrete model in the forced-guessing condition. The key difference is in the intermediate condition when responses do not reflect a mixture of detection and guessing, but sim-

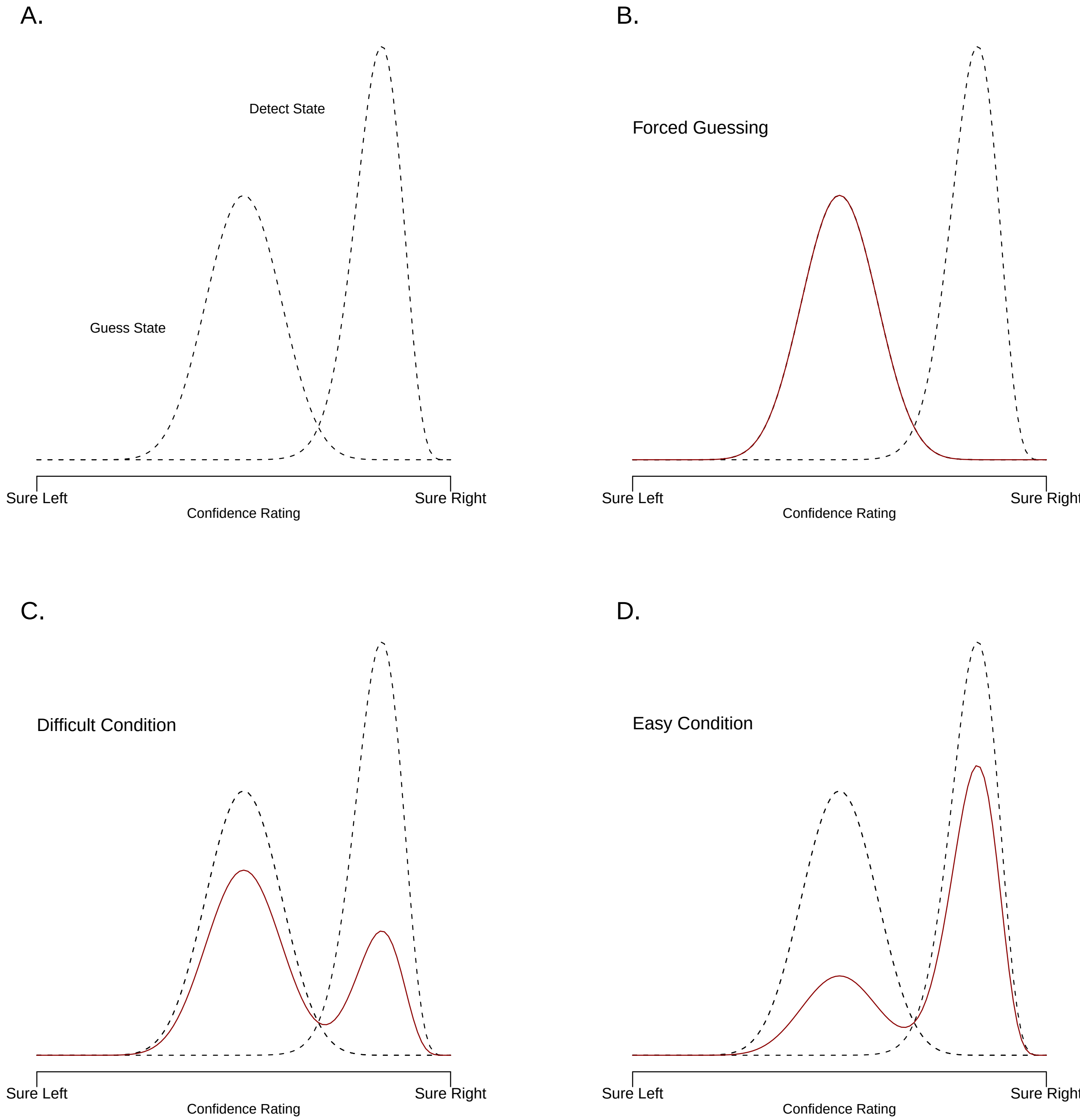

Figure 1.10: Discrete-state model predictions for continuous confidence ratings across three conditions. For ease of representation, it is assumed here that the target word is always on the right. Pure detection and guessing states are represented with dotted lines. Confidence rating distributions for each condition are shown with a solid red line.

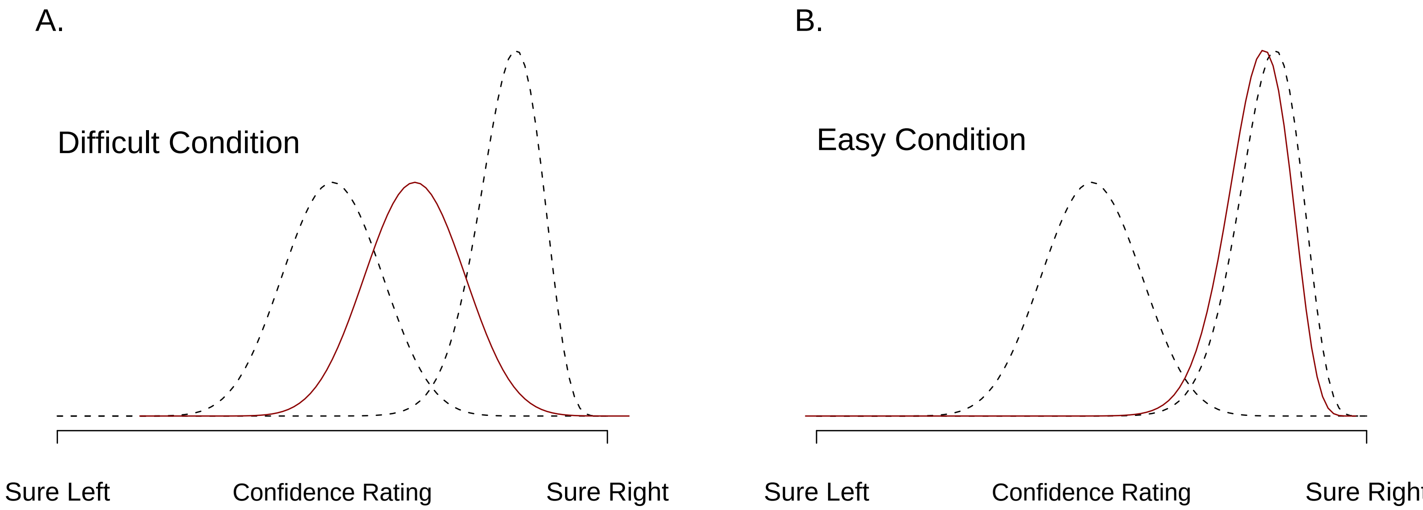

Figure 1.11: Latent-strength model predictions for continuous confidence ratings in two conditions. The target word is assumed always on the right. Hypothetical detect and guess states are represented with dotted lines. Responses are shown with a solid red line. **A.** Difficult condition predictions. **B.** Easy condition predictions.

ply a shift in perceptual strength along the confidence scale. The shift in confidence for a latent-strength model is illustrated in Figure 1.11.

Data Representation

Histograms of confidence ratings are an easy way to look for discrete or continuous patterns in data. In this section, I want to briefly describe the data representation I will use going forward for displaying participant histograms.

We first take all responses and group them by condition. We usually pool responses from the medium and difficult conditions together to simplify the plot. Plotting raw confidence ratings from each condition is shown in Figure 1.12A with forced guessing responses in yellow, medium and difficult responses in green, and easy responses in blue. We then flip all correct leftward responses to the right side, so that the x-axis

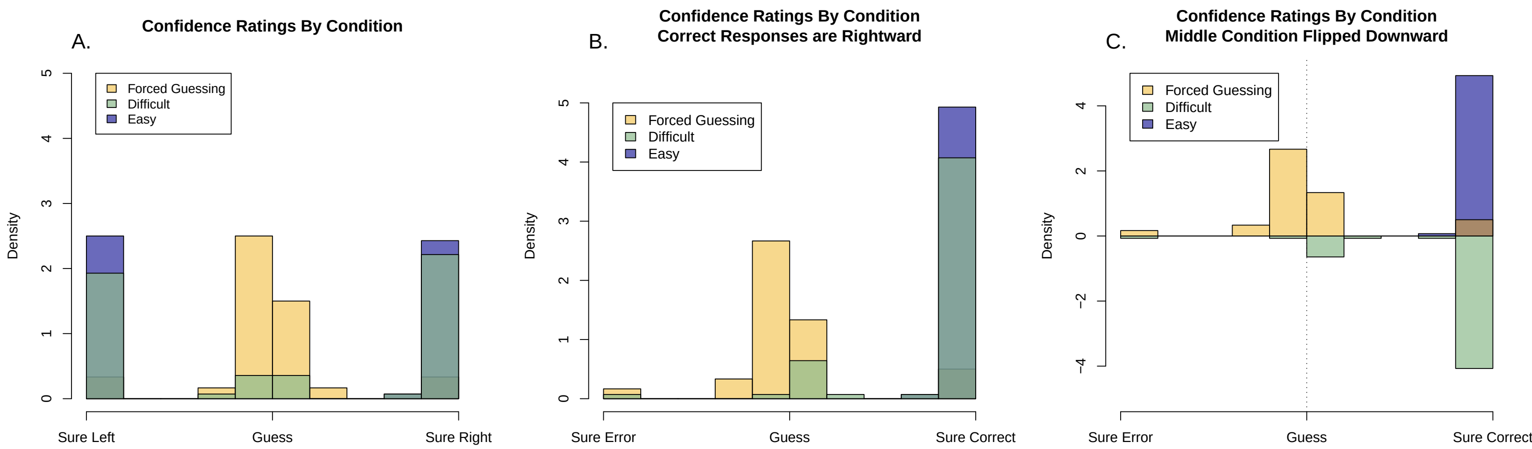

Figure 1.12: Construction of confidence ratings histograms for one participant's data. **A.** Raw confidence rating responses for all trial types and conditions (medium and difficult condition responses are pooled). **B.** Responses to leftward targets have been flipped such that all correct responses are now on the right and errors are on the left. **C.** Responses from the difficult condition are mirrored downward on the x-axis.

now ranges from high-confidence errors to high-confidence correct responses, shown in Figure 1.12B. Because overlapping conditions and colors can be unclear, responses from trials in the hard condition are then projected downward, which also gives a more direct comparison of responses in the three conditions. The final product is shown in Figure 1.12C.

