SIGNAL DETECTION THEORY: A PROPOSAL FOR A NONPARAMETRIC MODEL

A Thesis

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Arts in the

Graduate School of The Ohio State University

By

Brandon M. Turner, B.S.

* * * * *

The Ohio State University

2009

Master's Examination Committee:

Trisha Van Zandt, Adviser

Mike Edwards

Simon Dennis

Approved by

Adviser

Graduate Program in Psychology

ABSTRACT

Signal detection theory forms the basis of many current models of memory, choice, and categorization. However, little research has examined precisely how the decision-making process unfolds over time. In this paper, a new nonparametric, dynamic model is proposed with the intentions of ameliorating some long-standing issues in the signal detection framework and describing the changes in signal detection performance over time. The model uses a recursive kernel density estimation procedure that accumulates and stores experience across trials. I present the results of several simulations and show that the proposed model bypasses the rigid assumptions of prior internal representations of the sampling distributions and as a consequence, it allows the criterion location to shift to accommodate new information as it is presented.

I would like to dedicate this thesis to my wife, Katherine.

ACKNOWLEDGMENTS

I would like to thank my adviser, Trish Van Zandt, for her time and patience in this endeavor. Her continuous support enabled me to complete this project successfully. Her committment to this thesis merits more consideration than I am able to provide here.

VITA

November 18, 1985 ..... Born - Bolivar, Missouri

2008 ..... B.S. Mathematics,
Missouri State University, Missouri

2008 ..... B.S. Psychology,
Missouri State University, Missouri

2008-present ..... Graduate Teaching Associate,
The Ohio State University

FIELDS OF STUDY

Major Field: Quantitative Psychology

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

Signal detection theory (SDT) had its beginnings in the simple problem of deciding whether or not a signal was present in a noisy background. These problems originally dealt with radio signals, but since World War II, SDT has been extended to many applications. Over 50 years have passed since the first application of signal detection theory (SDT) to human sensory discrimination. Since then, SDT has been crucial to many important theories in cognitive psychology. Developed by electrical engineers and based on statistical decision theory, SDT is simply a framework for explaining choice performance. SDT assumes that the decision maker makes difficult perceptual judgments under conditions of uncertainty (Green and Swets, 1966).

SDT explains performance in two-choice decision-making tasks. In such tasks, an observer is presented with a series of trials in which either noise alone or a signal embedded in noise is presented. The observer is to say "yes" when a signal is presented and to say "no" when only noise is presented. This type of experiment leads to four possible outcomes (see Table 1.1). Of these four outcomes, we focus on the hit rate $H = P(\text{“}yes''|S)$ and false alarm rate $FA = P(\text{“}yes''|N)$, where S represents a signal and N represents noise.

1.0.1 Discriminability and Bias

The SDT framework assumes that the presentation of a stimulus gives rise to a perception of sensory effect. Noise in the perceptual system (or in the stimulus itself) results in varying effects over different stimulus presentations. Signals result in larger or more positive effects than noise alone. Variability in sensory effects is represented by two random variables or distributions, often equal-variance normal distributions, though this assumption is not strictly necessary. To make a decision, we assume observers place a criterion along the axis of effect. The "yes" or "no" response is determined by whether the perceived effect is above or below this criterion.

The observer may place the criterion in one of two ways. He or she may consider only the magnitude c of the perceived effect, or, given some perceived effect, he or she may consider the relative likelihood of a stimulus coming from either the signal or noise distribution, written as

$$\beta = \frac{f_S(c)}{f_N(c)}, \tag{1.1}$$

where f is the probability density function of the distribution of effect.

Most treatments of SDT assume that the criterion is placed according to the likelihood. The criterion will also depend on the perceived consequences of different outcomes. For example, if the consequences are costly for saying "yes" when the signal is absent (a false alarm), then an observer may adjust his or her criterion upward to reduce the frequency of "yes" responses. By contrast, if the consequences are more costly for failing to detect a signal when it is present (a miss), then an observer may adjust his or her criterion downward to increase the frequency of "yes" responses.

	Response			
	"No"		"Yes"	
Stimulus	Type	Payoff	Type	Payoff
Signal	Miss	C_M	Hit	V_H
Noise	Correct Rejection	V_{CR}	False Alarm	C_{FA}

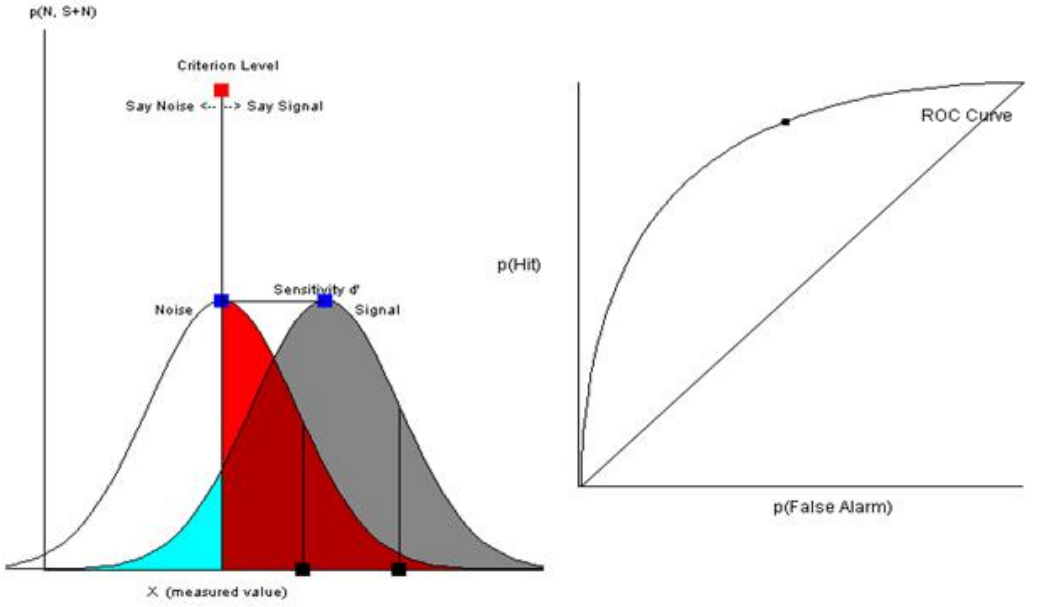
Table 1.1: The four possible outcomes of the two-choice decision task are shown with their associated cost or value.

Table 1.1 shows a payoff matrix for a signal detection task. Each type of response has associated with it a value (V_H or V_{CR}) if it is correct and a cost (C_M or C_{FA}) if it is wrong. An ideal observer will respond "yes" when the expected value of a "yes" response exceeds the expected value of a "no" response. This leads to an ideal criterion (in terms of the likelihood ratio)

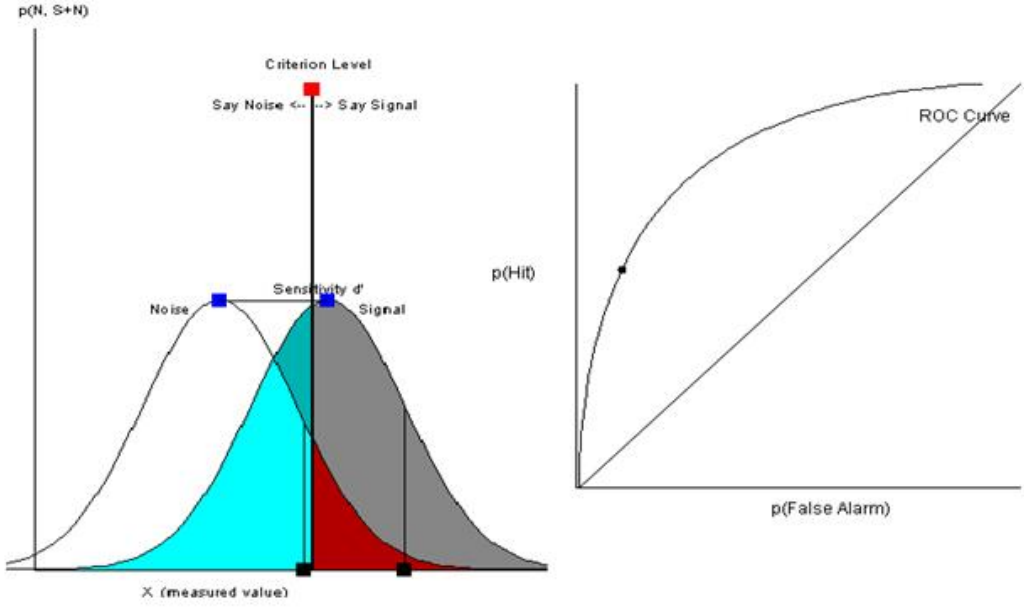
$$\beta_o = \frac{1-p}{p}\left(\frac{V_{CR}+C_{FA}}{V_H+C_M}\right), \tag{1.2}$$

where the costs and values are all positive, and $p = P(signal)$. Any tendency to be conservative (more likely to say "no") or liberal (more likely to say "yes") reflects observer bias. Bias is measured by the difference between an observer's criterion β and the optimal criterion β_o. A criterion β greater than β_o implies that the subject requires stronger evidence before saying that a signal is present. Figure 1.1 shows a very liberal subject ($\beta < 1$, or equivalently a low c value, Panel A) and a much more conservative subject ($\beta > 1$, Panel B).

We assume that signals give rise to greater sensory effects than noise. Sensitivity is a measure of how discriminable signals are from noise. The most commonly used measure of sensitivity is d-prime (d'), the standardized difference between the means of the N and S distributions. To estimate d' (under the Gaussian assumption), we



Panel A



Panel B

Figure 1.1: Two criterion placements. Panel A shows an observer with a liberal bias ($\beta < 1$) and Panel B shows an observer with a conservative bias ($\beta > 1$).

can calculate $d' = z(FA) - z(H)$ where FA and H are the false alarm and hit rates, respectively, and $z(FA)$ and $z(H)$ are the z-scores that correspond to the right-tail (normal) probabilities FA and H. High d' values indicate good discriminability, while d' values near zero indicate very low discriminability.

Figure 1.2 shows low ($d' = .40$, Panel A) and high ($d' = 2.0$, Panel B) discriminability. This figure (and Figure 1.1) also shows a receiver operating characteristic (ROC) curve (e.g. Egan, 1975, Swets, 1996, Wickens, 2002), which is obtained by plotting H as a function of FA. This curve shows how hit rate trades off against false alarm rate as an observer moves the criterion from higher to lower levels. The difference between the curve and the diagonal reflects the level of discriminability and the position of any point along the curve represents bias β. Thus, while we can't directly observe the distributions of sensory effect shown in Figures 1.1 and 1.2, we can observe their "footprints" in the form of the ROC.

1.0.2 Alternative Measures

Estimates of d' and β are the most popular statistics for characterizing detection behavior (Macmillan, 2002). Two alternative nonparametric measures, A' and B'' make no assumptions about the form of the stimulus representations, and may be used in place of d' and β when the normality or equal variance assumptions are violated.

The measure A' was developed by Pollack and Norman (1964), and it represents the average area between the ROC curves that maximize and minimize the hit rate (Pastore et al., 2003). The measure B'' was developed by Grier (1971) and is another ROC-based descriptive statistic reflecting the bias for a single data point. However,

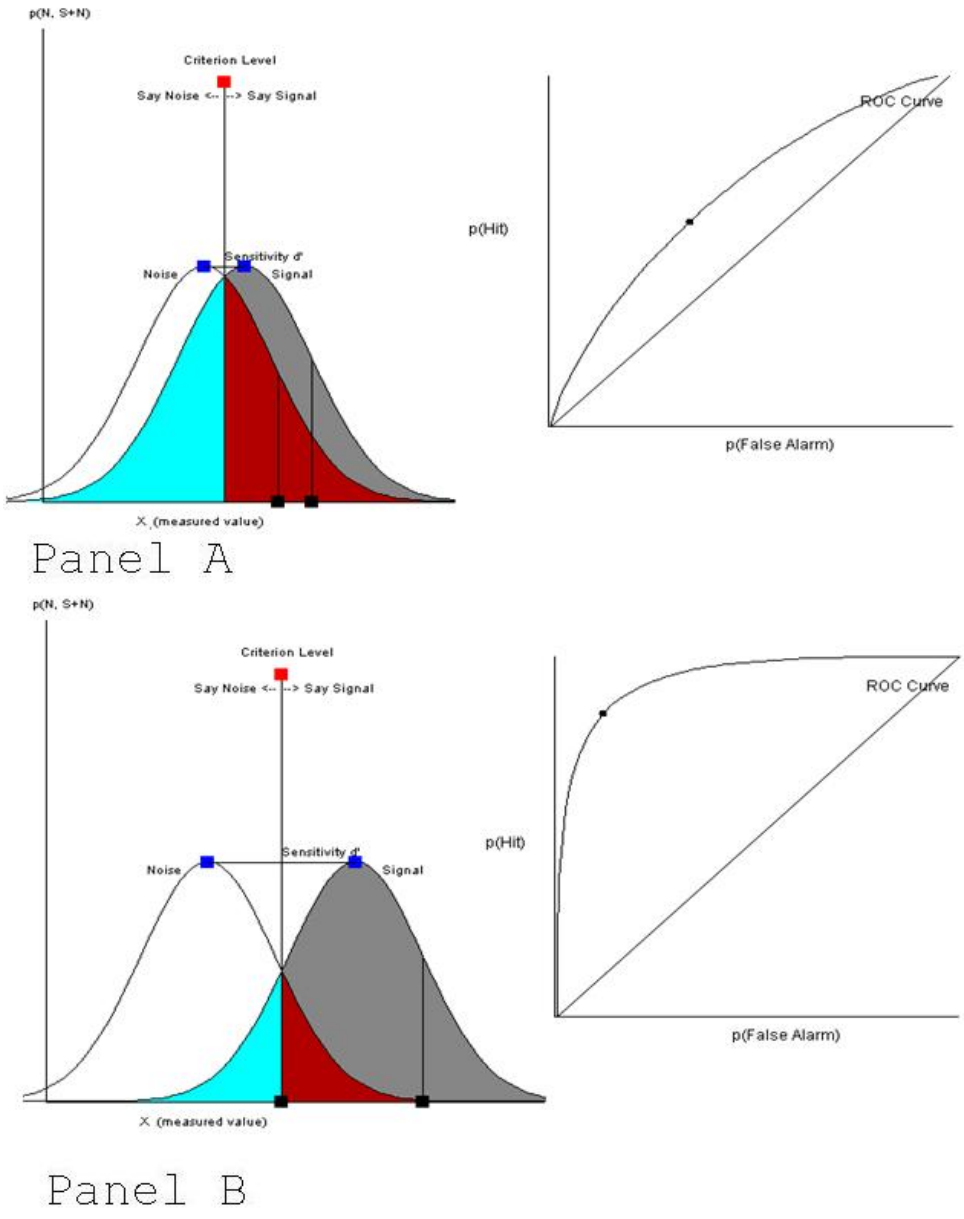


Figure 1.2: Two values of d'. Panel A shows low level of discriminability ($d' = .40$) and Panel B shows a high level of discriminability ($d' = 2.0$).

there are problems with these measures. Pastore et al. argue that A' and B'' are not independent descriptive statistics on the basis of calculations given by Snodgrass and Corwin (1988). This lack of independence destroys the ability to isolate stimulus effects from response effects. Furthermore, Pastore et al. claim that A' is not an accurate measure of area under the ROC curve when there is enough bias in initial criterion placement.

Balakrishnan (1998a,b, 1999) strenuously objects to the indiscriminant use of d' and β in the analysis of data from signal detection tasks. Balakrishnan has developed novel nonparametric measures that he claims are better justified than d' and β. He claims these measures are "rooted within the general framework of statistical decision theory, and hence are also consistent with earlier empirical results, including ROC curve data" (Balakrishnan, 1998b, p.69). However, Treisman (2002) has argued that Balakrishnan has overstated his case.

1.0.3 Statement of the Present Problem

The traditional SDT framework assumes the parameters (d' and β) are fixed. One implication of parameters that are fixed over trials is that decisions are independent of previous decisions and their outcomes. However, we know that this is not true. For example, Howarth and Bulmer (1956) asked subjects to perform a detection task in which an auditory signal was presented in white noise. They found strong positive sequential dependencies, such that subjects tended to repeat responses they had made on earlier trials. Treisman and Williams (1984) found similar effects in their own data.

These findings and others like them led to the development of several dynamic models that can explain sequential dependencies in signal detection tasks (e.g. Brown

and Steyvers, 2005, Treisman and Williams, 1984, Vickers and Lee, 1998, 2000). Each of these models have different explanations of behavioral changes over time, which I will describe below.

The purpose of this thesis is to propose a dynamic decision-making model which solves two of the long-standing problems in SDT models. First, it does not require *a priori* stimulus representations for signals and noise, but instead describes how these representations evolve over time. Second, and as a consequence, it allows the criterion location to shift over time to accommodate new information as it is presented. This in turn makes strong predictions concerning signal detection performance with changes in experience and with changes in the stimulus distributions.

To set the stage for this model, I will first introduce some of the literature in signal detection theory and kernel density estimation. Second, I will present a new model and justify its assumptions. Third, I will present the results from two experiments that support the justification of the model's assumptions. Fourth, I will present the results of several simulations to convince the reader that the model behaves appropriately. Finally, I will compare and contrast the model's predictions to the predictions of signal detection theory.

1.1 Literature Review

1.1.1 Signal Detection Theory

The two problems with traditional SDT, just described, create models of decision making in which successive decisions are independent. Several alternative models allow for movement of the criterion and therefore sequential dependencies (e.g. Kac, 1962, 1969, Rabbitt, 1981, Treisman and Williams, 1984, Vickers and Lee, 1998, 2000).

These models, called additive learning models, explain sequential dependencies in a number of different ways. Kac (1962, 1969) proposed an error-correcting model in which the observer's criterion could be altered from trial to trial - but only after an error. For example, if an observer's response was "yes" when a stimulus came from the N distribution, the criterion shifted upward to prevent this error in subsequent trials. Dorfman et al. (1975) expanded this model to allow for random fluctuations in the criterion from trial to trial (cf. Healy and Kubovy, 1977).

Kac's model may be classified as one that performs error-rate monitoring. Other adjustment mechanisms have been proposed on the basis of stimulus monitoring (Vickers and Lee, 1998, 2000) and change detection (Brown and Steyvers, 2005). Treisman and Williams (1984) proposed a mechanism for criterion placement based on response monitoring. I will focus on Treisman and Williams' and Brown and Steyvers' models, which are most closely related to the new model that I will present later.

Treisman and Williams (1984) generalized the additive learning models of Kac (1962), Dorfman and Biderman (1971), and Thomas (1973). They investigated three models: a linear additive learning (LAL) model, an exponential additive learning (EAL) model, and an independent trace (IT) model. The LAL and EAL models shift the criterion relative to a "reference" criterion depending on the response on each trial. The criterion shifts in such a way that the frequency of the response just made will increase or decrease, resulting in positive or negative dependencies in the response sequence.

The IT model is slightly different, but follows the same structure. Each response results in a memory "trace" for the criterion shift (positive or negative, depending

on the response). This trace decays toward zero with time. The criterion location at any point in time is a function of the summed memory traces. Like the LAL and EAL models, this model produces positive or negative dependencies in the response sequence.

More recently, Brown and Steyvers (2005) examined performance in two tasks typically treated within the SDT framework: a lexical decision task and a numerosity judgment task. Subjects performed one of these tasks with stimuli that were either easy or hard to discriminate. In the lexical decision task, difficulty was determined by the number of letters nonwords differed from words. In the numerosity judgment task, subjects were asked to determine, in a display of 10 symbols, which of two symbols was more frequent. Difficulty was determined by how many of each symbol was present in the display. Within blocks of trials or between blocks of trials, the task environment shifted from easy to hard (or vice versa).

Brown and Steyvers (2005) proposed a model of dynamic criterion adjustments in which two *a priori* stimulus representations (one set for the easy environment and one set for the hard environment) determine decision-making performance. Each representation also has a fixed criterion. The model assumes that, after switching from one environment to the other, the subject continues to use the criterion he or she was using before, until he or she recognizes that the task has changed. The critical parameter is the lag, the time it takes the subject to switch from one criterion to the other.

Brown and Steyvers (2005) study showed that subjects do adjust their criteria, and that it takes some time (12 trials on average) to do so. Their model is very similar to the models investigated by Treisman and Williams (1984), except that it

proposes the existence of two separate stimulus representations and fixed criteria for each. For all these models, even though the criterion may shift from trial to trial with experience, the problems we noted above for the classic SDT framework still exist. The dynamic criterion is determined by a fixed reference level and the stimulus representation (or representations) is also fixed and assumed to exist *a priori*.

