UNIVERSITY OF CALIFORNIA

SANTA BARBARA

Group Decision-Making Models for Cybernetic Systems

A Dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy
in Mechanical Engineering

by

Margot Kimura

Committee in charge:
Professor Jeffrey Moehlis, Chair
Professor Miguel Eckstein
Professor Frédéric Gibou
Professor Linda Petzold

September 2011

UMI Number: 3481996

UMI

Dissertation Publishing

UMI 3481996

ProQuest®

The dissertation of Margot Kimura is approved:

_____

Professor Miguel Eckstein

_____

Professor Frédéric Gibou

_____

Professor Linda Petzold

_____

Professor Jeffrey Moehlis, Committee Chair

September 2011

Group Decision-Making Models for Cybernetic Systems

iii

To Dad

Acknowledgments

I would like to acknowledge my committee for their time and care in reading this dissertation, and their support during my graduate study. I'd like to thank my advisor, Jeff, for introducing me to the SPRT, for his frankness, and for being so giving of his time. A heartfelt thank you goes out to my IGERT co-advisor, Linda, for her support and advice when I needed it most. I am very grateful to Miguel for mentoring me on the Psychophysics-based section of my dissertation and adopting me into his group, even though I technically was not one of his advisees. Thank you to Frédéric for his advice and for being a wonderful teacher. I am also grateful to Dr. Binh Pham for providing me with the data I used in Chapter 4, and the Eckstein Group for sharing their research in Vision, Cognitive Psychology, and the Brain Sciences with me.

I am extremely grateful to the National Science Foundation and the University of California at Santa Barbara Graduate Division for the IGERT and Doctoral Scholars Fellowships (respectively) that funded the bulk of my doctoral research. I am also very grateful to the Hawaii Community Foundation, particularly the Marion Maccarrell Scott Foundation and the William & Dorothy Bading Lanquist Fund, and the Mechanical Engineering Department for their generous scholarships. Your financial support allowed me to be more independent in my studies and ultimately helped me mature as a researcher in a way that nothing else could. My thanks also goes out to the administrative staff that helped make everything possible, especially

Roxanne, Stephanie, Jenny, and Laura.

Thank you to my labmates for your support, friendship, and understanding. Though we don't all work for the same professor, I've come to think of the DCR lab as one research group because we're always together. Thank you to my friends in the UCSB Aikido and Judo Clubs, Ota-sensei's ballroom dance class, and Sylvia's Lindy hop classes for helping me with stress relief and keeping things fun.

Finally, I am very grateful to my friends and family for their support. It is ironic that the most difficult parts of my graduate experience had nothing to do with the technical aspects of this dissertation, and I'm not sure I would have made it through without your support. I am particularly grateful to Euan, Joanne, Audrey, Sue Ann, Haomiao, Haruna, and Greta. A special thank you goes out to my Dad, for reminding me that there are many interesting things in life outside of school and that I should always keep things in perspective.

<center>

Curriculum Vitae

Margot C. Kimura
September 2011

</center>

Education

September 2011	Ph.D. in Mechanical Engineering, with an Emphasis in Computational Science & Engineering University of California, Santa Barbara
September 2007	M.S. in Mechanical Engineering University of California, Santa Barbara
June 2005	B.S. in Mechanical Engineering California Institute of Technology

Professional Experience

2006–2011	Graduate Student Researcher University of California, Santa Barbara
2005–2010	Teaching Assistant University of California, Santa Barbara
2003–2005	Undergraduate Student Researcher/SURF Fellow Jet Propulsion Laboratory
2004–2005	Teaching Assistant California Institute of Technology
Summer 2002	Summer Support, Sophomore Level Jet Propulsion Laboratory

Publications

- M. Kimura and J. Moehlis. "Group Decision-Making Models for Sequential Tasks." *Accepted by the SIAM Review.*

<center>

</center>

- M. Kimura and J. Moehlis. "Novel Vehicular Trajectories for Collective Motion from Coupled Oscillator Steering." *SIADS*, Vol. 7, Iss. 4, pp. 1191-1212, 2008.

- M. Kimura and J. Moehlis. "Novel Vehicular Trajectories for Collective Motion from Coupled Oscillator Steering." *M.S. Thesis*, September 2007.

- S. Zhang, M. Kimura and M. Eckstein. "Group Decision-Making." *In preparation.*

Presentations

- M. Kimura. "Group Decision-Making Models for Cybernetic Systems." *UCSB Ph.D. Dissertation Defense*, 8 February 2011.

- M. Kimura. "Analysis of Centralized Sequential Probability Ratio Test (SPRT)-Based Models for Group Decision-Making." *SIAM Annual Conference: Contributed Presentation*, 14 July 2010.

- M. Kimura. "Hierarchical Multi-Level Decision-Making for Cybernetic Systems." *UCSB Candidacy Exam Presentation*, 4 June 2008.

- M. Kimura. "Multi-Level Decision-Making for Cybernetic Systems." *UCSB NSF IGERT Seminar*, 3 June 2008.

- M. Kimura. "Novel Vehicular Trajectories for Collective Motion from Coupled Oscillator Steering Control." *SIAM Conference on Applications of Dynamical Systems: Mini-Symposium Presentation*, 31 May 2007.

- M. Kimura and A. Kolpas. "Coupled Oscillators, Collective Motion, and Synchrony." *UCSB Ecology Seminar*, 26 February 2007.

Awards and Honors

2011	Research featured in UCSB's NSF IGERT Annual Report
2010	University of California Doctoral Travel Grant
2005–2010	**University of California, Santa Barbara** **Doctoral Scholars Fellowship**

2007–2009 National Science Foundation
Integrative Graduate Education and Research
Traineeship (IGERT) Fellowship
in Computational Science & Engineering

2002–2009 Marion Maccarrell Scott Scholarship

2007 Mathematical Sciences Research Institute
Dynamical Systems Workshop Scholarship

2005–2007 William James & Dorothy Bading Lanquist Scholarship

2006 Grace Hopper Celebration Conference Scholarship

2006 Feedback and Dynamics in Nature Workshop Scholarship

2005 **University of California, Santa Barbara
ME Department Merit Scholarship**

2005 National Science Foundation Fellowship Finalist

2003–2004 **California Institute of Technology
Summer Undergraduate Research Fellowship**

2001–2005 Donald Beall Scholarship

2001 Cal Kawamoto Valedictorian Award

2001 Mamoru and Aiko Takitani Scholarship

2001 Hawaii Rotary Youth Foundation Scholarship

2001 Pearl City Lions Scholarship

2001 Shiela Nonaka Ellis Memorial Scholarship

2001 Gentry Neighborhood Fund Scholarship

2001 Mufi Hanneman Female Scholar-Athlete Award

2001 Sterling Scholar Music Category 2nd Runner-up

Abstract

Group Decision-Making Models for Cybernetic Systems

by

Margot Kimura

From law-making to the research publication process, we rely on groups to make our most important decisions. The purpose of this dissertation is to gain a better understanding of the performance of a group of decision-makers (DMs) in a simple decision-making task.

We begin the first part of this dissertation by analyzing the Sequential Probability Ratio Test (SPRT), which is optimal for choosing between two hypotheses, and models an individual DM in a two-alternative forced-choice task. We show its equivalence to a simple form of the Drift-Diffusion Model (DDM), and solve the resulting DDM for the performance of an observer using the SPRT. We verify our analytical results with simulation, and investigate how our simulations are related to various approximations.

We then use the individual model to develop intuitive mathematical models for a group performing the same task. Previous works in this area have dismissed a full characterization of the group's decision time as being too complicated, focused on simplified cases such as assuming that the group's members are identical, and/or

considered only mean values. We find explicit general solutions for the group decision rule under two previously proposed group rules (Race and Majority Total schemes) and a novel rule, the Majority First scheme. We then generalize the Majority Total scheme to the η-Total rule and the Majority First scheme to the η-First rule, and provide explicit solutions for the performance of a group under either family of rules. Most notably, our models are simple enough to solve completely and flexible enough to accommodate numerous hierarchical group topologies and related group rules. In addition, our models can be applied to groups of devices, groups of human or monkey observers, or cybernetic groups, with adequate experimental manipulation.

In the second part of this dissertation, we analyze data from a human group-based signal detection experiment. We first compare the performance of the individual observers and the group, then analyze eleven different group rules that can be used to aggregate the individuals' responses, to find which are optimal and which the observers actually used. We also investigate some common assumptions through simulation. These results are relevant to applications involving a visual task, such as cancer detection in mammograms or threat detection in baggage screening.

Our models and analysis aim to inform designing cybernetic decision-making systems. Our goal is to establish a way to objectively and quantitatively compare different group decision rules and model group performance in an intuitive manner that is accessible to a wide range of communities.

Contents

List of Figures

Chapter 1

Background

1.1 Introduction

Group decision-making is highly relevant to a wide range of subjects: from law-making to the research publication process, we rely on groups to make our most important decisions. As a result, group decision-making is studied in many fields under very different applications. In different contexts, "group decision-making" could refer to anything ranging from group motion coordination and pattern formation [24, 27], to determining the optimal (legislative or board) committee size [35, 47], to distributed detection (using sensors) [12], to reaching a consensus in animals and insects [22, 61], to optimal stopping [53], to studying the ability of groups of humans to combine information [102], to probabilistic search [20]. Each area of study focuses on its own problems and applications of interest, has its own

philosophy for approaching and defining a problem, and set of strategies for finding a solution. Our goal is to gain a better understanding of the performance of a group in a simple decision-making task. To this end, we have explored a number of different approaches and philosophies relevant to decision-making, and arrived at the subjects discussed in this dissertation.

Though we will present relevant background information where appropriate, either within the text or through a reference to the Appendix, we begin with a general introduction to basic concepts that are useful to all works presented in this dissertation. We intend to provide enough detail for one to understand the gist of some of the different philosophies that exist in areas related to decision-making. In Section 1.2, we discuss basic ideas from statistics, focusing on highlighting the differences between some common methods and philosophies. In Section 1.3, we introduce the Likelihood Ratio Test (LRT), which is the optimal fixed-sample test. We finish with Section 1.4, where we define some decision-making tasks from the psychophysical literature and describe the general type of experiment that we will consider, as well as discuss some basic concepts from Signal Detection Theory and psychophysics, which will be relevant to our models and data.

In Chapter 2, we introduce the Sequential Probability Ratio Test (SPRT), the optimal test for choosing between two hypotheses using the minimal (average) number of samples, which we will use to model an individual decision-maker (DM). We first discuss its relevance to decision-making in human and monkey observers, and

mention its historical uses. We then show its equivalence to a simple form of the Drift-Diffusion Model (DDM), and solve the resulting DDM to find the performance of an observer using the SPRT (or DDM). We verify our results with simulation, and discuss how the simulation is affected by various parameter values and approximations taken in the model.

We continue this style of analysis in Chapter 3, where we present some intuitive mathematical models for group decision-making in a sequential task, inspired by both human observer-based and device-based research. Most notably, our models are simple enough to solve completely and flexible enough to accommodate numerous hierarchical group topologies and related group rules. We derive explicit solutions for a group of individual DMs using one of three simple group rules: Race, Majority Total, and Majority First. We demonstrate the flexibility of our group models with an example, where we analytically derive the performance of a non-uniform group, and verify our results through simulation. We also discuss other group decision-making models from the literature where appropriate. Finally, we present two general families of group rules, the η-Total and η-First schemes, and provide explicit solutions for each. Our models are formulated with the understanding that they can be applied to groups of device-based DMs, human or monkey observers, or a group comprised of both, with adequate experimental manipulation. We provide evidence from various literatures to support this notion.

In Chapter 4, we explore a different facet of a very similar problem: we analyze

data from an experiment with a group of human observers using a Signal Detection Theory (SDT)-based approach. We show the performance of the individual observers and groups, discuss the efficiency of the group, and determine which decision aggregation rules are equivalent to the strategy that the group actually used. Our main interests lie in determining what rule(s) are optimal and what rule(s) the groups likely used. We also test some commonly-held assumptions on the "Ideal Group" through simulation.

We finish with Chapter 5, where we offer some concluding remarks.

1.2 Some Basic Ideas From Statistics

The following is a brief overview of some basic ideas from Statistics and Statistical Hypothesis Testing which will be useful for understanding the analysis presented later. We begin with a comparison between the Bayesian and Frequentist philosophies to illustrate why we consider Bayesian Statistics to be more appropriate for decision-making tasks. We then cover the differences between fixed-sample and sequential methods, and finish with a statement and proof of Bayes' theorem.

1.2.1 Bayesian vs. Frequentist Statistics

Two popular schools of thought in statistics are Bayesian Statistics, and Frequentist Statistics. The following will very briefly and somewhat biasedly highlight the philosophy behind each, then present an example which illustrates why Bayesian

statistics is an appropriate choice for the problems considered in this dissertation.[1]

Frequentist Statistics

Frequentists believe that the "probability" of an event should be interpreted as the long-run relative frequency of the event when the trial is repeated under "similar" conditions in a population. Also, parameters that are numerical characteristics of the population are considered fixed but unknown. Given this, probabilities between 0 and 1 are only allowed for truly random quantities – since any unknown parameters in a population are actually fixed, it is meaningless to make a probability statement about their value. To construct a probability statement about a random variable, the Frequentist draws a sample from the population and calculates the sample's statistic. This process is repeated until a probability distribution of a sample's statistic over all possible random samples from the population is determined. The resulting probability distribution is called the sampling distribution of the statistic. This distribution is then used to determine a confidence statement about the parameter in question. Thus, the confidence is based on the average behavior of the procedure under all possible samples [15]. For example, when estimating a random variable in the Frequentist approach, a 95% confidence interval is defined as an interval that is one result of a procedure that has a 95% chance of producing an interval that contains the parameter one is trying to estimate.

---

[1]The material in Section 1.2.1 is based on lectures by Professor John Hsu, from the Statistics and Applied Probability Department at the University of California, Santa Barbara, unless stated otherwise.

Frequentists also believe that events that happened in the past are no longer random, so it is meaningless to discuss their probabilities. It is also worthwhile to note that the Frequentist approach uses different statistical tests for different properties and situations.

Bayesian Statistics

In contrast, Bayesians believe that the "probability" of an event is a measure of a person's degree of belief that the event will occur. Therefore, probability is necessarily subjective and the calculated probability of an event may vary from person to person. Unknown parameters of a population are treated as random variables, and the rules of probability are used to directly make inferences about the parameters.

Bayesians use a "prior" probability when calculating the posterior (or a posteriori) probability of an event supported by data. The prior is the probability that the Bayesian would use if no data was available to calculate the posterior probability. After formulating a prior probability distribution on the possible parameter values, the Bayesian then takes samples of data from the population and uses these values to revise his or her beliefs about the parameters. In other words, Bayesian analysis uses the actual samples of data that occurred in calculating probability, not all possible data sets which could have occurred (but did not). The combination of the prior probabilities and the probabilities calculated from the data gives the

posterior probability distribution, which represents the weights one gives to each parameter value after analyzing the data [15]. Thus, Bayesian statistics provides a consistent way to model how beliefs about parameters in the data are modified through data analysis.

When estimating a random variable in the Bayesian approach, a 95% credible set (or credible interval or posterior probability interval) is an interval in which the Bayesian is 95% certain contains the parameter. This is more intuitive, and in this sense, Bayesian statistics are predictive. The Bayesian approach to statistics is centered around Bayes' Theorem, which will be discussed later.

Disadvantages and Criticisms of Bayesian Statistics

The Bayesian approach is conceptually very simple, but in practice, it is difficult to write a closed form expression for the posterior except in simple cases, such as when one has normal data samples with a normal prior. In other cases, one must integrate the data numerically, which rapidly becomes more difficult as the number of parameters in the system increases [15].

One heavily criticized aspect of Bayesian statistics stems from its use of a prior probability. The general argument is that there is no absolutely correct way to define the prior for any given experiment, which subsequently makes the result (the posterior) fundamentally flawed.

While this is a valid criticism, it should be noted that the effect of the prior

on the posterior decreases rapidly as more data as collected; therefore, the posterior only relies on the prior when there is an insufficient amount of data available. Bayesians also maintain that any prior probability used in a good Bayesian analysis must be sufficiently justified. In cases where a good prior is difficult to find, it is acceptable to use an uninformative prior, such as a uniform prior, to avoid unnecessarily biasing the posterior. These two considerations noticeably weaken the Frequentists' arguments that the use of a prior invalidates the results found using Bayesian methods.

Advantages of Bayesian Statistics

While it can be argued that Frequentist approaches are more "objective", it can also be argued that that objectivity comes at the price of disregarding any prior knowledge about the process being measured. Since it is common for there to be prior knowledge available about the process, throwing out the prior may be wasteful and unnecessary. In this sense, using a prior probability in Bayesian statistics is an advantage [15].

Another advantage of the Bayesian approach is that it allows one to make direct probability statements about parameters, which provides an intuition significantly more useful than the abstract confidence statements provided by Frequentist arguments. This is actually very important: even though a statistician understands the difference between the two, a client who has hired the statistician will often

interpret a Frequentist confidence interval as a Bayesian credible set. In many situations, is also more important to consider intervals that are based on data which actually occurred, rather than data which could have occurred but did not [15].