1.3.3 Previous Work

Province and Rouder (2012)

I believe discrete-state models can do an adequate job of accounting for recognition data. In recent research, my colleagues have found these models to perform as well as, or better than, leading signal detection and dual-process models in several memory tasks. For example, Province and Rouder (2012) found evidence for discrete-state processing in various recognition memory experiments. They utilized a two-interval design, with a continuous confidence rating scale. The stimulus manipulation was number of word repetitions at study. Words were studied either once, twice, four times, or not at all (in the forced-guessing condition). Support was found for discrete-state models via formal model fittings and comparison of log-likelihood ratios and G^2 scores. Visual inspection of participant confidence distributions also garnered support for discrete states. The authors displayed several such plots of individual confidence ratings across repetition conditions using a similar data representation as the one I just described, and I include one in Figure 1.13.

In Figure 1.13, responses have been adjusted so that correct responses were always on the right, and responses from the one-repetition and two-repetition conditions have been collapsed into one intermediate category. This participant was peculiar in her guessing responses: notice that instead of a single guessing distribution centered around the middle of the confidence scale, the participant was partial to responding off to the sides, slightly favoring the right side. Nonetheless, we can clearly see that her responses in the other conditions were obvious mixtures of her detect and guess states. The mass simply shifted out of the guessing distribution(s) and into the detect distribution as study repetition increased. This participant's plot also demonstrates why it is crucial to look at these plots individually by person. An aggregation across subjects just doesn't make sense in this kind of analysis since participants have differing response strategies.

The work by Province, Rouder, and others finding evidence for discrete states in cognition inspired the experimental work of this thesis. I wanted to replicate findings for recognition memory as well as extend the paradigm to word identification. Next I will introduce two new experiments designed for my thesis and subsequent findings supporting discrete-state models for a majority of participants.

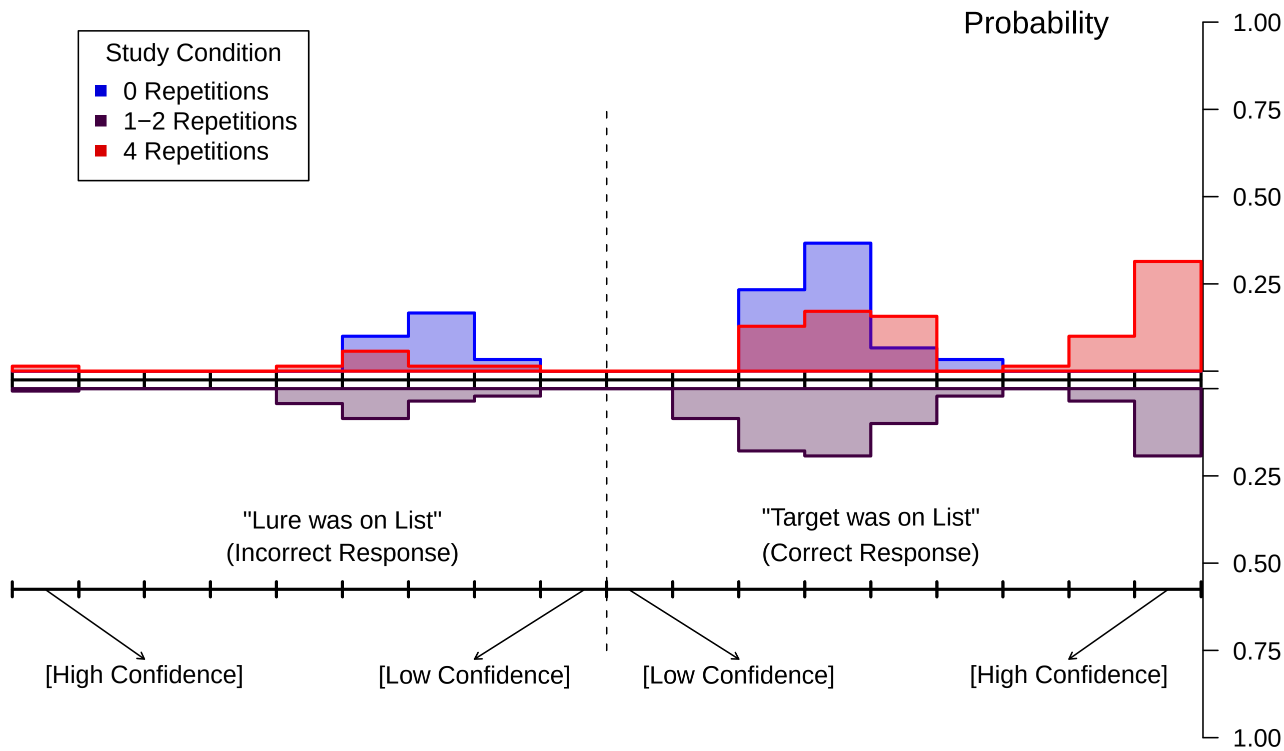

Figure 1.13: Histograms of confidence ratings across repetition conditions for a selected participant in the Province and Rouder (2012) study.

Chapter 2

Experiment 1: Analysis of Two-Interval Recognition Memory and Verbal Perception Tasks

2.1 Overview of Experiment

Experiment 1 was constructed to compare the processes underlying recognition memory and visual perception of verbal stimuli. As I previously mentioned, many modern models of recognition memory were born out of early signal detection studies which were perceptual in nature. Whereas classic signal detection studies utilized audio perception, I was interested in analyzing visual perception in the form of word identification. The discrete-state models I have discussed can be implemented in any domain where signal detection is used.

I constructed an experiment which included both a recognition memory task and a visual perception task. The recognition memory task consisted first of a word study phase with a repetition manipulation, followed by a standard two-interval test phase.

The perception task was designed to mirror the memory task in design. A word was flashed quickly on the screen followed by a test where participants had to choose which of two test items was the flashed word. The stimulus strength manipulation in the perception task was the length of time in milliseconds the target was flashed. Both tasks had four conditions of varying difficulty, including a forced-guessing condition.

The perception task has a few methodological benefits over memory tasks which makes it attractive. The forced-guessing condition I previously introduced is an extremely useful tool for estimating guessing states in discrete-state models; however, it can be difficult to implement in certain designs. Occasionally, high-performing participants will notice and become confused when they are presented with two new items. Forced guessing is more straightforward in a visual perception design. We can simply flash the target too quickly to be detected and follow through with a two-interval test. It would be entirely reasonable for certain flash trials to be too difficult to accurately respond to, and thus participants don't become suspicious. The perception task also works well as a one-interval design, which I will talk more about in the next chapter.

Since each participant completed both the perception and memory tasks, we can make comparisons across the two processes. It is especially useful to look at a participant's distribution of responses conditional on a given state for both tasks and see how they match. Analyzing how a person responds in a given state helps us answer certain questions: How are the underlying processes of recognition memory and visual perception similar or different? Can they both be represented with discrete-state models? Might both tasks be mediated by a common decision process?

2.2 Method

2.2.1 Participants

Thirty-nine University of Missouri students served as participants in exchange for course credit.

2.2.2 Stimuli

Two separate lists of 480 items each were used for the memory and perception tasks. In the perception task, each item was a six or seven letter noun. In the memory task, word length varied from four to nine characters. Both lists were constructed from the MRC Psycholinguistic Database (Coltheart, 1981). For each participant, a subset of 240 items per task were chosen at random to serve as the target items, and these items were randomly crossed with experimental factors.

2.2.3 Procedure

The experiment was divided into three phases. First, participants studied 240 words from the memory list. Seventy of these 240 items were studied once, another 70 were studied twice, and a third 70 were studied four times. Thirty items were studied zero times and reserved for the forced-guessing condition. Each item was studied for 2 seconds, with the exception of forced-guessing items which were skipped, and the entire study phase took about 15 minutes.

Following the study phase, participants performed the perception task. Half of the 480 words in this task served as targets; the other half served as lures. Seventy words were flashed for 117 ms (easy condition), 70 were flashed for 83 ms (moderate), 70 were flashed for 50 ms (hard), and 30 were flashed for 0 ms, which again comprised the forced-guessing condition. A visual mask of nine punctuation characters was presented immediately before and after each flashed target word to suppress iconic memory. After the post-stimulus mask, participants were presented with the target word and a lure item, one on the left and one on the right of the computer screen. Participants responded to the test items using a continuous confidence scale, weighted by *Sure left word* and *Sure right word*, with guessing responses in the middle. To motivate accurate performance and keep participants engaged, we introduced a scoring system wherein they could gain points for being correct on each trial, and lose points for being incorrect. The amount of points one could gain or lose increased with higher confidence responses. To discourage people from constantly responding with high confidence, we employed an exponential loss structure such that a person could lose up to 500 points for being wrong, and a simple additive scale for gains, with a maximum gain of 100 points. Participants controlled a slider along the horizontal confidence scale and made a response via mouse click. They received immediate feedback on their accuracy, and their points total was always visible.

After the perception task, participants completed the memory task for the studied items. All 240 memory test trials consisted of a simple two-interval test, with a studied word and new word presented on either side of the screen. Under the two items was a continuous confidence ratings scale ranging from *Sure left word* to *Sure right word*, with guessing responses in the middle. The same points structure was employed, and participants again moved a slider along a horizontal scale and made a response via mouse click.