1.1.2 An Alternative Approach

In a more recent paper, Brown and Steyvers (2009) use a particle filter in place of the static stimulus representation. This is one of the first examples of the use of a particle filter as a component of a psychological process. It is from this paper that I take much of my inspiration for the modeling that I will outline in this proposal. The particle filter, although it will not be appropriate for the standard signal detection paradigm, gives a hint as to how one might model a dynamic representation built up gradually from experience, which in turn makes strong predictions about criterion shifts and sequential effects in signal detection tasks.

Brown and Steyvers (2009) asked their subjects to perform a four-choice task. They were told that stimuli would be generated from one of four possible distributions and that they had to decide from which distribution a stimulus came. Unlike a signal detection task, the stimuli were not generated from distributions at random. Rather, the stimuli were generated from a single distribution, and after each trial, there was a small probability (8%, 16% or 32%) that the generating distribution would change. Brown and Steyvers found that as the rate of change increased, subjects' accuracies decreased and their proportion of response changes increased. Most interestingly,

subjects predicted changes more often as the rate of change was increased; they tended to vastly overestimate the likelihood of a change.

The particle filter model assumes the existence of a (possibly large) number of stimulus "traces" or particles. Each particle is assigned to or identifies one of the possible responses. Over time, particles assigned to unlikely responses become reassigned to the responses more appropriate for the stimuli that are presented. Thus, a coarse representation of the stimuli can be constructed by the frequency distribution over the different particles' assignments. This representation changes over time, depending on the stimuli that are presented and the responses that are made to those stimuli.

Brown and Steyvers (2009) presented simulations of the particle filter model and showed that the predicted data surface (within the parameter space of the model) conformed well to the subjects' data for both the identification task and for a second prediction task.

Brown and Steyvers' (2009) particle filter model is designed for a dynamic decision-making environment in which the stimulus-generating mechanism changes over trials. While Brown and Steyver's model inspired my current research, their approach cannot work for signal detection tasks in which the stimulus-generating mechanism is constant over trials. In the Brown and Steyver's paradigm, there is only a single stimulus representation that changes with changes in the stimulus-generation mechanism and the observer's responses to those stimuli. In standard signal detection tasks, two representations are needed, one for signals and one for noise. My approach makes use of a recursive kernel density estimation procedure which builds the signal and noise representations over time. Because this method produces representations

that change over time, it predicts that there will be strong sequential dependencies across trials.

Before I present the details of my modeling approach, in the next section I will discuss the fundamentals of kernel-based density estimates. There are, of course, many ways to estimate a density. These include splines, polynomial regression, Fourier transform approaches, among others. The kernel approach is (arguably) the simplest of these alternatives and so the easiest for constructing psychologically plausible models.

1.1.3 Kernel Density Estimation

Kernel density estimation is a process for estimating the density of a random variable from an independent and identically-distributed (*iid*) sample of data. The idea was first proposed by Fix and Hodges as a way of avoiding distributional assumptions in discriminant analysis (Silverman, 1986). Thus, it is a nonparametric approach to the estimation problem, which is one reason why it is attractive for modeling the evolution of stimulus representations. For an *iid* sample $\{X_1, X_2, \ldots, X_n\}$ a kernel estimator has the form

$$f_n(x; h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \tag{1.3}$$

The function $K(\cdot)$ is the kernel and h is a smoothing parameter called the bandwidth. The kernel is usually chosen to be unimodal and symmetric about zero. The kernel function can take many forms; however, in this paper, only the Gaussian and the rectangular kernels are used. The density estimate at a point x is obtained by centering the kernel over x and summing the individual contributions of each data point

x_i in the sample that is within the bandwidth. In doing so, the kernel (K) weights the observations so that the shape of the density estimate is determined jointly by the data and the bandwidth.

The bandwidth of the kernel determines how far away an observation can be from the point x and still contribute to the estimate. Large bandwidths will allow more distant observations to influence the estimate at x; smaller bandwidths will attenuate the effects of distant observations. As the bandwidth increases, the estimate will get smoother. If the bandwidth gets too large, the density estimate becomes impoverished and potentially important details in the data will be lost.

The choice of the bandwidth is an important decision and much work has been devoted to the optimization of the estimator based on this choice (see Wand and Jones, 1995). Other bandwidth selection methods are proposed by Taylor (1989) and are used in de Silva and Mukhopadhyay's (2004) application of kernel density estimation methods to wool-fiber data. Silverman (1986) gives a survey of methods for kernel estimation, then examines many areas in which density estimation can be applied. Additional real-world examples can be found in *Applied Sequential Methodologies: Real World Examples with Data Analysis* (de Silva and Mukhopadhyay, 2004).

1.2 A Nonparametric Model of Signal Detection

My new model assumes, like SDT, that decisions about signals and noise are determined by the likelihood of each for a particular perceived stimulus strength. The likelihoods are determined by a representation of the stimulus-generating mechanism that is constructed over time. The representation includes likelihoods for signals and for noise at a limited number of points along the stimulus continuum. In some

sense, then, the stimulus representations are "sparse"; the observer does not have an accurate representation of all the stimulus strengths he or she might perceive, but only the likelihoods at a few points.

In contrast to other SDT models (e.g. Kac, 1962, 1969, Rabbitt, 1981, Treisman and Williams, 1984) that focus on the location of a response criterion, I assume that the presentation of a stimulus causes the observer to recall the signal and noise likelihoods at those locations in the representation closest to the perceived stimulus strength. The observer then selects the response according to whichever likelihood is highest. In some ways this is similar to setting a criterion at the point where the likelihood ratio is equal to one. However, in my framework, the likelihood ratio is not monotonic with stimulus strength, depending as it does on a rough and sparse estimate of the signal and noise distributions. Therefore, there is no fixed criterion in the usual sense, and the criterion is not unique as assumed by traditional SDT (e.g. Egan, 1975, Green and Swets, 1966).

The representations are constructed trial by trial as stimuli are presented. Before the first trial, the observer has only a poor idea of what signals and noise look like. However, based on the experimenter's description of the task and the observer's previous experience with similar stimuli, the observer can construct a "prior" for the representation. For my purposes I assume a uniform prior in which all values have an equal chance of being from the S or the N distributions. At this point, the subject has no representation of the stimuli beyond this prior that permits a "criterion placement" in the usual sense.

As mentioned, I assume that the presentation of a stimulus causes the observer to recall the signal and noise likelihoods at those locations in the representation closest

to the perceived stimulus strength. These locations will be referred to as points. The assumptions are in Table 1.2. The points used by the observer play a critical role in the ability of the model to account for individual differences.

On the first trial, the observer is presented a stimulus and, based on the likelihoods at the point closest to the perceived stimulus strength, makes a decision about which distribution it came from (S or N). Because of the uniform prior over the points, this first choice is a guess. Following the guess, the observer is given feedback telling him or her whether the stimulus was a signal or noise. The observer then updates his or her stimulus representations at all points close to the presented stimulus. The next stimulus is then evaluated using the still-primitive updated representation. As the subject progresses through the experiment, the subject's stimulus representations begin to closely resemble the sampling distributions and consequently the subject's decisions become more and more optimal.

The updating procedure is a recursive density estimation procedure. I assume a rectangular kernel which influences equally the representation at all points within the bandwidth from the perceived strength of the presented stimulus. Concretely, if a sample y_n from the S distribution is presented on trial n, following 1.3, the S representation $\hat{f}_{S,n}$ at the point x is updated by

$$\hat{f}_{S,n}(x;h) = \frac{n-1}{n}\hat{f}_{S,n-1}(x;h) + \frac{1}{n}K\left(\frac{|y_n - x|}{h}\right), \tag{1.4}$$

and the N representation $\hat{f}_{N,n}$ at the point x decays by

$$\hat{f}_{N,n}(x;h) = \frac{n-1}{n}\hat{f}_{N,n-1}(x;h). \tag{1.5}$$

If the sample y_n were obtained from the N distribution, then the representation of the N distribution $\hat{f}_{N,n}$ would be updated and the representation of the S distribution $\hat{f}_{S,n}$ would decay at the points closest to x. The model assumes equal weighting to both correct and incorrect decisions. That is, there is no variation in the estimation procedure to allow for a reaction to a mistake.

As a consequence of 1.4 and 1.5, the representations are not true probability density functions. However, in the limit (as $n \to \infty$), the representations approach true joint probability density functions and simultaneously approach the distributions of the signal and noise stimuli.

Stimuli that are "ambiguous" (close to the point where an unbiased fixed criterion might be appropriate) will influence both the S and N representations equally. Stimuli that are extreme will only influence the S (or N) representation.

As I just mentioned, kernel estimators smooth the contribution of each point within a specified width h of a particular stimulus. The farther away a point is from the stimulus, the less impact that point will have on the likelihood of that stimulus. As a consequence, the set of points can be spread out well beyond the S and N distributions; only those points close to the sampled stimuli will be relevant for the decision process. This means that the range of the representations is not important as long as there are enough points within the stimulus range to effectively update the density estimates. For the simulations ahead, the points are selected from a uniform distribution with support $L = [0, 100]$. Thus, each point in L is equally likely to be chosen as a part of the set.

Table 1.2 summarizes the model assumptions. The first, number of points, determines how sparse the representation will be. Smaller numbers of points will result in

List of Assumptions	Values
Number of points	10
Distribution of points	$U[0, 100]$
Specified range (denoted L)	$[0, 100]$
Prior distribution	$U[0, 100]$
Kernel used	Rectangular
Bandwidth	10

Table 1.2: List of assumptions made for implemetation of the model.

more sparse representations, which in turn will lead to more variable, less accurate choice behavior. The distribution of these points will also influence performance: points clustered where the S and N distributions overlap may yield different behavior than when points are more disperse. The model assumes that the placement of points is not of any importance to the subject. Therefore, I will uniformly distribute the points across the possible range of stimuli (L). The model also assumes that the subject establishes a prior idea of how the stimuli are distributed on L. I assume a uniform prior over L making all initial decision rules equally likely. The kernel used is rectangular, but Gaussian (or other symmetric, decreasing kernels) may also be useful, as they represent a decreasing weight as a function of distance from the stimulus (see Simulation 3 for the effects of a Gaussian kernel assumption). The last assumption is the bandwidth. This parameter involves answering the question, "How far away must a point be to still have an effect on the subject's internal representation?" This feature is undoubtedly a question of individual differences, which will be further analyzed in Simulation 7. However, to observe the effects of other assumptions, I hold the bandwidth constant at 10.

The next chapter presents simulations to identify effects that certain parameters have on the model or to test the model in an environment similar to the one the subjects experience. The model is tested in many versions of the same basic simulation, while only one assumption is modified per simulation. The first simulation is a control: all subsequent simulations had default values equal to the values initialized in this simulation. The second simulation increased the number of points from 10 to 40. The third simulation used a Gaussian kernel for the density estimate instead of the rectangular density estimate. The fourth simulation increased the separation from 1 to 2 standard deviations. The fifth and sixth simulation examined the performance of the model when the probability of the stimulus coming from the S distribution was .75 and when it was .25 respectively. This manipulation should result in effects similar to those observed in changing the value of β. Simulation 7 doubles the bandwidth from 10 to 20. Finally, Simuluation 8 mimics the setup of Experiment 2; that is, Simulation 8 is designed to test the model in an environment where the parameters shift unannounced.

For each simulation, I evaluate the performance of the model in three ways. First, I show graphs of the internal represenations (both S and N) at particular trials (10, 30, 50, 70, and 100) with the corresponding sampling distribution. Second, I present a table for the cumulative average of the accuracy, hit rate, and false alarm rate of 100 simulated subjects for each condition at particular trials. Finally, I plot the final false alarm rate against the final hit rate for 1000 simulated subjects producing a cluster which unveils a receiver operator curve.

Chapter 3 presents the results of two experiments used to test the model. The first experiment is a simple signal detection task with two conditions. The two conditions

differ by the separation between the S and N distributions. The subject is given a stimulus, makes a decision regarding its identity (signal or noise), and the subject subsequently updates his or her internal representation of the S and N distributions as given by 1.4 and 1.5. Given this type of recursive updating, the subject's internal representation should more closely resemble the S and N distributions as the number of trials increase. The first condition had a separation of 3 standard deviations and the second condition had a separation of 2 standard deviations. The second experiment is similar, except I examine the effects of an experimentally induced criterion shift. The shift is forced by a unannounced change in the frequency of stimuli from the S distributions. The shift occurs four times during the experiment, each occuring on 100 trial intervals. The first, third and fifth 100 trials have a $P(signal) = .5$ and the second and fourth 100 trials have a $P(signal) = .8$.

CHAPTER 2

SIMULATIONS

The objective of the simulations is to generate the predictions of the model with respect to the number of points, the type of kernel, a decrease in d', an increase and a decrease in the frequency of the presentation of a signal, the bandwidth, and Experiment 2. For each simulation I created an environment similar to the environment the subjects in Experiment 1 and 2 experience. The N distribution was normal with a mean of 40 and a standard deviation of 20. The S distribution was also normal with a mean of 60 and a standard deviation of 20. I manipulated several variables with the intentions of showing that the model performed similarly to human subjects in Experiment 1 and Experiment 2, and to generate the model's unique predictions. In each simulation, only one parameter was manipulated to better indentify these effects.

For each simulation, the performance of the model is evaluated in four ways. First, I examine the model's internal representation of both the N and S distributions together with the sampling distributions. The internal representations evolve to a close approximation of the sampling distribution. This process is plotted for each simulation. Error bars are included in the plots and they represent the 95% confidence interval of the variance for 100 simulations. Second, I examine the cumulative hit rate,

false alarm rate, and accuracy for each version of the model (Table 2.2, Table 2.3, and Table 2.4 respectively). Third, I show the final false alarm rate plotted against the final hit rate for 1000 simulations. This produces a cluster near the theoretical receiver operator curve, which is included in the plots. Fourth, I conduct an assesment of the performance of the model. To do this, I computed the root mean square for the difference between both representations and the sampling distributions after each trial. The root mean squared difference is calculated by

$$\mathrm{x}_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \phi_i)^2},$$

where x_i represents the value of the representation at point i, ϕ_i represents the value of the sampling distribution at point i, and n represents the number of points. I performed 1000 simulations and averaged the results for each trial. I then plotted this calculation for trials 1 through 150. These figures are provided in the appendix.

The first simulation serves as a control, in which a set of parameters are established (Table 2.1), which can then be systematically manipulated to produce subsequent simulations. In the second simulation, the number of points was increased to 40. In the third simulation, a Gaussian kernel was used for updating the internal representation. The Gaussian kernel is more efficient than the rectangular kernel used in the control simulation (Silverman, 1986). In the forth simulation, the standard deviations of the sampling distributions were changed from 20 to 10, increasing S and N discriminability from 1 to 2 standard deviations.

The fifth simulation increased the probability of the stimulus coming from the S distribution from .5 to .75. The sixth simulation decreased the probability of the stimulus coming from the S distribution from .5 to .25. These two simulations are

Control Simulation Assumptions	
Assumption	Assignment
Number of Points	10
Specified Range	$(0, 100)$
Distribution of Points	$U(0, 100)$
Prior Distribution	$U(0, 100)$
Kernel	rectangular
Bandwidth	10

Table 2.1: The assumptions of the control (first) simulation are shown.

intended to generate predictions of the model under different experimental setups similar to those I propose for Experiment 2. Finally, Simulation 7 modifies the bandwidth of the rectangular kernel from 10 to 20. This modification is intended to find the effects of the bandwidth on the accuracy of the simulated subject. The eighth simulation will show the effects of a change in the experimental parameters, namely an increase in the probability of the stimulus being a signal. This simulation is similar to Simulations 5 and 6, but the increase in frequency of a signal happens during the experiment as it did in Experiment 2. Another important matter is the different rates of adjustment by the simulated subject. From Experiment 2 (Figure 3.3), we should see a decay in the sensitivity of the simulated subject to the experimental parameters which results in poorer accuracy following later experimental parameter manipulations.

2.1 Simulation 1 - The Control Simulation

Simulation 1 serves as the basis of the other simulations. This simulution will be similar to Experiment 1 (particularly condition 2); however, here $d' = 1$ whereas in Condition 1, $d' = 3$ and in Condition 2, $d' = 2$. (A more suitable model simulation of

the second condition may be found in Simulation 4.) The optimal criterion location can be found by Equation 1.2 to be 50. Consequently, the internal representations of the S and N distributions of the optimal observer should intersect at that point uniquely.

The evolution of the internal representation of a simulated subject is given in Figure 2.1. This figure shows a single simulation superimposed on the mean and 95% confidence interval for the standard deviations of 100 simulations. The hollow circles represent the mean likelihoood representation for 100 simulated subjects at the points 5, 15, 25, ..., 95. The error bars with the solid lines represent the 95% confidence interval for the N representation and the error bars with the dashed lines represent the 95% confidence interval for the S representation. As the number of trials increase, the variance in the represenations decrease for both N and S. A "good" representation is one that closely resembles the respective sampling distribution. To avoid a subjective interpretation of "goodness", the root mean squares are calculated and plots of these calculations are shown in the Appendix. As we can see from Figure A.1, a relatively "good" representation of the sampling distributions is established by trial 100. The convergence to the final value (.00038) is nearly identical for both representations. This convergence is slow and undoubtedly contributes to the poor accuracy found in the simulations: the average final accuracy is 66.41% with a standard deviation of 3.26%.

Figure 2.2 shows a cluster of final hit/false alarm rates generated by 1000 simulations. The cluster is centered on the ideal curve for $d' = 1$. Some simulations exhibit a high degree of bias due to variance in the stimuli. However, the mean of the cluster falls on the point of no bias. The mean hit rate was .649 and the mean false alarm

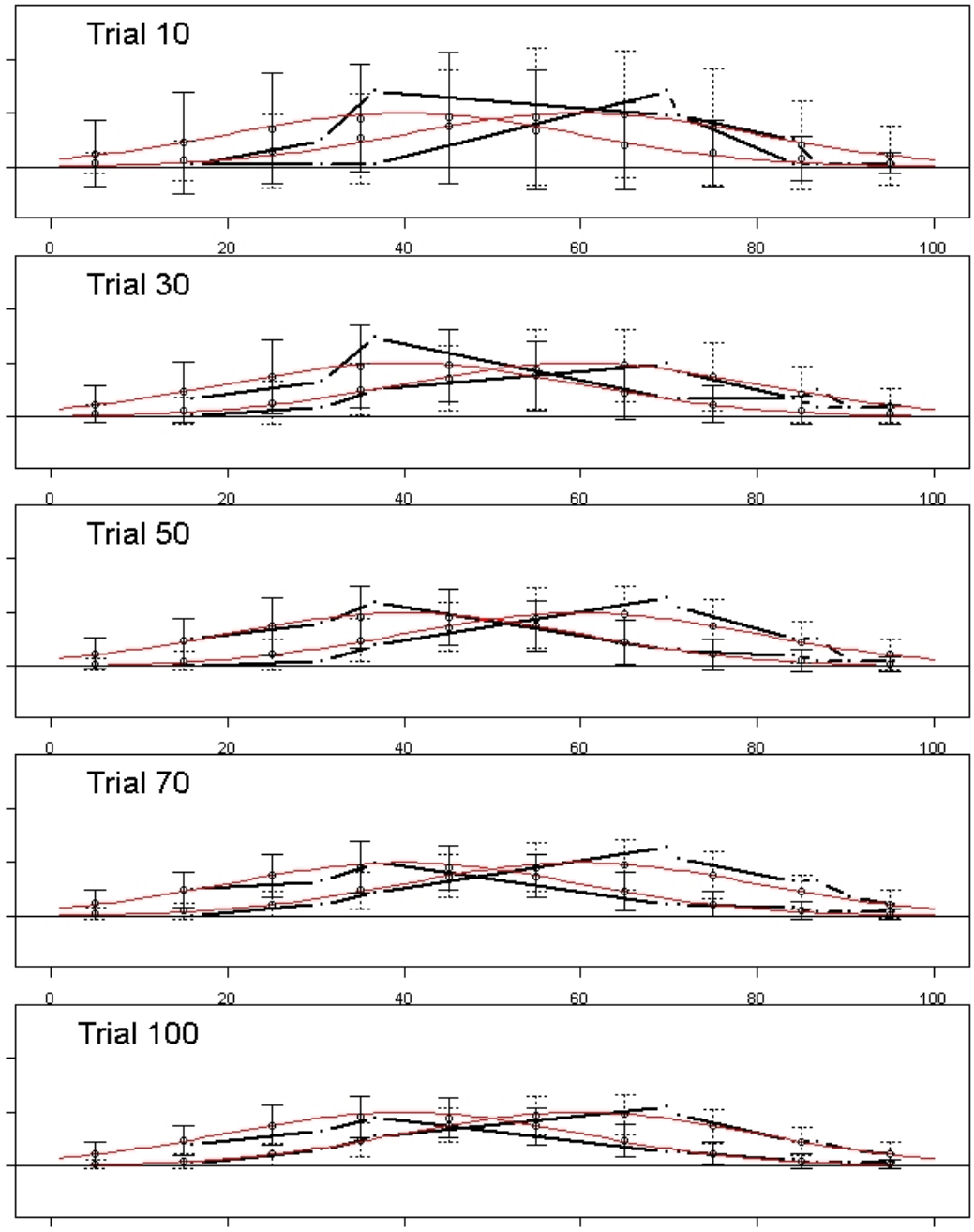


Figure 2.1: Evolution of Internal Representation in Simulation 1. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

rate was .326 (with standard deviations of .112 and .116 respectively). From Egan (1975), we see that the ROC cluster closely resembles the predictions made by the traditional SDT model.