Bayesian statistics is conceptually very simple: it is based around a single statistical tool: Bayes' Theorem. In contrast, Frequentist analysis requires many different statistical techniques. Bayes' theorem also provides a way to find the predictive distribution of future observations. This is not always easily done in the Frequentist framework.

Bayesian methods also have a more straightforward way of dealing with "nuisance parameters". These are parameters which one is not interested in, but one must account for because they affect the analysis of the parameters of interest. In the Bayesian approach, these parameters are always marginalized out of the joint posterior distribution [15].

Lastly, and perhaps most significantly, it has been claimed that Bayesian methods often outperform Frequentist methods, even when judged by Frequentist criteria [15]. This source also claims that (relatively) recent improvements in computation has further progressed the Bayesian approach, by providing a means by which one can draw an (approximate) random sample from the posterior distribution without having to completely evaluate it, via the Gibbs Sample or Metropolis-Hasting algorithms. In this way, one can approximate the posterior distribution to any desired accuracy by taking a large enough random sample from it, which removes

a significant difficulty in Bayesian statistics – now one can use Bayesian statistics in practice for problems with many parameters, as well as for distributions from general samples and from general priors. For more details, see [15].

A Caveat: The Dangers of Misunderstanding Statistics

A humorous example illustrating the dangers of misunderstanding the difference between Frequentist and Bayesian statistics is shown in Figure 1.1. The hikers calculate the probability of an American being killed by lightning using a Frequentist-like approach, then interpret the resulting value in a Bayesian manner, thinking that the probability calculated over the entire American population applies to them individually. They neglect to use the calculated prior with additional data (as a Bayesian would), or to use a proper population for their statistic (as a Frequentist would), and as a result, they neglect to take into account the fact that being outside in a lightning storm is a significantly higher risk factor than being American when calculating the probability of death by lightning.

An Argument For Using Bayesian Statistics in Decision-Making

Frequentist and Bayesian approaches often return the same result, and there are diabolical examples which disprove one approach or the other; however, this dissertation focuses on the performance of groups in a decision-making task. Since decision-making is necessarily a subjective process, we will be utilizing the Bayesian approach to statistics. A standard (but biased) example is provided below to il-
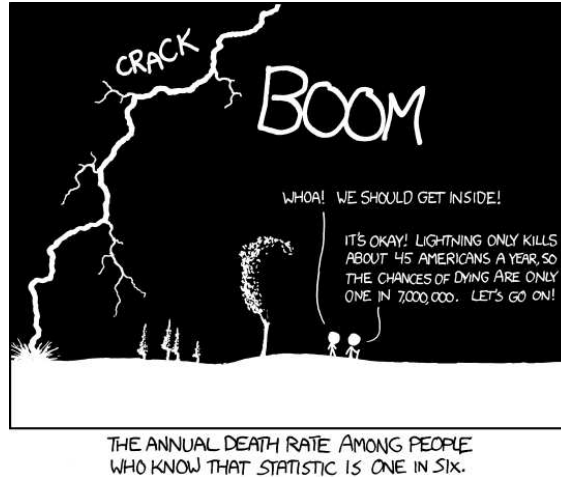
Figure 1.1: An humorous example of the dangers of misunderstanding statistics. Reprinted with permission from xkcd.com.

lustrate why it makes sense to use a Bayesian approach to statistics in a decision-making task.

A Biased Example   Suppose Joe sees an advertisement in the newspaper for free HIV testing at a local hospital and he decides to take advantage of the offer. He is told that the test is 90% reliable, in the following sense: if a person has HIV, the probability that the test will be positive is 0.9, and if the person does not have the disease, there is a 0.1 probability that the test will give a positive response (false positive). Now, suppose his test returns positive. Should he be worried? What is the probability that he actually has HIV?

The Bayesian approach is relatively intuitive. If Joe's personal medical history is unknown, a Bayesian would use a non-informative prior, such as the probability that a randomly selected person in the United States has HIV, to calculate the

posterior probability that Joe has HIV, given the results of the medical test. If that fraction is $\frac{1}{10,000}$, then the Bayesian would use Bayes' Theorem to conclude that Joe has a 0.0008993 probability of having HIV, given the positive test result. Therefore, Joe should not be too worried yet, but he should take a second test to be sure. Subsequent tests provide more information and update the posterior probability that Joe has HIV. If Joe's personal medical history is known, it is taken into account through the prior, which may increase or decrease the previously-calculated posterior probability.

On the other hand, a Frequentist would say that trying to assign a probability between 0 and 1 is meaningless. Joe either does or does not currently have HIV, so the probability is either 0 or 1, no matter how many tests he takes. It would not make sense to have any other probability, because probabilities between 0 and 1 represent long-run relative frequencies in repeated trials, which is not possible for a single person.

While a probability calculated from averaging over a population of people "similar" to Joe may provide a good general statistic for the population, it does not tell Joe what he wants to know: the probability that **he** has HIV given the results of the test. Thus, despite that there is logic to the Frequentist's analysis, it clearly does not help Joe make a decision. One could argue that a Bayesian analysis is prone to error because it uses a prior; however, we note that Joe can take repeated observations (i.e., further tests) to decrease the effect of the non-informative prior

and thereby arrive at a more accurate posterior probability. The above example can also be applied to other realistic applications, such as trying to detect a contaminant or intruder, and quality control. We rest our case for using a Bayesian approach in our work here, and move on to another distinctive divide in statistics below.

1.2.2 Sequential vs. Fixed-Sample Methods

In addition to different philosophies about what probability means, there are also different ways to relate statistical analysis to the experiment that provides the data. Statistical analysis is often thought of as a process that begins once all of the data has been collected; however, this is true only for fixed-sample methods. Below we contrast these fixed-sample tests with sequential methods. For an in-depth discussion, see [38].

Fixed-Sample Methods

Fixed-sample statistical methods are very commonly used, and widely known. The name comes from the fact that the statistical analysis is performed on a data set of fixed size, which means that the experiment that produces the data finishes before the statistical analysis begins. These methods include ways to estimate the number of trials required to reach a given level of certainty in the results, and are typically favored because they are usually conceptually and mathematically simpler

13

to perform, and can be done after the experiment is over. This makes experimental design convenient as well, since the experimenter can know beforehand how many samples or trials are required, run them, go home, and then do the statistical analysis at her leisure.

Fixed-sample methods are most appropriate for applications in which the data is intrinsically fixed and finite in size and time. An example is breast cancer detection: for a given patient, there usually is a fixed set of mammograms for the doctors to examine in order to decide whether or not the patient should get a biopsy. The x-rays are all taken in one appointment, and it does not matter to the doctor which order the technician took the images or which order the doctor analyzes the different views, since they are meant to be 2-dimensional representations of a single snapshot (in time) of a 3-dimensional image.

However, there are also applications that are intrinsically sequential in nature because they produce observations in a spatial and/or temporal order. For these experiments, fixed-sample methods are not optimal, and one may need to consider sequential methods.

Sequential Methods

Sequential methods correspond to experiments that integrate statistical processing into the procedure to determine the stopping condition. These methods may be more complicated to calculate, and make the experiment more difficult to

design and perform because it is not known beforehand how many trials or data samples are required. While sequential methods may increase the complexity of the experiment, they often reach conclusions with the same certainty as a fixed-sample method in a fewer number of trials on average. Additionally, sequential methods sometimes can provide a solution where fixed-sample procedures fail. For further details, see [38].

1.2.3   Introduction to Bayesian Statistics

Bayes' Theorem is attributed to Reverend Thomas Bayes: the theorem was first written up in "An Essay Towards Solving a Problem in the Doctrine of Chances", which was found after Bayes' death. It was published posthumously in "Philosophical Transactions of the Royal Society" in 1763 by his friend, Richard Price [5]. Bayes showed how inverse probability could be used to calculate the probability of an antecedent event from the occurrence of the consequent event. In other words, given the knowledge of the present, one could calculate the probability that a particular related event occurred in the past. These methods were adopted by Laplace and other scientists in the 19th century, but fell out of favor in the early 20th century. In the middle of the 20th century, De Finetti, Jeffreys, Savage, and Lindley renewed interest in these methods when they developed a complete method of statistical inference based on Bayes' Theorem [15].

Notation

We begin with a list of notation conventions for clarity.

- $P(A)$: This is the prior probability of event A (i.e., the probability that event A will occur without taking any data into account). It is often referred to simply as the "prior".

- $P(A|B)$: This is the "posterior", "a posteriori", or "conditional" probability of event A given the data that event B occurred. Conditional probability is an idea central to Bayesian statistics. This value can be calculated via Bayes' Theorem, which is discussed below.

- $\text{LR}(A|B)$: This is the likelihood ratio that event A occurred, given that event B occurred. The likelihood ratio test uses the maximum probability of a result under two different hypotheses (here denoted A and A^C) to decide which hypothesis is more likely. Specifically, if $\text{LR}(A|B) > 1$, then it is more likely that event A occurred. Otherwise, it is more likely that A did not occur, which is denoted as A^C, or the complement of A. Likelihoods can be combined via multiplication if the data is independent and identically distributed. The likelihood ratio is defined as follows:

$$\text{LR}(A|B) = \frac{P(A|B)}{P(A^C|B)}. \tag{1.1}$$

16

It is simpler to visually see A^C as the complement of event A, as shown in Figures 1.2 and 1.3. In deciding between two hypotheses, H_0 and H_1, if event A is defined as H_0 being true, then event A^C is defined as H_1 being true.

Bayes' Theorem

Bayes' Theorem is typically stated as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{1.2}$$

In words, Bayes' Theorem states that the posterior probability of event A occurring when we know that B occurred is equal to the probability of event B occurring given that A has occurred, multiplied by the prior probability that A occurs and normalized by the probability that B occurs.

Bayes' Theorem is also commonly found in a number of equivalent forms, including

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)}, & (1.3)\\
&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}. & (1.4)
\end{aligned}
$$

Note that the probability that B occurs (regardless of whether or not A occurs) can be expanded into the following forms:
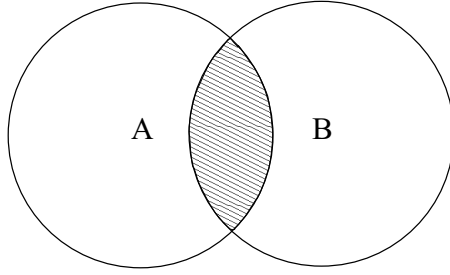
Figure 1.2: Proof of Bayes' Theorem. The shaded area can be defined as both $P(B|A)P(A)$ and $P(A|B)P(B)$. This gives Equation (1.7), which can be rearranged to give Bayes' Theorem, Equation (1.2).

$$P(B) = P(B \cap A) + P(B \cap A^C) = P(B|A)P(A) + P(B|A^C)P(A^C). \qquad (1.5)$$

An informal way of thinking about Bayes' Theorem is given by

$$P(A|B) = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}. \qquad (1.6)$$

In practice, the denominator $P(B)$ in Equation (1.2) is frequently ignored because it functions simply as a normalization factor, and is a constant in calculating and comparing the probabilities of various events B_i. Since it is the relative probabilities that are important in these cases, normalizing the probabilities being compared (by the same normalization factor) is often unnecessary.

Proof of Bayes' Theorem The proof of Bayes' Theorem is most easily given visually.

Given two partially overlapping circles, one labeled A and the other labeled B as shown in Figure 1.2, suppose the area of each circle is related to the probability of each event occurring. Thus, we can describe the region of overlap as $P(A \cap B)$. As shown above, this value can equivalently be expressed as either $P(B|A)P(A)$ or $P(A|B)P(B)$, so

$$P(A|B)P(B) = P(B|A)P(A). \tag{1.7}$$

This can be arranged to give

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

which is Bayes' Theorem. $\square$

Explanation of $P(B \cap A)$ This is best illustrated in terms of Bayes' Theorem involving multiple disjoint events. Let $A_1, \ldots, A_r$ be r disjoint events, and suppose that the probabilities $P(A_i)$ for $i \in \{1, \ldots, r\}$ are specified. B is an event that may or may not occur. Suppose further that $P(B|A_i)$ is also available for each i.

Now, suppose B occurs, and we wish to see how this additional information affects the initial probabilities $P(A_i)$. (See Figure 1.3 for an illustration of the problem description.) This is given by:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)},$$
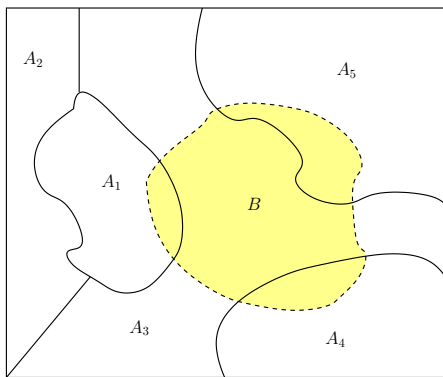
Figure 1.3: Suppose the A_i are r disjoint events, where $i = 1, \ldots, r$. In this picture, the area assigned to each A_i represents the probability of that event occurring. One way to see this is to randomly select a point within the picture to determine which event occurred - whichever region the point falls into corresponds with the event that actually occurred. The randomly selected point is more likely to lie within one of the larger regions. Now suppose we know that event B occurred. Equation (1.8) shows how the additional information that B occurred affects the probabilities each A_i occurred. Note that in this case, since B occurred, $P(A_2|B) = 0$.

where

$$P(A \cap B) = P(B|A_i)P(A_i),$$

and

$$P(B) \;= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_r \cap B)$$

$$= P(B|A_1)P(A_1) + \cdots + P(B|A_r)P(A_r).$$

So, given that event B occurred, the new probability of each (disjoint) event A_i occurring is equal to the probability of B occurring if A_i has occurred, multiplied by the probability that B_i occurs, normalized by the probability that B occurs.

This gives us a discrete formula for Bayes' Theorem,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum\limits_{n=1}^{r} P(B|A_n)P(A_n)}. \tag{1.8}$$

We now introduce the optimal fixed-sample test, the LRT.

1.3 Optimal Fixed-Sample Test: Likelihood Ratio Test (LRT)

For a set number of observations N (i.e., for a fixed-sample experiment), the optimal test for deciding which hypothesis is true is given by the LRT. The LRT is similar to the SPRT, which we will use in our mathematical group decision-making models. It is also generally assumed that human observers utilize a LRT-like test in making a decision in a Signal Detection task; therefore, we introduce this method first. We primarily consider simple or point hypothesis tests in our work: we wish to decide between two simple hypotheses, H_0 and H_1. The Neyman-Pearson Lemma [62] states that the uniformly most powerful (UMP) test for fixed-sample simple hypothesis testing is the LRT, which we define below. (See the Appendix, Section A.2, for the definition of a UMP test.)

Let $Y = y_1, y_2, \ldots, y_N$ be a random sequence of N independent and identically distributed (iid) observations. For a two-alternative forced-choice (2AFC) experi-

ment, we wish to determine which probability distribution these observations were drawn from. To do this, we calculate the likelihood ratio (LR):

$$\text{LR} = \frac{p_1}{p_0} = \frac{p_1(y_1)p_1(y_2)\ldots p_1(y_N)}{p_0(y_1)p_0(y_2)\ldots p_0(y_N)}, \tag{1.9}$$

and compare it to a constant, K:

$$\text{If LR} < K, \quad \text{choose } H_0 \tag{1.10}$$

$$\text{If LR} > K, \quad \text{choose } H_1. \tag{1.11}$$

The constant K is a free parameter, and can provide some bias towards one hypothesis: if $K > 1$, the probability of choosing H_0 increases, and if $K < 1$, the probability of choosing H_1 increases. If $K = 1$, the test simply determines which hypothesis is more likely (i.e., it is a maximum likelihood test), and the procedure guarantees the smallest overall error rate. Note that for the LRT, the order in which the data is analyzed does not matter (because it is a fixed-sample test), and in experiments whose data will be analyzed using this test, the observers typically have a fixed amount of time to analyze the data, in addition to having only a fixed amount of data available.

1.4 Experiment Basics

In this section, we introduce some different types of decision-making experiment tasks, as well as different answering paradigms one can adopt in each type of experiment. We will also provide a general framework for how an experiment is set up and run, and discuss error and measures of performance.

1.4.1 Basic Tasks

There are several categories of experiment tasks that are used to measure the perception and decision-making capabilities of decision-makers (DMs). These tasks are typically very simple, so experimenters add noise (typically Gaussian white noise) to each image to make the task more difficult for the DM. Sometimes the prior for each hypothesis is available to the DM, and sometimes it is not, depending on what the experimenter is testing. As a note, when running and analyzing the results of an experiment with human observers, it is convention to refer to individual observers by their initials. We first list some common types of tasks to illustrate the differences among them and to provide some intuition on what kinds of tasks our decision-making models may be useful or adapted for, then detail some common response paradigms.