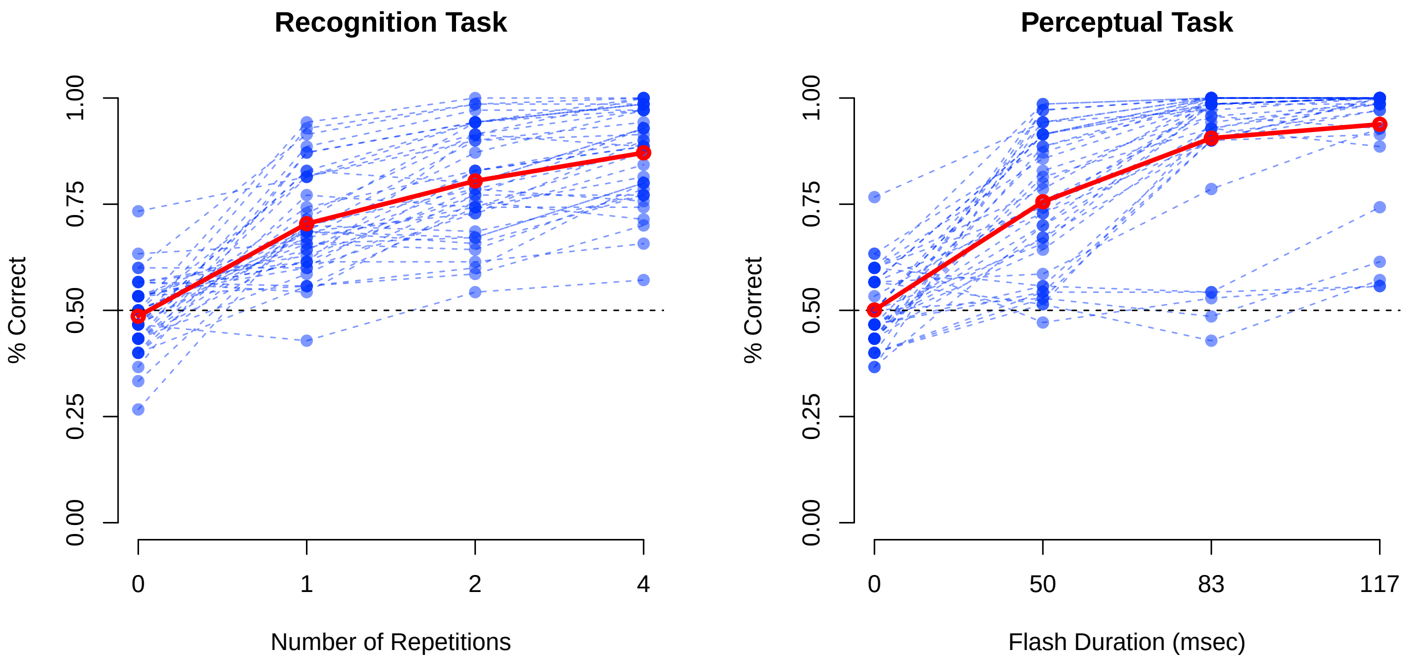

Figure 2.1: Average accuracy scores for all participants across all conditions in both the memory and perception tasks. The solid red line indicates the grand mean across conditions.

2.3 Results

The scoring system proved useful in keeping participants from extreme responding, and many reported the points helped keep them motivated throughout the task. Overall mean accuracies per individual across stimuli conditions for both tasks is shown in Figure 2.1. Average accuracy in the memory task increased in an additive fashion from one repetition condition to the next for the majority of participants, as shown in the left panel of Figure 2.1, which proves the manipulation was successful. Results for the perception task, however, showed a different pattern. The majority of participants reached near-ceiling performance in the intermediate and easy conditions, whereas a small group of participants barely made it above chance performance, even in the easiest condition. These results illustrate the rather surprising degree of

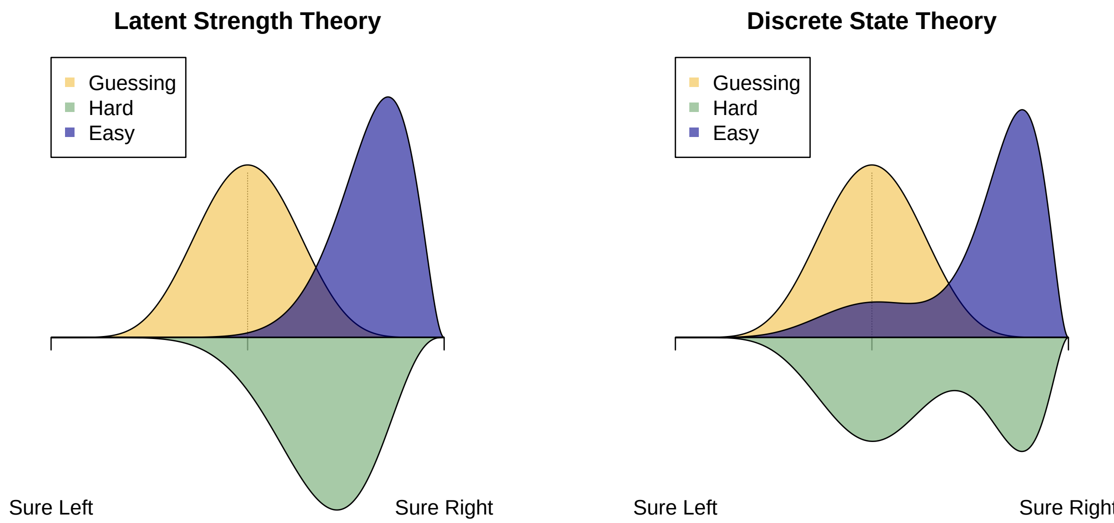

Figure 2.2: Confidence rating predictions for latent strength and discrete state models using the new data representation.

variability in visual perception ability.

2.3.1 Confidence Rating Histograms

Despite the differing data patterns, analysis of the data led to support for discrete-state models in both tasks. One way to see this is by observing participant confidence rating histograms, looking for a discrete-state pattern. I will be using the same data representation I described previously. Responses in the middle and difficult conditions have been pooled into one difficult category. In these participant histograms, we are looking for the latent strength and discrete state patterns shown previously in Figures 1.10 and 1.11. The two models' predictions, using the new data representation, are shown in Figure 2.2.

The participant in Figure 2.3 consistently responded in the middle of the scale (around the *guess* response) when she was unsure of the answer. In the perception

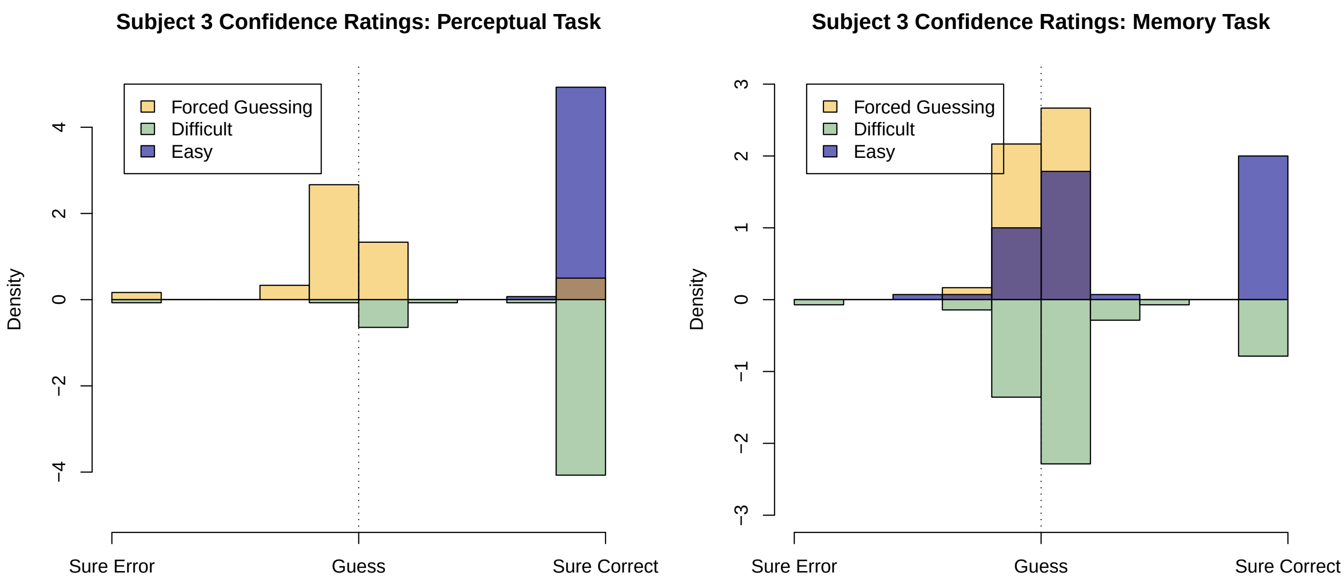

Figure 2.3: Confidence rating histograms for Subject 3, both tasks. This pattern shows evidence of discrete states.

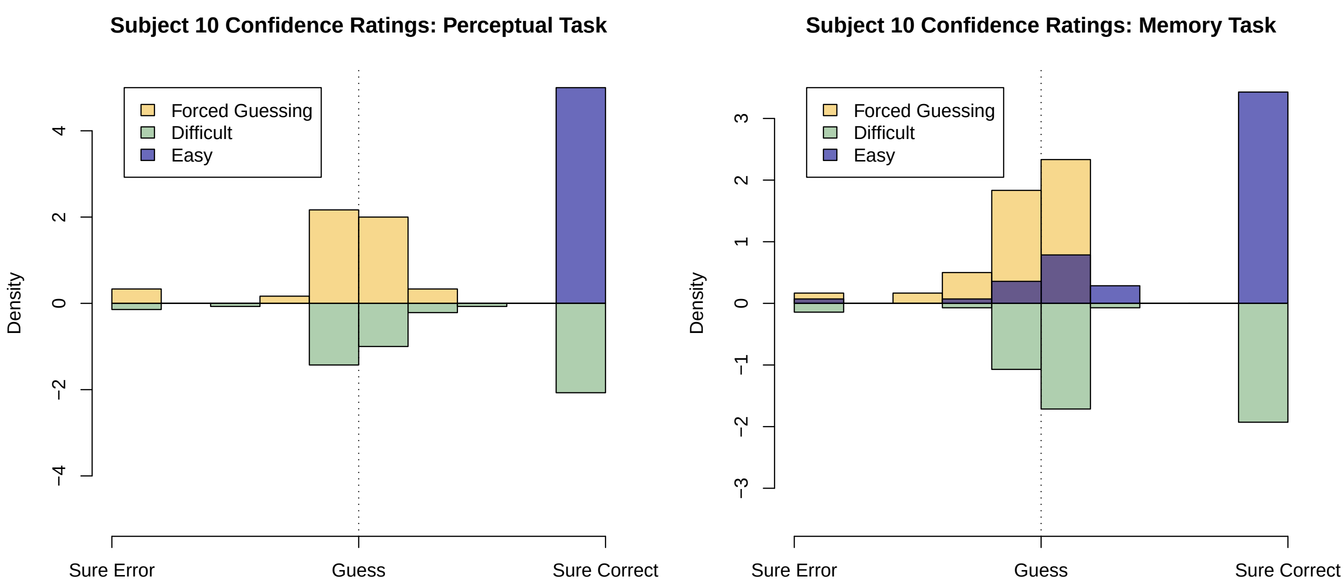

Figure 2.4: Confidence rating histograms for Subject 10, both tasks. This pattern also shows evidence of discrete states.