2.2 Simulation 2 - Increasing the Number of Points

In Simulation 2, the number of points was increased from 10 to 40. This should allow for better estimates of the S and N distributions because more points are contributing to the estimate. Accuracy should therefore increase faster and as a byproduct, the final accuracy will be higher.

Figure 2.3 shows the evolution of the model's internal representation. Figure A.2 shows the internal representation closely estimates the S and N distributions by trial 100. Additionally, the convergence point has not been reduced (.00037). The buildup of the representations is further reflected in the performance changes over time, as demonstrated in Tables 2.2, 2.3, and 2.4. There is almost no difference between the performance of this model and that of Simulation 1.

The results of Simuluation 2 show that the number of points has little effect on the final accuracy of the representation. When the number of points is increased, the model has only a 1.11 percentage unit superiority over Simulation 1. While small, this performance enhancement is still important. It can best be explained by the dispersion of points. Because estimates of the density at a particular stimulus are acquired by the comparison of the likelihoods of the closest points, the superiority of these simulated subjects can be attributed to likelihood estimates that are closer to the stimulus and to the more detailed estimate of the S and N distributions. Because there are more points, there is a higher probability of a point being nearer to the

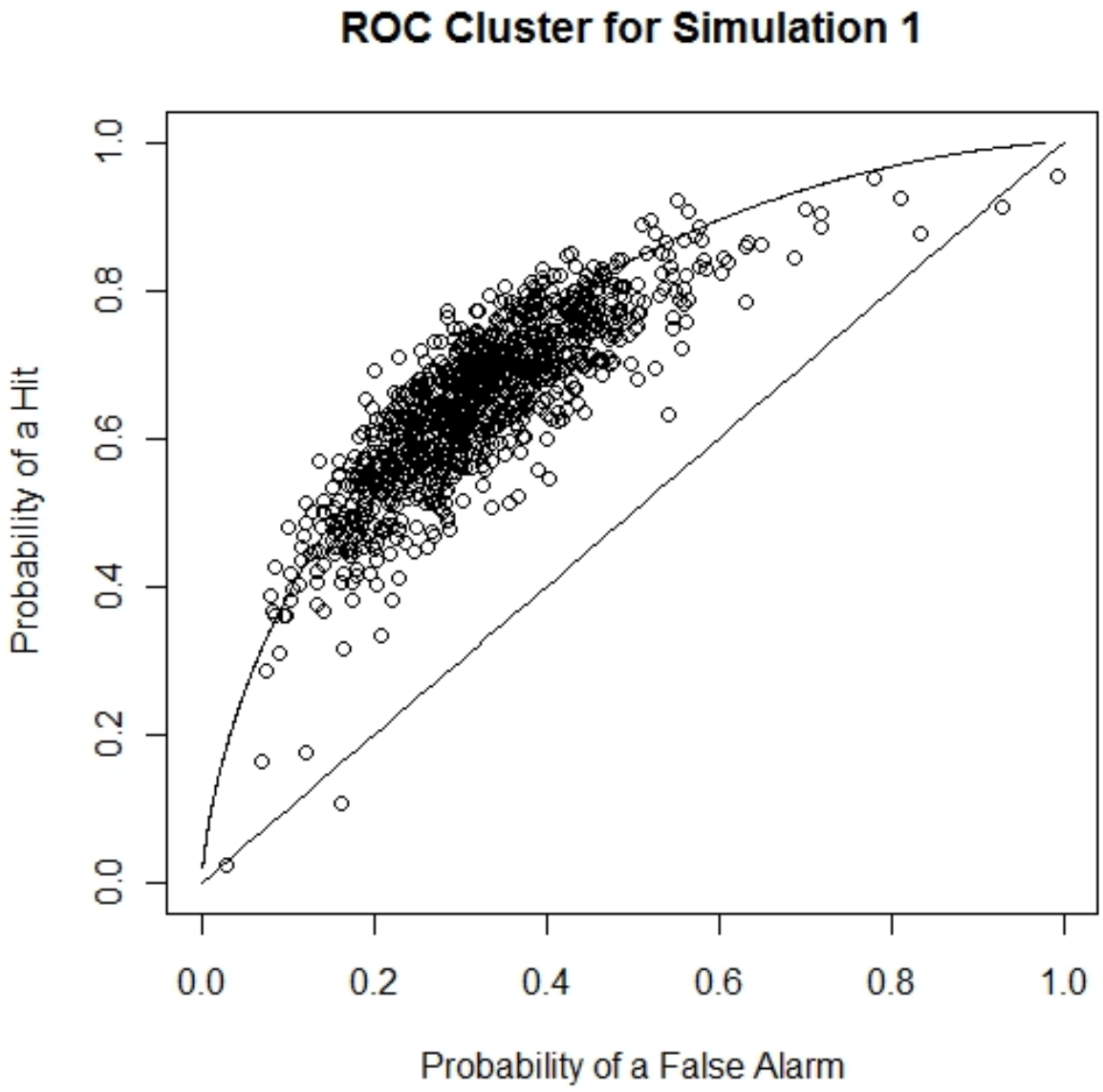


Figure 2.2: ROC Cluster of Simulation 1 ($N = 1000$). The curve represents a d' of 1.

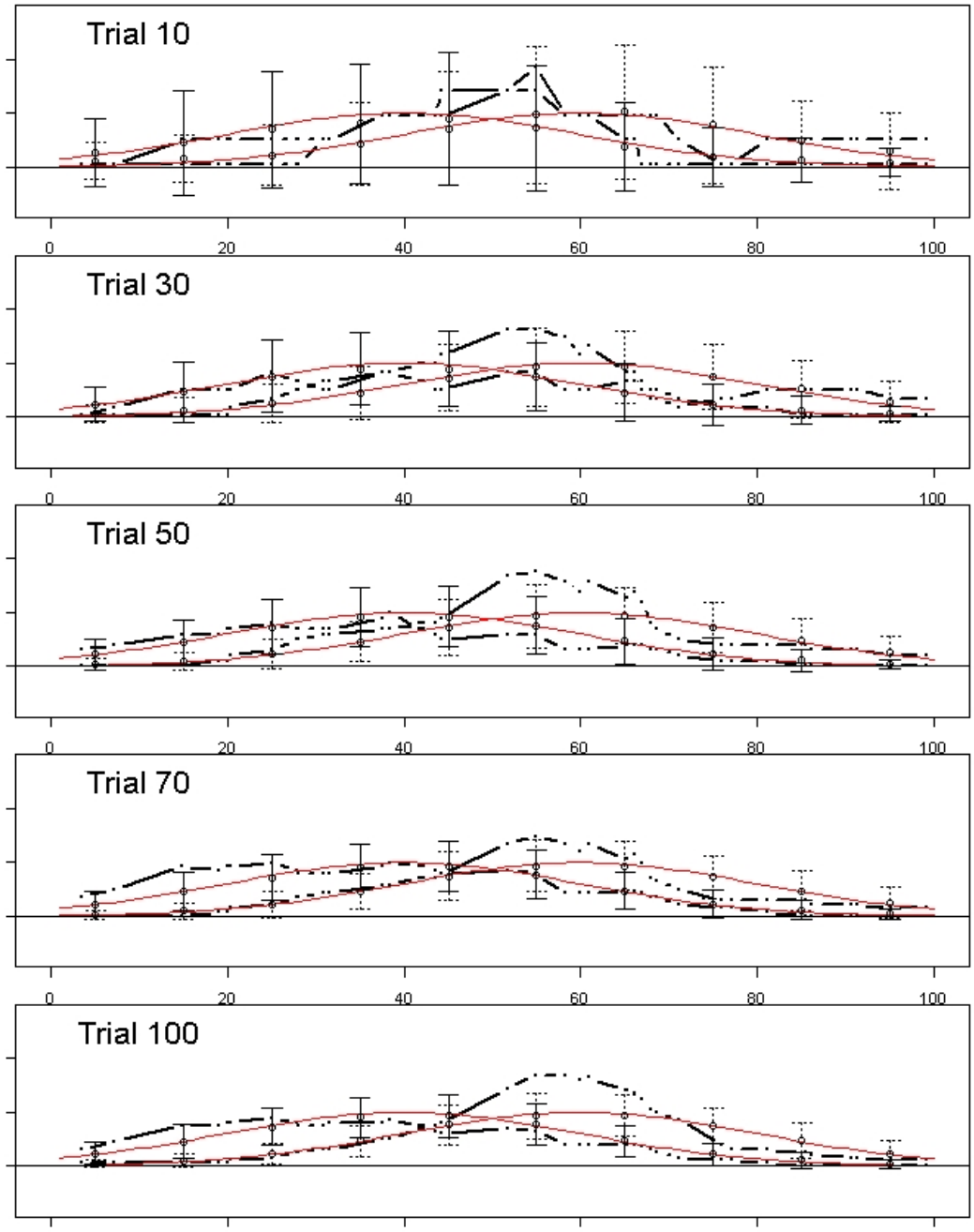


Figure 2.3: Evolution of Internal Representation in Simulation 2. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

particular stimulus than in Simulation 1. Additionally, the number of points reduces the standard deviation of the final cumulative accuracy (2.81 in this simulation).

The ROC plot is similar to the ROC plot in Simulation 1 (see Figure 2.4). There is a much smaller variance in the dispersion of the cluster compared to Simulation 1: the standard deviations of the hit rate and false alarm rate are .081 and .082, respectively.

2.3 Simulation 3 - The Gaussian Kernel

Although the Gaussian kernel has better efficiency as a density estimator (see Silverman, 1986), it may not be psychologically plausible as a basis for the representations. If the subjects update their internal representations of the data with a Gaussian kernel, this means that subjects use a decreasing weight as a function of distance from the stimulus to estimate the likelihoods of the choice alternatives; which may make sense, but such a computation may require more cognitive effort. However, it also means they have a Gaussian representation available for the computation, which may not be reasonable. In this simulation, the assumption of a rectangular kernel is replaced with a Gaussian kernel.

From Figure 2.5, we see that the model's internal representation more closely resembles the S and N distributions than the rectangular kernel did in Simulation 1. This is reflected in Figure A.3, where the root mean squares decrease rapidly. However, it is important to note that the final root mean square has not changed. This representation does not result in higher response accuracy as shown in Tables 2.2, 2.3, and 2.4. Here the mean final cumulative accuracy is 66.96% with a standard deviation of 2.92%.

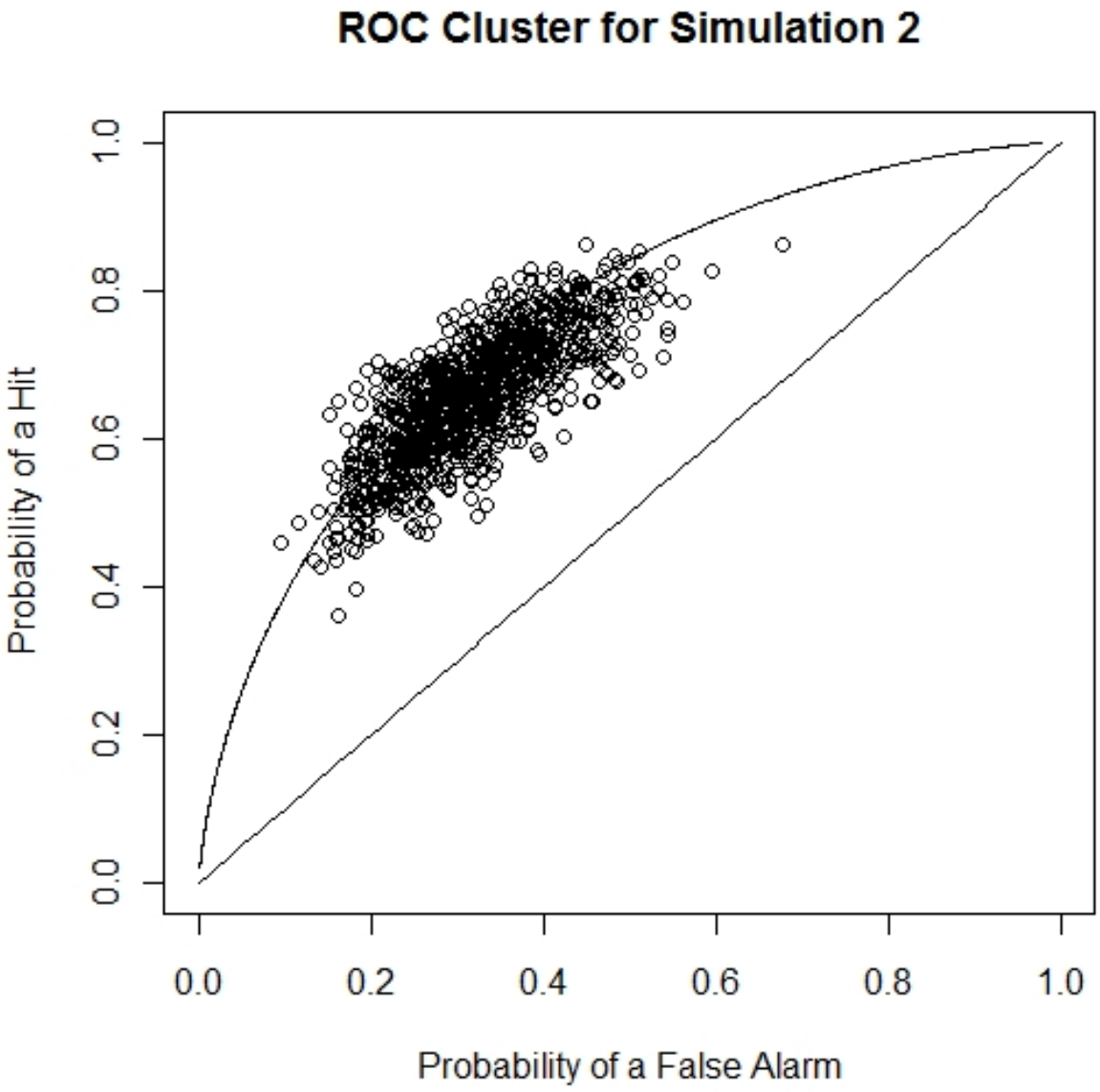


Figure 2.4: ROC Cluster of Simulation 2 ($N = 1000$). The curve represents a d' of 1.

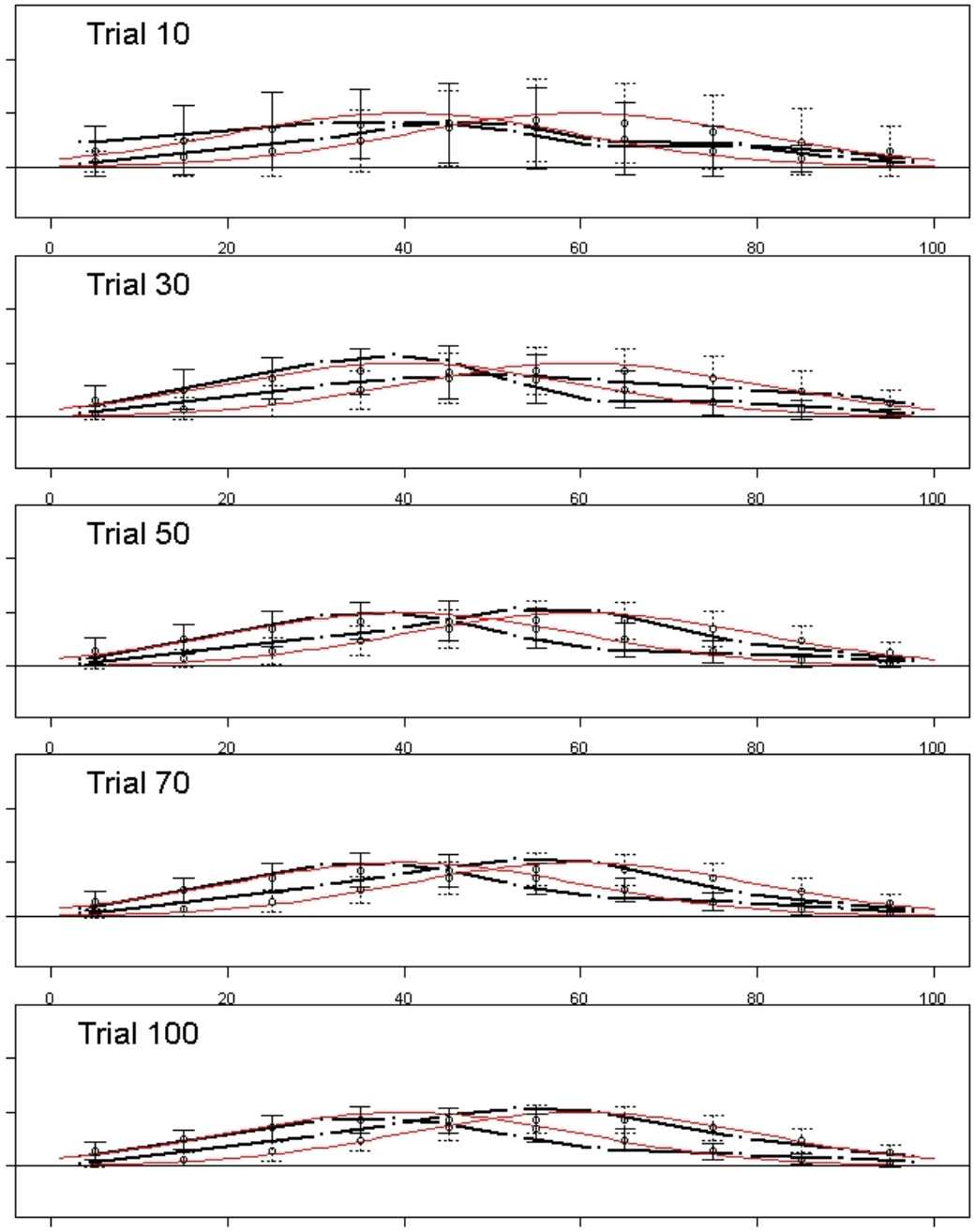


Figure 2.5: Evolution of Internal Representation in Simulation 3. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

The ROC plot of this simulation is similar to that of the Simulation 1. From Figures 2.2 and 2.6, we can see very little differences in the clusters of the two simulations of 1000 subjects. From this we can conclude that these three different types of internal representations have similar levels of discriminability (d') and bias (β).

From the analysis, we see that the representations are not more accurate. In this simulation, we see that the hit rate has increased (.663 with a standard deviation of .114), but without the sacrifice of a higher false alarm rate that would be seen if the simulated subject had become more liberal as we will see in Simulation 5 (mean of .321 with a standard deviation of .114). However, the Gaussian kernel did decrease the root mean square faster, indicating an increased rate of convergence.

2.4 Simulation 4 - Increasing d'

In this simulation, the standard deviations of the sampling distributions are reduced, doubling discriminability. With a higher level of discriminability, the model should have a lower false alarm rate, a higher hit rate, and overall higher accuracy (Macmillan, 2002). Furthermore, the ROC cluster should be centered around the intersection of a nonbiased ($\beta = 1$) subject in an environment with a d' of 2.

From Figure 2.7, we see that the model has a good estimate of the S and N distributions by trial 50. Figure A.4 shows that the root mean square for this model is much higher than that of the previous three simulations. This can be attributed to the increase in height of the sampling distributions. Additionally, the small number of points hinders the ability of the model to approximate the sampling distributions. It seems that the number of points must be adjusted to accommodate for changes in the sampling distribution.