Types of Tasks

Detection: The DM's task is to detect if the signal is present or absent. In the signal absent case, there is only noise on the display; thus, this is sometimes referred to simply as the "Noise" case. In the signal present case, there is a signal that is embedded in noise. For simplicity, we refer to this as the "Signal" case, but it is understood that the both the signal and noise are displayed. This is sometimes also referred to as a "Yes/No" task (i.e., the answer to the question "Do you detect the signal in this sample?").

Discrimination: The DM's task is to decide if the stimulus shown is Stimulus A or Stimulus B. Loosely, one could interpret this as a Signal Detection Task, if one considered the stimulus of interest (say, A) to be Signal and the other stimulus (B) to be Noise. Conversely, one could consider the detection task to be a discrimination task between a signal present response and a signal absent response. However, in Psychology, this is generally considered to be a fundamentally different task, even if the same or very similar math can be used to analyze both signal detection tasks and discrimination tasks. The tasks we consider fall into either the discrimination or detection categories. We primarily state the following tasks for completeness.

Identification: This is typically a memory-based task. The DM is shown a set of stimuli in one session, and asked to remember the items in the set. In a later session, which may happen minutes or even weeks later, the DM is shown a set of

stimuli that is partly composed of stimuli from the first session and partly composed of stimuli not previously presented. The DM's task is to identify each stimulus as being either "new" (i.e., was not in the original set) or "old" (i.e., was in the original set).

Scaling: The DM's task is to judge how much of something is present, with respect to a reference that is usually not present during the task. The quality (i.e., size, angle, frequency) being scaled depends on the experiment's specific stimulus.

Matching: This task is very similar to the "Scaling" task; however, in "matching" tasks, the observer is given a means by which he/she can adjust the experimental stimulus until it appears to match the reference that is present during the task. While the experimental stimulus and the reference may be identical when aligned, the background and position of each may be different.

Search: This generically refers to any task where the observer is not certain where the Signal (target) will appear (if present), and the observer must determine in which location the target appeared (if present). It sometimes includes searching an entire (continuous) image, or checking several possible (discrete) locations where the target could appear. In some variations of the task, the observer is given the prior probability that the target is at each location, and in some tasks, the observer must infer these on his own.

Types of Response Paradigms

Free-Response: The observer is allowed to observe the stimulus for as long as the observer wishes before choosing one alternative. This paradigm is frequently used in experiments where one is interested in measuring the reaction or decision time (DT) of the observer. Typically, the goal of experiments using this response paradigm is to look at how DT is related to performance or accuracy [13].

In some other classifications, this sometimes refers to a test in which the observer can respond freely, and is not limited to a fixed list of responses.

Interrogation: The observer is allowed to look at the information for a fixed amount of time, after which he/she is prompted to respond.

Forced-Choice: The observer is given a set number of alternatives to choose among, only one of which is the correct answer [59].

1.4.2 Basic Decision Making Problem Formulation

Consider the basic two-alternative forced-choice (2AFC) task, where one wants to choose between the null hypothesis (H_0) and the alternative hypothesis (H_1). This can generically be applied to most of the Psychophysical tasks described above. Here we set up the general procedure for a decision-making experiment that occurs for each observation, regardless of the number of observations taken per decision. We borrow some terminology from [94]. In this dissertation, we will generically

26

associate a Signal response with H_1 and a Noise response with H_0, and use the two sets of labels interchangeably.

Truth

In this first step, the "truth" of the experiment is laid out. This involves specifying the values which accurately represent the alternatives in the task, and selecting which is correct. In practice, selecting the correct hypothesis is typically done using a pseudo-random number generator, weighted by prior probabilities that may or may not be available to the observer doing the task. An example of this is shown in Figure 1.4. Suppose the 2AFC task is to decide whether or not an intruder is present in a remote room, and the observation given is the temperature of the room. Then, in Figure 1.4, the red line shows the true temperature of the room when the intruder is not present, and the green line shows the true temperature of the room when the intruder is present. The Truth step represents an idealistic situation, where there is no noise.

Probabilistic Transition Mechanism

All measurements are affected by noise. In this step, noise smears out the "true" values into distributions. It is often assumed that the noise is Gaussian, but other distributions can be used to characterize the noise. The Probabilistic Transition Mechanism selects a sample at random from the correct distribution, and returns this value to the observer. In an experiment, this is equivalent to taking a sample
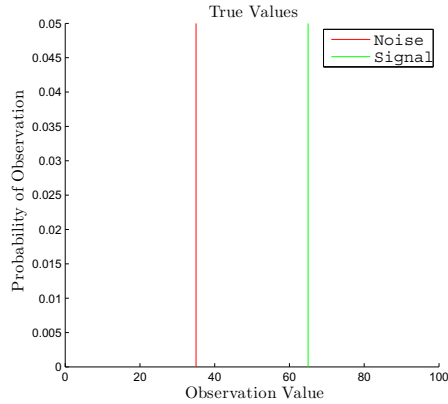
Figure 1.4: Two true values in an example of a 2AFC task. The red value corresponds to "Noise" (H_0) being correct, and the green value corresponds to "Signal" (H_1) being correct.

of data – the observer does not have direct access to the truth, but does have access to a sensor (e.g., the observer's eyes, a device such as a thermometer, etc) that can access the system containing the truth. The sensor returns a noisy sample from the correct distribution.

Continuing the previous example, Figure 1.5 shows the full distributions of possible observations that could be given to the observer after the true values were processed by the Probabilistic Transition Mechanism.

Observation Space

The observer may not exactly know the full distributions produced by the Probabilistic Transition Mechanism, but she knows the observation space, or the set of possible observations that the Probabilistic Transition Mechanism can produce. Furthermore, it is assumed that the observer has an internal representation of the distributions that is sufficiently accurate to perform the task. It is a fine point to

28
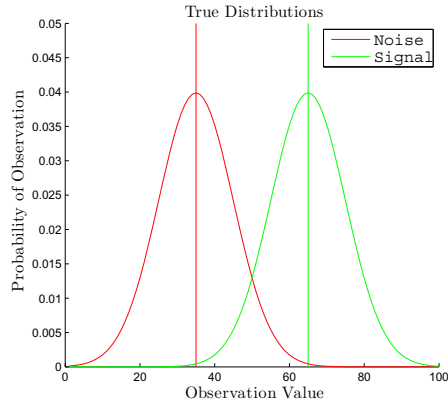
Figure 1.5: True distributions of the two alternatives in a 2AFC task. This represents the observation space after the Probabilistic Transition Mechanism.

note that there are effectively two observation spaces: the first which represents the observations put out by the system, and the second which represents the observations as received by the observer. Though it is often assumed that the two are the same (and we will use this assumption because it is reasonable for our experiments), it is important to note that they are not necessarily (and generally not) the same.

If the Signal distribution is correct, the observer could be given an observation like the one shown in blue in Figure 1.6. The green dot on the blue line shows the probability that the sample came from the Signal distribution and the red dot shows the probability that the observation came from the Noise distribution. These values are typically processed using a LRT to determine which is more likely. Generically, in a simple situation, such as the one in Figure 1.6, the distribution with a higher value at the given observation value is more likely (i.e., has a higher probability of occurring).

29
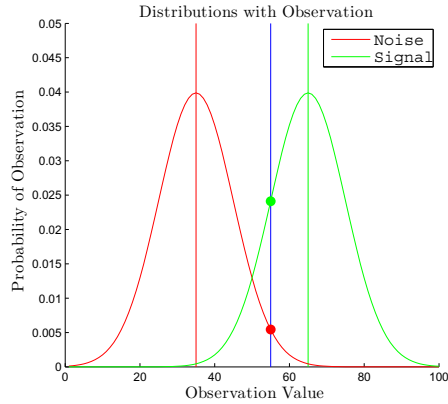
Figure 1.6: The observation space, with the sample observation value shown in blue. In our example, the correct distribution is Signal, but this need not necessarily be true for this observation. The green dot on the observation line shows the probability that the observation came from the Signal distribution, and the red dot shows the probability that the observation came from the Noise distribution. In this case, the Signal distribution is more likely.

Decision Rule

Here, the observer partitions her observation space with a decision rule, dividing the space into regions corresponding to each alternative, then applies her decision rule to the observation. In the sequential case, the observer calculates the evidence from the observation, adds it to the decision variable, then checks if one of the termination conditions has been satisfied. If the termination condition has been satisfied, the observer selects the appropriate hypothesis, and if it has not, the observer takes another sample of data. In the fixed-sample case, the observer compares the observation to the decision rule and returns a decision.

An example decision rule for the fixed-sample 2AFC case is shown in Figure 1.7. The black dashed line represents the decision rule, against which the observer compares the observation (blue). In this example, the observer would respond

30
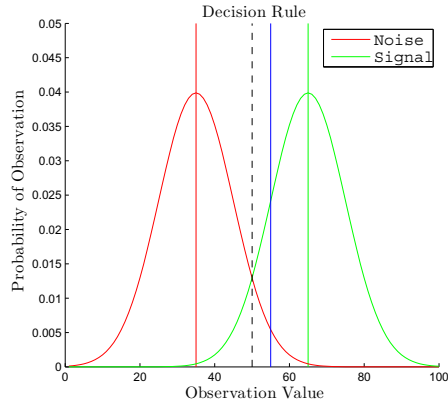
Figure 1.7: The observation space, with the sample observation value shown in blue and the decision rule shown in black (dashed). Since the observation is to the right of the decision rule, the observer selects the Signal response.

Signal because the observation is to the right of the decision rule (and therefore, Signal is more likely to be true).

1.4.3   Measures of Performance

Now that we have established some types of experiments that are frequently considered and how they are set up and run with respect to the observation space, we can discuss the measures of performance that we will be using to rate the abilities of the observer.

Error Rate (ER)

There are four possible outcomes in a 2AFC experiment, as shown in Table 1.1: either H_0 or H_1 is true, and the observer selects either H_0 or H_1. If the observer selected H_1 and it is true, that is considered a hit. Similarly, if the observer selected

31

	H_1 is true	H_0 is true
Decide H_1	Hit, or Correct Detection	False Alarm or Type I Error (α_0)
Decide H_0	Miss or Type II Error (α_1)	Correct Rejection

Table 1.1: All possible outcomes of a 2AFC task.

H_0 and it is true, that is considered a correct rejection. Both of these values contribute to the Percent Correct (PC). If the observer selects H_1 when H_0 is true, then the trial is considered a false alarm (or false positive, or Type I error), and if the observer selects H_0 when H_1 is true, the trial is considered a miss (or false negative, or Type II error). Both of these errors contribute to the general Error Rate (ER).

A false alarm rate is typically denoted by α in the literature and a miss rate by β; however, in our mathematical models, we will use the convention that the false alarm rate is denoted by α_0 and the miss rate is denoted by α_1. Our notation for the error should be interpreted as "α_i represents the probability of rejecting H_i when it is true."

When we do not differentiate between the two types of error, we will refer to the error rate generically as the ER for a given observer or sensor. This is equivalent to considering the two types of error to be equally undesirable. When appropriate, we may specifically distinguish between the minimum error rate allowed (a theoretical value called the "set ER") and the actual error rate found by simulation (actual ER), but in general, we will drop the qualifiers when the context is clear. Similarly, when we do not differentiate between the two types of correct responses, we will

refer to the observer's performance using the PC.

Cost Functions Often, there are costs or benefits associated with the different possible outcomes described in Table 1.1, and these can be taken into account via a cost function. A cost function formally states the specific cost of actions related to making a decision. Cost functions typically include the cost associated with taking data and/or time, and the cost associated with making each type of error. A number of different cost function-related strategies are discussed in [94]. While defining a cost function allows one to definitively state that one's decision-making scheme is optimal (with respect to that particular cost function), we will not be focusing on cost functions in this work. To declare a strategy as optimal, one must characterize all possible situations and outcomes and formulate a clearly defined cost function; this leads to a relatively narrow definition of reality. In many situations, the costs of different outcomes is known only approximately, because the exact value is not known in advance (i.e., unless it happens). We also argue that for many good strategies, it is likely possible to post-facto construct a reasonable cost function that makes that strategy optimal. In this sense, focusing on cost functions seems slightly artificial for a system that is likely to encounter an unpredictable set of situations.

Sensitivity, d'

In characterizing the capabilities of an observer, it is natural to discuss the observer's "sensitivity". PC provides information on how well the observer performs in the task, and thereby provides some indication of the observer's ability at the task; however, the most commonly used measure comes from Signal Detection Theory, and is known as d' ("dee-prime"). d' is a better measure than PC, because it is invariant to changes in other factors, such as criterion [59]. Though there are several ways to calculate this value, the most common and simplest definition is

$$d' = z(\text{HR}) - z(\text{FA}), \tag{1.12}$$

where z is the inverse of the normal distribution function [59], and it is assumed that the distributions in the observation space are normal. From the definition, we can see that an observer with a high hit rate (HR) and high false alarm rate (FA) can potentially have the same d' as an observer with a low HR and low FA. This is reasonable: if the observers have the same ability at the task and differ only in their tendency to respond Signal, they should have the same d' value. This will be explained in terms of the observation space below.

In theory, d' ranges in value from 0 to infinity. Zero values correspond to an observer who randomly guesses at the solution, and infinite values correspond to an observer with perfect performance. In practice, zero and infinite values are avoided by introducing the following corrections for perfect performance:

$$
\begin{aligned}
\widehat{\mathrm{FA}} \ &= \frac{1}{2N}, \\
\widehat{\mathrm{HR}} \ &= 1 - \frac{1}{2N},
\end{aligned}
\tag{1.13}
$$

where $\widehat{\mathrm{FA}}$ and $\widehat{\mathrm{HR}}$ are the corrected values, and N is the number of trials in the study.

Though Equation (1.12) is frequently used, it implicitly includes some assumptions on the underlying observation space. It is sufficient for our models in Chapters 2 and 3; however, we will need a more general definition in our data analysis, which we will introduce in Chapter 4.

Performance and Bias in the Observation Space

The observation space provides more information than just the distributions used to generate the observations – it also contains information about the observer's ERs and d'. In Figure 1.8, we use the same observation space as above, with the same optimal criterion, to illustrate how to calculate the observer's performance from the observation space. Figure 1.8(a) shows the HR, which is found by integrating the area under the Signal distribution to the right of the criterion. This is the HR because it represents the total probability of answering Signal (since the distribution is to the right of the criterion) when Signal is the correct answer (since the Signal distribution is only used to determine the experiment's outcome when Signal is the correct answer). The FA is shown in Figure 1.8(b), and is found by integrating over the Noise distribution to the right of the criterion. This is the FA

because it is the total probability that the observer answers Signal when the Noise distribution is correct. Similar intuition applies to Figures 1.8(c) and (d), which represent the miss rate and correct rejection rate, respectively.

This representation also makes intuitive sense: if H_1 is true, then the observer can get only either a hit or a miss, so one can find the probability that H_1 is true by adding the observer's hit and miss rates. The same holds for H_0, with the false alarm and correct rejection rates.

The criterion is optimal because it maximizes the HR while minimizing the FA rate. Moving the criterion changes both the HR and the FA; however, it does not change the observer's d'. This is more easily seen by comparing Figure 1.8 to Figure 1.9, which shows an observer with the same d' as the one shown in Figure 1.8, but with a different, non-optimal criterion. In this example, the observer is biased towards responding Noise. This results in a lower HR, but also a significantly lower FA rate.

The general consensus is that observers are not able to move their distributions (i.e., change their sensitivity), but are able to shift their criterion [53, 88], which is another reason why d' is considered a better measure of performance than PC: PC changes as the criterion is moved, while d' does not. We will focus on this measure in our fixed-sample analysis, since decisions are made there for each observation. In our sequential analysis, d' enters solely through the separation between the two distributions that are used to generate the observations in the experiment. Since

(a) Hit Rate

(b) False Alarm Rate

(c) Miss Rate

(d) Correct Rejection Rate
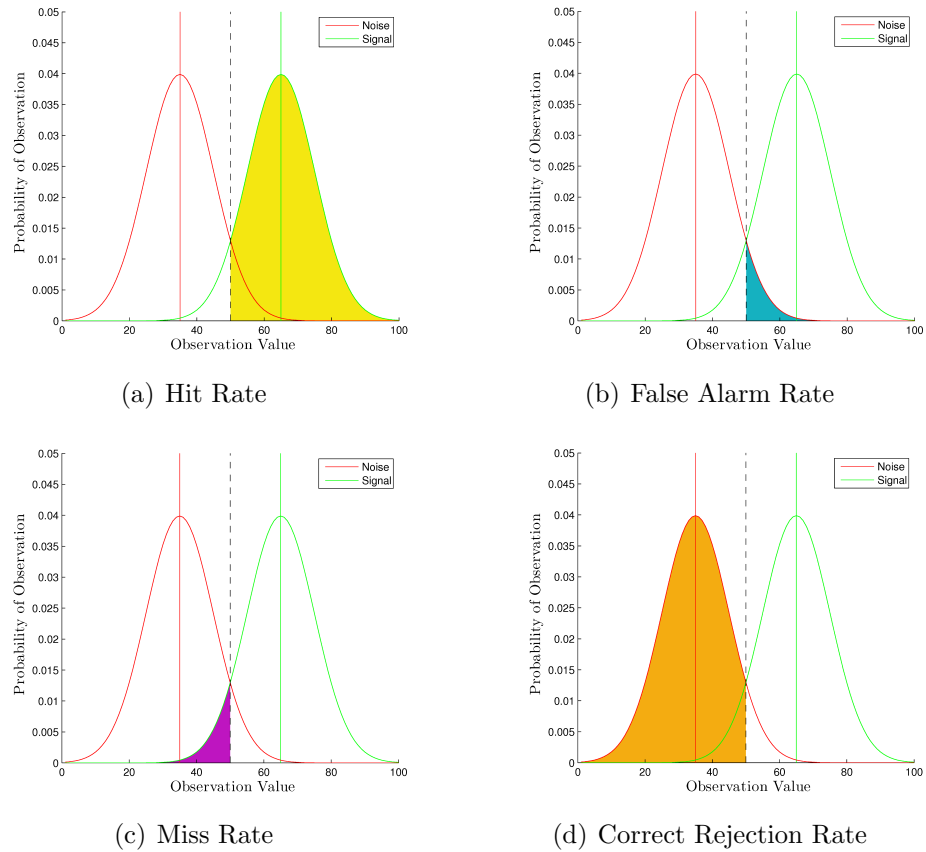
Figure 1.8: The observation space, used to show the observer's performance and ERs with an optimal criterion. (a) The HR is the total probability that the observer will select Signal when Signal is true. This is equivalent to integrating the area under the Signal distribution curve (because it is true) to the right of the criterion (which indicates that the observer selects this response), which is shown as a shaded region. The same idea applies to the (b) FA, (c) miss rate, and (d) correct rejection rate.

we use two distributions that differ only in mean and have a standard unit deviation, the d' of the observer is equal to the difference between the means of the distributions.