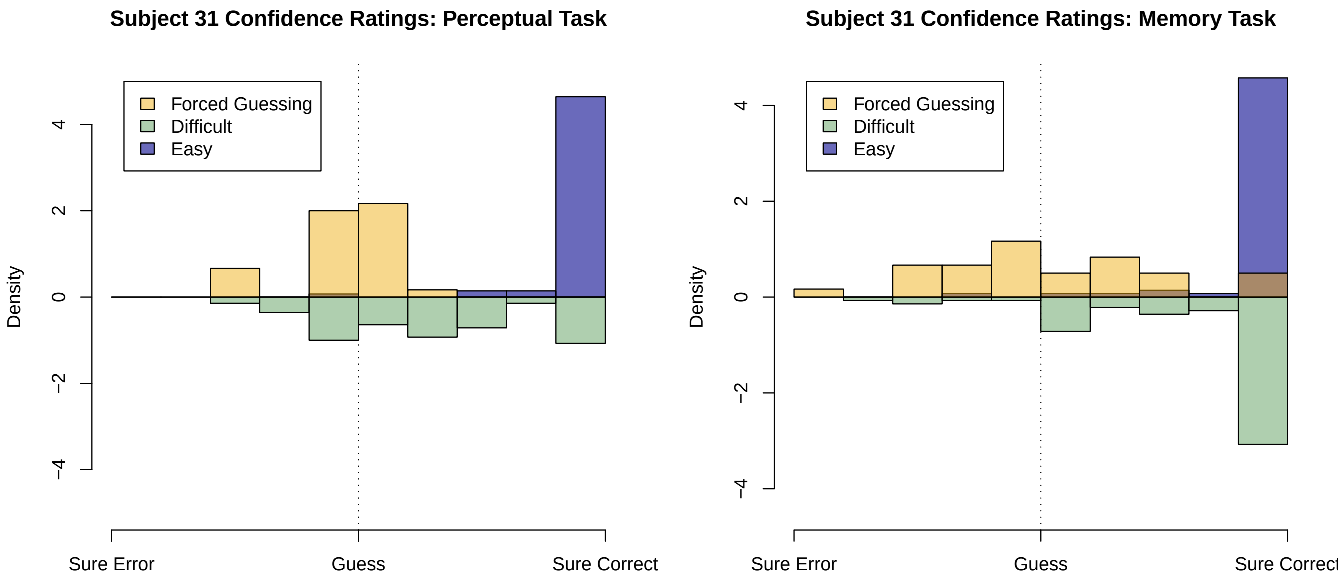

Figure 2.5: Confidence rating histograms for Subject 31, both tasks. This participant shows greater support for a latent-strength model. Notice especially in the memory task how responses in the hard condition are shifted to the right of the forced guessing distribution.

task, we see the mixture of the participant's guessing and detect distributions very clearly in the 50 ms (difficult) condition. This participant was less likely to enter a detect state in the memory task, but still responds from a clear mixture of guessing and detect states in the difficult one-repetition condition.

Figure 2.4 shows another example of a participant with clear discrete-state patterns of responding in both tasks. Figure 2.5, however, shows responses from a participant who was not well fit by a discrete-state model. Looking at his responses particularly in the perception task (left panel), we see more of a shift in responses in the hard condition rather than the mixture of states characteristic of discrete models.

2.3.2 Model Fittings

Observing histograms of confidence ratings is a nice way to visually search for evidence of discrete states or latent strength, but further testing is required to confirm one model outperforms the other. I therefore constructed a ten-parameter discrete-state model and a ten-parameter latent-strength model which are generalized from the models described in the first chapter. We compared models with the same number of parameters because we feel it makes the comparison more straightforward and convincing.

Because of the nature of two-interval designs, a natural symmetry arises in the formation of models. Manipulating whether the target item is presented on the left or right side of the screen should have no effect on detection. That is, the probability of detecting the target on the left in a certain condition is equal to the probability of detecting a target on the right in the same condition. Therefore, we only need three detection parameters (one for each condition) per individual.

When a continuous confidence scale is used, the researcher can partition the scale into bins to analyze the models. I partitioned the scale into eight bins, with four on each side. This model has three bias parameters instead of just one, to account for all possible detect responses. The fourth bias parameter is constrained to equal $1 - b_1 - b_2 - b_3$. Similarly, the design has eight possible guessing responses, which can be accounted for by adding another type of bias parameter (denoted S) to the guessing state. This parameter estimates how likely a person, once guessing, responds leftward or rightward. With the addition of S, only three guessing parameters are needed (because again $g_4 = 1 - g_1 - g_2 - g_3$). The ten-parameter discrete-state model is illustrated in Figure 2.6.

The latent-strength model extends naturally from the signal detection model pictured in Figure 1.3 to ten parameters. Similarly to the discrete-state model, we expect

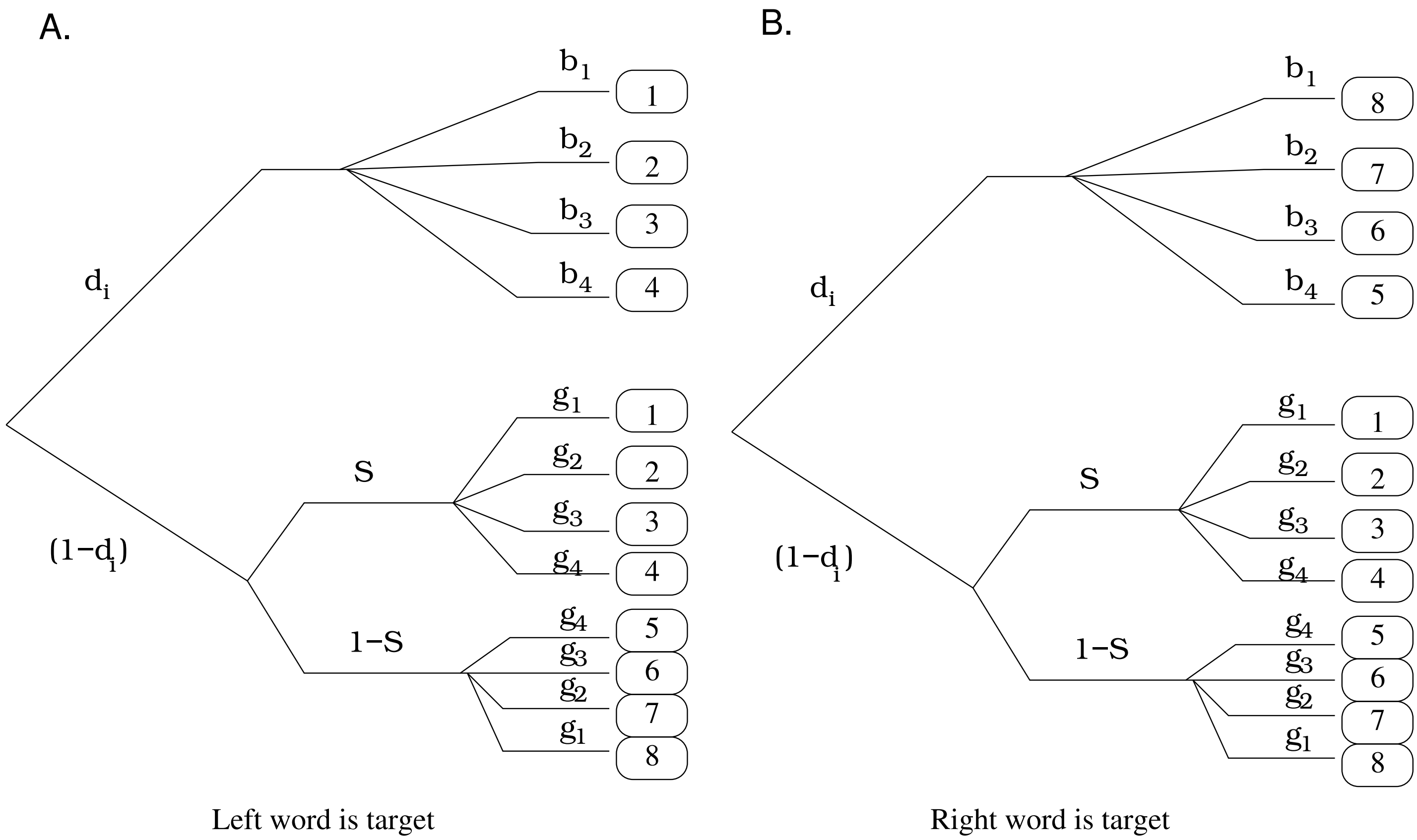

Figure 2.6: Ten-parameter discrete-state model. The numbers 1 to 8 represent confidence ratings from sure left to sure right, respectively. **A.** Model given target was presented on left. **B.** Model given target was presented on right.

the d' values given the target was on the left or right to be equally distanced from zero (i.e. $d'_{iL} = -d'_{iR}$) within each condition. We now have three d' values, one for each condition, and seven free criterion parameters. This is quite a flexible model because the criteria lines can fall virtually anywhere along the latent-strength axis.

I was then able to estimate parameters for both models and each participant. I used maximum likelihood estimation to obtain the best fits and to produce two deviance values for each participant and each model. Subtracting the discrete-state model likelihood from the latent-strength model likelihood for each participant gives a difference score indicating which model produced a better fit. Because I compared models with equal numbers of parameters, this goodness-of-fit test is comparable to AIC or BIC. In this case, negative values show greater support for the latent-strength model. Models were fit separately for the memory and perception tasks.

Figure 2.7 shows model deviance scores for each participant and both tasks in a cumulative density function format. Each point represents a participant's deviance score. Model comparisons revealed that the discrete-state model outperformed the latent-strength model for 26 of 39 participants (67%) in the memory task, and for 30 of 39 participants (77%) in the perception task. This is fairly strong evidence for a discrete-state model, especially for perception.

I conclude from Experiment 1 that discrete-state models are at least as viable as latent-strength models for two-interval recognition memory and verbal perception tasks. My interest to further explore perception and one-interval designs led to the development of Experiment 2.