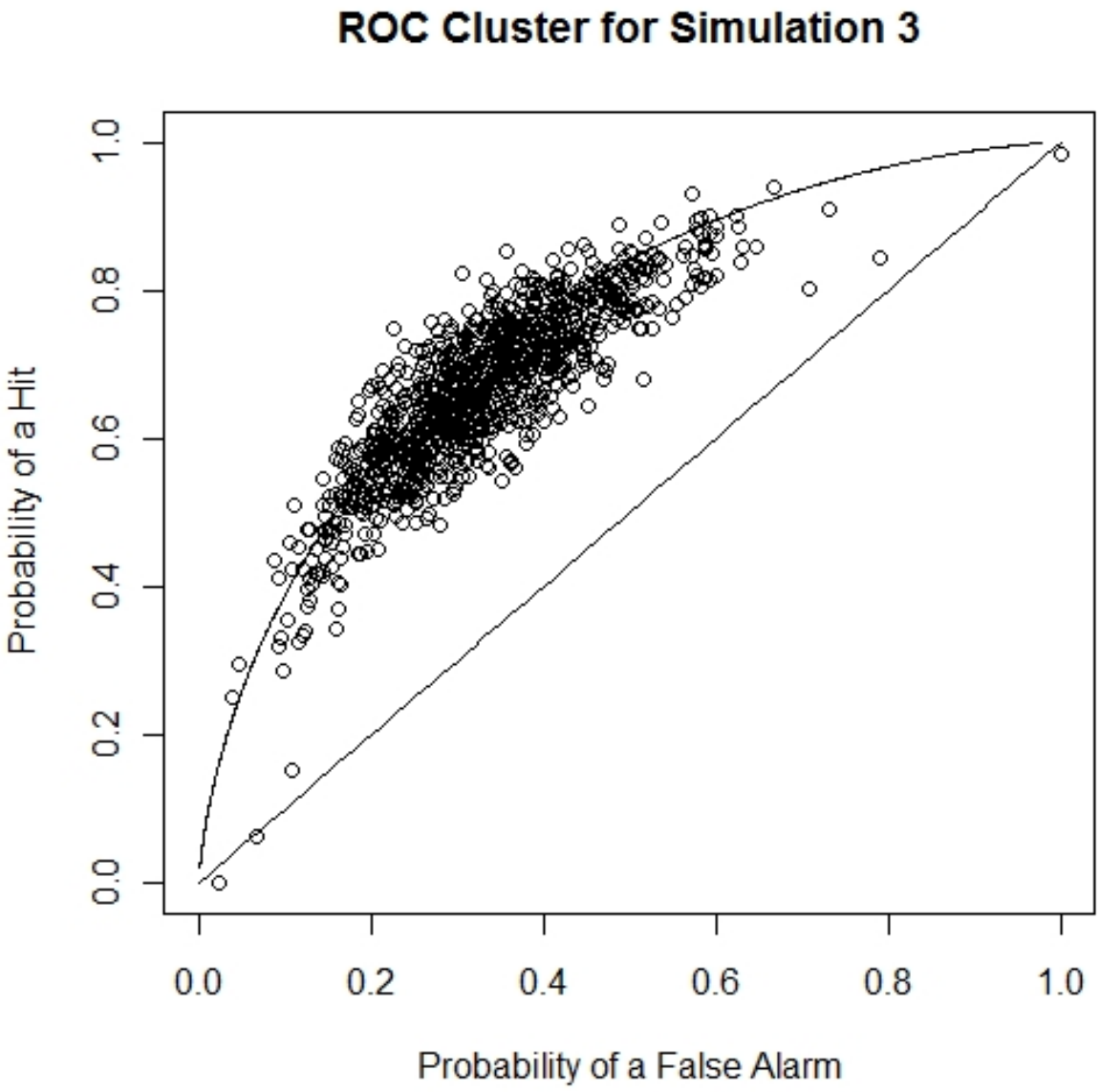


Figure 2.6: ROC Cluster of Simulation 3 ($N = 1000$). The curve represents a d' of 1.

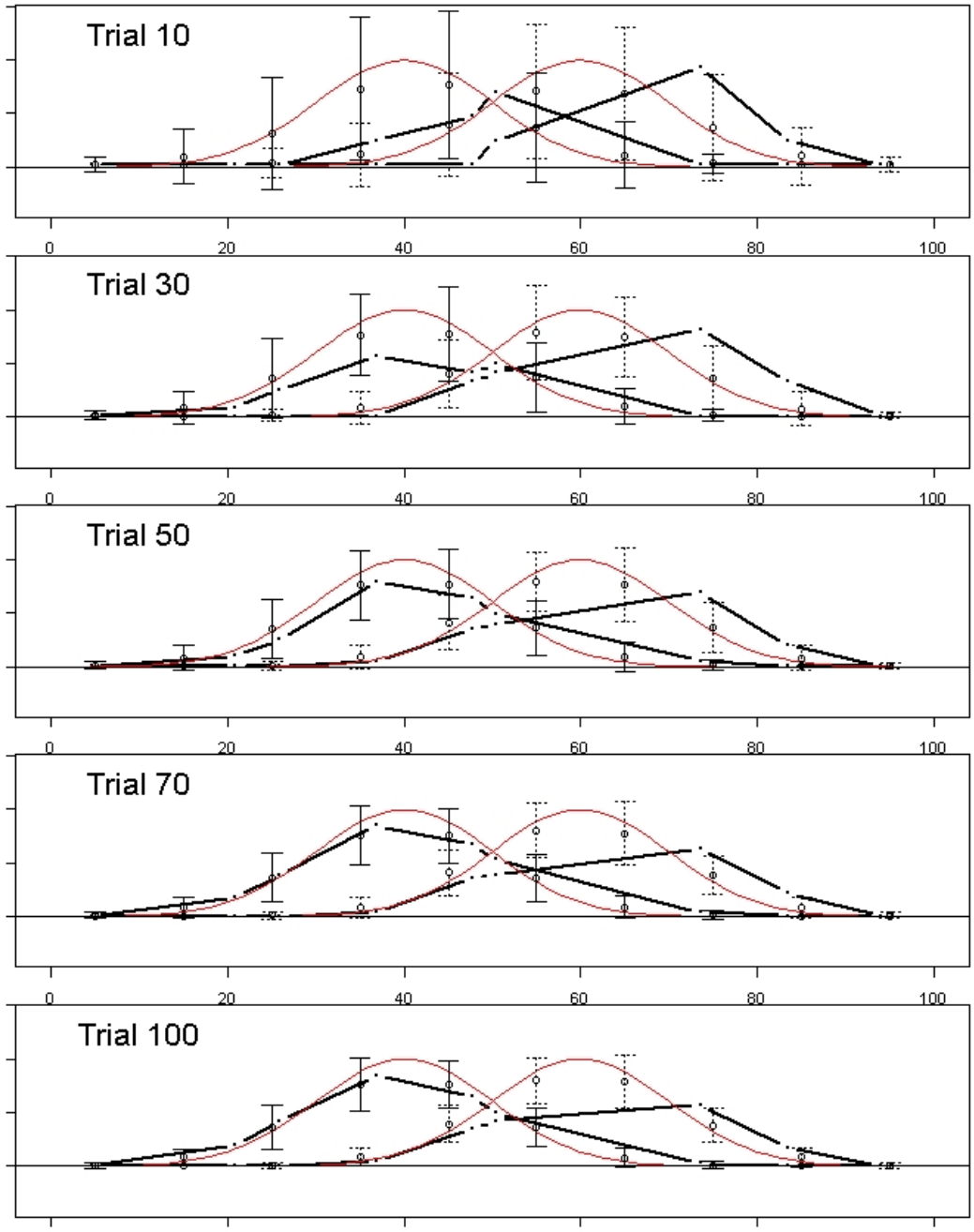


Figure 2.7: Evolution of Internal Representation in Simulation 4. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

Table 2.4 shows that the model's performance more closely mimics the accuracy of human subjects. The model's final cumulative accuracy was 79.44% on trial 340. This is close to the results of the second condition of Experiment 1. The model exhibited a very high hit rate (.663 with a standard deviation of .146) and a low false alarm rate (.329 with a standard deviation of .146). The increase in d' has affected the simulated ROC statistics and the cumulative accuracy in the way we expected.

The ROC cluster falls in the expected area. From Figures 1.2 and 2.8, we see that the simulated subjects are unbiased ($\beta = 1$) and the simulated hit and false alarm rates fall in the region of the ROC generated by a d' value of 2. The variability of this cluster is greater than in the previous simulations. This result is reflected in the variance of the standard deviation of the mean final accuracy (5.56%).

2.5 Simulation 5 - A Liberal Bias

This simulation is intended to show the response of the model to changes in the probability of a stimulus being a signal. For this simulation, I increased the probability of a signal from .50 to .75. This type of manipulation should result in effects similar to those observed by increasing β (reflecting a liberal bias). The model should identify more stimuli as signals to account for the increased frequency of a signal, similar to the way a liberally biased subject would say "yes" more often than "no" ("signal" more often than "noise"). However, this results from changes in the representation, not changes in the criterion.

Table 2.4 shows that the model performed only slightly better than chance (75.6% accuracy with a standard deviation of 2.26%). This finding is not surprising given the extreme overlap of the S distribution as illustrated in Figure 2.9. The model

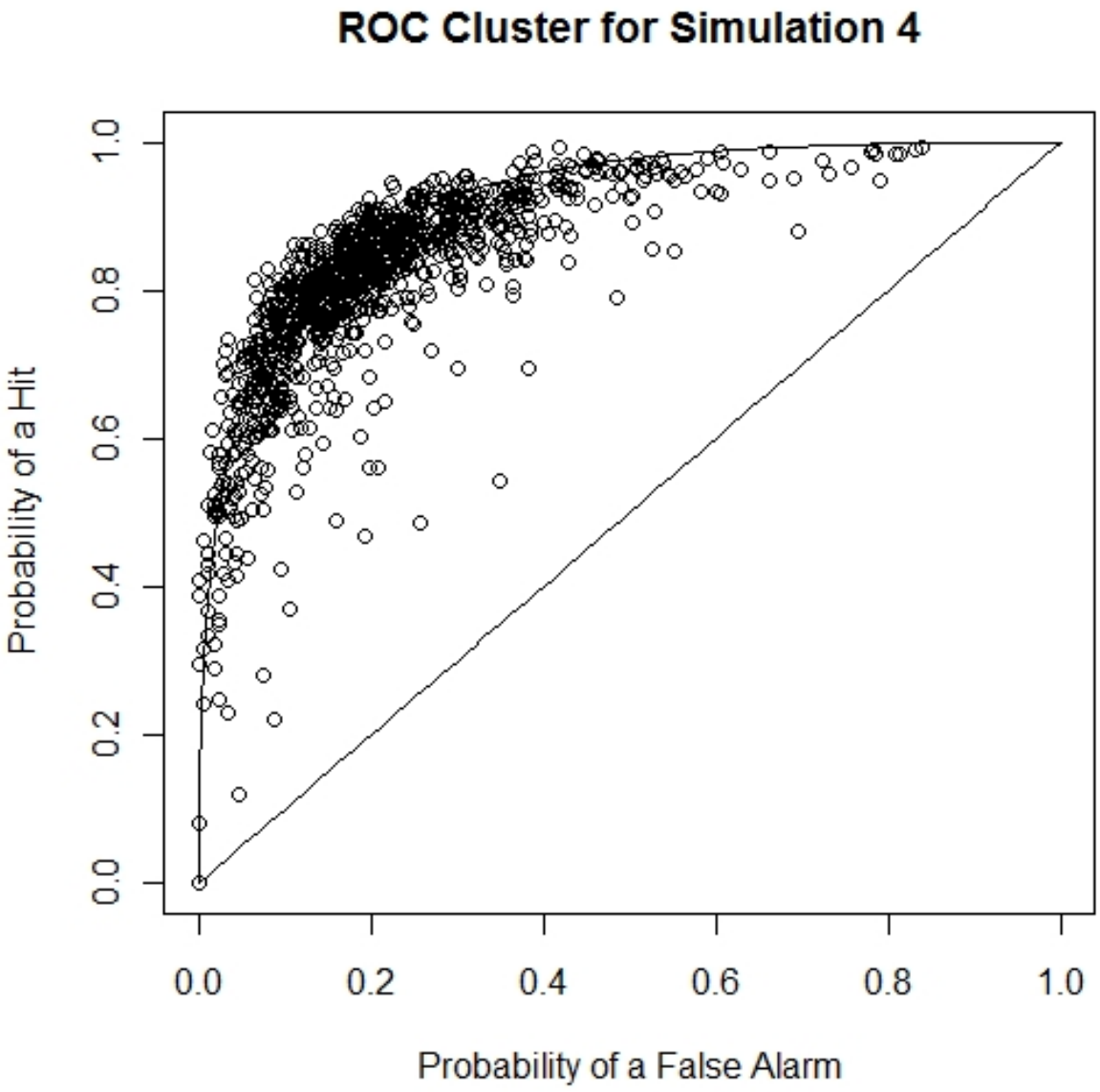


Figure 2.8: ROC Cluster of Simulation 4 ($N = 1000$). The curve represents a d' of 2.

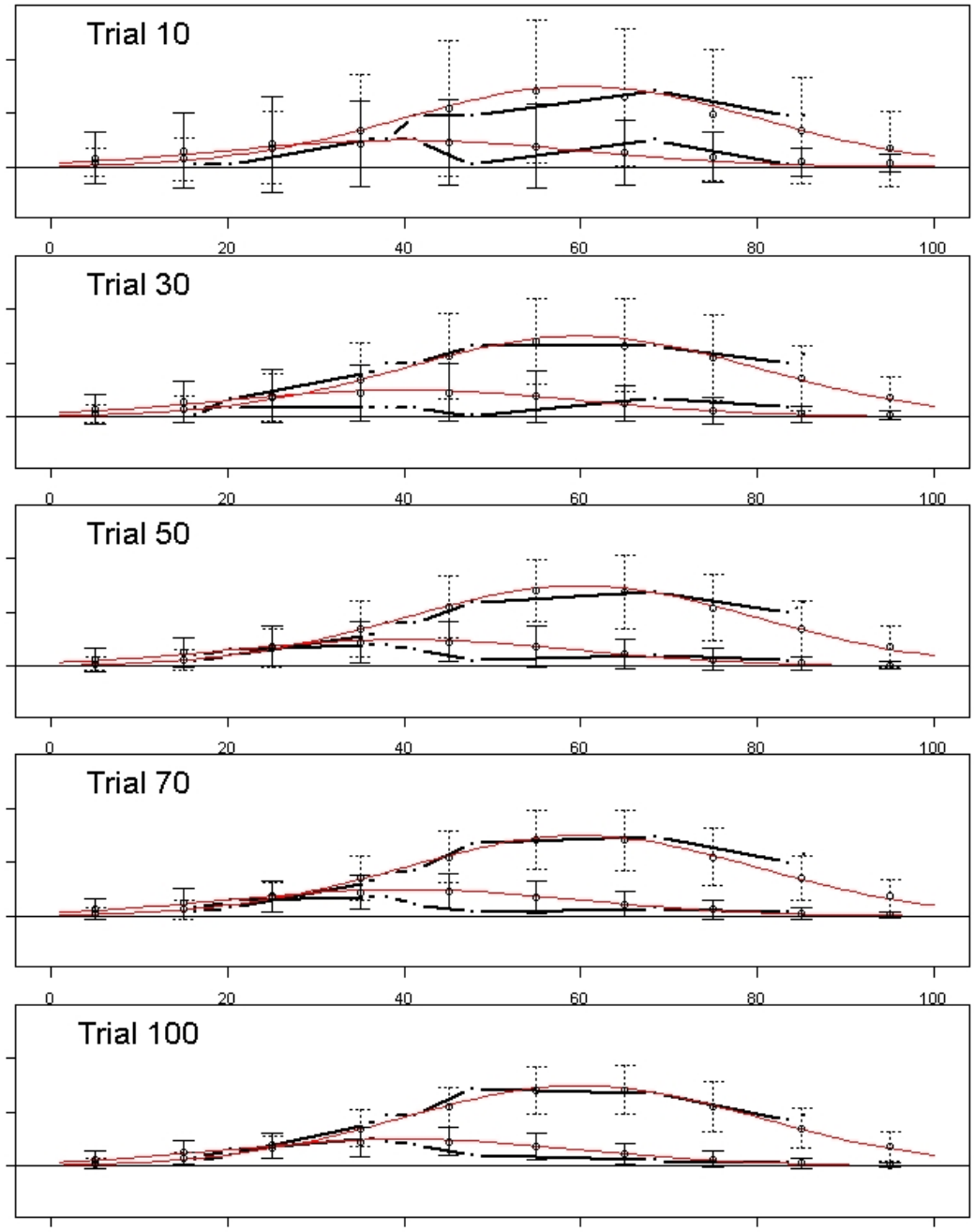


Figure 2.9: Evolution of Internal Representation in Simulation 5. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

corrected for the change in the stimulus probabilities as the optimal subject should and consequently, the representation of the S distribution is consideraby larger than the N distribution. Furthermore, a good estimate of the S distribution was obtained by trial 70 as illustrated in Figure A.5. The model did not obtain a good estimate of the N distribution, most likely due to the relative infrequecy of "noise" stimuli and the decay that corresponds to it given by Equation 1.5.

The cluster does fall on the $d' = 1$ curve as it should (see Figure 1.1). The ROC cluster also shows that the model performed like a liberally biased subject (responds "yes" more often). This liberal bias is reflected in the ROC statistics. The mean hit rate was .917 with a standard deviation of .041 and the mean false alarm rate was .734 with a standard deviation of .13.

2.6 Simulation 6 - A Conservative Bias

This simulation, in combination with Simulation 5, is intended to show the response of the model to changes in stimulus probability. In this simulation, the probability of a signal is .25. This type of manipulation is similar to decreasing β (reflecting a conservative bias). Consequently, the model should identify more stimuli as noise to account for the decreased frequency of a signal, similar to the way a conservatively-biased subject would say "no" more often than "yes" ("noise" more often than "signal").

Table 2.4 shows that the model performed again only slightly better than chance (76.16% accuracy). However, this time the model converged to this percentage quickly. It took only 40 trials for the accuracy to converge to the final value. The model corrected for the change in stimulus probabilities as the optimal subject should

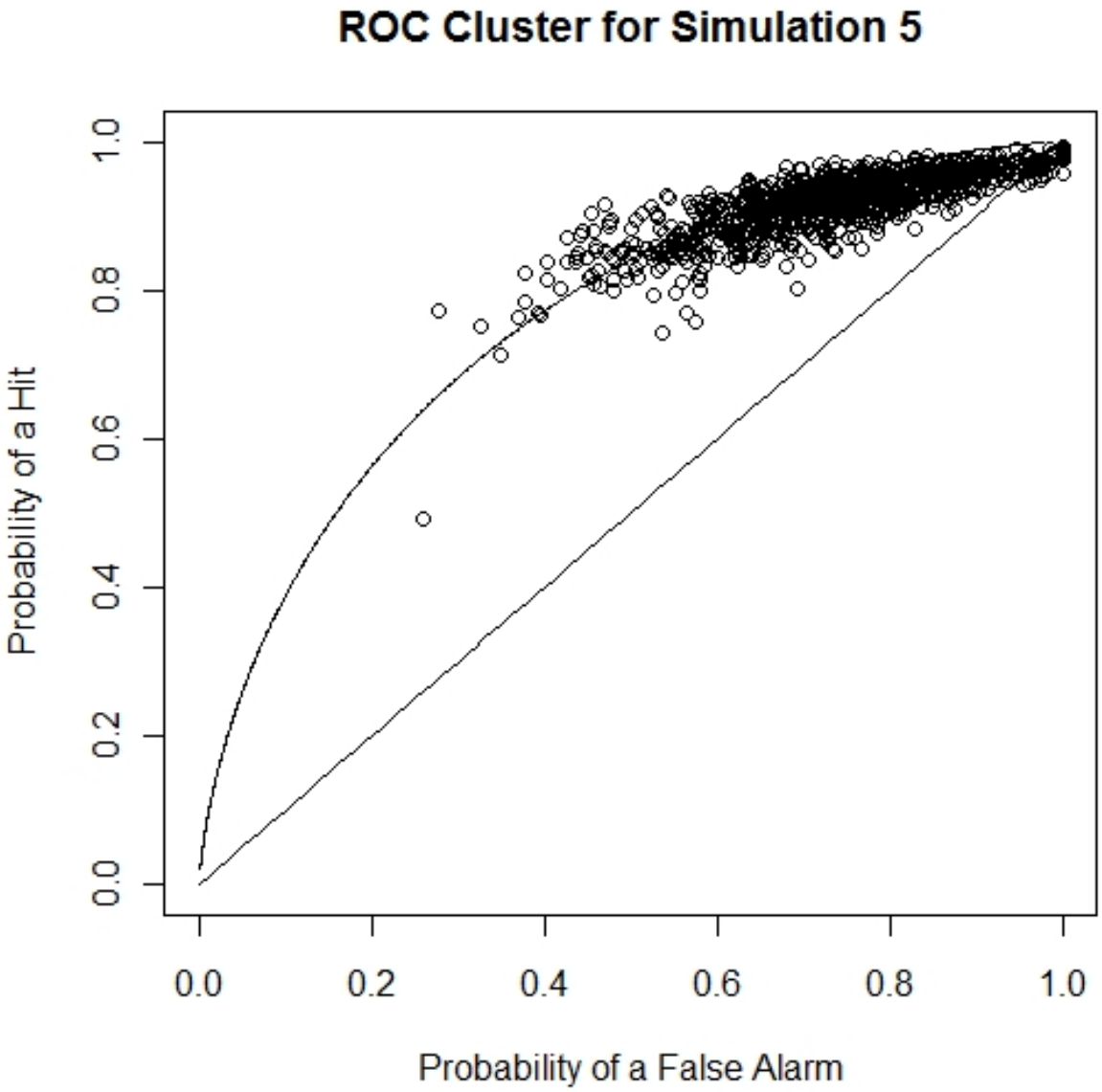


Figure 2.10: ROC Cluster of Simulation 5 ($N = 1000$). The curve represents a d' of 1.