In the two examples in Figures 1.8 and 1.9, the observers' d' values can be thought of in terms of the distance between the two distributions, since the two distributions differ only in mean. If the observer's observation space is more com-
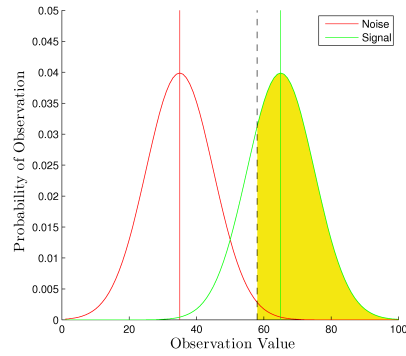
(a) Hit Rate
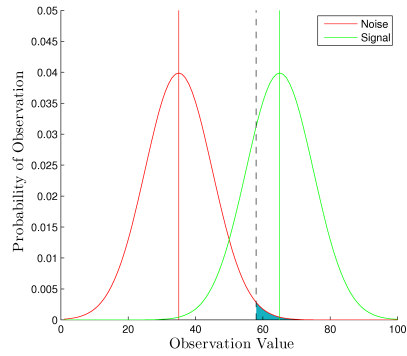
(b) False Alarm Rate
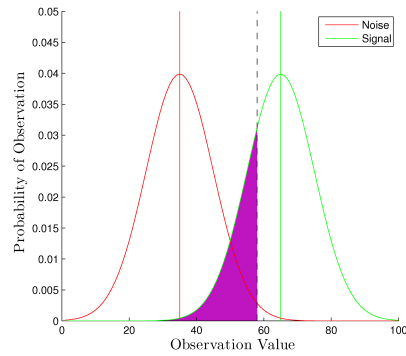
(c) Miss Rate
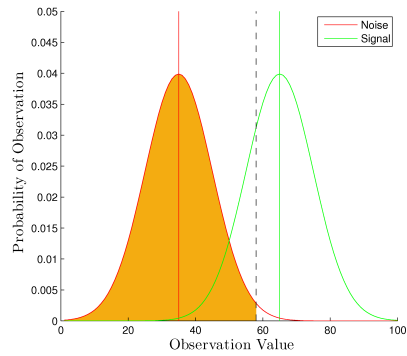
(d) Correct Rejection Rate

Figure 1.9: The observation space, showing the performance of an observer using a non-optimal criterion. The observer shown here has the same d' as the observer shown in Figure 1.8, but it is clear that this observer has a different (a) PC, (b) FA, (c) miss rate, and (d) correct rejection rate.

plicated (e.g., the distributions have different standard deviations or at least one distribution is bimodal, etc.), then another formula and possibly multiple criteria must be used. One can generically think of d' as being represented by the area of overlap between the two distributions, since this sets a limit on the group's error rates: with the optimal criterion, a greater area of overlap results in a higher level of error and a lower d', and a smaller area of overlap results in a lower level of error and a higher d'.

An example of a more complicated observation space with the optimal criteria is shown in Figure 1.10. In this case, the optimal decision criterion requires more than one criterion. To the left of the first dashed black line at about $x = 53$, the Noise response is more likely. Directly to the right of that, up to the second dashed black line near $x = 58$, Signal is more likely. Between that criterion and the last criterion near $x = 70$, Noise is again more likely, and to the right of the last criterion, Signal is more likely. See [29, 59] for more details on complicated observation spaces.

Decision Time (DT)

One last measure of performance that we will focus on in our mathematical models is decision time (DT). We primarily focus on DT for our sequential method-based mathematical models, since one of the main benefits of sequential methods is its average decision speed, which can equivalently refer to DT or the number of samples required to reach a decision. Our fixed-sample analysis is relevant to
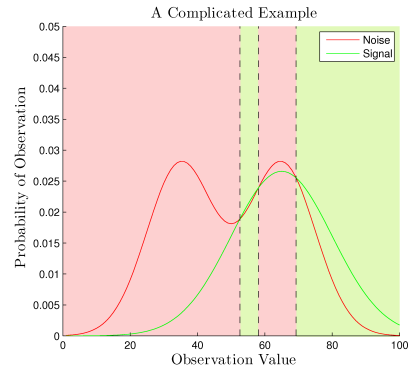
Figure 1.10: An example of a more complicated observation space with the optimal criteria. In this case, the optimal decision criterion requires more than one criterion. The optimal criteria partition the observation space into regions where one hypothesis is more likely than the other. The light red-colored regions show the observation values for which Noise is more likely, and the light green-colored regions show the observation values for which Signal is more likely. This is equivalent to dividing the observation space into regions in which one curve has a higher value; therefore, the optimal criteria are defined by the locations at which the two distributions cross.

applications where DT is not a large concern, so DT is not considered an important

performance parameter for that work.

Chapter 2

Individual Model: The

SPRT-Based DDM

2.1   Introduction

The Sequential Probability Ratio Test (SPRT) is a hypothesis testing procedure from sequential analysis that is optimal in the sense that it minimizes the average number of samples required to reach a decision while not exceeding specified error rates. This holds true for all tests, whether sequential or fixed-sample, whose error probabilities are at most equal to those of the SPRT [98, 99].

There are many situations in which the sample size is not fixed in advance or when one wants to minimize the number of samples. For example: in ammunition quality control, tested samples are unusable; in clinical trials, there is a moral

41

obligation to use the minimal number of test subjects to achieve the desired error rate; and in situations where the goal is to maximize a reward rate in time, where there may be a time penalty for incorrect answers.

We primarily consider the solution to a two-alternative forced-choice (2AFC) task. Though there are multiple definitions of what constitutes a 2AFC task, we follow the convention laid out in [13]: A 2AFC task is one in which the decision-maker (DM) must choose between two specific hypotheses, while making the following three assumptions: (i) evidence favoring each alternative is integrated over time, (ii) the process that provides the DMs with observations is subject to random fluctuations, and (iii) a decision is made when sufficient evidence has accumulated favoring one of the alternatives. An example of a 2AFC task is deciding if there is a signal (e.g., intruder, contamination, etc.) in a given area.

We first provide a brief history of the SPRT, then discuss its relevance in recent literature. We then present the SPRT, demonstrate its optimality, and show how it is related to the Drift-Diffusion Model (DDM), on which we will base our analytical results. We first solve the DDM for general average performance indices, mean Error Rate (ER) and mean Decision Time (DT), and later specialize those results to the constant-drift constant-diffusion case. We then solve the SPRT-based DDM for the full probability distribution function (pdf) of DTs and verify our results using simulation. We finish by investigating some issues that arose during simulation.

2.1.1 A Brief History of the SPRT

Below we briefly describe the development of the SPRT and related methods, then discuss the SPRT's relevance in current literature.

The first formalized sequential procedure was the "double-sample" inspection method by H. F. Dodge and G. G. Romig in 1929 [81]. This scheme differed from previous methods in that it was possible for results from the first sample to make it unnecessary to take the second sample. While W. Bartkey generalized this to a multi-stage procedure in 1943, and other multi-stage procedures were proposed, it was not until the last two years of World War II that sequential methods such as the SPRT and Banburismus were systematically developed [81].

Banburismus was developed by Alan Turing and his colleagues at Bletchley Park in the early 1940s. The process was named in honor of the town of Banbury, where the special sheets of paper used to perform computations for the process were printed. Turing's group was originally founded for the purpose of statistical quality control, but is most famous for using Banburismus to break the supposedly unbreakable Enigma code used by the German Navy [39]. Banburismus allowed the codebreakers to infer details about the configuration of the Enigma machines from the intercepted messages [23]. In other words, Turing's group searched for evidence in the intercepted messages to support or refute various hypotheses about the encoding scheme, and to find messages that were encoded by Enigma machines with identical configurations. Banburismus has three main features: first, it provides a

means to quantify the weight of evidence provided by individual clues. Second, it provides a way to update that quantity to account for new information. Lastly, Banburismus provides a decision rule to determine when there is sufficient evidence to render judgment [39].

Around the same time, the Statistical Research Group (SRG) based at Columbia University developed the SPRT as a means to reliably test the quality of items such as artillery while saving on sampling costs [81]. The SRG was supported by the Applied Mathematics Panel of the National Defense Research Committee, which was a part of the Office of Science Research and Development. The group was started by Warren Weaver and Harold Hotelling. On a social note, despite being assembled via "the old boy network", the SRG was a socially diverse group, consisting of 18 principles and about 60 others, including about 30 young women, two blacks, and two "persons with severe physical handicaps" [100]. The original idea that led to the development of the SPRT largely came from Milton Friedman and W. Allen Wallis, who, lacking a formal background in statistics, initially approached Jacob Wolfowitz for help in developing the idea. Wolfowitz was uninterested, however, so Wallis and Friedman gave the idea to Abraham Wald. Though initially uninterested in the problem, Wald quickly gained enthusiasm as he discovered its potential [38, 100]. Wald's theoretical results on the SPRT were presented in two of SRG's restricted reports in 1943 and 1944, and later in a book in 1947 [98]. Wald and Wolfowitz proved the optimality of the SPRT in 1948 [99].

44

Though Banburismus and the SPRT were developed mostly independently, they share many characteristics, and essentially are equivalent tests. We focus on the SPRT in our models because it is the most familiar, and is more relevant in the recent literature we consider, as we will detail below.

Recent Applications in Neuroscience

The DDM, a continuum-limit of the SPRT, has been used to model oculomotor decision-making in the brain, supported by experiments with *in vivo* recordings from monkeys performing the random dynamic dot task, in which a proportion of dots on a computer screen move coherently either to the right or left, with speed and number determined by given distributions, and the rest of the dots move randomly [44]. The experimenters implanted electrodes in the monkeys' brains to monitor neural activity in relevant areas. The 2AFC task associated with this experiment is to decide if more of the coherent dots were moving to the left or to the right. The DDM has been demonstrated to be a reasonable model for predicting macaques' response time using data from their middle temporal (MT) and lateral intraparietal (LIP) areas [16], or frontal eye field [34]. We explain the relevance of the SPRT to decision making in these regions of the brain below.

The MT is the first area in the visual pathway where a majority of neurons are directionally selective, so one would expect the firing rate of the neurons selective for rightward motion to be slightly higher if more of the correlated dots were moving

to the right, and likewise if the majority of correlated dots were moving to the left. However, the actual output from the MT is very noisy, and for a task in which the correlation is small, it is very difficult to decide which direction to choose based on information from the MT alone. The LIP area is involved in eye control – there are neurons that are selective for leftward saccades and neurons selective for rightward saccades. In the experiment, measurements from the neurons in the LIP indicated that the neurons were actually integrating the small differences in the firing rates in the MT. A nice correlation was found between the time when the monkey made its decision and when the neuron firing rate in the LIP area passed a certain threshold value [16], similar to the SPRT and DDM. A cartoon example of this is shown in Figure 2.1 [13]. In the MT, shown in Figure 2.1(a), it is not immediately clear which direction is favored: after close inspection, the leftwards firing generally seems slightly higher, but it is not always higher, and when it is, it is not higher by much. In contrast, it is immediately clear which direction is preferred in the LIP response, shown in Figure 2.1(b).

Similar results were found using data from the superior colliculus in different tasks [71, 74]. A number of DDM-related models were also used to fit psychophysical data from monkeys learning a decision-making task [30]. A nice review of several mathematical models used in this area is given in [13, 85]. These works have inspired the development of reward-maximizing neural network models [17, 83].

A DDM-based model by Ratcliff et al. called Ratcliff's Wiener (RW) diffusion

| (a) MT Response | (b) LIP Response |
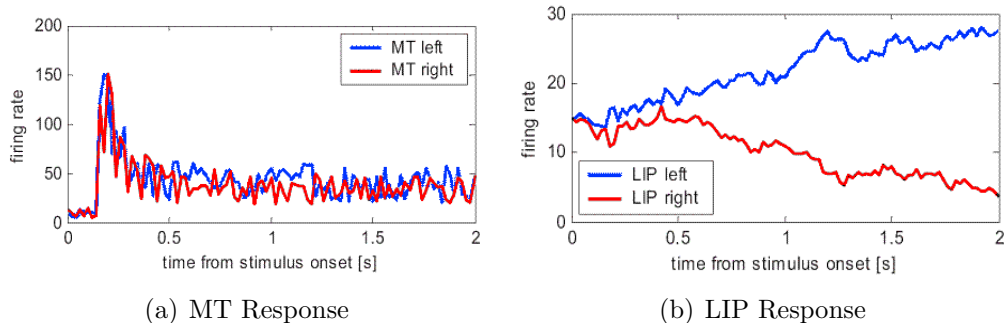
Figure 2.1: Cartoon showing the output of the MT and the LIP of a monkey performing the dynamic random dot experiment, from [13]. Though it is not immediately clear which direction the brain selects for in the MT response, it is very clear from the LIP response. The LIP is thought to integrate the difference in signal from the MT. The rightward response decreases towards zero, while the leftward response continues to climb. Similar to the SPRT or DDM, the monkeys were found to respond when the signal in the LIP reached a threshold value. Thus, the SPRT is a relevant model for modeling decision making in monkeys.

model was shown to be relevant for human decision-making in psychophysical tasks in a large body of work [73, 76]; see [75] for a comprehensive review of relevant literature. The RW model has been used to model individual differences and correlations between model parameters and data, the effects of aging [70], IQ [77], sleep deprivation [72] and dysphoria [101] on observers performing 2AFC tasks. A number of popular sequential sampling models were summarized and compared in [78], and evaluated by their performance in modeling data from three different 2AFC experiments. The models discussed include RW diffusion, Ornstein-Uhlenbeck (O-U) diffusion, accumulator, Poisson counter, leaking competing accumulator, leaky accumulator with relative criteria, and Usher and McClelland models. It was found that there was a substantial overlap among the different models, but that in general, the RW diffusion model, O-U diffusion model with small-to-moderate decay,

and the two leaky accumulator models provided the best performance. In [55], a similar study was done for several sequential sampling models using racing diffusion processes to perform a multiple-alternative task, and some general issues in formulating multiple-alternative models were raised. A method for averaging the performance of a group of individuals was demonstrated in [69]. We note that the individual model we derive below is equivalent to the RW model's single diffusion process [68]; however, no derivation of Ratcliff's single diffusion process solution was provided in [68]. We provide our independently-formulated derivation below for completeness. We also note that the RW model was intended for matching tasks, and is therefore fundamentally different from our group models.

A reward-based study at Princeton University with undergraduates performing the random dynamic dot task was presented in [14]. The results indicate that the constant-drift, constant-diffusion DDM (called the "pure" DDM) does not accurately account for all empirical phenomena. As a result, they developed the "extended" DDM to include variable drift rates, non-decision-making-related delay times, and initial bias. The extended DDM was found to account for a wider range of empirical phenomena but was significantly more difficult to analyze. A main finding relevant to our work is the result that the human observers' performance could be modeled reasonably well by a pure DDM: introducing variable drift rates tended to slightly decrease the predicted DT for a given ER, while varying the initial bias tended to slightly increase the predicted DT; thus, accounting for both

resulted in performance similar to accounting for neither. It was additionally found in [84] that grouped data matched well with a DDM optimized for maximum reward rate.

Based on these studies, we present the following solution to a pure DDM as a model which could be used to fit the performance of a human observer or to design a detector. Though our solutions are arguably a special case of the models presented in [13], we have not yet seen the specific treatment we present.