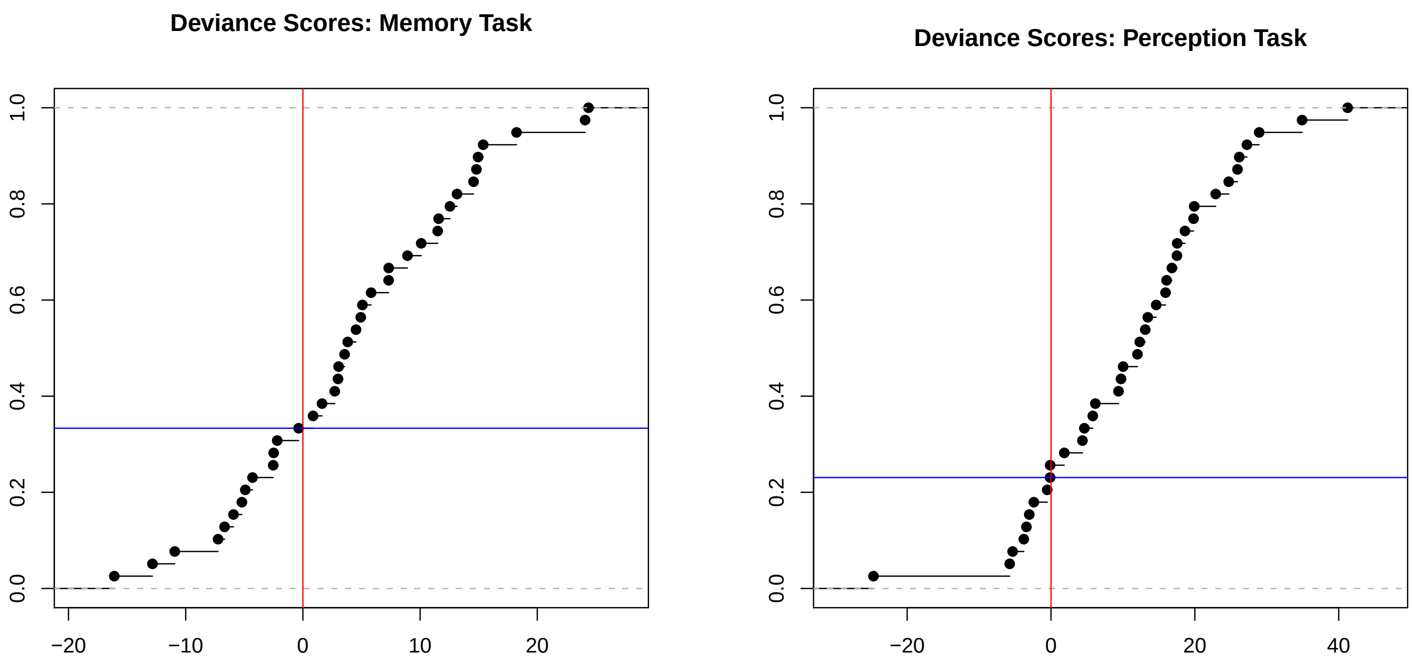

Figure 2.7: Deviance difference scores from model comparisons for the memory and perception task, per subject. Positive values on the x-axis support the discrete-states model.

Chapter 3

Experiment 2: Comparison of One-Interval and Two-Interval Performance in Verbal Perception

3.1 Overview of Experiment

We believe visual perception for verbal information could be a discrete process. Because this view is out of favor, I wanted more evidence to support our claim; Experiment 2 was proposed for this reason. My second experiment is a word identification study utilizing both one-interval and two-interval designs. One-interval designs generally produce a different pattern of results than two-interval designs, so I wanted to look for discrete-state evidence in both. I also wanted to utilize a one-interval perception design because its unique structure allows the possibility of a forced-guessing condition that isn't possible in memory experiments. In a one-interval recognition memory test, the participant sees a test word and decides old or new. If the participant does not recognize the word, they simply respond *new*. It is impossible to tell whether the item was studied zero times, and thus reserved for a forced-guessing trial, or is simply an item from the lure list. In a perception task, however, we can flash a word for 0ms and immediately present a one-interval test item. The essence of forced-guessing is still intact because we know for sure they did not detect the flashed stimulus; therefore, a participant will respond in a manner that reflects their guessing state on these trials. The ability to carry out forced-guessing in a one-interval test is an attractive property of the perception task.

The design of Experiment 2 closely resembles the perception task of Experiment 1. The flash portion remains the same as in Experiment 1, but the test portion was augmented so that half of the trials were one-interval tests and half were two-interval tests. That is, in two-interval tests, the participant saw the target word and a lure and selected the target word, and in one-interval tests, participants saw either the target or a lure, and decided if it was the target. The participants made their decision with confidence ratings on a continuous scale, anchored by *sure* ratings, with *guess* responses in the middle. See Figure 3.1 for an illustration of the word identification paradigm, including both types of tests. I ran the first 25 participants using the same points system from Experiment 1; I refer to this as Experiment 2A. Because I had some questions about how this scoring system might be affecting response patterns, I also ran 25 participants with no points system, using standard confidence ratings. These 25 participants comprise Experiment 2B.

A big problem I encountered in the perception task from Experiment 1 was the wide range of variability in performance on the task. The word flash duration was clearly too fast for some, leaving them around chance performance in all conditions. Others reached near-perfect performance in the middle and easy conditions. To address these ceiling and floor effects, I incorporated an adaptive pre-test into the design of Experiment 2 to find individualized flash durations so that performance would be

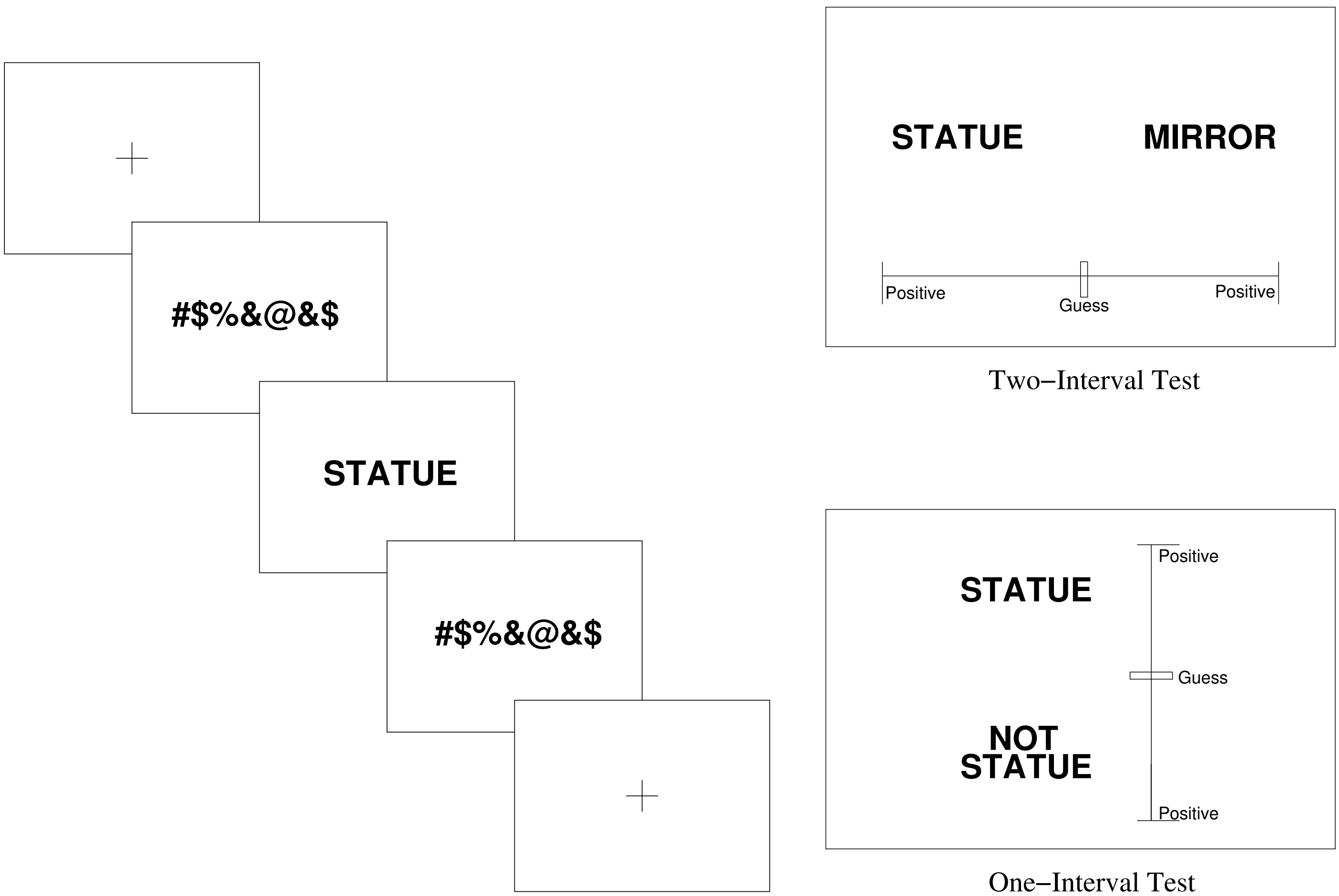

Figure 3.1: Perceptual test paradigm. The study sequence is shown on the left, and at the right are one- and two-interval test screens.

at neither chance nor ceiling. The pre-test served as a calibration phase to determine each individual's ability at the task. The pre-test utilized an adaptive staircasing method, meaning the difficulty of the task adjusted after every few trials based on the participant's performance. We chose a transformed up-down method called the 2-up/1-down, or 2-step rule (Treutwein, 1995): the flash duration decreased by 10ms after two correct responses, making the task more difficult, and the duration increased by 10ms after every incorrect response, making the task easier. The final average duration gathered from the 2-step process in the pre-test was then used as the medium-difficulty condition in the main experiment, with faster and slower durations chosen for the harder and easier conditions, respectively.

The main goal of Experiment 2 remains to explore the structure of visual perception by comparing models of discrete states versus latent strength in varying experimental designs.

3.2 Method

3.2.1 Participants

Twenty-five University of Missouri students served as participants in Experiment 2A in exchange for course credit. Additionally, twenty-five University of Missouri students participated in Experiment 2B in exchange for course credit.

3.2.2 Stimuli

A list of 680 nouns was compiled from the MRC Psycholinguistic Database (Coltheart, 1981). All words were six or seven letters long. The 680 items were randomly broken down into the following lists: 40 pre-test target words, 40 pre-test lures, 150 one-interval target words, 150-one interval lures, 150 two-interval target words, and 150-two interval lures. After the initial randomization, all participants in Experiments 2A and 2B saw the same presentation of stimuli.

3.2.3 Procedure

The experiment was divided into three phases: pre-test, two-interval task, and one-interval task. First, participants completed a 40-trial two-interval adaptive pre-test. In each trial, participants saw a word flashed briefly on a computer screen, masked before and after with a random string of 9 characters. They then saw two words on the screen, the target word and a lure, and responded either *Left Word Flashed* or *Right Word Flashed* via button press with no confidence ratings. The flash duration for the first trial was set to 100ms. The adaptive 2-up/1-down procedure then followed. For every two successive correct responses, the flash duration decreased by 10 ms. For every incorrect response, the flash duration increased by 10ms. Starting with the 15th trial, the flash duration for the current trial was stored and an average duration was computed. The average duration for the last 25 trials was then recorded and passed to the main program.