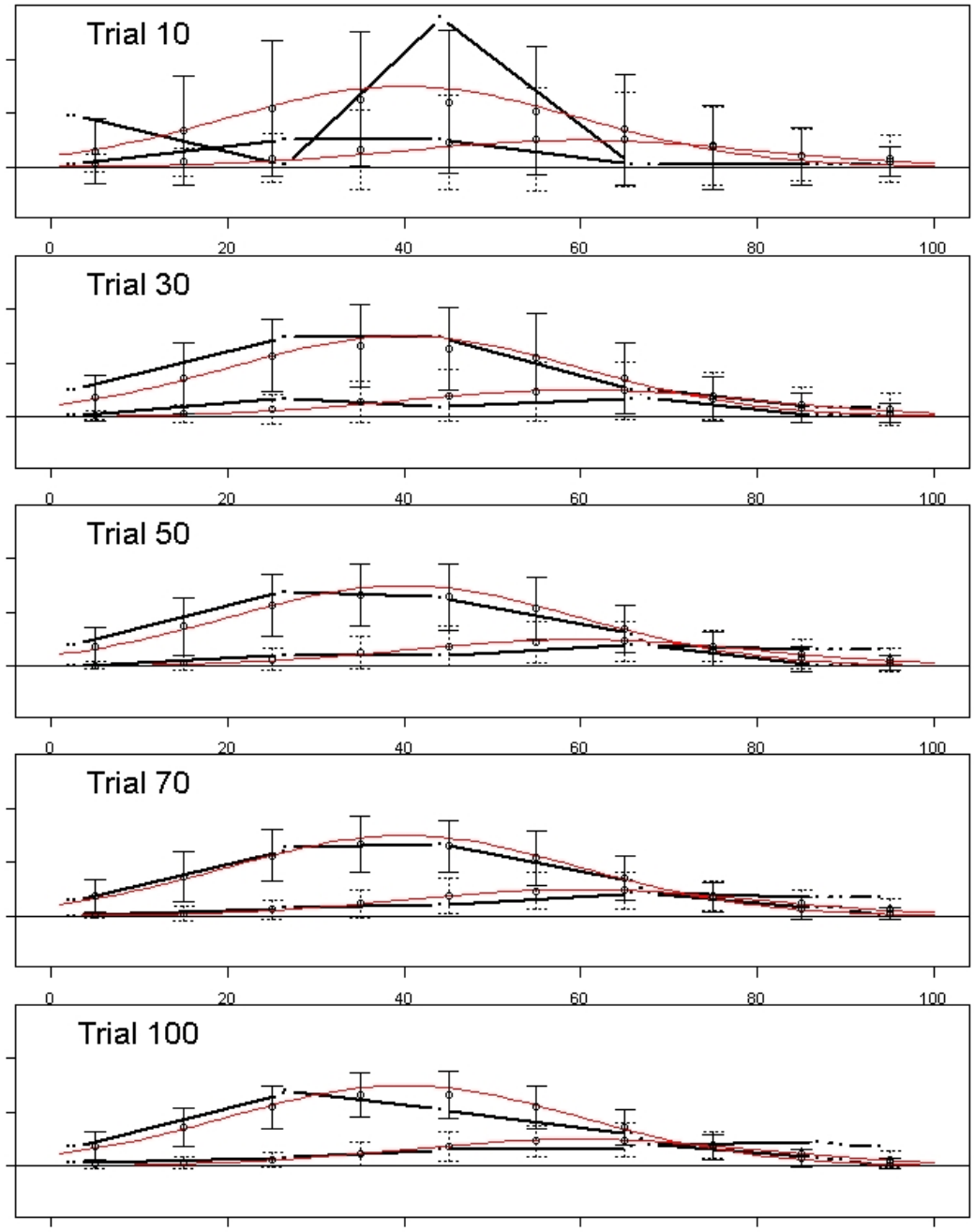


Figure 2.11: Evolution of Internal Representation in Simulation 6. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

and consequently, the representation of the N distribution is consideraby larger than the S distribution. Furthermore, a good estimate of the N distribution was obtained by trial 70 (see Figures 2.11 and A.6). Again, the model did not obtain a good estimate of the S distribution, most likely for the same reasons given in Simulation 5.

The cluster falls on the $d' = 1$ curve as it should (see Figure 1.1). The ROC cluster also shows that the model performed like a conservatively-biased subject (responds "no" more often). This bias is reflected in the ROC statistics. The mean hit rate was .224 with a standard deviation of .124 and the mean false alarm rate was .061 with a standard deviation of .041. Notice that these values are the opposite in magnitude of those in Simulation 5. Both the mean hit rate and mean false alarm rate are small due to the infrequency of "yes" responses.

2.7 Simulation 7 - Increasing the Bandwidth

In this simulation, I demonstrate the effect of the bandwidth for a rectangular kernel on the performance of the model. This parameter will play a large role in fitting the model to real data. For this simulation, I increased the bandwidth from 10 to 20.

Figure 2.13 shows the model's internal representations of the S and N distributions. Figure 2.13 also shows that two representations are not well separated (the error bars overlap a great deal). The representations are much more smooth than they were in Simulation 1. However, the mean final accuracy of the model has not improved (66.91% at trial 340). Figure A.7 also shows that the root mean squares are similar to that of Simulation 1. One important difference between Simulation 1 and

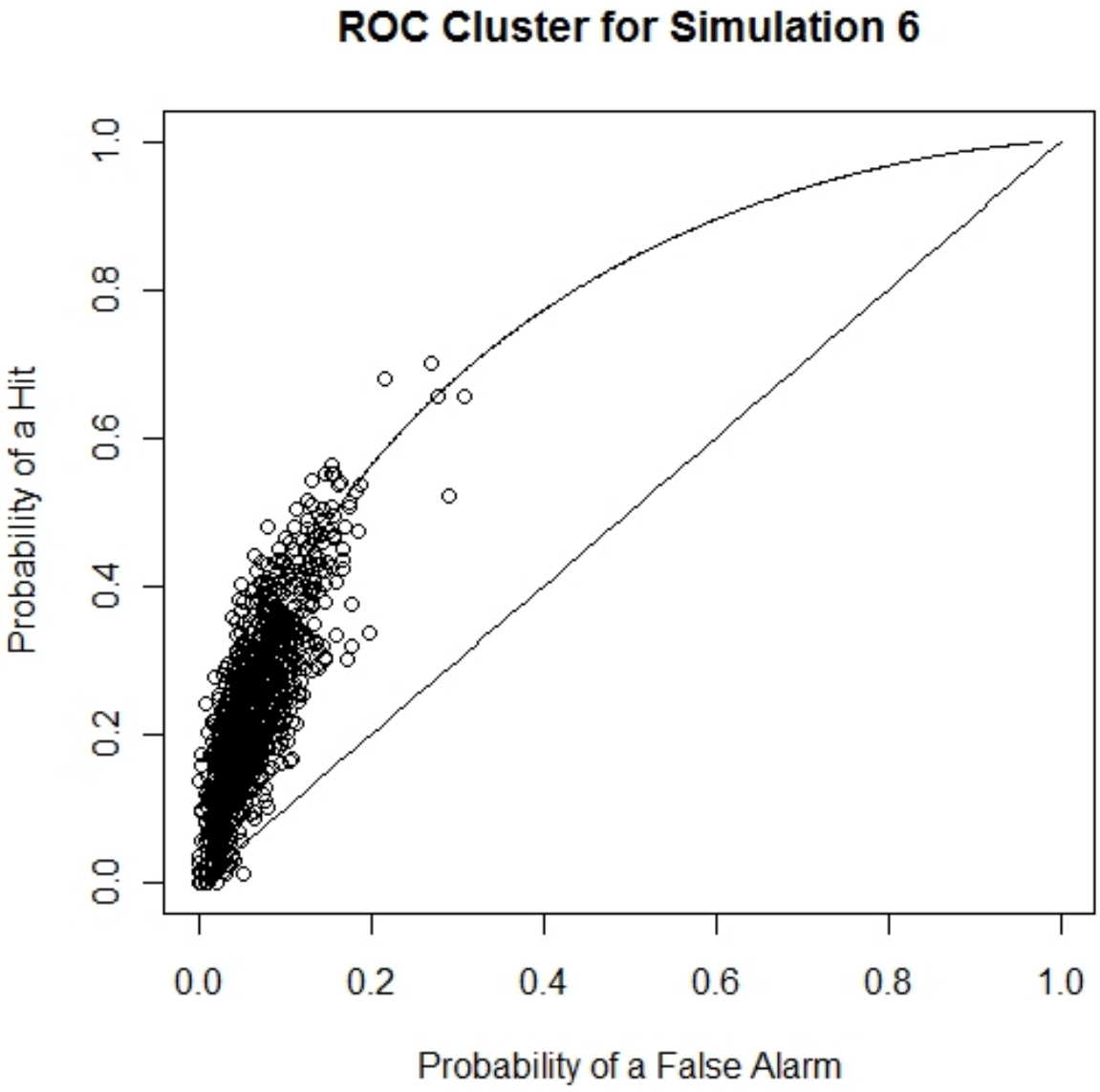


Figure 2.12: ROC Cluster of Simulation 6 ($N = 1000$). The curve represents a d' of 1.

Simulated Cumulative Hit Rate

Trial	1	2	3	4	5	6	7
2	.355±.28	.232±.25	.413±.31	.444±.28	.461±.23	.131±.26	.466±.28
10	.385±.24	.305±.22	.465±.27	.499±.25	.493±.19	.150±.24	.475±.26
20	.414±.21	.418±.20	.534±.24	.583±.22	.647±.15	.164±.21	.486±.25
30	.470±.19	.471±.18	.568±.22	.647±.20	.723±.12	.176±.20	.526±.23
40	.504±.18	.507±.18	.588±.21	.682±.18	.769±.10	.184±.19	.554±.21
50	.529±.17	.531±.16	.600±.19	.704±.17	.797±.09	.189±.19	.571±.20
60	.544±.16	.549±.16	.611±.18	.720±.17	.817±.08	.197±.18	.585±.19
70	.559±.16	.565±.15	.619±.18	.732±.16	.833±.08	.201±.17	.597±.18
80	.571±.15	.576±.14	.624±.17	.740±.16	.844±.07	.204±.17	.607±.17
90	.581±.15	.586±.14	.631±.16	.748±.16	.854±.07	.205±.16	.614±.17
100	.591±.14	.594±.13	.635±.16	.754±.16	.862±.07	.206±.16	.620±.16
150	.616±.13	.621±.11	.646±.14	.770±.16	.887±.06	.214±.14	.641±.14
200	.630±.12	.636±.10	.653±.13	.779±.15	.899±.05	.218±.14	.650±.13
250	.639±.11	.646±.09	.658±.12	.785±.15	.908±.05	.221±.13	.657±.13
300	.646±.11	.653±.09	.662±.11	.790±.15	.914±.04	.223±.13	.661±.12
340	.649±.11	.657±.08	.663±.11	.791±.15	.917±.04	.224±.12	.664±.12

Table 2.2: The mean cumulative hit rates are shown with the mean standard deviations for 1000 simulations in each simulated condition.

Simulation 7 is that the highest root mean square value of this simulation is .0002 smaller than that of Simulation 1. It seems that a higher bandwidth decreases the root mean square initially, but does not seem to improve the end result.

The ROC cluster is slightly less variable than that of Simulation 1. Tables 2.2 and 2.3 show that the mean hit rate was .664 (standard deviation of .112) and the mean false alarm rate was .331 (standard deviation of .114).

2.8 Simulation 8 - Induced Criterion Shift

In this simulation, I will show the effects of an experimentally induced criterion shift similar to that experienced by subjects in Experiment 2. I will change the probability of a stimulus coming from the S distribution from .5 (Trials 1-100) to .2

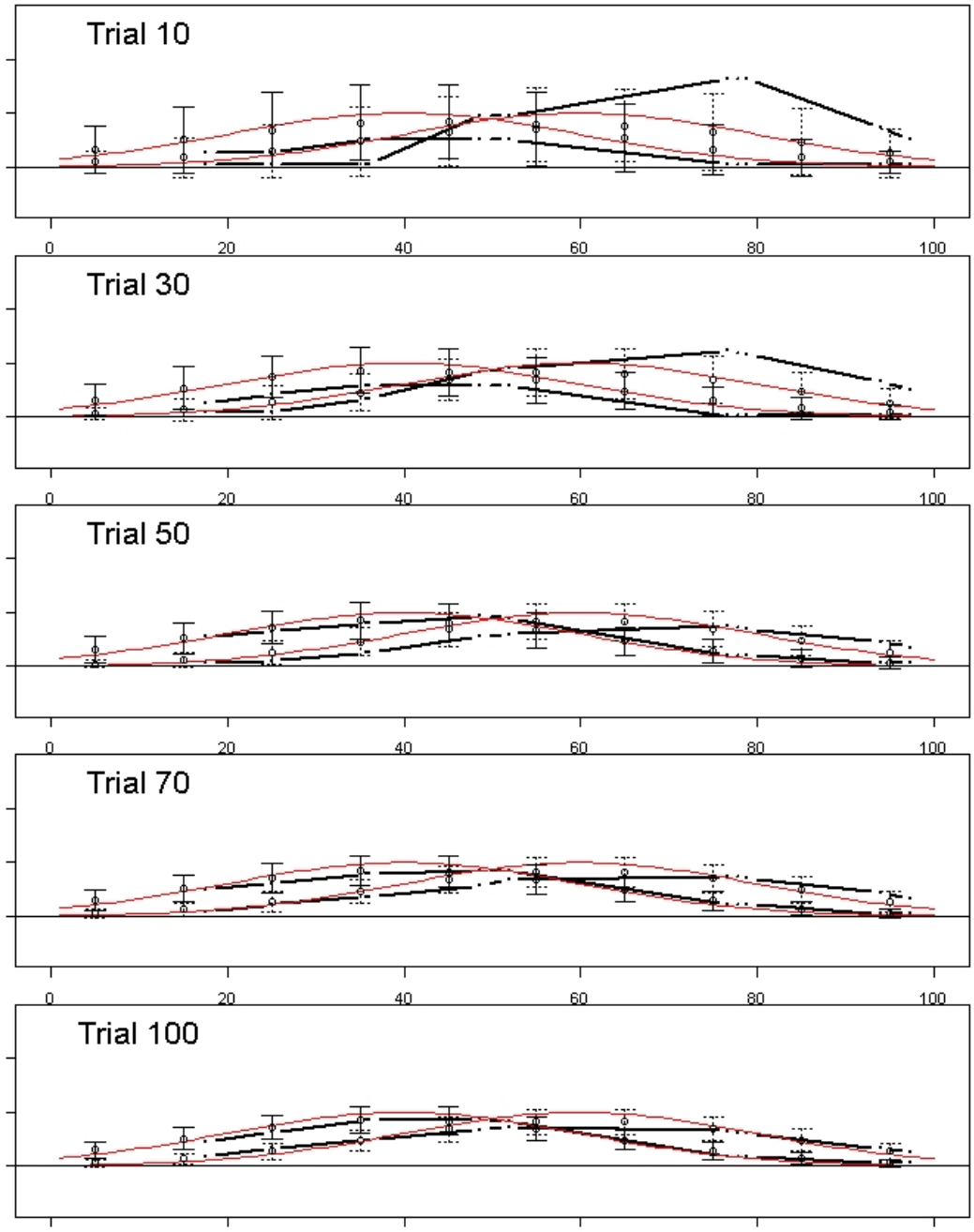


Figure 2.13: Evolution of Internal Representation in Simulation 7. The hollow circle represents the mean of simulated subjects at that point. The error bars represent the 95% confidence interval. A solid line indicates a confidence interval from the N representation, while a dashed line indicates a confidence interval from the S representation.

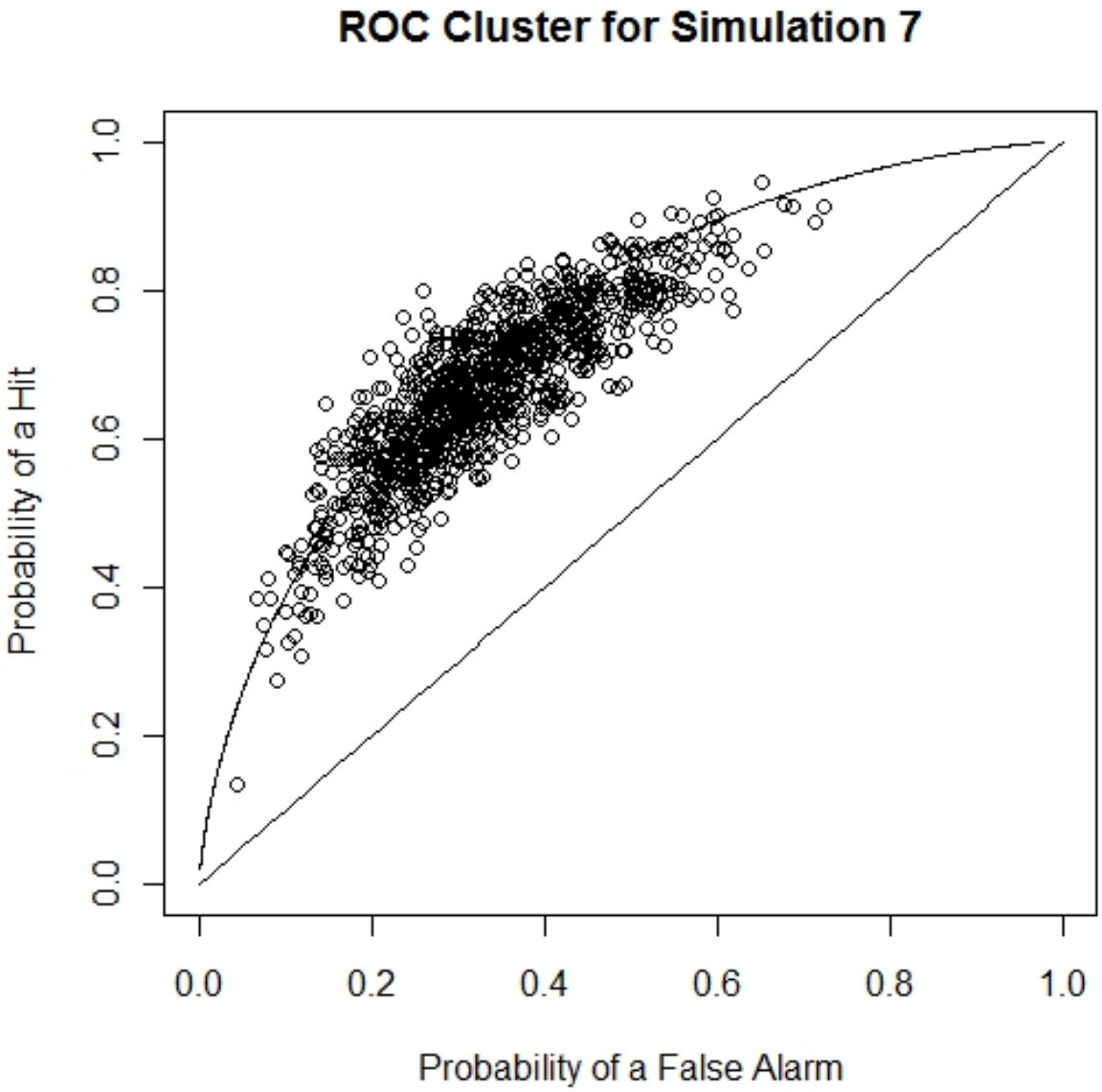


Figure 2.14: ROC Cluster of Simulation 7 ($N = 1000$). The curve represents a d' of 1.