2.1.2 The Sequential Probability Ratio Test (SPRT)

We begin with the standard definition of the SPRT, with likelihood ratios, so we can show a derivation of the boundary conditions, establish ways to account for bias, and show the expected number of samples required to reach a decision while being consistent with previous works. We then introduce the log-likelihood-based SPRT, which we will use in our models.

Standard Definition of the SPRT

Suppose Y is a random variable with unknown pdf p over the observation space. The SPRT chooses between H_0:$p = p_0$ and H_1:$p = p_1$ with ERs no greater than a specified α_0 and α_1. The pdfs p_0 and p_1 are assumed to be known.

Since we assume that the observations are iid, the general procedure for the nth stage of the SPRT is similar to the LRT discussed in Section 1.3: first, take the nth

observation of Y, denoted y_n. Then, update the stage-n likelihood ratio, given by

$$\mathrm{LR}_n = \prod_{i=1}^{n} \frac{p_1(y_i)}{p_0(y_i)}. \tag{2.1}$$

After processing observation y_n, then select one of three actions, based on the value of LR_n:

$$b_0 < \mathrm{LR}_n < b_1 : \text{Take another sample of data,}$$

$$\mathrm{LR}_n \le b_0 : \text{Choose } H_0,$$

$$\mathrm{LR}_n \ge b_1 : \text{Choose } H_1.$$

The SPRT (and experiment) continues as long as $b_0 < \mathrm{LR}_n < b_1$, where b_0 and b_1 are constants and $b_0 < 1 < b_1$. The latter condition ensures that the test selects the more likely hypothesis.

Relationship between Boundaries and Error Rates It is intuitive that the boundaries b_0 and b_1 are related to the error rates α_0 and α_1: boundaries that are further away from the decision variable's initial condition provide more accurate decisions because the decision variable is less susceptible to being driven across a boundary by noise. (It is implicitly assumed that $\mathrm{LR} = 1$ is the initial condition.) Here, we make this idea concrete.

Let T be the number of observations required for the SPRT to finish. Note that T is not a predetermined number; however, by assumption, T is a finite number.

Let $\mathbb{Y}_{1T}$ be the set of all T-length sequences of observations $(\mathcal{Y}_T)$ such that the SPRT chooses H_1 at step T. In other words, for any $\mathcal{Y}_T \in \mathbb{Y}_{1T}$,

$$\frac{p_1(y_1)p_1(y_2)\ldots p_1(y_T)}{p_0(y_1)p_0(y_2)\ldots p_0(y_T)} \geq b_1,$$

$$\Leftrightarrow p_1(y_1)p_1(y_2)\ldots p_1(y_T) \geq b_1 p_0(y_1)p_0(y_2)\ldots p_0(y_T).$$

We now integrate both sides over $\mathbb{Y}_{1T}$ and sum over all possible T to get

$$\sum_{T=1}^{\infty} \int_{\mathbb{Y}_{1T}} p_1(y_1)p_1(y_2)\ldots p_1(y_T) \geq b_1 \sum_{T=1}^{\infty} \int_{\mathbb{Y}_{1T}} p_0(y_1)p_0(y_2)\ldots p_0(y_T),$$

which simplifies to

$$p_1(\mathbb{Y}_1) \geq b_1 p_0(\mathbb{Y}_1), \tag{2.2}$$

where $p_j(\mathbb{Y}_1)$ is the probability of choosing H_1 given that H_j is true. By the definition of the error rates α_0 and α_1 we have

$$p_0(\mathbb{Y}_1) \leq \alpha_0,$$

$$p_1(\mathbb{Y}_0) \leq \alpha_1.$$

Since $p_i(\mathbb{Y}_1) + p_i(\mathbb{Y}_0) = 1$ for $i = 0, 1$, this implies that

51

$$p_0(\mathbb{Y}_0) \geq 1 - \alpha_0,$$

$$p_1(\mathbb{Y}_1) \geq 1 - \alpha_1.$$

To define the boundary, we take $p_0(\mathbb{Y}_1) = \alpha_0$ and $p_1(\mathbb{Y}_1) = 1 - \alpha_1$. Plugging this into Equation (2.2) we have

$$
\begin{aligned}
1 - \alpha_1 &\geq b_1 \alpha_0, \\
\Rightarrow b_1 &\leq \frac{1 - \alpha_1}{\alpha_0}.
\end{aligned}
\tag{2.3}
$$

Similarly, let $\mathbb{Y}_{0T}$ be the set of T-length sequences of observations $(\mathcal{Y}_T)$ such that the SPRT chooses H_0 at step T. So, for any $\mathcal{Y}_T \in \mathbb{Y}_{0T}$,

$$
\begin{aligned}
\frac{p_1(y_1)p_1(y_2)\dots p_1(y_T)}{p_0(y_1)p_0(y_2)\dots p_0(y_T)} &\leq b_0, \\
\Leftrightarrow p_1(y_1)p_1(y_2)\dots p_1(y_T) &\leq b_0 p_0(y_1)p_0(y_2)\dots p_0(y_T).
\end{aligned}
$$

Integrating over $\mathbb{Y}_{0T}$ and summing over all possible T gives

$$\sum_{T=1}^{\infty} \int_{\mathbb{Y}_{0T}} p_1(y_1)p_1(y_2)\dots p_1(y_T) \leq b_0 \sum_{T=1}^{\infty} \int_{\mathbb{Y}_{0T}} p_0(y_1)p_0(y_2)\dots p_0(y_T),$$

which simplifies to

52

$$p_1(\mathbb{Y}_0) \leq b_0 p_0(\mathbb{Y}_0). \tag{2.4}$$

Again, to define the boundary, we take $p_1(\mathbb{Y}_0) = \alpha_1$ and $p_0(\mathbb{Y}_0) = 1 - \alpha_0$. Plugging these values into Equation (2.4), we get

$$
\begin{aligned}
\alpha_1 &\leq \ b_0(1 - \alpha_0), \\
\Rightarrow b_0 &\geq \ \frac{\alpha_1}{1 - \alpha_0}.
\end{aligned}
\tag{2.5}
$$

The inequalities in Equations (2.3) and (2.5) become equalities only if the decision variable hits a boundary exactly (rather than overshooting the boundary). In practice, it is customary to apply the small-overshoot assumption [54, 98], which allows us to define the boundaries by the (slightly) conservative values

$$b_0 = \ \frac{\alpha_1}{1 - \alpha_0}, \tag{2.6}$$

$$b_1 = \ \frac{1 - \alpha_1}{\alpha_0}. \tag{2.7}$$

For now, we will take the small-overshoot assumption. We will later revisit this assumption in greater detail in our simulations in Section 2.6.2.

Earlier, in our definition of the SPRT, we introduced the condition $b_0 < 1 < b_1$. Under the above relations, this condition becomes

$$\frac{\alpha_1}{1 - \alpha_0} < 1 < \frac{1 - \alpha_1}{\alpha_0}. \tag{2.8}$$

Thus, we require that $\alpha_1 < 1 - \alpha_0$ and $\alpha_0 < 1 - \alpha_1$. This is very reasonable: it is equivalent to requiring that the probability of choosing the correct hypothesis be greater than the probability of choosing the incorrect hypothesis.

Bias Let π_0 (resp., π_1) be the observer's prior bias that H_0 (resp., H_1) is correct. There are two ways in which we can take prior bias into account: we can either shift the likelihood ratio's initial condition, LR_0 and use the previously-calculated boundaries, or we can use the initial condition $LR_0 = 1$ and shift the location of the boundaries. These are equivalent ways of accounting for bias.

Shifted Initial Condition In cases where it is convenient to set the location of the boundaries using Equations (2.6) and (2.7), we can account for the observers' prior bias through the decision variable's initial condition. In this case, we set

$$LR_0 = \frac{\pi_1}{\pi_0}, \tag{2.9}$$

to shift the initial condition towards the boundary corresponding to the more likely hypothesis. Note that we require $\alpha_0 < \min(\pi_0, \pi_1)$ and $\alpha_1 < \min(\pi_0, \pi_1)$.

Shifted Boundaries In cases where it is convenient to have a unity initial condition, we can account for the observer's prior bias by shifting the boundaries.

In this case, the new boundaries are

$$\hat{b}_0 = \frac{\pi_1}{\pi_0}b_0, \tag{2.10}$$

$$\hat{b}_1 = \frac{\pi_1}{\pi_0}b_1. \tag{2.11}$$

Expected Number of Samples Wald [98] also gives approximate expressions for the expected numbers of observations $E_i[N]$, $i = 0, 1$. Using the relations in Equations (2.6) and (2.7), they are

$$E_1[N] \approx \frac{\alpha_1 \log(b_0) + (1 - \alpha_1)\log(b_1)}{E_1\left[\log\left(\frac{p_1(y)}{p_0(y)}\right)\right]} \tag{2.12}$$

$$E_0[N] \approx \frac{(1 - \alpha_0)\log(b_0) + \alpha_0 \log(b_1)}{E_0\left[\log\left(\frac{p_1(y)}{p_0(y)}\right)\right]}. \tag{2.13}$$

The SPRT with Log-Likelihoods

Here, we establish the form of the SPRT that we will use in the rest of this dissertation.

We introduce the decision variable x that tracks position in the one-dimensional decision space (B_0, B_1) after receiving observation y_n. Since the observations are iid, we can use the log-likelihood ratio to update the location of the decision variable. After the nth observation is processed, the decision variable is

55

$$x_n = \log(\text{LR}_n) + x_0 = \sum_{i=1}^{n} \log \left[\frac{p_1(y_i)}{p_0(y_i)} \right] + x_0. \tag{2.14}$$

We will account for bias in the initial condition:

$$x_0 = \log \left(\frac{\pi_1}{\pi_0} \right), \tag{2.15}$$

where $B_0 < x_0 < B_1$. After processing this observation, we then select one of the following actions:

$$B_0 < x_n < B_1 : \quad \text{Take another sample of data,}$$

$$x_n \leq B_0 : \quad \text{Choose } H_0,$$

$$x_n \geq B_1 : \quad \text{Choose } H_1,$$

where

$$B_0 = \log \left(\frac{\alpha_1}{1 - \alpha_0} \right), \tag{2.16}$$

$$B_1 = \log \left(\frac{1 - \alpha_1}{\alpha_0} \right). \tag{2.17}$$

As before, we require $B_0 < 0 < B_1$ to ensure that the test chooses the hypothesis that is more likely.

If we assume $\alpha_0 = \alpha_1 = \varepsilon$, which is reasonable for the case where we want to

56

have a minimum amount of any type of error, then the boundaries are

$$B_0 = \log\left(\frac{\varepsilon}{1-\varepsilon}\right) \equiv -z, \qquad (2.18)$$

$$B_1 = \log\left(\frac{1-\varepsilon}{\varepsilon}\right) \equiv z. \qquad (2.19)$$

(Note that here, $-B_0 = B_1$). We also require that $\varepsilon < \min(\pi_0, \pi_1)$.

2.1.3 Optimality of the SPRT

Let $P(\text{reject } H_i | H_i)$ be the probability that hypothesis H_i is true but has been rejected, for $i = 0, 1$. Let $E_i[N]$ be the expected number of observations required to reach a decision when hypothesis H_i is true. Then, we have the following theorem:

SPRT Optimality Theorem Among all tests (fixed sample or sequential), for which

$$P(\text{reject } H_i | H_i) \leq \alpha_i, \quad i = 0, 1 \qquad (2.20)$$

and for which $E_0[N]$ and $E_1[N]$ are finite, the SPRT with error probabilities $P(\text{reject } H_i | H_i) = \alpha_i$, $i = 0, 1$ minimizes both $E_0[N]$ and $E_1[N]$ [99].

This theorem was proven by Wald and Wolfowitz in 1948; however, we provide a sketch of two simpler proofs from [13], the first of which follows a proof by Lehmann [54]; also see [37].

A Sketch of the Proof of the SPRT Optimality Theorem

Let w_0 be the loss from choosing H_1 when H_0 is true, and w_1 be the loss from choosing H_0 when H_1 is true. H_0 is true with prior probability π_0 and H_1 is true with prior probability $\pi_1 = 1 - \pi_0$. Let κ be the cost of taking one observation. For H_i true, we define the risk R_i as the sum of the expected loss due to an incorrect decision and the expected cost of reaching a decision:

$$R_i = \alpha_i w_i + \kappa E_i[N], \quad i = 0, 1.$$

Since we are looking at expected values, we assume that these values are calculated over a large number of trials. The total average risk of a decision procedure φ is given by

$$
\begin{aligned}
r(\pi_0, \pi_1, \kappa, w_0, w_1, \varphi) &= \pi_0 R_0 + \pi_1 R_1 \\
&= \pi_0(\alpha_0 w_0 + \kappa E_0[N]) + \pi_1(\alpha_1 w_1 + \kappa E_1[N]).
\end{aligned}
$$

Now we note that the SPRT minimizes $r(\pi_0, \pi_1, \kappa, w_0, w_1, \varphi)$ in the following sense: for any SPRT with $b_0 < 1 < b_1$ and any $0 < \pi_1 < 1$ (since this also makes $0 < \pi_0 < 1$ true), there exist positive constants κ, w_0, and w_1 such that the SPRT minimizes $r(\pi_0, \pi_1, \kappa, w_0, w_1, \varphi)$ for those values of π_1, κ, w_0, and w_1 [37, 54].

Consider a different decision procedure φ^*, with error probabilities $\alpha_i^* \leq \alpha_i$ and

finite expected sample size $E_i^*(N)$, $i = 0, 1$. For this decision procedure, the average risk is given by

$$r(\pi_0, \pi_1, \kappa, w_0, w_1, \varphi^*) = \pi_0(\alpha_0^* w_0 + \kappa E_0^*[N]) + \pi_1(\alpha_1^* w_1 + \kappa E_1^*[N]).$$

Since we can choose κ, w_0 and w_1 so that the SPRT minimizes the average risk for any π_1, for those values, we have

$$
\begin{aligned}
r(\pi_0, \pi_1, \kappa, w_0, w_1, \varphi) \;\leq\; & r(\pi_0, \pi_1, \kappa, w_0, w_1, \varphi^*) \\
\pi_0(\alpha_0 w_0 + \kappa E_0[N]) + \pi_1(\alpha_1 w_1 + \kappa E_1[N]) \;\leq\; & \pi_0(\alpha_0^* w_0 + \kappa E_0^*[N]) \\
& + \pi_1(\alpha_1^* w_1 + \kappa E_1^*[N]) \\
\leq\; & \pi_0(\alpha_0 w_0 + \kappa E_0^*[N]) \\
& + \pi_1(\alpha_1 w_1 + \kappa E_1^*[N]).
\end{aligned}
$$

This gives

$$\pi_0 E_0[N] + \pi_1 E[N] \leq \pi_0 E_0^*[N] + \pi_1 E_1^*[N].$$

Because this expression must hold for any π_1, we can conclude that

$$E_0[N] \leq E_0^*[N] \quad \text{and} \quad E_1[N] \leq E_1^*[N]. \;\square$$

SPRT Optimizes Reward Rate

In [13], it was also shown that the SPRT optimizes the reward rate introduced by Gold and Shadlen [39]. The proof requires that the ERs be allowed to vary and that the number of observations required be minimized.

As before, we assume that the ERs are equal, $\alpha_0 = \alpha_1 = \varepsilon$, and that the mean DT is the average time needed to make a decision. We define the reward rate (RR) as the probability of a correct response divided by the average time between responses. In calculating the average time between responses, we include an imposed time delay $D \geq 0$ between trials, and a time penalty $D_p \geq 0$ for incorrect responses. Thus,

$$\text{RR} = \frac{1 - \varepsilon}{\text{DT} + D + D_p \varepsilon}.$$

For this proof, we follow a similar line of reasoning as before: for a given decision procedure φ^*, suppose that the RR is maximized for $\varepsilon = \varepsilon^*$. The SPRT Optimality Theorem gives that the SPRT decision procedure φ has the property that $\text{DT}(\varphi, \varepsilon^*) \leq \text{DT}(\varphi^*, \varepsilon^*)$. Thus,

$$\text{RR}(\varphi^*, \varepsilon^*) = \frac{1 - \varepsilon^*}{\text{DT}(\varphi^*, \varepsilon^*) + D + D_p \varepsilon^*} \leq \frac{1 - \varepsilon^*}{\text{DT}(\varphi, \varepsilon^*) + D + D_p \varepsilon^*} = \text{RR}(\varphi, \varepsilon^*).$$

In other words, for any decision procedure φ^*, one can always find an SPRT

with a RR at least as high. Furthermore, there is a particular value of ε which maximizes the reward rate of the SPRT. This value of ε is generally difficult to determine in practice because different ERs generally correspond to different DTs.