The two-interval perception task followed the pre-test. The task was split into three conditions (easy, medium, hard) plus a forced guessing condition. There were 150 trials total: 44 of each type plus 18 forced guessing trials. The design was counterbalanced such that the target word was on the left and right at test an equal amount. The flash duration values for each participant were based off pre-test performance. The aforementioned average duration from the pre-test was used for the medium-difficulty condition, the average minus 10ms was used for the hard condition, and the average plus 10ms was the duration for the easy condition trials. Again, in forced guessing trials, only the masks were flashed. All participants saw the same order and pairing of stimuli. At test, participants were presented with the target word and a lure item, one on the left of the screen and one on the right. Participants responded using a continuous confidence scale, weighted by *Sure left word* and *Sure right word*, with guessing responses in the middle. Participants in Experiment 2A had the opportunity to gain or lose points on every trial. Participants in Experiment 2B just responded with pure confidence ratings. All participants made a response by moving a slider along a horizontal confidence scale and clicking, and all received immediate feedback on their accuracy for each trial.

After all two-interval trials, participants completed 150 one-interval trials. The flash durations and breakdown of trial type were the same as the two-interval design. The difference came at test: instead of seeing two words, participants saw only one and had to decide if it was a target or lure. The design was counterbalanced such that participants saw the target or lure at test an equal amount. The confidence scale was placed vertically on the screen, weighted by *Sure target* and *Sure not target* (see Figure 3.1). Participants made a response by sliding a cursor along the scale and making a mouse click. Again the only difference between Experiments 2A and 2B was the inclusion of a scoring system in 2A.

We ended up finding no significant differences in the data or response patterns from participants in Experiments 2A and 2B. Because of this, I will refer to them jointly as Experiment 2 going forward. Combining the two sets of data makes little difference because analysis is done on an individual level.

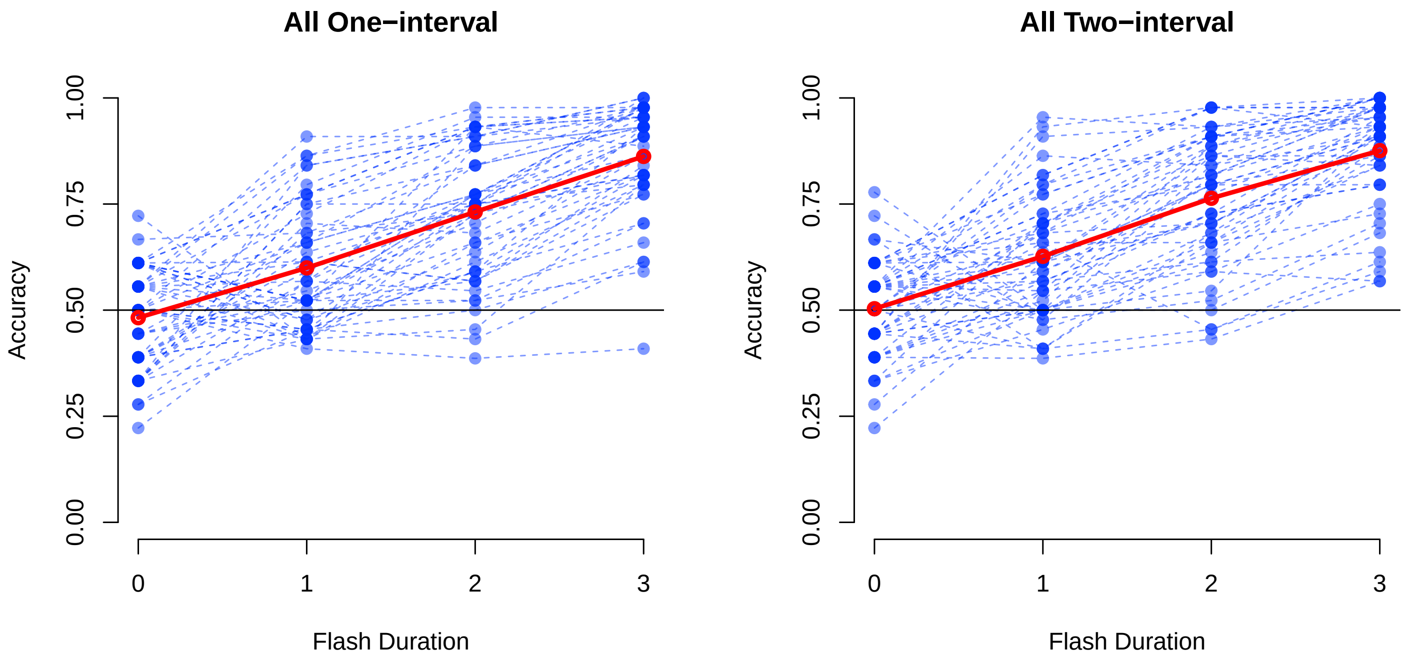

Figure 3.2: Average accuracy scores for all 50 participants across all conditions in both the one-interval and two-interval tasks. The solid red line indicates the grand mean across conditions.

3.3 Results

Performance on the pre-test varied greatly from person to person. Average flash durations ranged from 20ms to 60ms, with the majority of participants (19 of 50) averaging 40ms, 13 participants averaging 30ms, and 11 participants averaging 50ms. The pre-test proved to be very effective in adjusting the difficulty of the main experiment to each participant's ability. This fact is illustrated in the accuracy plots of Figure 3.2. Average accuracy in both tasks increased in an additive fashion from one condition to the next for the majority of participants. Comparing these plots to Figure 2.1 shows just how much of a difference the pre-test made on overall accuracy per individual.

As a side note, I was pleasantly surprised with the results of the pre-test. I never would have guessed that performance on such a basic visual perception task would vary so widely from one person to the next. I know other researchers often run into problems with ceiling and floor effects on tasks such as working-memory span, and I think an adaptive pre-test could be greatly beneficial in these domains.

3.3.1 Confidence Rating Histograms

Analysis of Experiment 2 again led to support for a discrete-state model in both the one-interval and two-interval tasks. As I have shown in previous chapters, one way of analyzing patterns in the data is by observing participant confidence rating histograms. I have shown how each model makes specific predictions as to how a person will distribute their responses, particularly in a mid-difficulty condition. Discrete models predict a clear mixture of guessing and detection states, whereas latent-strength models predict a simple shift in confidence from one condition to the next. Here I provide several examples of plotted confidence ratings from Experiment 2, using the same data representation as before.

Figure 3.3 shows confidence rating histograms for one subject and both tasks. This data reveals an obvious discrete-state pattern in both panels. The two-interval response histogram on the right is especially interesting; the participant's guessing distribution is not centered around the middle, but rather split into a bimodal pattern of riskier guessing. His responses in the difficult condition, however, show a direct reflection of this bimodal guessing pattern and a high confidence detection state.

Figure 3.4 provides two more examples, one from each task, of participants who showed discrete-state response patterns. Though a discrete model was the best match for the majority of participants, there were a few strange response patterns that could not be accounted for. An example of confidence ratings better fit by latent-strength

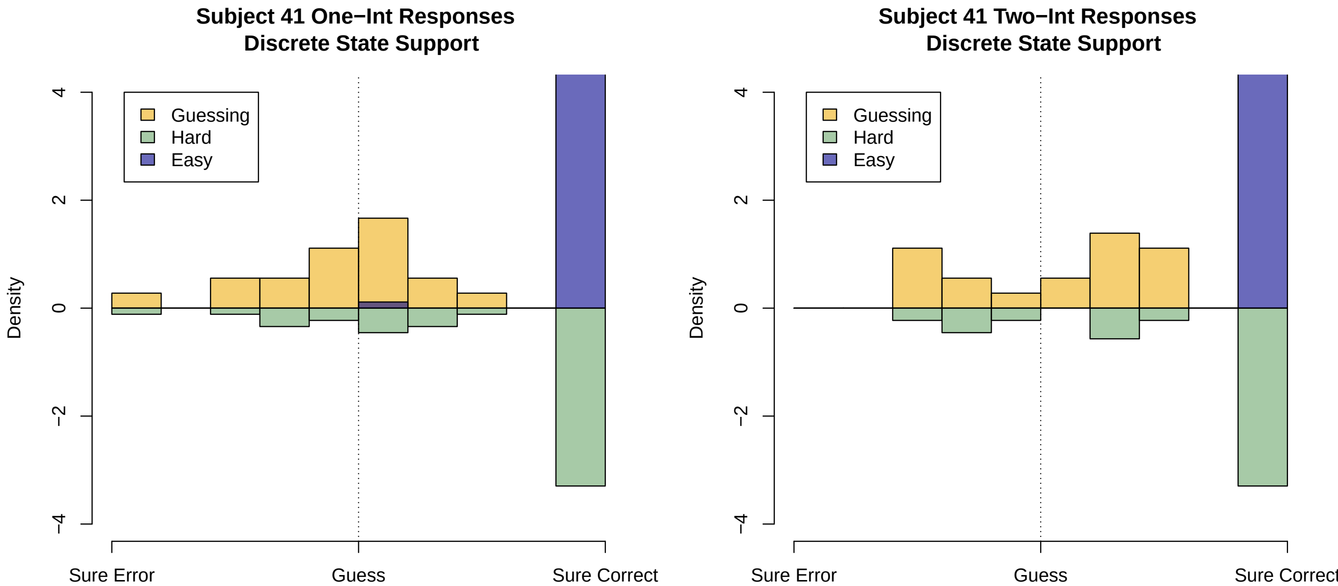

Figure 3.3: Confidence rating histograms for Subject 41, both tasks. This participant shows support for a discrete-state model.