Simulated Cumulative False Alarm Rate

Trial	1	2	3	4	5	6	7
2	.311±.28	.289±.26	.493±.31	.288±.24	.500±.35	.141±.22	.378±.33
10	.333±.26	.299±.24	.467±.28	.256±.22	.567±.31	.123±.19	.372±.29
20	.333±.24	.331±.22	.429±.26	.227±.21	.599±.30	.114±.14	.368±.27
30	.342±.21	.338±.19	.408±.21	.222±.18	.615±.26	.106±.12	.365±.24
40	.346±.19	.341±.18	.396±.20	.219±.17	.626±.23	.099±.10	.359±.22
50	.348±.18	.339±.17	.389±.19	.215±.16	.645±.21	.094±.09	.355±.21
60	.347±.17	.338±.16	.377±.18	.215±.16	.657±.20	.089±.08	.354±.20
70	.344±.17	.337±.15	.370±.17	.214±.16	.668±.19	.086±.08	.350±.19
80	.341±.16	.335±.14	.366±.16	.213±.16	.679±.18	.083±.07	.348±.18
90	.341±.15	.333±.14	.360±.16	.212±.15	.687±.18	.080±.07	.348±.17
100	.340±.15	.333±.13	.356±.14	.212±.15	.692±.17	.078±.06	.348±.17
150	.335±.13	.328±.12	.345±.13	.210±.15	.711±.15	.071±.05	.342±.14
200	.331±.12	.324±.10	.338±.12	.208±.15	.722±.14	.067±.05	.338±.13
250	.328±.11	.322±.09	.334±.12	.207±.14	.728±.14	.064±.04	.335±.12
300	.326±.11	.322±.09	.330±.11	.207±.14	.732±.13	.063±.04	.333±.12
340	.326±.11	.321±.08	.329±.11	.205±.15	.734±.13	.061±.04	.331±.11

Table 2.3: The mean cumulative false alarm rates are shown with the mean standard deviations for 1000 simulations in each simulated condition.

Simulated Cumulative Accuracy

Trial	1	2	3	4	5	6	7
2	51.6±34.6	49.4±35.4	51.4±35.1	54.1±33.3	30.0±29.8	75.1±33.5	50.6±34.4
10	54.6±16.0	53.4±15.9	56.6±15.9	64.8±15.0	52.6±14.1	71.1±17.1	56.8±16.2
20	56.6±12.1	56.5±11.5	58.8±11.7	69.9±11.1	61.3±10.8	71.9±12.2	59.3±11.8
30	58.3±10.0	58.5±9.8	60.4±10.1	72.6±9.4	65.6±9.2	72.5±9.8	60.6±9.8
40	59.6±8.9	59.9±8.6	61.5±8.8	74.2±8.5	67.9±7.9	73.1±8.6	61.8±8.6
50	60.5±8.1	60.9±7.8	62.1±7.8	75.3±7.9	69.3±7.1	73.4±7.4	62.4±7.7
60	61.1±7.4	61.8±7.2	63.0±7.1	75.9±7.5	70.4±6.4	73.7±6.7	63.0±7.0
70	61.8±6.9	62.4±6.5	63.6±6.5	76.5±7.2	71.3±5.9	74.0±6.1	63.6±6.6
80	62.4±6.5	63.0±6.1	64.0±6.0	76.9±6.9	71.9±5.4	74.3±5.7	64.0±6.0
90	62.9±6.1	63.4±5.8	64.5±5.6	77.3±6.7	72.3±5.1	74.5±5.2	64.3±5.7
100	63.4±5.8	63.8±5.5	64.8±5.3	77.5±6.6	72.7±4.8	74.6±4.9	64.6±5.3
150	64.6±4.7	65.2±4.5	65.6±4.3	78.3±6.1	74.0±4.1	75.3±4.0	65.6±4.3
200	65.4±3.9	66.0±3.8	66.2±3.8	78.8±5.9	74.6±3.4	75.6±3.4	66.1±3.7
250	65.9±3.5	66.5±3.4	66.6±3.4	79.1±5.8	75.1±3.1	75.8±3.0	66.5±3.4
300	66.2±3.3	66.8±3.0	66.9±3.1	79.3±5.7	75.4±2.7	76.1±2.7	66.7±3.1
340	66.4±3.0	67.5±2.8	67.0±3.0	79.4±5.6	75.6±2.6	76.2±2.5	66.9±2.9

Table 2.4: The mean cumulative accuracies are shown with the mean standard deviations for 1000 simulations in each simulated condition.

(Trials 101-200) to .5 (Trials 201-300) to .2 (Trials 301-400) and back to .5 (Trials 401-500). This simulation will generate predictions that can be compared to the data from Experiment 2.

Figure 2.15 shows a moving average filter of 16 simulations under identical conditions. The moving accuracy for trial $i + w$ is calculated by summing the number of correct responses from trial i to trial $i + w$ and dividing by w (where w represents the width of the moving average window). This technique gives the reader a clear view of the frequency of correct responses by the subject. The vertical lines represent the onset of the change in the stimulus probabilities. We should see that once the plot crosses these lines, the accuracy should decrease. Figure 2.15 shows that this is usually, but not always the case. Overall, accuracy is slowly trending upward as predicted.

Figure 2.15 shows that the model's accuracy tends to decrease when it crosses the "shift point." After the first shift, the model took about 20 trials to accommodate the increased probability (an increase of .30) of a signal. After the second shift, the model took a little longer to recover (32 trials). The model did not seem to be sensitive to the third shift. The fourth shift has the smallest length of decrease, lasting 12 trials. One surprising finding is the consistency of the variance across trials. As I will show, this finding does not seem to be the general trend found in Experiments 1 and 2.

2.9 Summary

Simulations 2, 3, and 7 are intended to show the effects that the number of points, the type of kernel and the bandwidth have on the performance of the model. Simulation 2 shows that if the number of points is increased to 40, the simulations become

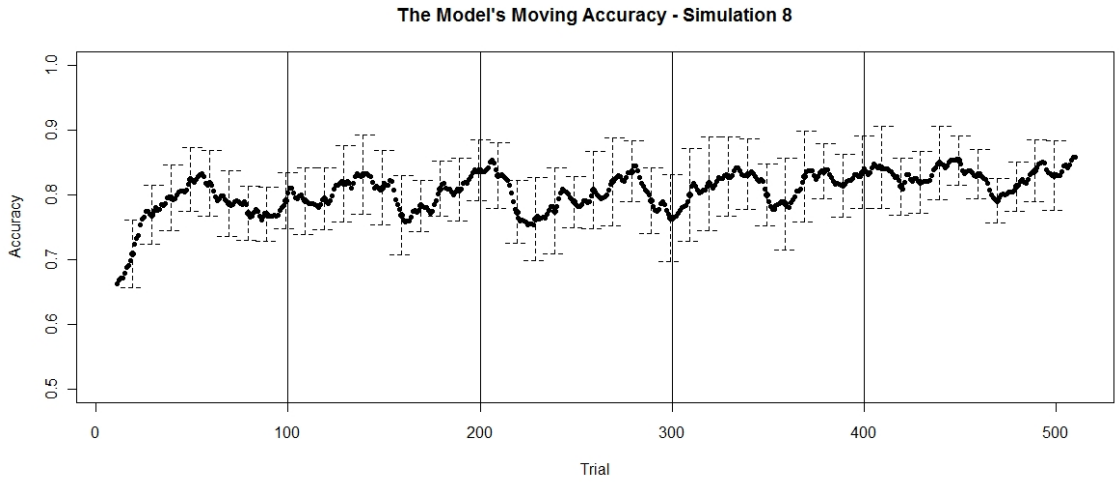


Figure 2.15: Moving Average of Simulated Subjects in Simulation 8. This figure shows the moving mean accuracy of 50 simulated subjects in Simulation 8. A single point represents the average accuracy from that point to 20 trials after that point for all 50 simulated subjects. The vertical lines represent the onset of the calculations that include the new parameters. The error bars represent the 95% confidence interval.

less variable, but overall accuracy increases only slightly. Simulation 3 shows that if a Gaussian kernel is used, performance is enhanced by 1.4% and the root mean squares are more quickly minimized. Simulation 7 shows that increasing the bandwidth to 20 increases the performance of the model by 1.5%. This simulation also showed that an increase in bandwidth decreases the root mean square initially, but does not seem to improve the end result.

The results of Simulations 1, 4, 5, and 6 show that the model behaves as a subject does. In Simulation 1, when $d' = 1$, the model's accuracy is 65.91% (the maximum possible is 69.15%). In Simulation 4, when $d' = 2$, the model's accuracy increases to 79.14% (the maximum possible is 84.13%). The underperformance of the model relative to human performance can be attributed to my choice of parameter values.

Simulations 5 and 6 show that a liberal or a conservative bias can be created by changing the frequency of signals. If $P(signal)$ increases, the model makes responses consistent with a liberal bias. If $P(signal)$ is decreased, the model makes responses consistent with a conservative bias.

Simulation 8 shows that the model is sensitive to within-sequence changes in stimulus probabilities as shown in Figure 2.15.

CHAPTER 3

EXPERIMENTS

3.1 Experiment 1

This experiment was designed to elicit simple yes-no choices from decision makers in a controlled setting. Novel stimuli were used so that participants could not draw on previous experience or memory associations to perform the assigned tasks. In all cases, the subjects chose one of two possible responses after a stimulus was presented.

This experiment was originally conducted by Van Zandt and Jones (2009) for a different purpose.

3.1.1 Subjects

Seventy-six subjects from the OSU psychology subject pool served in this experiment in exchange for course credit. Subjects were randomly assigned to one of two conditions (44 subjects in the first and 32 in the second). All subjects reported normal or corrected-to-normal vision. Naive to the purpose of the experiments but informed of their role, they were tested individually after receiving both oral and written instructions.

3.1.2 Apparatus

The display of stimuli used a CRT video monitor with timing of presentation controlled by an IBM PC/SX desktop computer. Using Ratcliff's COGSYS software, they recorded data from the subjects, who made responses on a standard NMB keyboard with all nonessential keys covered by a white piece of cardboard.

3.1.3 Procedure

Subjects were told that a new disease was raging through the community and that their job was to screen patients who may have been exposed to the disease. They were presented with a single two-digit number and told it was the result of a blood assay. Patients who were ill had higher assay numbers (mean 60) than patients who were not (mean 40). Subjects were given these means. Subjects were to decide whether or not each patient should be treated for the disease. If a sick patient were not treated, s/he would die, and if a well patient were treated, s/he would die.

The number of each assay was drawn from one of two normal distributions with means 40 and 60. Subjects participated in one of two discriminability conditions determined by the standard deviation of each distribution ($\sigma = 6.667$ or $\sigma = 10$).

Each trial began with a random inter-trial interval which followed an exponential distribution with mean 400 ms and shift 200 ms. The assay was presented for 100 ms. The subjects pressed either the "Z" or "/" key on the computer keyboard (marked as "Treat" or "Do Not Treat," response assignment counterbalanced across subjects). Feedback (whether the patient died or not) was presented (for 100 ms) immediately upon the subject's response.

Subjects completed 5 blocks of 68 trials within 30 minutes. After each block, subjects were told how many of their patients lived and how many died and were encouraged to take a short rest.

3.1.4 Experimental Results

Figure 3.1 shows the cumulative accuracy of all 44 subjects over time for the easy condition ($d' = 3$). The cumulative accuracy for trial i is found by summing the number of correct answers from trial 1 to trial i and dividing by i. Figure 3.2 shows the cumulative accuracy of all 32 subjects over time for the hard condition ($d' = 2$). A single point on the graph represents the cumulative accuracy at that point. The horizontal line represents the maximum possible performance on this task. The error bars represent the 95% confidence interval for every tenth trial.

Both figures show that accuracy increases over time. These findings support the claim that as a subject gains more information (experience), the subject becomes more accurate over time. However, towards the end of the experiment, the accuracy curves trend downward. This is most likely due to subject fatigue. In both figures, we also see that the error bars surrounding the accuracy become smaller as the number of trials increase. This effect indicates that while individual variability is high initially, as experience is gained, the subjects approach the same final cumulative accuracy.

The sequential dependencies in the data were also analyzed. A positive sequential dependency (on responses) is given by an increase in the probability of a response R_i given a response R_{i-1} on the previous trial. In contrast to Treisman and Williams (1984) and Howarth and Bulmer (1956), the subjects did not show any sequential dependencies for Condition 1 ($\chi^2(1) = .0003, p = .987$) or Condition 2 ($\chi^2(1) =$

$.0006, p = .980$) at the $\alpha = .05$ level. The effects of sequential dependencies for two trials back were also analyzed; that is, the probability of a response R_i given a response R_{i-2} and R_{i-1} on the previous two trials. Again in contrast to Treisman and Williams (1984) and Howarth and Bulmer (1956), the subjects did not elicit any sequential dependencies for either condition.

From Figure 3.1 and Figure 3.2, we see that the final accuracy for the hard condition is lower than for the easy condition. The final cumulative accuracy of the easy condition is 88.09% with a standard deviation of .023. (The maximum possible performance accuracy is 93.32%.) The final cumulative accuracy of the second condition is 75.75% with a standard deviation of .044. (The maximum possible performance accuracy is 84.13%.) These results support the general finding in SDT: as d' decreases, an observer is less capable of discriminating between the S and N distributions and consequently will show lower accuracy. The effect of the changing representations is still prevelent in this experiment.

3.2 Experiment 2

This experiment was designed to show changes in response frequency with changes in the frequency of signals. The model predicts that changes in stimulus probability lead to changes in the internal representations of the S and N distributions. If evidence of representational changes were found, this experiment would not support the contention of traditional signal detection theory that the subject's internal representations of the stimuli are fixed. Similar to Experiment 1, subjects were asked to make simple yes-no choices in a controlled setting. Novel stimuli were used so that participants could not draw on previous experience or memory associations to perform the

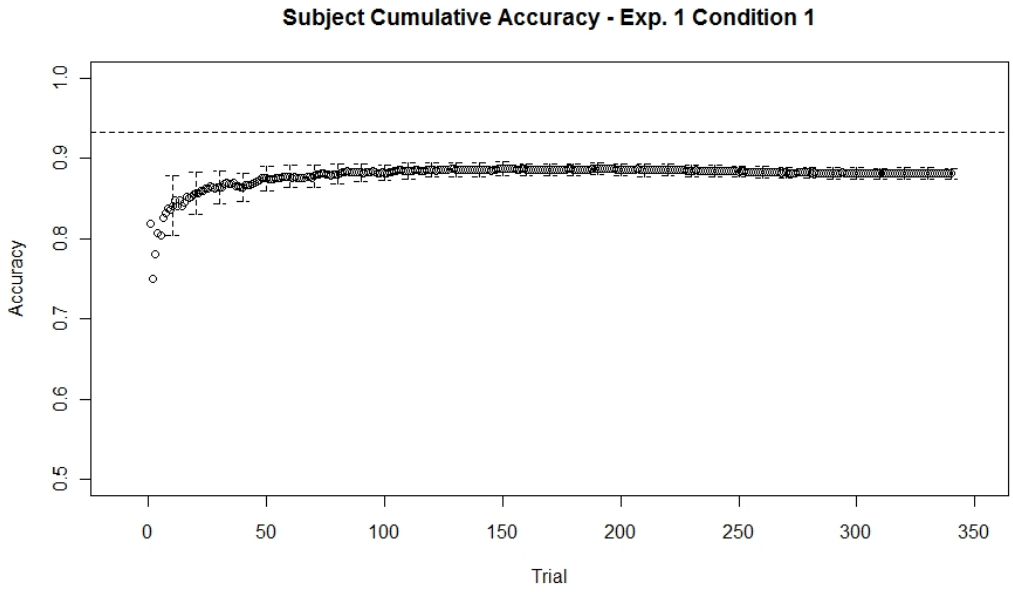


Figure 3.1: Cumulative Accuracy of Experiment 1 Condition 1 ($d' = 3$). This figure shows the average cumulative accuracy of the 44 subjects across trials. The error bars represent the 95% confidence interval for every tenth trial.

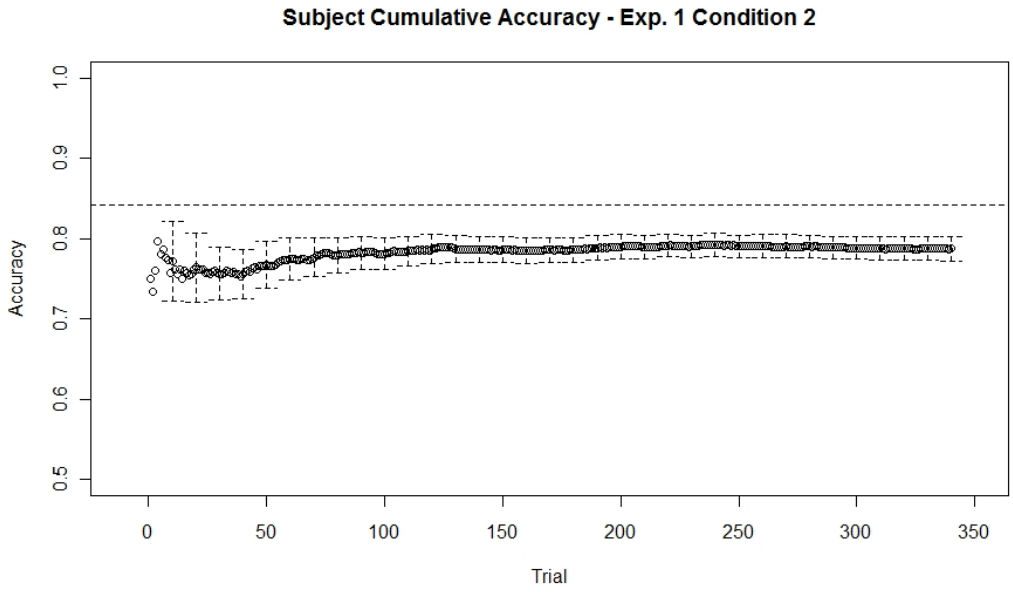


Figure 3.2: Cumulative Subject Accuracy of Experiment 1 Condition 2 ($d' = 2$). This figure shows the average cumulative accuracy of the 32 subjects across trials. The error bars represent the 95% confidence interval for every tenth trial.

assigned tasks. In all cases, the subjects chose one of two possible responses after a stimulus was presented.

3.2.1 Subjects

Sixteen subjects from the OSU undergraduate subject pool served in this experiment in exchange for course credit. All subjects reported normal or corrected-to-normal vision and normal basic motor skills. Naive to the purpose of the experiments but informed of their role, they were tested individually after receiving both oral and written instructions.

3.2.2 Apparatus

The apparatus was the same as that of Experiment 1.

3.2.3 Procedure

The second experiment was similar to the first, except the *a priori* stimulus probabilities shifted across several blocks of trials. These data allow us to model criterion shifts over time. We were also interested in the fidelity of the stimulus representations. The stimulus representation should change to approximate the sampling distributions after each change in the experimental parameters. Each "assay" was drawn from one of two normal distributions with means 40 and 60 (N and S distributions respectively; subjects were given the means of the distributions).

Subjects completed 5 blocks of 100 trials. For the first, third, and fifth block the probability of the "assay" being drawn from the S distribution was .5. For the second and fourth block the probability of the "assay" being drawn from the S distribution was .8.

3.2.4 Experimental Results

The purpose of this experiment (along with Simulation 8) is to show the sequential dependencies that the model predicts from the recursive nature of the construction of the stimulus representations. Simulation 8 showed that as the model gained more experience, the time (or number of trials) required to shift the criterion decreased. The model makes strong predictions for the adoption of and biased use of decision rules. From the results of Simulation 8, we should see a decrease in accuracy for each subject at the onset of each shift. This decrease in accuracy should decrease in magnitude as time progresses.

The mean accuracy for the subjects in the first block was 79.41%, 87.79%, 87.50%, 88.69% and 84.58% for Blocks 1, 2, 3, 4 and 5 respectively. These findings show that the accuracy is still trends upward and is consistent with the claim that the representations are approximating the sampling distributions as time progresses.