$\square$

2.1.4 Continuum Limit

Rather than deriving the continuum limit, we provide a sketch of the proof shown in [13]. Some care is required to preserve the variability of x_n when taking the limit. Up to an unimportant scale factor relating the "time steps" n and the continuous time t, the limiting procedure is as follows:

Let Δx_r have mean m and variance D^2, which is assumed to be finite. Now define the family of random functions of $t \in [0, T]$ indexed by $M = 1, 2, ...$ (where T is some large time) as

$$x^M(t) = \frac{1}{\sqrt{M}} \sum_{r-1}^{k} (\Delta x_r - m) + \frac{1}{M} \sum_{r=1}^{k} \Delta x_r, \qquad (2.21)$$

where $k = \lfloor \frac{Mt}{T} \rfloor$, the largest integer smaller than $\frac{Mt}{T}$. Note the different normalization factors in the equation for x^M: this reflects the different rates at which fluctuations and means accumulate as the random increments are summed.

For any value of M, $x^M(t)$ has mean $m\lfloor t/T \rfloor$ and variance $D^2\lfloor t/T \rfloor$ (e.g, for $x_n = \sum_{r=1}^{n} \log \left[\frac{p_1(y_r)}{p_0(y_r)} \right]$, $x^M(t)$ has mean mn and variance D^2n).

Donsker Invariance Principle: Under some conditions, the distribu-

tion of a functional of normalized sums of iid random variables converges to the distribution of this functional of the Wiener process. (See Theorem 37.8 in [9]; also [10]).

The Donsker Invariance Principle, together with the Law of Large Numbers, implies that as $M \to \infty$,

$$x^M(t) \to DW(t) + mt \equiv x(t), \tag{2.22}$$

where $W(t)$ is a Wiener process (see the Appendix, Section B.1) and the convergence of the random functions $x^N(\cdot)$ is in the sense of distributions.

In other words, the limiting process $x(t)$ satisfies the stochastic differential equation (SDE)

$$dx = mdt + DdW, \tag{2.23}$$

$$x(0) = 0, \tag{2.24}$$

with boundaries $B_0 < 0 < B_1$. Below, we solve for m and D in terms of the pdfs.

Drift and Diffusion Constants for the SPRT-based DDM

The drift m and variance D^2 of Δx_r in Equations (2.23) - (2.24) depend on the pdfs $p_i(y_r)$. For example, for Gaussians with $\mu_1 > \mu_0$ and $\sigma_0 = \sigma_1 \equiv \sigma$, applied to the rth observation, we have

$$p_1(y_r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_r - \mu_1)^2}{2\sigma^2}},$$

$$p_0(y_r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_r - \mu_0)^2}{2\sigma^2}}.$$

The corresponding Δx_r, the distance traveled by the decision variable on the rth time step, is:

$$
\begin{aligned}
\Delta x_r &= \log\left[\frac{p_1(y_r)}{p_0(y_r)}\right], \\
&= \log\left[\frac{\frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(y_r - \mu_1)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_r - \mu_0)^2}{2\sigma^2}}}\right], \\
&= \frac{\mu_0^2 - \mu_1^2 + 2y_r\mu_1 - 2y_r\mu_0}{2\sigma^2}, \\
&= \frac{y_r(\mu_1 - \mu_0)}{\sigma^2} - \frac{(\mu_1 + \mu_0)(\mu_1 - \mu_0)}{2\sigma^2},
\end{aligned}
$$

$$\Delta x_r = \frac{\mu_1 - \mu_0}{\sigma^2}\left(y_r - \frac{\mu_1 + \mu_0}{2}\right). \tag{2.25}$$

We can now find the statistics of Δx_r. If H_i is true, then the expected value and variance of observation y_r are

$$E(y_r) = \mu_i, \tag{2.26}$$

$$Var(y_r) = \sigma^2. \tag{2.27}$$

Given this, we can find the expected value of Δx_r:

$$
\begin{aligned}
E[\Delta x_r] &= E\left[\frac{\mu_1 - \mu_0}{\sigma^2}y_r - \frac{(\mu_1 - \mu_0)(\mu_1 + \mu_0)}{2\sigma^2}\right], \\
&= \frac{\mu_1 - \mu_0}{\sigma^2}E[y_r] - \frac{(\mu_1 - \mu_0)(\mu_1 + \mu_0)}{2\sigma^2}, \\
&= \frac{-\mu_1^2 + \mu_0^2 + 2(\mu_1 - \mu_0)\mu_i}{2\sigma^2}.
\end{aligned}
\tag{2.28}
$$

If H_1 is true, $\mu_i = \mu_1$:

$$
\begin{aligned}
E[\Delta x_r | H_1 \text{ true}] &= \frac{-\mu_1^2 + \mu_0^2 + 2\mu_1^2 - 2\mu_0\mu_1}{2\sigma^2}, \\
&= \frac{\mu_1^2 - 2\mu_0\mu_1 + \mu_0^2}{2\sigma^2}, \\
&= \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}.
\end{aligned}
\tag{2.29}
$$

If H_0 is true, $\mu_i = \mu_0$:

64

$$E[\Delta x_r | H_0 \text{ true}] = \frac{-\mu_1^2 + \mu_0^2 + 2\mu_0\mu_1 - 2\mu_0^2}{2\sigma^2},$$

$$= \frac{-\mu_0^2 + 2\mu_0\mu_1 - \mu_1^2}{2\sigma^2},$$

$$= -\frac{(\mu_0 - \mu_1)^2}{2\sigma^2}. \tag{2.30}$$

In summary,

$$E[\Delta x_r] \equiv m = \begin{cases} \frac{(\mu_1 - \mu_0)^2}{2\sigma^2} \text{ for } H_1 \text{ true,} \\ \\ -\frac{(\mu_1 - \mu_0)^2}{2\sigma^2} \text{ for } H_0 \text{ true.} \end{cases} \tag{2.31}$$

By definition, the variance is

$$Var(\Delta x_r) = E[(\Delta x_r - m)^2] = E[(\Delta x_r)^2 - 2\Delta x_r m + m^2],$$

$$= E[(\Delta x_r)^2] - \underbrace{E[2\Delta x_r m - m^2]}_{2mE[\Delta x_r] - m^2 = (E[\Delta x_r])^2},$$

$$= E[(\Delta x_r)^2] - (E[\Delta x_r])^2. \tag{2.32}$$

First we find $E[(\Delta x_r)^2]$. We assume that hypothesis H_i is true, and H_k is not true.

$$
\begin{aligned}
E[(\Delta x_r)^2] &= E\left[\left(\frac{\mu_1 - \mu_0}{\sigma^2}\left\{y_r - \frac{\mu_1 + \mu_0}{2}\right\}\right)^2\right], \\
&= E\left[\frac{(\mu_1 - \mu_0)^2}{\sigma^4}\left(y_r^2 - (\mu_1 + \mu_0)y_r + \frac{(\mu_1 + \mu_0)^2}{4}\right)\right], \\
&= \frac{(\mu_1 - \mu_0)^2}{\sigma^4}\underbrace{(E[y_r^2] - (\mu_1 + \mu_0)E[y_r])}_{} + \frac{(\mu_1 - \mu_0)^2(\mu_1 + \mu_0)^2}{4\sigma^4},
\end{aligned}
$$

$$
E[y_r^2] - \underbrace{\mu_i}_{E[y_r]}E[y_r] - \mu_k\underbrace{E[y_r]}_{\mu_i},
$$

$$
\begin{aligned}
&= \underbrace{E[y_r^2] - (E[y_r])^2}_{\sigma^2} - \underbrace{\mu_k\mu_i}_{\mu_0\mu_1} \\
&= \frac{(\mu_1 - \mu_0)^2}{\sigma^4}(\sigma^2 - \mu_0\mu_1) + \frac{(\mu_1 - \mu_0)^2(\mu_1 + \mu_0)^2}{4\sigma^2}, \\
&= \frac{(\mu_1 - \mu_0)^2}{\sigma^4}\left(\sigma^2 - \mu_0\mu_1 + \frac{1}{4}(\mu_1 + \mu_0)^2\right), \\
&= \frac{(\mu_1 - \mu_0)^2}{\sigma^4}\left(\sigma^2 + \frac{1}{4}(\mu_1 - \mu_0)^2\right). \quad\quad\quad (2.33)
\end{aligned}
$$

Now, to calculate the variance, substitute Equations (2.31) and (2.33) into (2.32):

$$
\begin{aligned}
Var(\Delta x_r) &= \frac{(\mu_1 - \mu_0)^2}{\sigma^4}\left(\sigma^2 + \frac{1}{4}(\mu_1 - \mu_0)^2\right) - \frac{(\mu_1 - \mu_0)^4}{4\sigma^4}, \\
&= \frac{(\mu_1 - \mu_0)^4}{\sigma^2} + \frac{(\mu_1 - \mu_0)^4}{4\sigma^4} - \frac{\mu_1 - \mu_0)^4}{4\sigma^4},
\end{aligned}
$$

$$
\boxed{Var(\Delta x_r) = \frac{(\mu_1 - \mu_0)^2}{\sigma^2} \equiv D^2.} \quad\quad\quad (2.34)
$$

The total distance that the decision variable has traveled after n observations is given by

$$x_n = \sum_{r=1}^{n} \Delta x_r = \sum_{r=1}^{n} \log \left[\frac{p_1(y_r)}{p_0(y_r)} \right]. \qquad (2.35)$$

In logarithmic variables, the trajectory $\overrightarrow{x_n}$ is a discrete-time, biased random walk (with $x_0 = 0$), since the increments Δx_r are iid. As the time between arriving increments of information goes to zero (equivalently, as sampling becomes continuous), this process approaches the continuous-time process $x(t)$. In this limit, the SPRT approaches the DDM, which is of interest because it has been demonstrated to be a biologically relevant model for decision making, and because it can be converted to a Kolmogorov or Fokker-Planck Equation (FPE), as discussed in the following section, which we can solve.

2.2 DDM $\rightarrow$ Fokker-Planck Equation (FPE)

We are interested in converting the DDM to a FPE because we can solve it to find expressions for the measures of performance that we are interested in. To see the equivalence between the FPE and the DDM (SDE), consider the time development of an arbitrary function $f[x(t)]$. If we use Ito's Formula (Equation (B.11) in the Appendix), we get

$$
\begin{aligned}
\frac{\langle df[x(t)]\rangle}{dt} &= \left\langle \frac{df[x(t)]}{dt} \right\rangle = \frac{d}{dt} \langle f[x(t)]\rangle \\
&= \left\langle a[x(t),t]\partial_x f + \frac{1}{2}(b[x(t),t])^2 \partial_x^2 f + b[x(t),t]\partial_x f\, dW(t) \right\rangle \\
&= \left\langle a[x(t),t]\partial_x f + \frac{1}{2}(b[x(t),t])^2 \partial_x^2 f \right\rangle,
\end{aligned}
$$

since $\langle dW(t)\rangle = 0$. Note that for a discrete random variable $X(\omega)$, we denote its

mean value by

$$
\langle X\rangle = \sum_{\omega} X(\omega)P(\omega).
$$

In the discrete case, ω represents all (countably specifiable) basic events, and $P(\omega)$ is

the probability of the set containing only the single event ω. If $X(\omega)$ is a continuous

variable, then

$$
\langle X\rangle = \int_{\omega=\Omega} X(\omega)p(\omega)d\omega.
$$

We know that $x(t)$ has a conditional probability density function $p(x,t|x_0,t_0)$.

Using this, we find

$$\frac{d}{dt}\langle f[x(t)]\rangle = \left\langle \frac{df[x(t)]}{dt} \right\rangle = \int \frac{df[x(t)]}{dt} p(x,t|x_0,t_0)dx$$

$$= \int a[x(t),t]p(x,t|x_0,t_0)\partial_x f dx$$

$$+\frac{1}{2}\int (b[x(t),t])^2 p(x,t|x_0,t_0)\partial_x^2 f dx. \qquad (2.36)$$

Now we integrate the first term of Equation (2.36) by parts:

Let $\qquad u = a[x(t),t]p(x,t|x_0,t_0) \qquad\qquad dv = \partial_x f[x(t)]dx$

$$du = \partial_x\left\{a[x(t),t]p(x,t|x_0,t_0)\right\}dx \qquad v = f[x(t)]$$

$$\int a[x(t),t]p(x,t|x_0,t_0)\partial_x f[x(t)]dx = uv|_{\text{boundary}} -$$

$$\int f[x(t)]\partial_x\left\{a[x(t),t]p(x,t|x_0,t_0)\right\}dx.$$

The boundary term disappears because we will be considering a system with absorbing boundaries; see Section 2.3 below.

Similarly, integrating the second term by parts twice, we get

$$\int f[x(t)]\partial_x^2\left\{(b[x(t),t])^2 p(x,t|x_0,t_0)\right\}dx.$$

69

Note that

$$\frac{d}{dt} \langle f[x(t)] \rangle = \frac{d}{dt} \int f(x) p(x, t|x_0, t_0) dx = \int f(x) \partial_t p(x, t|x_0, t_0) dx.$$

Combining results, we get

$$\int f[x(t)] \partial_t p(x, t|x_0, t_0) dx = \int f[x(t)] \left(-\partial_x \left\{ a[x(t), t] p(x, t|x_0, t_0) \right\} + \frac{1}{2} \partial_x^2 \left\{ (b[x(t), t])^2 p(x, t|x_0, t_0) \right\} \right) dx.$$

Since $f[x(t)]$ is arbitrary, we now have the relation

$$\partial_t p(x, t|x_0, t_0) = -\partial_x \left\{ a[x(t), t] p(x, t|x_0, t_0) \right\} + \frac{1}{2} \partial_x^2 \left\{ (b[x(t), t])^2 p(x, t|x_0, t_0) \right\},$$

$$(2.37)$$

which is the forward Fokker-Planck Equation (fFPE) for a 1-dimensional system.

2.3   Some General Solutions to the 1-Dimensional FPE

For our stochastic decision variable $x(t)$, consider the SDE

70

$$dx = a(x)dt + \sqrt{b(x)}dW$$

$$x(0) = x_0.$$

(2.38)

Suppose that H_1 is correct, so the first passage of x through B_0 corresponds to an incorrect decision. We now solve for general expressions for the average performance using the fFPE, Equation (2.37), which is a second-order parabolic partial differential equation (PDE). We have an initial condition, but also require boundary conditions for the system to be solvable.

For our system, we will use absorbing boundary conditions. Absorbing boundary conditions mean that as soon as the decision variable touches the boundary, it is removed from the system; hence the boundary absorbs. Therefore, the probability of being on the boundary is zero, and the boundary conditions are

$$p(x, t) = 0 \quad \text{for } x \in (B_0, B_1).$$

(2.39)

See Section B.5 in the Appendix for more detail on boundary conditions.

Now that we have appropriate boundary conditions for our system, we can solve for the mean DT and probability of exit through each boundary for the general problem. We modify the analysis in [36] to solve for our results.

2.3.1   Mean Decision Time

Let T be the time that the decision variable exits (B_0, B_1): in other words, the time when the DM reaches a decision. We assume that the decision variable begins at x_0 at time $t = 0$, with $B_0 < x_0 < B_1$. We want to find the mean DT as a function of initial position, $\mathcal{T}(x_0)$. This is equivalent to a first passage time problem for a particle undergoing drift-diffusion in an interval.

To do this, we erect absorbing boundaries at B_0 and B_1. Thus, if $x \in (B_0, B_1)$, it has never left (or touched the boundaries of) that interval. Given the above, the probability that $x \in (B_0, B_1)$ at time t is given by integrating the probability that the decision variable is at a given location at time t given that it started at x_0 at time $t = 0$, over the interval. We define this to be $g(x_0, t)$:

$$\int_{B_0}^{B_1} p(\xi, t | x_0, 0) d\xi \equiv g(x_0, t).$$

Note that $g(x_0, t) = P(T \geq t)$.

Since the system is time homogeneous (the transition probability density is invariant to translations in time),

$$p(\xi, t | x_0, 0) = p(\xi, 0 | x_0, -t).$$

In first passage time problems, it is useful to use the backward Fokker-Planck Equation (bFPE). For the bFPE, solutions exist for times $\tau \leq t$, and the bFPE

expresses the development of the system in τ, or in backwards time. In this sense,

$$p(\xi, 0|x_0, 0) = \delta(x_0 - \xi)$$

is more of a final condition rather than an initial condition. The bFPE is equivalent to the fFPE, but provides a more convenient form.