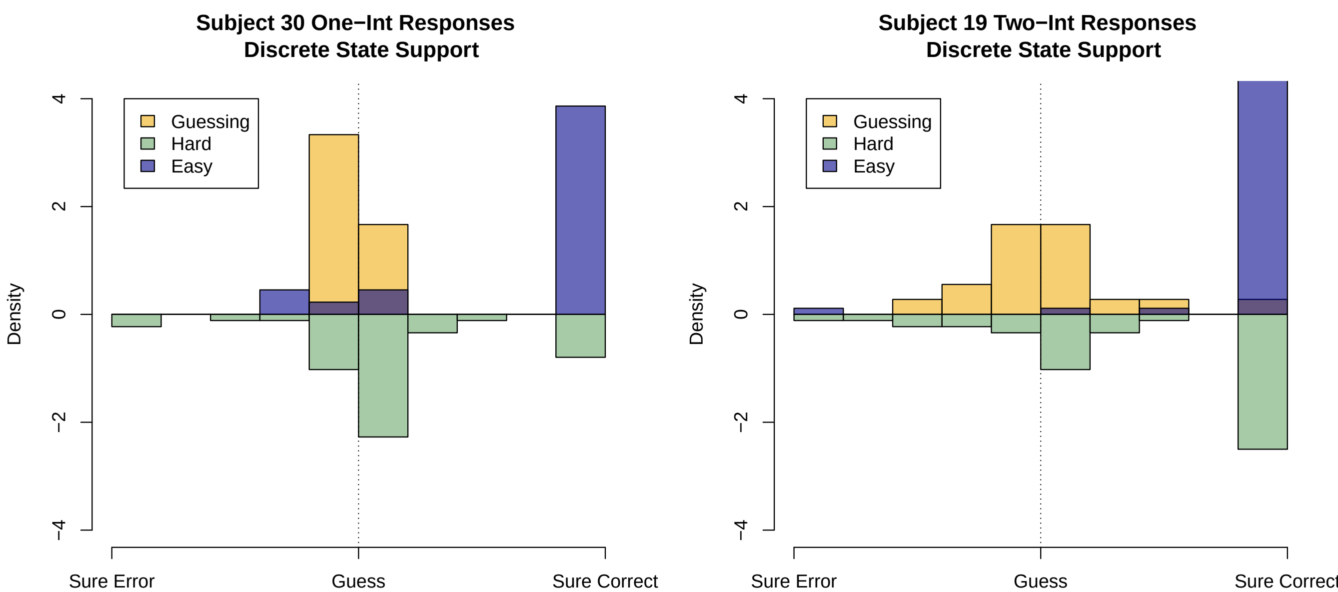

Figure 3.4: Confidence rating histograms for two other participants who show support for a discrete-state model.

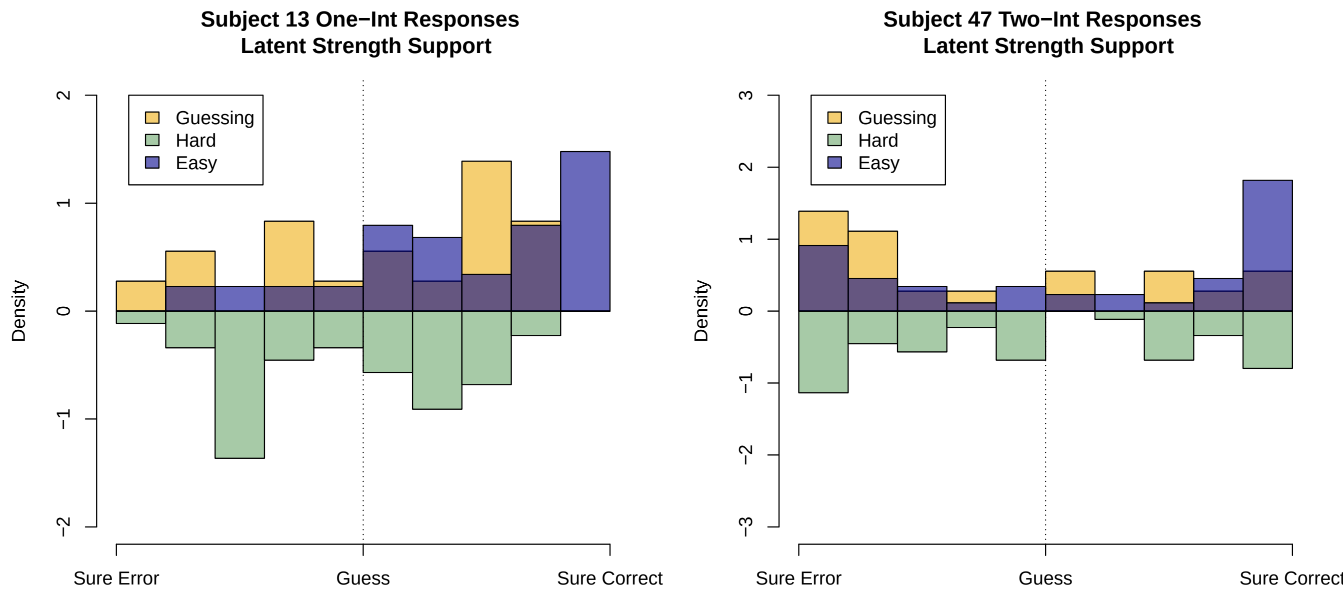

Figure 3.5: Confidence rating histograms for two different participants who showed greater support for a latent-strength model.

models is seen in Figure 3.5. The fact that a latent-strength model could fit these seemingly random data patterns illustrates just how overly-flexible these models can be.

3.3.2 Model Fittings and Comparisons

I again sought to bolster the strength of my analysis by carrying out proper model fittings and parameter estimation. Discrete-state and latent-strength models were constructed separately for the two-interval and one-interval data. Since the two-interval task was a near replication of the perception task from Experiment 1, the same ten-parameter models were used. The discrete-state model is illustrated in Figure 2.6.

Parameters for each model were estimated utilizing maximum likelihood estima-

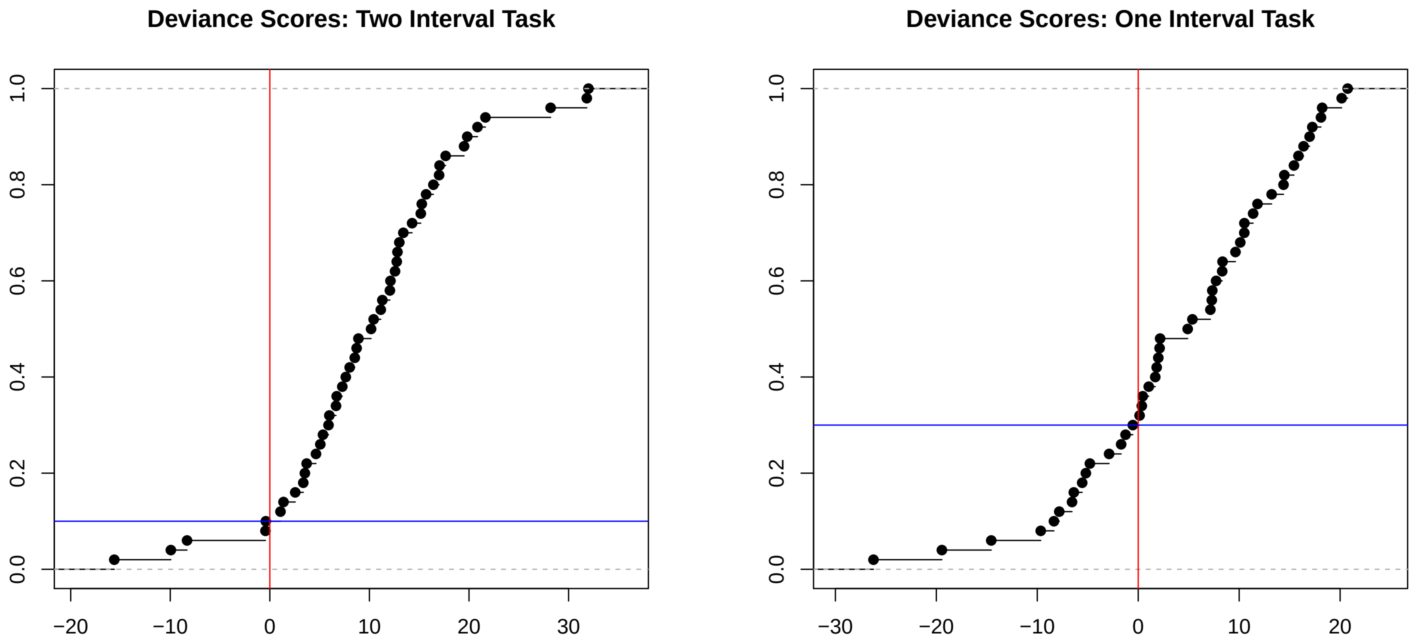

Figure 3.6: Deviance difference scores from model comparisons for the two-interval and one-interval task, per subject. Positive values on the x-axis support the discrete-states model.

tion. Parameters were optimized and a likelihood value was computed for each participant and both models. The likelihoods for each model were then compared by subtracting the discrete-state score from the latent-strength score to create a deviance difference value wherein negative values show a better fit for the latent-strength model. The left panel of Figure 3.6 shows difference scores for all participants in the two-interval task. Clearly, the discrete-state model is an obvious winner over the latent-strength model, with 90% of participants favoring discrete states.

In a two-interval model, we assume performance is equal when the target is presented on the left or on the right, and we can thus benefit from the symmetry in the model. Specifically, we can assume the probability of detecting a target given it is on the left is equal to the probability of detecting a target given it is on the right

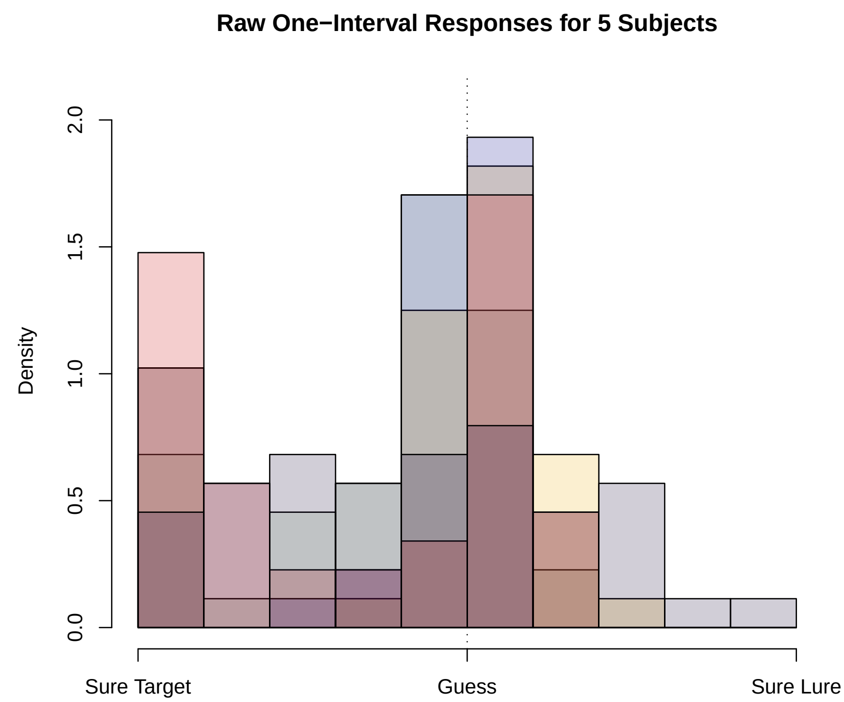

Figure 3.7: A compilation of five participants' raw confidence ratings from the easiest condition of the one-interval task. These responses are from both target and lure trials, and are not flipped or mirrored as in other histogram plots.