Figure 3.3 shows the moving accuracy of all the subjects in Experiment 2. The moving accuracy was calculated in the same way as Simulation 8. This technique gives the reader a view of the frequency of correct responses by the subjects over time. The vertical lines again represent the onset of the change in experimental parameters. Figure 3.3 shows that the subjects were more sensitive to the shift in experimental parameters than the model was. The first shift has little effect on the subjects, partly due the low accuracy prior to the initialization of the shift. The second shift clearly shows a long decrease in the plot, lasting 60 trials. The third shift is not as dramatic, a decrease in accuracy lasting only 36 trials. Finally, the fourth shift exhibits a decreasing accuracy for 25 trials. The subjects still show a steady increase in accuracy and are more variable in the beginning of the experiment

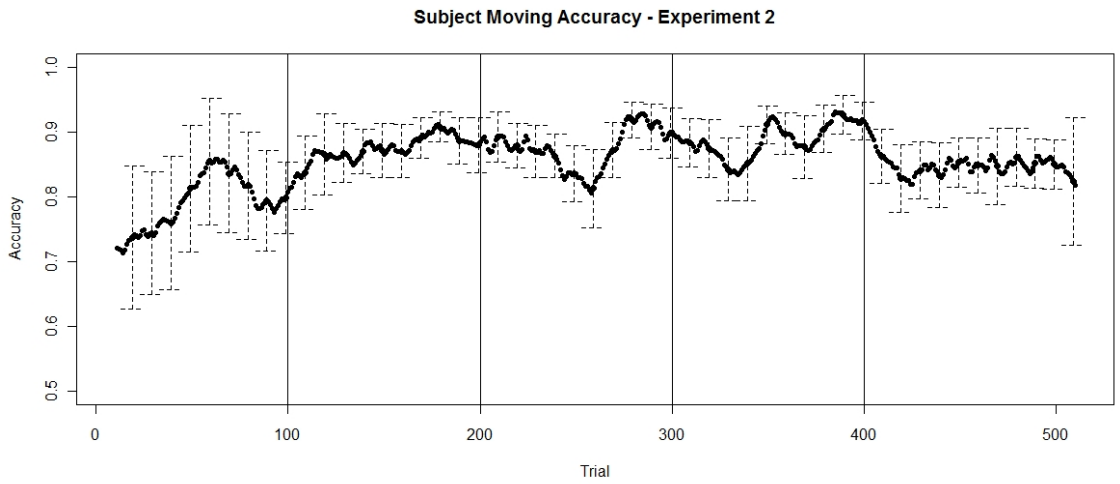


Figure 3.3: Moving Average of Subjects in Experiment 2. This figure shows the moving accuracy of all the subjects in Experiment 2. A single point represents the average accuracy from that point to 20 trials after that point. The vertical lines represent the onset of the calculations that include the new parameters. The error bars represent the 95% confidence interval for the subjects.

than they were in the later trials. Similar effects were found in the simulations and in the results of Experiment 1.

3.3 Summary

Figures 3.1 and 3.2 show the cumulative accuracies for Experiment 1. The final cumulative accuracy of the easy condition was 88.09% and of the hard condition was 75.75%. The maximum possible performance accuracy for these conditions was 93.32% and 84.13%, respectively. This result shows that while the subjects are not optimal, they perform reasonable well, and as d' increases, subjects' accuracy increases. This finding is consistent with many other published findings in the SDT literature (e.g. Egan, 1975, Green and Swets, 1966, Macmillan, 2002, Swets, 1996).

Both figures show that accuracy increases over time. These findings support the claim that as a subject gains more information (experience), the subject becomes more accurate over time. Both figures also indicate that while individual variability is high initially, as experience is gained, the variability decreases.

The second experiment shows that the subjects were more sensitive than the model to changes in stimulus probability. The subject's accuracy typically decreased following a shift in the experimental parameters. This shift decreased in magnitude as trials progressed. This indicates that the subjects were becoming less sensitive to the changes in probability. They behaved more variably in the beginning of the experiment than in later trials. The subjects also exhibited a slow progressing upward, which indicates that they were getting better at the task.

CHAPTER 4

DISCUSSION

The traditional SDT framework assumes the stimulus representations and parameters (d' and β) are fixed from the first trial of a detection task. One implication of parameters that are fixed over trials is that decisions are independent of previous decisions and their outcomes. However, we know that this is not true (e.g. Howarth and Bulmer, 1956, Treisman and Williams, 1984); subjects show sequential effects across trials in a detection task.

I proposed and tested a dynamic decision-making model which solves these two long-standing problems. First, it does not require *a priori* stimulus representations for signals and noise, but instead describes how these representations evolve over time. As a consequence of this evolution, this model is not dependent on a criterion location. Instead, the decision is based on the likelihood ratio. This allows the decision rule to shift over time to accommodate new information as it is presented. This in turn makes strong predictions concerning signal detection performance with changes in experience and with changes in the stimulus generating mechanism.

I first introduced the basic concepts of signal detection theory and kernel density estimation. Second, I presented the new model and justified its assumptions. Third, I presented the results of several simulations to demonstrate that the model behaves

appropriately and to generate novel predictions. Then, I presented the results from two experiments and showed that they justified the model's assumptions.

Experiment 1 showed effects of d' and β similar to the traditional SDT literature (e.g. Egan, 1975, Green and Swets, 1966, Macmillan, 2002, Swets, 1996). However, my analysis of the cumulative accuracies describe the role of experience on task performance. From Experiment 2 and Simulation 8, I have eliminated the need for a shifting criterion to account for unannounced changes in the presentations of the stimuli. Instead, the likelihood ratio at a single point is the ultimate decision rule by the observer. This experiment was designed to show that subjects do change their decision rules even when they are not instructed to do so. From Simulation 8, we see that the proposed model can show this result using the likelihood ratio as a decision rule based on the perception of the sensory effect of the available stimuli. This finding rejects the assumption of a fixed criterion location as a decision rule (e.g. Green and Swets, 1966, Treisman and Williams, 1984).

4.1 Model Comparison

Treisman and Williams (1984) investigated three models: a linear additive learning (LAL) model, an exponential additive learning (EAL) model, and an independent trace (IT) model. The LAL and EAL models shift the criterion relative to a reference criterion depending on the response on each trial. For the IT model, a memory trace resulting from the response decays toward zero with time. The criterion location at any point in time is a function of the summed memory traces. These types of adjustments make strong predictions concerning sequential dependencies. Treisman and Williams found significant positive sequential effects in all of their experiments

such that responses tended to be repeated. They suggest that the decision at hand is best determined by the subject's most recent responses weighted by their immediate relevance.

Treisman and Williams (1984) found significant positive sequential dependencies in all of their experiments and in the data provided by Howarth and Bulmer (1956). In contrast to Treisman and Williams (1984), using one set of parameter values, my model did not generate any sequential dependencies. The effect of sequential dependencies was not evident in either condition of my Experiment 1, nor were they in the simulations. Sequential dependencies were not significant in a simulation of the easy condition ($\chi^2(1) = 2.1401, p = .144$) or the hard condition ($\chi^2(1) = 3.6665, p = .055$) at the $\alpha = .05$ level. The effects of sequential dependencies for two trials back were also not significant. A possible explanation for these findings is the experimental design. The experiments in both Treisman and Williams (1984) and Howarth and Bulmer (1956) did not provide the subject with feedback. However, the experiments conducted here provided feedback after every trial.

Brown and Steyvers (2005) proposed a model of dynamic criterion adjustments in which two *a priori* stimulus representations determine decision-making performance. Additionally, each representation also has a fixed criterion. The model assumes that subjects continue to use a past criterion despite a change in the experimental parameters. In their experiments, the switch is from an easy task environment to a hard task environment. Eventually, the subject recognizes that the task has changed and uses the other decision-making criterion.

The key difference between my model and the model used by Brown and Steyvers (2005) is the rate in which the decision rule is modified. Brown and Steyvers (2005)

assume that subjects switch criteria immediately after recognizing a change in the experimental parameters. My model assumes that a change in the decision rule is not immediate. Because the decision rule in my model is based on the observed stimuli, the rate of change is dependent upon the number of stimuli that have been presented. As the number of trials increase, the model accumulates more stimuli and as a consequence, the rate at which a subject's internal representations adapt slows. This can be seen in Equations 1.4 and 1.5. Because the ratio $\frac{n-1}{n}$ converges to 1 and $\frac{1}{n}$ converges to 0 as $n \to \infty$, the weight on the previous representation eventually slows the adaptation of the model to new experimental parameters. The assumption in Brown and Steyvers (2005) that this change is immediate is not supported here. Subject accuracy does decrease and then increase after a change in experimental parameters, but this change is not immediate nor precise. Furthermore, as shown by Figure 3.3 in Experiment 2, subjects are less capable of acclimatizing to new decision rules as time progresses.

4.2 Future Directions

The question of how subjects adapt to changes in stimulus frequency was not originally part of this project. However, it seems that this concept is a fundamental difference between the three models considered here. Due to the precision of these models in the standard two-choice task, a more exotic type of stimulus context must be considered. To distinguish between these models, an experiment in which a decision rule that is enveloped inside another decision rule is necessary. Consider an experiment in which the signal distribution has a mean of 50 with a standard deviation of 5. Assume that the noise distribution also has a mean of 50, but has a

standard deviation of 20. This would lead to a non-monotonic decision rule. Values very near 50 would be most likely from the signal distribution, whereas the values further from 50 would be more likely from the noise distribution.

The recursive nature of my model allows for these types of frameworks after a small adjustment. A point must be distributed near the mean of the two distributions (50) in order for my model to adequately represent both the S and N distributions. If the representations are appropriate (i.e. have low sums of squares), then the model will assess the likelihood in the usual way without difficulty. However, the Treisman and Williams (1984) model could only succeed in this framework if there were two decision criteria. The Brown and Steyvers (2005) model could only succeed if there were four decision criteria. Given the difficulty that the subjects exhibited in Brown and Steyvers (2005) in the detection of a change in decision-making environments, this model might not be capable of modeling the accuracy of the subjects in this framework.

The main purpose of my thesis is to present a possible model that can be used in cognitive modeling in the yes-no decision task of SDT. The results of the experiments and the simulations have not presented any potential difficulties in the standard SDT task. The model has behaved very well and it is well-suited for human cognitive modeling.

APPENDIX A

ROOT MEAN SQUARES PLOTS

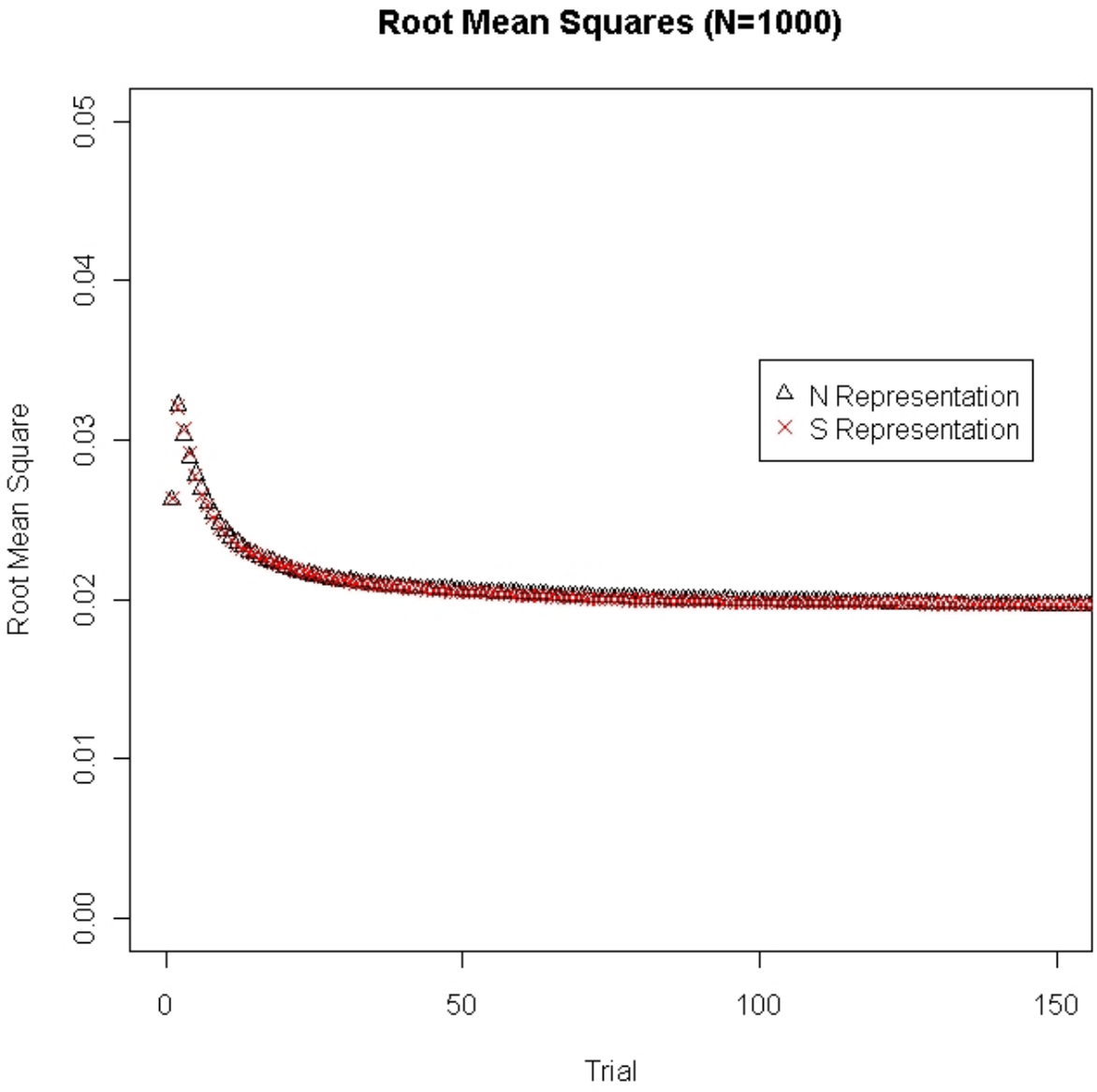


Figure A.1: Mean Sums of Squares for Simulation 1. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

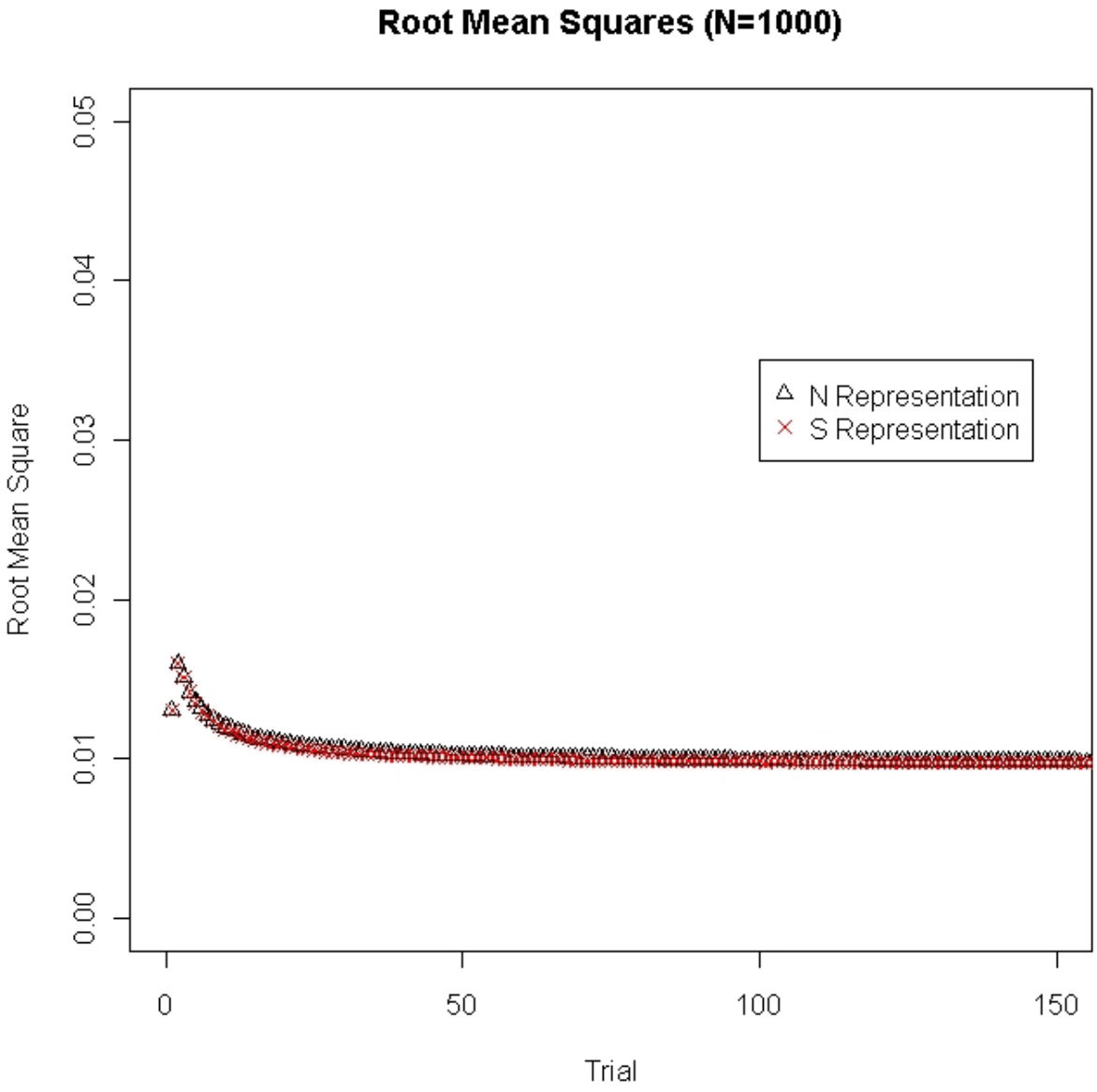


Figure A.2: Mean Sums of Squares for Simulation 2. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

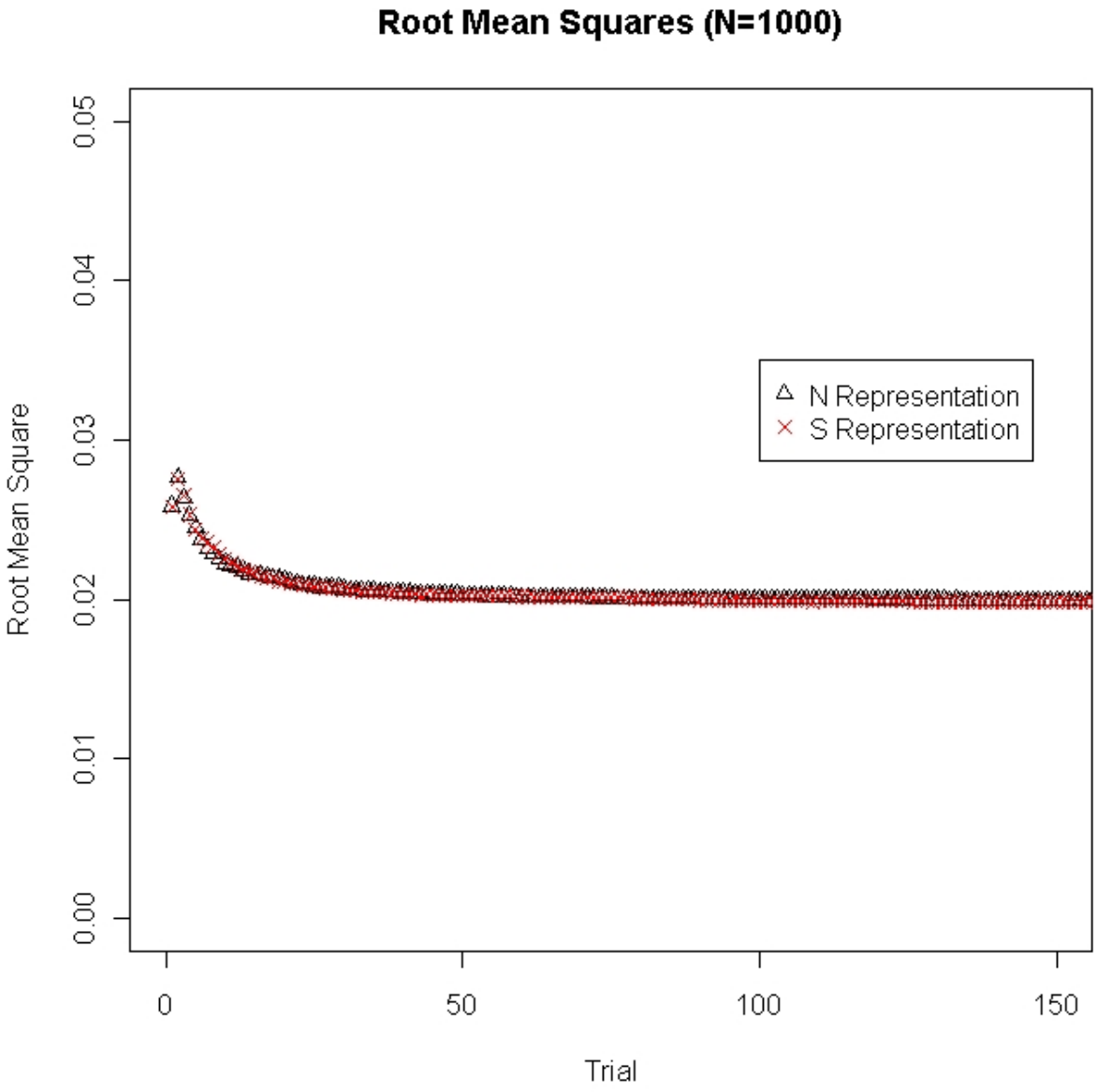


Figure A.3: Mean Sums of Squares for Simulation 3. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