The bFPE is

$$\partial_t p(\xi, t|x_0, 0) = a(x_0)\partial_{x_0} p(\xi, t|x_0, 0) + \frac{1}{2}b(x_0)\partial_{x_0}^2 p(\xi, t|x_0, 0), \qquad (2.40)$$

with

$$p(\xi, 0|x_0, 0) = \delta(x_0 - \xi). \qquad (2.41)$$

We can integrate Equation (2.40) over the interval (B_0, B_1) to get an equation in terms of $g(x_0, t)$:

$$\int_{B_0}^{B_1} \partial_t p(\xi, t|x_0, 0)d\xi = \int_{B_0}^{B_1} \left(a(x_0)\partial_{x_0} p(\xi, t|x_0, 0) + \frac{1}{2}b(x_0)\partial_{x_0}^2 p(\xi, t|x_0, 0) \right) d\xi,$$

$$\underbrace{\partial_t \int_{B_0}^{B_1} p(\xi, t|x_0, 0)d\xi}_{g(x_0,t)} = \underbrace{a(x_0)\partial_{x_0} \int_{B_0}^{B_1} p(\xi, t|x_0, 0)d\xi}_{g_0(x_0,t)} + \underbrace{\frac{1}{2}b(x_0)\partial_{x_0}^2 \int_{B_0}^{B_1} p(\xi, t|x_0, 0)d\xi}_{g_0(x_0,t)},$$

$$\partial_t g(x_0, t) = a(x_0)\partial_{x_0} g(x_0, t) + \frac{1}{2}b(x_0)\partial_{x_0}^2 g(x_0, t), \qquad (2.42)$$

73

with

$$g(x_0, 0) = \begin{cases} 1 & B_0 < x < B_1, \\ 0 & \text{elsewhere.} \end{cases} \tag{2.43}$$

If $x_0 = B_0$ or B_1, the decision variable is immediately absorbed, so

$$g(B_0, t) = g(B_1, t) = 0. \tag{2.44}$$

Since $g(x_0, t)$ is equivalent to $P(T \geq t)$, the mean of a function $f(T)$ is

$$\langle f(T) \rangle = -\int_0^\infty f(t) dg(x_0, t).$$

Let $f(T) = T$. Then we have that the mean decision time of the decision variable is $\mathcal{T}(x_0) = \langle T \rangle$, where

$$\mathcal{T}(x_0) = \int_0^\infty t \partial_t g(x_0, t) dt. \tag{2.45}$$

Note that $\mathcal{T}(x_0) = \int_0^\infty g(x_0, t) dt$, by integration by parts:

$$\text{Let } u = t, \quad dv = \partial_t g(x_0, t) dt,$$

$$du = dt, \qquad v = g(x_0, t).$$

$$\mathcal{T}(x_0) \;=\; -\left(uv\big|_0^\infty - \int_0^\infty v\,du \right),$$

$$\phantom{\mathcal{T}(x_0)} \;=\; -tg(x_0,t)\big|_0^\infty + \int_0^\infty g(x_0,t)dt,$$

$$\mathcal{T}(x_0) = \int_0^\infty g(x_0,t)dt, \tag{2.46}$$

since $g(x_0,\infty)=0$ (we assume that the decision is made in finite time).

We can derive an ordinary differential equation for $\mathcal{T}(x_0)$:

$$\underbrace{\int_0^\infty \partial_t g(x_0,t)dt}_{=g(x_0,\infty)-g(x_0,0)=-1} \;=\; \int_0^\infty \left(a(x_0)\partial_{x_0}g(x_0,t) + \frac{1}{2}b(x_0)\partial_{x_0}^2 g(x_0,t) \right) dt,$$

$$ \;=\; a(x_0)\partial_{x_0}\underbrace{\int_0^\infty g(x_0,t)dt}_{\mathcal{T}(x_0)} + \frac{1}{2}b(x_0)\partial_{x_0}^2 \underbrace{\int_0^\infty g(x_0,t)dt}_{\mathcal{T}(x_0)},$$

so we have

$$a(x_0)\partial_{x_0}\mathcal{T}(x_0) + \frac{1}{2}b(x_0)\partial_{x_0}^2\mathcal{T}(x_0) = -1 \tag{2.47}$$

with the boundary conditions

$$\mathcal{T}(B_0) = \mathcal{T}(B_1) = 0. \tag{2.48}$$

75

Similarly, for $\mathcal{T}^{(n)}(x_0) = \langle T^n \rangle$, we have that

$$\mathcal{T}^{(n)}(x_0) = \int_0^\infty t^{n-1} g(x_0, t) dt$$

and

$$-n\mathcal{T}^{(n-1)}(x_0) = a(x_0) \partial_{x_0} \mathcal{T}^{(n)}(x_0) + \frac{1}{2} b(x_0) \partial_{x_0}^2 \mathcal{T}^{(n)}(x_0), \qquad (2.49)$$

so moments of the decision time can be found through repeated integration.

Letting

$$\psi(\xi) = \exp\left[\int_{B_0}^\xi \frac{2a(\omega)}{b(\omega)} d\omega\right], \qquad (2.50)$$

$$f(\xi) = \frac{2}{\psi(\xi)} \int_{B_0}^\xi \frac{\psi(s)}{b(s)} ds, \qquad (2.51)$$

$$F_{x_1}^{x_2} = \int_{x_1}^{x_2} \frac{1}{\psi(\xi)} d\xi, \qquad (2.52)$$

$$G_{x_1}^{x_2} = \int_{x_1}^{x_2} f(\xi) d\xi, \qquad (2.53)$$

the mean DT is [13, 36]

$$\mathcal{T}(x_0) = \frac{F_{B_0}^{x_0} G_{x_0}^{B_1} - F_{x_0}^{B_1} G_{B_0}^{x_0}}{F_{B_0}^{B_1}}. \qquad (2.54)$$

76

2.3.2 Probability of Exit Through Each Boundary

Let $\Pi_i(x_0)$ be the probability with which the DM chooses hypothesis H_i for a given initial condition x_0. In trials where H_1 is true, then $\Pi_0(x_0)$ sets the ER and $\Pi_1(x_0)$ sets the PC.

Probability of Exit Through B_0

The total probability that the decision variable exits through B_0 after time t, given that it started at x_0, is given by the time integral of the probability current (flux) through B_0. We define this probability as

$$
\begin{aligned}
g_0(x_0, t) \;&=\; -\int_t^\infty J(B_0, s|x_0, 0)ds, & (2.55)\\
&= \int_t^\infty \left(-a(B_0)p(B_0, s|x_0, 0) + \frac{1}{2}\partial_{B_0}\left[b(B_0)p(B_0, s|x_0, 0) \right] \right) ds. & (2.56)
\end{aligned}
$$

The negative sign indicates that the flux points left.

Given that the decision variable exits through B_0, the probability that it exits after time t is

$$
P(T_0 > t) = \frac{g_0(x_0, t)}{g_0(x_0, 0)}, \tag{2.57}
$$

where we assume that T_0 is the time at which the decision variable exits through B_0, conditioned on the DM choosing H_0.

Since $p(B_0, t|x_0, 0)$ satisfies a bFPE, we have that

$$a(x_0)\partial_{x_0}g_0(x_0, t) + \frac{1}{2}b(x_0)\partial_{x_0}^2 g_0(x_0, t) = -\int_t^\infty \partial_s J(B_0, s|x_0, 0)ds, \quad (2.58)$$

$$= J(B_0, t|x_0, 0), \quad (2.59)$$

$$= \partial_t g_0(x_0, t). \quad (2.60)$$

Given that the decision variable exits through B_0, the mean exit time for a decision variable starting at x_0 and exiting through B_0 is $\mathcal{T}_0(x_0)$:

$$\mathcal{T}_0(x_0) = -\int_0^\infty t\partial_t P(T_0 > t)dt \quad (2.61)$$

$$= \int_0^\infty \frac{g_0(x_0, t)}{g_0(x_0, 0)}dt. \quad (2.62)$$

Let $\Pi_0(x_0)$ be the probability that the decision variable starting at the initial condition x_0 exits through B_0. Thus,

$$\Pi_0(x_0) = g_0(x_0, 0). \quad (2.63)$$

We can find an expression for this by integrating Equation (2.60) with respect to t:

$$\int_0^\infty \partial_t g_0(x_0, t) dt = \int_0^\infty \left[a(x_0) \partial_{x_0} g_0(x_0, t) + \frac{1}{2} b(x_0) \partial_{x_0}^2 g_0(x_0, t) \right] dt,$$

$$g_0(x_0, \infty) - g_0(x_0, 0) = a(x_0) \partial_{x_0} \left[\int_0^\infty g_0(x_0, t) dt \right] +$$
$$\frac{1}{2} b(x_0) \partial_{x_0}^2 \left[\int_0^\infty g_0(x_0, t) dt \right]. \tag{2.64}$$

Note that $g_0(x_0, \infty) = 0$ and

$$\int_0^\infty g_0(x_0, t) dt = \int_0^\infty \frac{g_0(x_0, t)}{g_0(x_0, 0)} g_0(x_0, 0) dt,$$
$$= g_0(x_0, 0) \int_0^\infty \frac{g_0(x_0, t)}{g_0(x_0, 0)} dt,$$
$$= \Pi_0(x_0) \mathcal{T}_0(x_0). \tag{2.65}$$

Then Equation (2.64) becomes

$$-\Pi_0(x_0) = a(x_0) \partial_{x_0} \left[\Pi_0(x_0) \mathcal{T}_0(x_0) \right] + \frac{1}{2} b(x_0) \partial_{x_0}^2 \left[\Pi_0(x_0) \mathcal{T}_0(x_0) \right]. \tag{2.66}$$

The corresponding boundary conditions are

$$\Pi_0(B_0) \mathcal{T}_0(B_0) = \Pi_0(B_1) \mathcal{T}_0(B_1) = 0. \tag{2.67}$$

The first term indicates that it takes zero time for a decision variable starting on

B_0 to reach the B_0 boundary, and the second term indicates that the probability of exiting through B_0 when the decision variable starts at B_1 is zero.

If we let $t \to 0$ in Equation (2.58), we see that $J(B_0, 0|x_0, 0)$ must vanish if $B_0 \neq x_0$, since our initial condition is $p(B_0, 0|x_0, 0) = \delta(x_0 - B_0)$. Therefore, the right-hand side of Equation (2.60) goes to zero. On the left-hand side of the equation, $g_0(x_0, t) \to g_0(x_0, 0) = \Pi_0(x_0)$, and we have

$$a(x_0)\partial_{x_0}\Pi_0(x_0) + \frac{1}{2}b(x_0)\partial_{x_0}^2\Pi_0(x_0) = 0,$$

$$\Pi_0(B_0) = 1,$$

$$\Pi_0(B_1) = 0,$$

$$\Pi_0(x_0) + \Pi_1(x_0) = 1.$$

The last condition comes from our requirement that the system reaches a decision in finite time.

From [13], we find that the solution can be given in terms of the functions in Equations (2.50) - (2.53). The solution is

$$\Pi_0(x_0) = \frac{F_{x_0}^{B_1}}{F_{B_0}^{B_1}}. \tag{2.68}$$

Note that the solution given by Equations (5.2.189) - (5.2.190) in [36] for this equation does not check out in Mathematica, whereas the solution above (from

[13]) does.

Probability of Exit Through B_1

By definition, $\Pi_1(x_0)$ is given by $\Pi_1(x_0) = 1 - \Pi_0(x_0)$; however, this can also be found the same way we found $\Pi_0(x_0)$: let $g_1(x_0, t)$ be the total probability that the decision variable exits through B_1 after time t. Since the flux now points to the right, we have

$$g_1(x_0, t) = \int_t^\infty J(B_1, s | x_0, 0) ds, \tag{2.69}$$

$$= \int_t^\infty \left(a(B_1) p(B_1, s | x_0, 0) - \frac{1}{2} \partial_{B_1} \left(b(B_1) p(B_1, s | x, 0) \right) \right) ds. \tag{2.70}$$

Let $\mathcal{T}_1(x_0)$ be the mean exit time for a decision variable starting at x_0 and exiting through B_1, given that the decision variable will eventually exit through B_1. Note that

$$P(T_1 > t) = \frac{g_1(x_0, t)}{g_1(x_0, 0)},$$

$$\mathcal{T}_1(x_0) = \int_0^\infty \frac{g_1(x_0, t)}{g_1(x_0, 0)} dt,$$

$$\int_t^\infty p(B_1, s | x, 0) ds = g_1(x_0, t).$$

Like before, $p(B_1, t | x_0, 0)$ satisfies a bFPE:

$$\partial_t p(B_1, t | x_0, 0) = a(x_0)\partial_{x_0} p(B_1, t | x_0, 0) + \frac{1}{2}b(x_0)\partial_{x_0}^2 p(B_1, t | x_0, 0), \qquad (2.71)$$

which can be integrated with respect to t to give

$$a(x_0)\partial_{x_0} g_1(x_0, t) + \frac{1}{2}b(x_0)\partial_{x_0}^2 g_1(x_0, t) = \int_t^\infty \partial_s J(b, s | x_0, 0)ds, \qquad (2.72)$$

$$= -J(B_1, t | x_0, 0), \qquad (2.73)$$

$$= \partial_t g_1(x_0, t). \qquad (2.74)$$

This in turn can be integrated with respect to t to get

$$\Pi_1(x_0) = a(x_0)\partial_{x_0}\left[\Pi_1(x_0)\mathcal{T}_1(x_0)\right] + \frac{1}{2}b(x_0)\partial_{x_0}^2\left[\Pi_1(x_0)\mathcal{T}_1(x_0)\right], \qquad (2.75)$$

with $\Pi_1(B_0)\mathcal{T}_1(B_0) = \Pi_1(B_1)\mathcal{T}_1(B_1) = 0$. Then, taking the limit as $t \to 0$, we get

$$a(x_0)\partial_{x_0}\Pi_1(x_0) + \frac{1}{2}b(x_0)\partial_{x_0}^2\Pi_1(x_0) = 0$$

$$\Pi_1(B_0) = 0$$

$$\Pi_1(B_1) = 1$$

$$\Pi_1(B_0) + \Pi_0(B_1) = 1.$$

This has the solution

$$\Pi_1(x_0) = \frac{F_{B_0}^{x_0}}{F_{B_0}^{B_1}}. \tag{2.76}$$

2.4 Solutions for the Constant-Drift Constant-Diffusion Case

We now specialize the above general solutions to the SPRT-based DDM, which has constant coefficients. This allows us to simplify the expressions found in Section 2.3.

Consider the system described by Equations (2.23) and (2.24): here, the boundaries at B_0 and B_1 are fixed at $-z$ and $+z$, respectively, and bias is captured in the initial condition x_0.

83

Non-Dimensionalizing the System

It is sometimes simpler to do computations with a non-dimensionalized system, which we solve for here. This follows a simplification step shown in [13]. We can scale Equation (2.23) by $\frac{1}{m}$ to get

$$\frac{1}{m}dx = dt + \frac{D}{m}dW. \tag{2.77}$$

We now introduce new variables:

$$\chi = \frac{1}{m}x \Rightarrow d\chi = \frac{1}{m}dx, \tag{2.78}$$

$$\zeta = \frac{1}{m}z > 0, \tag{2.79}$$

$$\vartheta = \left(\frac{m}{D}\right)^2 > 0. \tag{2.80}$$

Substituting Equations (2.78) - (2.80) into Equations (2.23) - (2.24), we now have the non-dimensionalized system

$$d\chi = dt + \frac{1}{\sqrt{\vartheta}}dW, \tag{2.81}$$

$$\chi(0) = \frac{1}{m}x_0 \equiv \chi_0, \tag{2.82}$$

with boundaries at $\pm\zeta$.

Some Useful Functions

For the non-dimensionalized, constant-drift, constant-diffusion problem defined by Equations (2.81) - (2.82), we get relatively simple expressions for the mean ER and mean DT, since $a(x) = 1$ and $b(x) = \frac{1}{\vartheta}$. Substituting these into Equations (2.50) - (2.53), we get

$$\psi(\xi) = e^{2\vartheta(\xi+\zeta)}, \qquad f(\xi) = (1 - e^{-2\vartheta(\xi+\zeta)}), \qquad F_{-\zeta}^{\chi_0} = \frac{1}{2\vartheta}(1 - e^{-2(\zeta+\chi_0)\vartheta}), \quad (2.83)$$

$$F_{\chi_0}^{\zeta} = \frac{e^{-2\zeta\vartheta}}{2\vartheta}(e^{-2\chi_0\vartheta} - e^{-2\zeta\vartheta}), \qquad F_{-\zeta}^{\zeta} = \frac{1}{2\vartheta}(1 - e^{-4\zeta\vartheta}), \quad (2.84)$$

$$G_{-\zeta}^{\chi_0} = \zeta + \chi_0 + \frac{1}{2\vartheta}(e^{-2(\zeta+\chi_0)\vartheta} - 1), \quad (2.85)$$

$$G_{\chi_0}^{\zeta} = \zeta - \chi_0 + \frac{1}{2\vartheta}e^{-2\zeta\vartheta}(e^{-2\zeta\vartheta} - e^{-2\chi_0\vartheta}). \quad (2.86)$$

We can plug Equations (2.84) - (2.86) into the general formulas for the probability of exit through each boundary and mean DT to get the solutions for the constant case.