($d_{left} = d_{right}$) for all conditions. The two test types in the one-interval task, however, are inherently different. Is the probability of detecting a target given the target is presented equal to the probability of detecting a lure given a lure is presented? I believed they would not be equal, and inspecting the data confirmed this notion. Histograms of confidence ratings showed participants were much less likely to respond *Sure lure* on lure trials than *Sure target* on target trials.

To show what this looks like, I have included Figure 3.7 in which I plotted raw confidence ratings from five different subjects in the easiest condition of the one-interval task. Responses on the left are from target trials, and rightward responses are from lure trials (they are not flipped as in other histogram plots). As you can see,

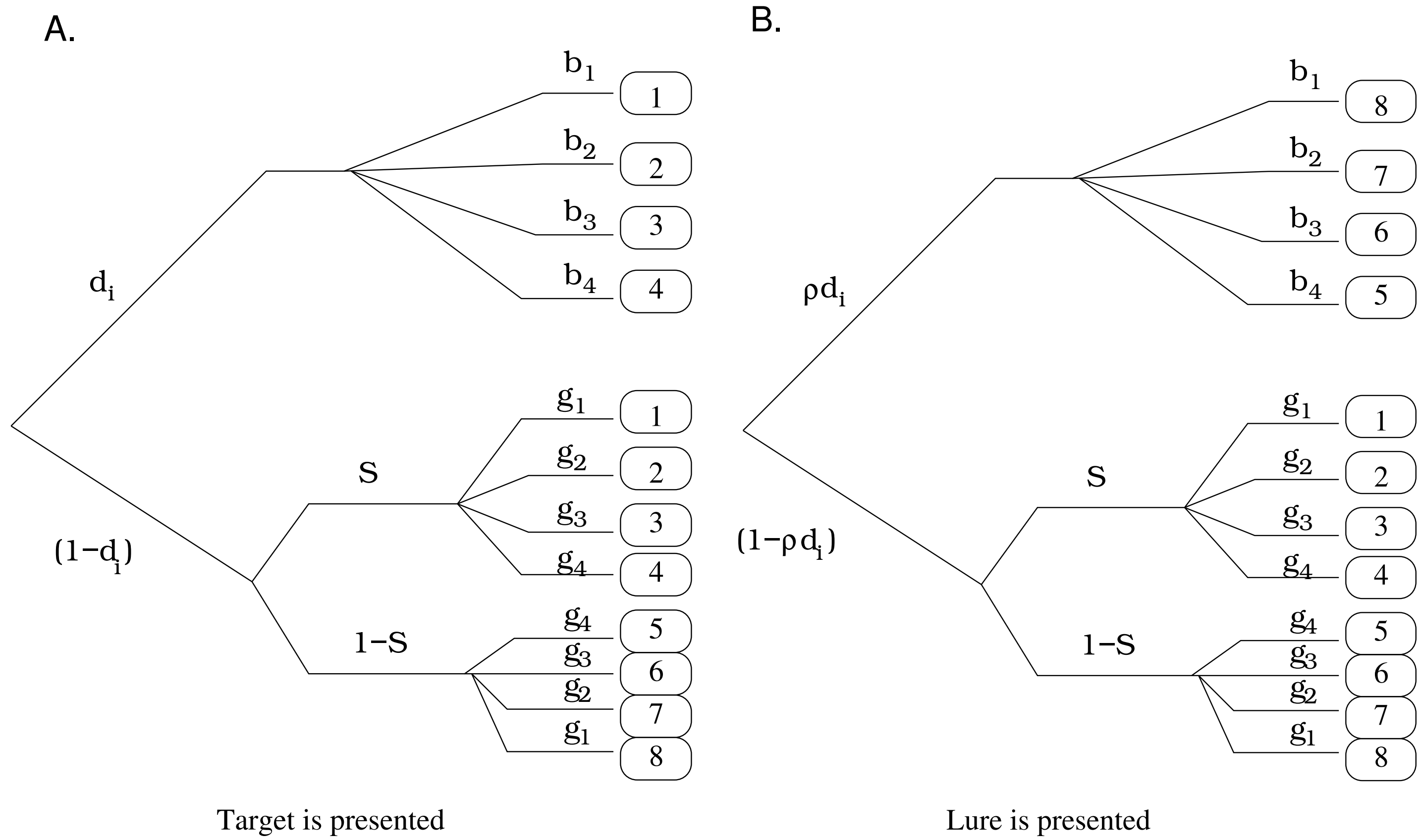

Figure 3.8: Eleven-parameter discrete-state model for the one-interval task. **A.** Model given the target was presented (same as previous ten-parameter model). **B.** Model given lure was presented.

there are very few responses near the *Sure lure* response from these subjects, even in the easiest condition. I therefore figured that participants were simply less likely to enter a detect state on lure trials ($d_{lure} < d_{target}$). I was then able to construct a more accurate model with the addition of only one new parameter. The new parameter, ρ, is a fractional value, equal across all conditions, such that $d_{lure} = \rho * d_{target}$. This maintains the concept of conditional independence; responses are determined solely by which state one is in. The probability of detection is still the only parameter changing value from one condition to the next. The new eleven-parameter model is illustrated in Figure 3.8.

The additional parameter has a straightforward extension to the latent-strength model as well. The parameter ρ acts as a fraction on d', such that the distribution for lure items is a fractional distance away from zero as the distribution for target items (i.e. $d'_{iT} = -\rho\ d'_{iL}$).

Parameters were estimated and optimized in the same manner as before for all participants, both models. Deviance scores were computed and compared for each participant. This comparison is shown in the right panel of Figure 3.6. While the evidence for discrete states is weaker in the one-interval task than the two-interval task, the discrete-state model still proved a better fit for 70% of participants.

The eleven-parameter models described here are the only ones I fit to the data, but I am interested to tweak other parameters to see if we are getting the full picture on one-interval designs. The overwhelming evidence for discrete states in the two-interval design was especially impressive. I once again showed with Experiment 2 that discrete-state models can account for most participants' performance on verbal perception tasks, especially in two-interval designs.

Chapter 4

Summary and Concluding Remarks

4.1 Summary of Experimental Findings

Results from Experiments 1 and 2 consistently favored a discrete-state model over a comparable latent-strength model. Experiment 1 tested discrete-state assumptions in the recognition memory domain, inspired by Province and Rouder's (2012) evidence for discrete states in recognition. Experiment 1 also included a second task testing visual perception for word identification. I thought word identification, an even simpler task than recognition memory, might be more likely to show evidence of continuous processing, but again we found strong support for the discrete-state model. Most modern recognition memory and visual perception researchers lean largely in favor of latent strength or dual-process models, but our evidence for discrete states suggests the debate is far from over.

Because I found the word identification results from Experiment 1 somewhat surprising and sought to strengthen the discrete-state evidence, I designed Experiment 2 and extended the verbal perception paradigm to both a two-interval and one-interval design. I also wanted to do a better job of tailoring the perceptual task to all participants' skill levels, and was able to do so with the addition of the adaptive pre-test in Experiment 2. Analyzing participant confidence ratings and comparing likelihood values of discrete and continuous models again led to strong evidence for discrete-state processing in verbal perception. Evidence was especially strong in the two-interval task.

4.2 Concluding Remarks

I hope I was able to demonstrate the viability of discrete-state models for cognition in this thesis. Basic discrete-state models with certainty assumptions do a poor job of predicting performance; however, with the addition of an extra detection parameter, discrete models can predict any pattern of behavior within a condition. Discrete-state models impose strong and testable constraints on data across multiple stimuli conditions. These constraints arise from a mixture of guessing and detect states and are largely met in the data, making discrete-state models viable for recognition memory and verbal perception. The evidence for discrete states I and my colleagues have recently found stands in opposition to the conventional continuous-processing viewpoint in the cognitive literature. We are interested in extending the paradigm to other domains such as lower-level perception (e.g., angle orientation), working memory, and higher-order functions (e.g., tumor detection in radiology).

Finally, I want to again reinforce the different levels of interpretation that evidence for discrete states brings forth. A very strict discrete-state interpreter would claim that processing is discrete even on the sensory (i.e. retinal) level. Some information is perceived, some information is never processed. I find this to be an overly simplistic idea and don't blame those who are put off by discrete-state models with this kind of interpretation. A more subtle discrete interpretation leaves room for some continuous processing. Information is processed continuously at the sensory level, but when higher-order processing is necessary, the information reaches a kind of threshold or bottleneck, wherein the finely graded representation is lost and discretized. This bottleneck may not necessarily be present in all domains. We can only make inference about these processes through experimentation, and we do so by observing the structure of data. Our evidence for discrete-state processing in recognition memory and word identification show that these tasks are beyond the continuous/discrete threshold, but we are welcome to and intrigued by the idea of finding a task with true continuous processing.

Bibliography

[1] A. Allin. Recognition. *Psychological Review*, 3(5):542–545, 1896.

[2] H. R. Blackwell. Neural theories of simple visual discriminations. *Journal of the Optical Society of America*, 53:129–160, 1963.

[3] D. E. Broadbent. Two-state threshold model and rating-scale experiments. *Journal of the Acoustical Society of America*, 40:244–245, 1966.

[4] M. Coltheart. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A:497–505, 1981.

[5] M. Glanzer, K. Kim, A. Hilford, and J. K. Adams. Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25:500–513, 1999.

[6] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.

[7] W. Kintsch. Memory and decision aspects of recognition learning. *Psychological Review*, 74:496–504, 1967.

[8] David H. Krantz. Threshold theories of signal detection. *Psychological Review*, 76(3):308 – 324, 1969.

[9] R. S. Lockhart and B. B. Murdock. Memory and the theory of signal detection. *Psychological Bulletin*, 74:100–109, 1970.

[10] T. E. Parks. Signal-detectability theory of recognition-memory performance. *Psychological Review*, 73:44–58, 1966.

[11] J. M. Province and J. N. Rouder. Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences*, 2012.

[12] J. Qin, C. L. Raye, M. K. Johnson, and K. J. Mitchell. Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27:1110–1115, 2001.

[13] D.L. Schacter and E. Tulving. What are the memory systems of 1994? In D.L. Schacter and E. Tulving, editors, *Memory Systems 1994*, pages 1–38. MIT Press, Cambridge, MA, 1994.

[14] Jr. Tanner, W. P and T. G. Birdsall. Definition of d' and η as psychophysical measures. *Journal of the Acoustical Society of America*, 30:922–928, 1958.

[15] B. Treutwein. Adaptive psychophysical procedures. *Vision Research*, 35:2503–2522, 1995.

[16] A. P. Yonelinas. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:1341–1354, 1994.