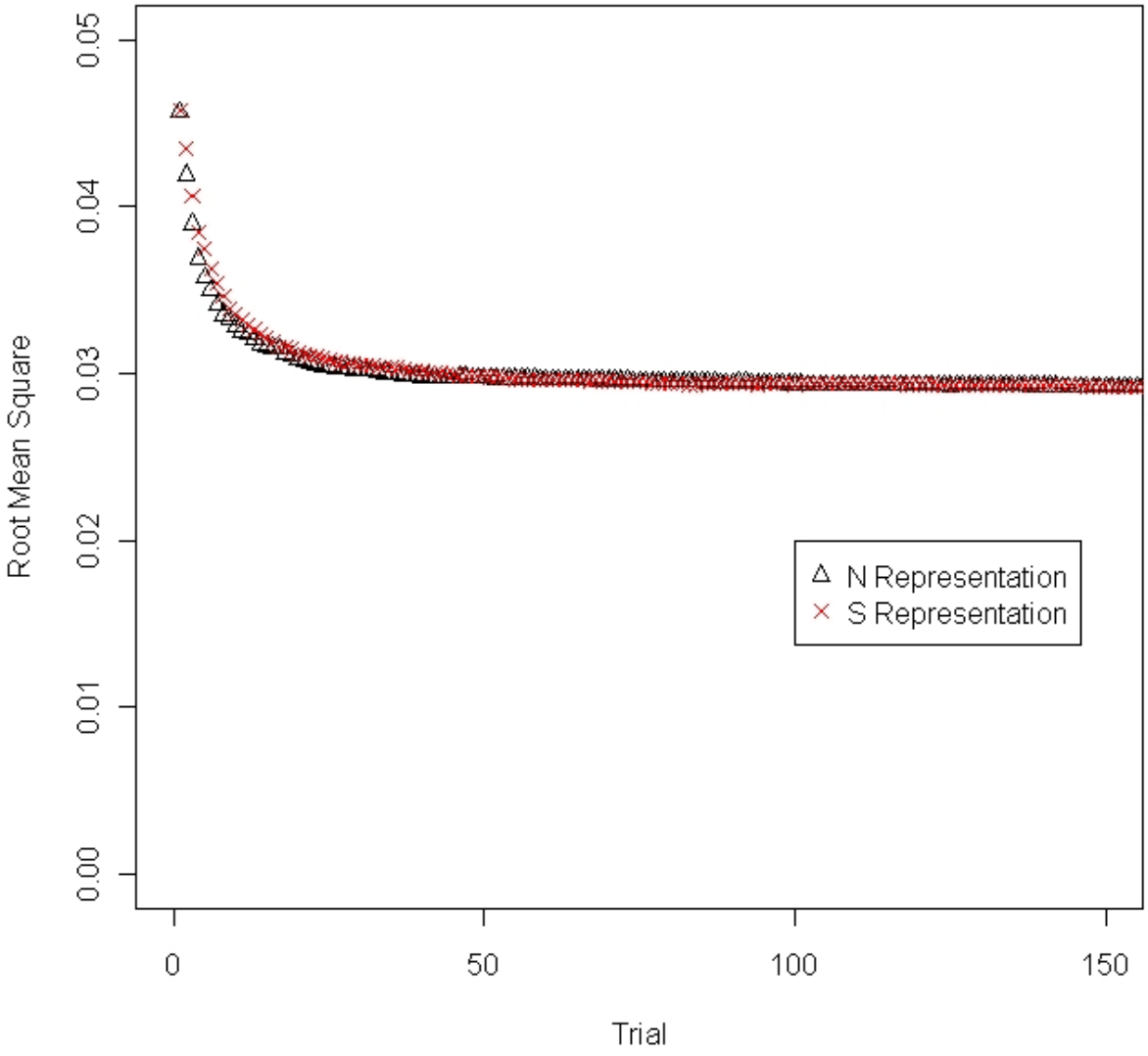


Figure A.4: Mean Sums of Squares for Simulation 4. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

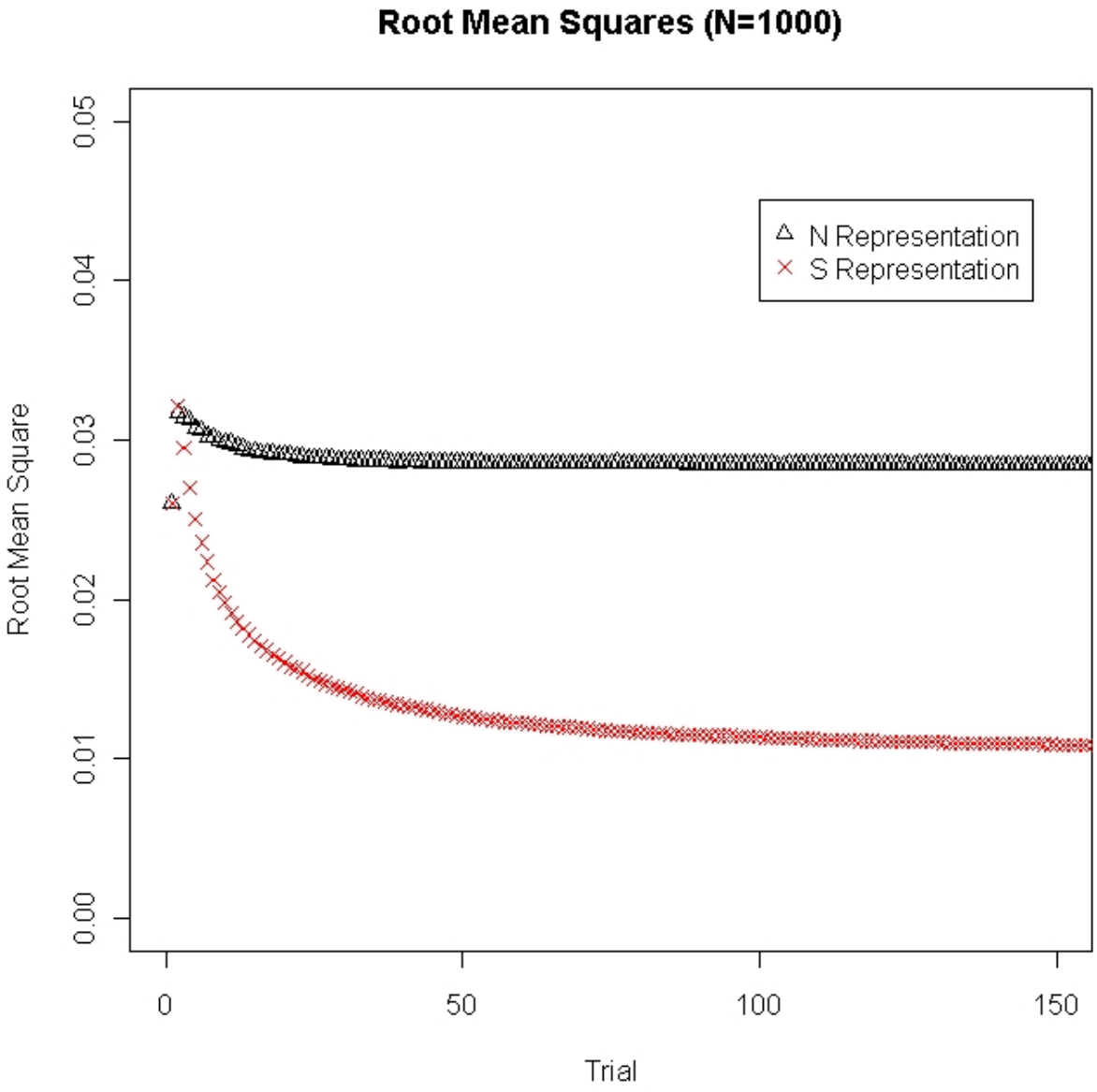


Figure A.5: Mean Sums of Squares for Simulation 5. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

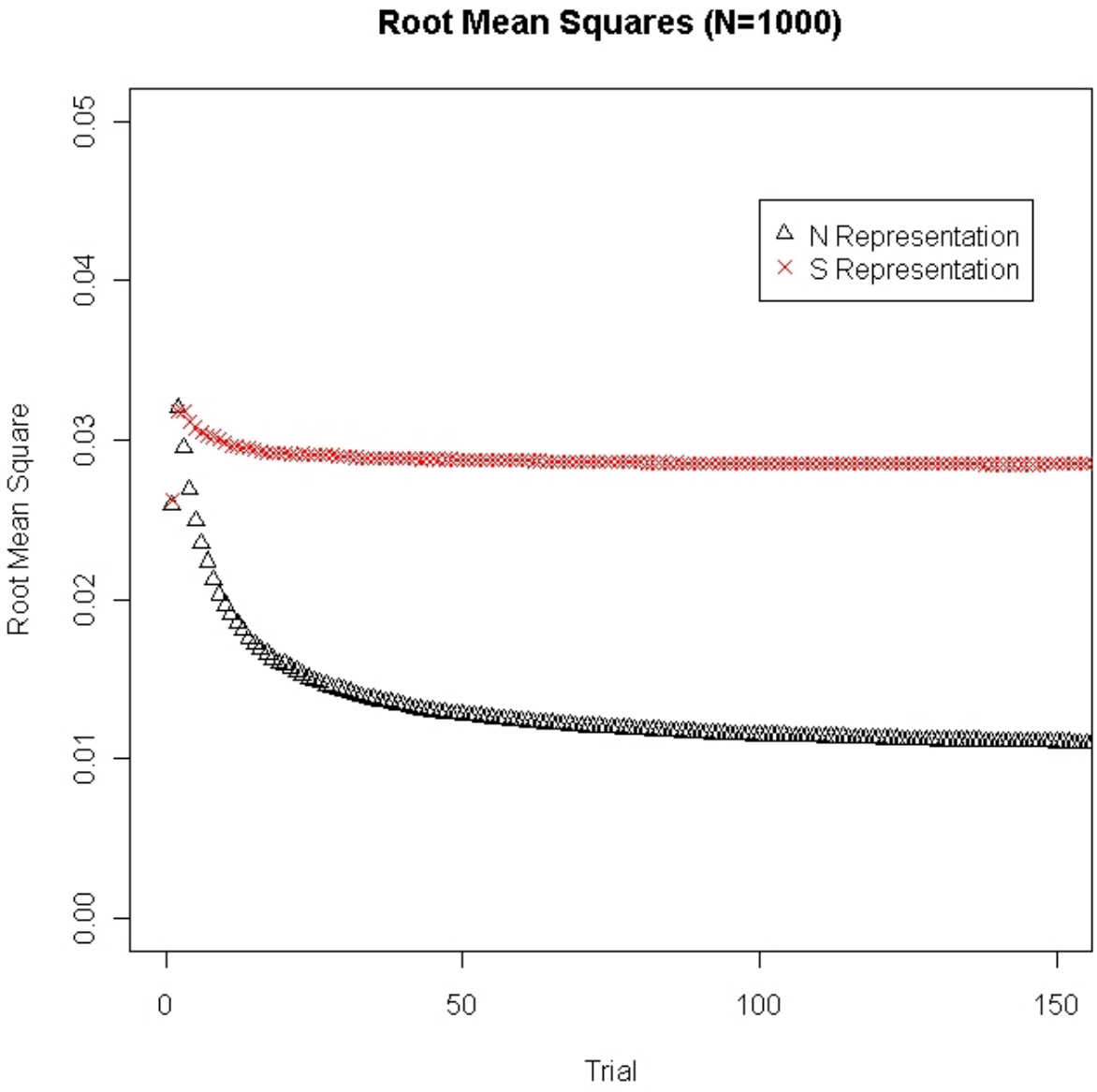


Figure A.6: Mean Sums of Squares for Simulation 6. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

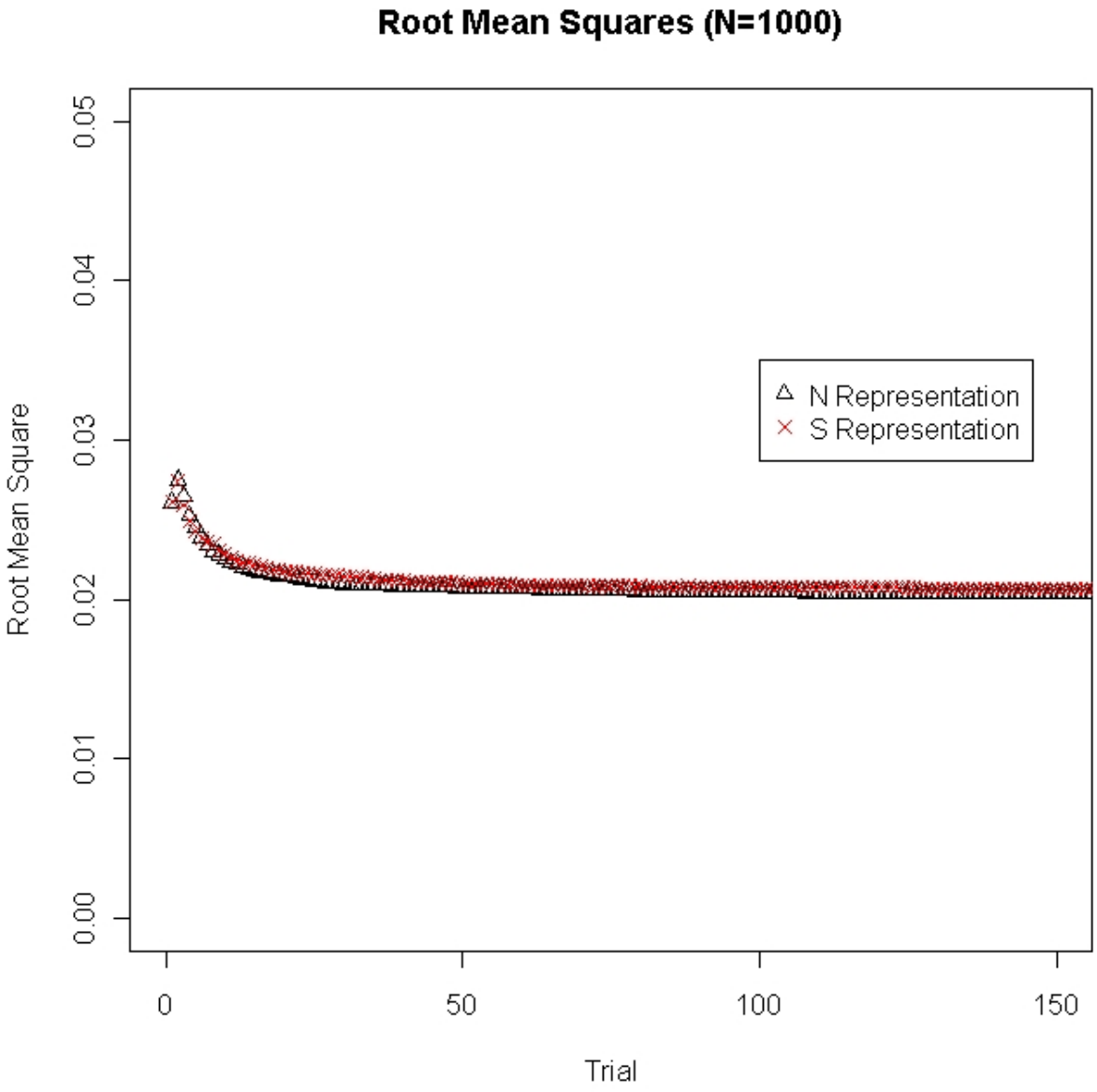


Figure A.7: Mean Sums of Squares for Simulation 7. The total sums of squares (averaged across 1000 simulations) is plotted for the N representation (triangles) and the S representation (x).

BIBLIOGRAPHY

J. Balakrishnan. Measures and interpretations of vigilance performance: Evidence against the detection criterion. *Human Factors*, 40:601–623, 1998a.

J. Balakrishnan. Some more sensitive measures of sensitivity and response bias. *Psychological Methods*, 3:68–90, 1998b.

J. Balakrishnan. Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance*, 25:1189–1206, 1999.

S. Brown and M. Steyvers. The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31:587–599, 2005.

S. Brown and M. Steyvers. Detecting and predicting changes. In Press, 2009.

B. de Silva and N. Mukhopadhyay. Kernel density estimation of wool fiber diameter. In S. Datta N. Mukhopadhyay and S. Chattopadhyay, editors, *Applied Sequential Methodologies: Real World Examples with Data Analysis*, pages 141–170. New York: Marcel Dekker, 2004.

D. Dorfman and M. Biderman. A learning model for a continuum of sensory states. *Journal of Mathematical Psychology*, 8:264–284, 1971.

D. Dorfman, C. Saslow, and J. Simpson. Learning models for a continuum of sensory states reexamined. *Journal of Mathematical Psychology*, 12:178–211, 1975.

J. Egan. *Signal Detection Theory and ROC Analysis*. New York: Academic Press, 1975.

D. Green and J. Swets. *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

J. Grier. Nonparametric indices for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 98:185–199, 1971.

A. Healy and M. Kubovy. The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology*, 106:427–446, 1977.

C. Howarth and M. Bulmer. Non-random sequences in visual threshold experiments. *The Quarterly Journal of Experimental Psychology*, 8:163–171, 1956.

M. Kac. A note on learning signal detection. *IRE Transactions on Information Theory*, pages 126–128, 1962.

M. Kac. Some mathematical models in science. *Science*, 166:695–699, 1969.

N. Macmillan. Signal detection theory. In H. Pashler and J. Wixted, editors, *Stevens' Handbook of Experimental Psychology: Methodology in Experimental Psychology*, pages 43–90. John Wiley and Sons: New York, 3 edition, 2002.

R. Pastore, E. Crawley, M. Berens, and M. Skelly. 2003.

I. Pollack and D. Norman. A non-parametric analysis of recognition experiments. *Psychonomic Science.*, 1:125–126, 1964.

P. Rabbitt. *Sequential Reactions. In D. H. Holding (Ed.).* London: Wiley, 1981.

B. Silverman. *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1986.

J. Snodgrass and J. Corwin. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117: 34–50, 1988.

J. Swets. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics*. Mahwah: Lawerence Erlbaum, 1996.

C. Taylor. Bootstrap choice of the smoothing assumption in kernel density estimation. *Biometrika*, 1989.

E. Thomas. On a class of additive learning models: Error-correcting and probability matching. *Journal of Mathematical Psychology*, 10:241–264, 1973.

M. Treisman. Is signal detection theory fundamentally flawed? a response to balakrishnan (1998a, 1998b, 1999). *Psychonomic Bulletin and Review*, 9:845–857, 2002.

M. Treisman and T. Williams. A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, 91:68–111, 1984.

D. Vickers and M. Lee. Dynamic models of simple judgments: i. properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences*, 1998.

D. Vickers and M. Lee. Dynamic models of simple judgements: ii. properties of a self-organizing pagan (parallel, adaptive, generalized accumulator network) model for mulit-choice tasks. *Nonlinear Dynamics, Psychology, and Life Sciences*, 2000.

M. Wand and M. Jones. *Kernel Smoothing*. Florida: Chapman and Hall, 1995.

T. D. Wickens. *Elementary signal detection theory*. New York: Oxford University Press, 2002.