2.4.1 Mean Decision Time

Substituting into Equation (2.54), the general expression for the mean DT, gives

$$\mathcal{T}(\chi_0) = \frac{F_{-\zeta}^{\chi_0} G_{\chi_0}^{\zeta} - F_{\chi_0}^{\zeta} G_{-\zeta}^{\chi_0}}{F_{-\zeta}^{\zeta}},$$

$$= \frac{\chi_0 - e^{4\vartheta\zeta}\chi_0 + (1 + e^{4\vartheta\zeta} - 2e^{2\vartheta(\zeta-\chi_0)})\zeta}{e^{4\vartheta\zeta} - 1}. \qquad (2.87)$$

If $\chi_0 = 0$, the expression simplifies considerably:

$$\mathcal{T}(\chi_0 = 0) = \frac{\zeta(1 + e^{4\vartheta\zeta} - 2e^{2\vartheta\zeta})}{e^{4\vartheta\zeta-1}},$$

$$= \zeta\frac{(e^{2\vartheta\zeta} - 1)^2}{e^{4\vartheta\zeta} - 1},$$

$$= \zeta\frac{e^{2\vartheta\zeta} - 1}{e^{2\vartheta\zeta} + 1},$$

$$= \zeta\tanh(\vartheta\zeta).$$

This gives a convenient formula for the constant case:

$$\mathcal{T}(\chi_0) = \zeta\tanh(\zeta\vartheta) + \left(\frac{2\zeta(1 - e^{-2\chi_0\vartheta})}{e^{2\zeta\vartheta} - e^{-2\zeta\vartheta}} - \chi_0\right). \qquad (2.88)$$

For systems with unbiased initial data and $\chi_0 = 0$, the terms in the parentheses in Equation (2.88) disappear.

2.4.2 Probability of Exit

Probability of Exit Through B_0

Substitution gives

$$
\begin{aligned}
\Pi_0(\chi_0) = \frac{F_{\chi_0}^{\zeta}}{F_{-\zeta}^{\zeta}} &= \frac{\frac{e^{-2\zeta\vartheta}}{2\vartheta}\left(e^{-2\chi_0\vartheta} - e^{-2\zeta\vartheta}\right)}{\frac{1}{2\vartheta}\left(1 - e^{-4\zeta\vartheta}\right)}, \\
&= \frac{e^{-2\zeta\vartheta}\left(e^{-2\chi_0\vartheta} - e^{-2\zeta\vartheta}\right)}{1 - e^{-4\zeta\vartheta}}.
\end{aligned}
\tag{2.89}
$$

If $\chi_0 = 0$, the above expression simplifies slightly:

$$
\begin{aligned}
\Pi_0(\chi_0)_{\chi_0=0} &= \frac{e^{-2\zeta\vartheta}\left(1 - e^{-2\zeta\vartheta}\right)}{1 - e^{-4\zeta\vartheta}}, \\
&= \frac{e^{2\vartheta\zeta} - 1}{e^{4\vartheta\zeta} - 1}, \\
&= \frac{1}{1 + e^{2\zeta\vartheta}}.
\end{aligned}
$$

Therefore,

$$
\Pi_0(\chi_0) = \frac{1}{1 + e^{2\zeta\vartheta}} - \left(\frac{1 - e^{-2\chi_0\vartheta}}{e^{2\zeta\vartheta} - e^{-2\zeta\vartheta}}\right),
\tag{2.90}
$$

where the terms in the parentheses disappear when $\chi_0 = 0$. Note that it is possible to make $\Pi_0(\chi_0)$ arbitrarily small, by choosing ζ (location of the boundary)

87

sufficiently large.

Probability of Exit Through B_1

Since $\Pi_0(\chi_0) + \Pi_1(\chi_0) = 1$,

$$\Pi_1(\chi_0) = \frac{e^{2\vartheta\zeta}}{1 + e^{2\vartheta\zeta}} + \left(\frac{e^{2\vartheta\zeta}(1 - e^{-2\chi_0\vartheta})}{e^{4\vartheta\zeta} - 1} \right), \tag{2.91}$$

where the terms in the parentheses disappear when $\chi_0 = 0$.

2.5 Distribution of Decision Times

Let $g(\chi_0, t)$ be the total probability that the decision variable starting at χ_0 will exit through a particular boundary after time t. We will initially drop the extra subscripts for notational simplicity, and then later specialize g to g_0 and g_1, for exit through B_0 and B_1, respectively. This means that the cumulative distribution function (cdf) for choosing H_i, given the initial condition χ_0, is given by

$$q_i(t) = \Pi_i(\chi_0) - g_i(\chi_0, t). \tag{2.92}$$

We denote the cdf as a function of time only, because we assume that the initial condition χ_0 is a constant.

As shown in Equation (2.60), $g(\chi_0, t)$ satisfies a bFPE,

$$\partial_t g(\chi_0, t) = \partial_{\chi_0} g(\chi_0, t) + \frac{1}{2\vartheta} \partial^2_{\chi_0} g(\chi_0, t), \qquad (2.93)$$

for the non-dimensionalized constant system. Now suppose that $g(\chi_0, t)$ can be found using separation of variables, i.e., $g(\chi_0, t) = \phi(t)\gamma(\chi_0)$. Substituting this into the bFPE and rearranging, we have

$$\frac{\partial_t[\phi(t)]}{\phi(t)} = \frac{\partial_{\chi_0}[\gamma(\chi_0)] + \frac{1}{2\vartheta}\partial^2_{\chi_0}[\gamma(\chi_0)]}{\gamma(\chi_0)} = -\lambda.$$

Since the left-hand side is only a function of t and the right-hand side is only a function of χ_0, the only way for the equation to hold true is if they are both equal to $-\lambda$, where λ is a constant. This leads to two sets of equations: the space equation

$$\frac{d^2\gamma(\chi_0)}{d\chi_0{}^2} + 2\vartheta\frac{d\gamma(\chi_0)}{d\chi_0} = -2\vartheta\lambda\gamma(\chi_0), \qquad (2.94)$$

and the time equation

$$\frac{d\phi(t)}{dt} = -\lambda\phi(t). \qquad (2.95)$$

We solve these separately below.

Space Equation

The space equation can be transformed into Sturm-Liouville form (see Section B.6 in the Appendix) using the integrating factor $e^{2\vartheta\chi_0}$. This gives

$$-\frac{d}{d\chi_0}\left[e^{2\vartheta\chi_0}\gamma'(\chi_0)\right] = 2\vartheta\lambda e^{2\vartheta\chi_0}\gamma(\chi_0). \tag{2.96}$$

Thus, the weighting function, which we will denote as $\varpi(\chi_0)$, is

$$\varpi(\chi_0) = e^{2\vartheta\chi_0}. \tag{2.97}$$

Assuming that the solution of Equation (2.94) is of the form $\gamma(\chi_0) = e^{r\chi_0}$, we can substitute into the above and factor out $e^{r\chi_0}$ to get

$$r^2 + 2\vartheta r + 2\vartheta\lambda = 0.$$

We can solve this equation for r via the quadratic formula:

$$r = -\vartheta \pm \sqrt{\vartheta^2 - 2\vartheta\lambda}.$$

For notational simplicity, let

$$\beta = \sqrt{\vartheta^2 - 2\vartheta\lambda}. \tag{2.98}$$

This indicates that $\gamma(\chi_0)$ is of the form

$$\gamma(\chi_0) = C_1 e^{(-\vartheta+\beta)\chi_0} + C_2 e^{(-\vartheta-\beta)\chi_0}. \tag{2.99}$$

Now we apply the absorbing boundary conditions. We know that $g(B_0, t) = g(B_1, t) = 0$: a decision variable starting at either boundary is considered to have already exited the system. Since $\phi(0) \neq 1$ and $g(B_0, 0) = g(B_1, 0) = 0$, this means that

$$\gamma(B_0) = \gamma(B_1) = 0.$$

We can plug this into Equation (2.99):

$$\gamma(B_0) = \gamma(-\zeta) = C_1 e^{(-\vartheta+\beta)(-\zeta)} + C_2 e^{(-\vartheta-\beta)(-\zeta)} = 0,$$

$$\gamma(B_1) = \gamma(\zeta) = C_1 e^{(-\vartheta+\beta)\zeta} + C_2 e^{(-\vartheta-\beta)\zeta} = 0.$$

These equations can be solved to yield the relations

$$C_1 = -C_2 e^{-2\beta\zeta} = -C_2 e^{2\beta\zeta}.$$

Thus, $2\beta\zeta = ik\pi$, and we can solve for an equivalent expression for β:

$$\beta = \frac{ik\pi}{2\zeta}, \tag{2.100}$$

for $k \in \mathbb{Z}$. Since we assume that the location of the boundary (ζ) is constant, β is

a function of k only. Thus, we have that

$$C_1 = \begin{cases} -C_2 & k \text{ even,} \\ \\ C_2 & k \text{ odd.} \end{cases} \tag{2.101}$$

We can solve for λ by equating the two values found for β in Equations (2.98) and (2.100):

$$\lambda_k = \frac{1}{2}\left(\vartheta + \frac{\pi^2 k^2}{4\vartheta\zeta^2}\right), \tag{2.102}$$

where we write λ as λ_k to indicate its dependence on k.

For simplicity, let $\nu = C_1 = \pm C_2$. This gives

$$\gamma_k(\chi_0) = \nu e^{-\vartheta\chi_0}\left(e^{\frac{i\pi k}{2\zeta}\chi_0} + (-1)^{k+1}e^{-\frac{i\pi k}{2\zeta}\chi_0}\right). \tag{2.103}$$

Note that this is equivalent to

$$\gamma_k(\chi_0) = \begin{cases} 2\nu\cos(\frac{\pi k}{2\zeta}\chi_0)e^{-\vartheta\chi_0} & \text{for } k \text{ odd,} \\ \\ 2i\nu\sin(\frac{\pi k}{2\zeta}\chi_0)e^{-\vartheta\chi_0} & \text{for } k \text{ even.} \end{cases} \tag{2.104}$$

This is a solution for any $k \in \mathbb{Z}$.

Now we can solve for ν. We want $\langle\gamma_k(\chi_0), \gamma_\ell(\chi_0)\rangle = \delta_{k\ell}$, where $\delta_{k\ell}$ is the Dirac delta function, since we would like these functions to form an orthonormal basis

with the weighting function $\varpi(\chi_0)$. We choose ν such that this relationship holds. We also require that $\gamma_k(\chi_0)$ be real, since it represents a probability, and thus define the inner product accordingly. In particular,

$$\langle \gamma_k, \gamma_\ell \rangle = \int_{-\zeta}^{\zeta} \gamma_k \cdot \gamma_\ell \cdot \varpi(\chi_0) d\chi_0 = 0, \quad k \neq \ell. \tag{2.105}$$

This results in

$$\nu = \pm \frac{i^{-k+1}}{2\sqrt{\zeta}}. \tag{2.106}$$

Both the (+) and (-) solutions of ν lead to a $\gamma_k(\chi_0)$ that is always real. We will denote the (+) solution by ν_+ and the ($-$) solution by ν_-. One can verify that for both values of ν, the eigenfunctions γ_k are orthonormal with the weighting function $\varpi(\chi_0)$ defined in Equation (2.97). We also find that this value of ν leads to a $\gamma(\chi_0)$ that is always real, and thus is a reasonable solution.

We now have an equation for $\gamma_k(\chi_0)$:

$$\gamma_k(\chi_0) = \begin{cases} \pm \frac{1}{\sqrt{\zeta}} \cos(\frac{\pi k}{2\zeta} \chi_0) e^{-\vartheta \chi_0} & \text{for } k \text{ odd}, \\ \\ \mp \frac{1}{\sqrt{\zeta}} \sin(\frac{\pi k}{2\zeta} \chi_0) e^{-\vartheta \chi_0} & \text{for } k \text{ even}, \end{cases} \tag{2.107}$$

where the sign of each solution for each case are given in Table 2.5.

Solution	ν_+	ν_-
k odd, $k = 1 + 2m$	m even: -	m even: +
	m odd: +	m odd: -
k even, $k = 2m$	m even: +	m even: -
	m odd: -	m odd: +

Table 2.1: The signs of the different cases of Equation (2.107) are shown here as a quick reference. Calculating $\gamma_k(\chi_0)$ for each possible value of k (using $m \in \mathbb{R}$) verifies that $\gamma_k(\chi_0) \in \mathbb{R}$.

By linearity of solutions, we now have a general form for $\gamma(\chi_0)$:

$$\gamma(\chi_0) = \sum_k c_k \gamma_k(\chi_0), \tag{2.108}$$

where c_k are constant coefficients whose value depends only on k. To find exact expressions for these coefficients, we solve for $g_i(\chi_0, t)$, the total probability that the decision maker, who will (eventually) decide H_i, has not yet reached a decision $(i = 0, 1)$.

By definition, $g_i(\chi_0, 0) = \Pi_i(\chi_0)$, so we can solve for the coefficients c_{ik}:

$$g_i(\chi_0, 0) = \sum_k c_{ik} \gamma_k(\chi_0) = \Pi_i(\chi_0),$$

$$\sum_k c_{ik} \gamma_k(\chi_0) \gamma_\ell(\chi_0) \varpi(\chi_0) = \Pi_i(\chi_0) \gamma_\ell(\chi_0) \varpi(\chi_0),$$

$$\int_{-\zeta}^{\zeta} \sum_k c_{ik} \gamma_k(\chi_0) \gamma_\ell(\chi_0) \varpi(\chi_0) d\chi_0 = \int_{-\zeta}^{\zeta} \Pi_i(\chi_0) \gamma_\ell(\chi_0) \varpi(\chi_0) d\chi_0,$$

$$\sum_k c_{ik} \underbrace{\int_{-\zeta}^{\zeta} \gamma_k(\chi_0) \gamma_\ell(\chi_0) \varpi(\chi_0) d\chi_0}_{\delta_{k\ell}} = \int_{-\zeta}^{\zeta} \Pi_i(\chi_0) \gamma_\ell(\chi_0) \varpi(\chi_0) d\chi_0,$$

94

which gives

$$c_{ik} = \int_{-\zeta}^{\zeta} \Pi_i(\chi_0)\gamma_k(\chi_0)\varpi(\chi_0)d\chi_0. \tag{2.109}$$

We can now calculate exact expressions for the c_{ik} for each boundary, $i = 0, 1$.

Time Equation

It is clear from the form of the time equation that the solution is an exponential, so let

$$\phi(t) = ce^{-\lambda t},$$

where c is a constant. Since $g_i(\chi_0, 0) = \Pi_i(\chi_0) = \phi(0)\gamma(\chi_0)$, without loss of generality, we take $\phi(0) = 1 = c$. Substituting the value we found for λ_k in Equation (2.102), we have

$$\phi_k(t) = e^{-\frac{1}{2}\left(\vartheta + \frac{\pi^2 k^2}{4\vartheta\zeta^2}\right)t}. \tag{2.110}$$

2.5.1 Distribution of Decision Times for Passage Through B_0

Here we solve for $g_0(\chi_0, t)$, the total probability that the DM, who will eventually decide H_0, has not yet reached a decision.

95

Using Equation (2.90) for $\Pi_0(\chi_0)$, we can solve for c_{0k}:

$$c_{0k} = \frac{2(-1)^{k+1}e^{-\vartheta\zeta}k\pi\sqrt{\zeta}}{k^2\pi^2 + 4\vartheta^2\zeta^2}. \tag{2.111}$$

General Solution for Passage Through B_0

We now have

$$
\begin{aligned}
g_0(\chi_0, t) &= \sum_k c_{0k}\gamma_k(\chi_0)\phi_k(t) \tag{2.112}\\
&= \sum_k c_{0k}\nu\left[e^{\left(-\vartheta+\frac{i\pi k}{2\zeta}\right)\chi_0} + (-1)^{k+1}e^{\left(-\vartheta-\frac{i\pi k}{2\zeta}\right)\chi_0}\right]e^{-\frac{1}{2}\left(\vartheta+\frac{\pi^2 k^2}{4\vartheta\zeta^2}\right)t}\\
&= \sum_k \frac{-i^{k+1}e^{-\vartheta\zeta}k\pi}{k^2\pi^2 + 4\vartheta^2\zeta^2}\left[e^{\left(-\vartheta+\frac{i\pi k}{2\zeta}\right)\chi_0} + (-1)^{k+1}e^{\left(-\vartheta-\frac{i\pi k}{2\zeta}\right)\chi_0}\right]e^{-\frac{1}{2}\left(\vartheta+\frac{\pi^2 k^2}{4\vartheta\zeta^2}\right)t}.
\end{aligned}
$$

The cdf of exit times through B_0 is

$$q_0(t) = \Pi_0(\chi_0) - g_0(\chi_0, t) \tag{2.113}$$

and the pdf of exit times through B_0 is given by

$$
\begin{aligned}
p_0(t) \ &= \ \frac{d}{dt}\left[q_0(t)\right] = -\frac{d}{dt}g_0(\chi_0, t), \\
&= \ \sum_k \left(\frac{-i^{k+1}k\pi}{8\vartheta\zeta^2}\exp\left[\left(\frac{k^2\pi^2 t}{8\vartheta\zeta^2} - \frac{ik\pi\chi_0}{2\zeta} - \frac{1}{2}\vartheta\left[t + 2(\chi_0 + \zeta)\right]\right)\right] \right. \\
&\qquad \left. \cdot \left((-1)^k - \exp\left[\frac{ik\pi\chi_0}{\zeta}\right]\right) \right).
\end{aligned}
\tag{2.114}
$$

Mean Decision Time for Choosing H_0

We defined the mean DT for choosing B_0 in Equation (2.62). Now that we have a formula for $g_0(\chi_0, t)$ which is separated into space and time, we can explicitly find a formula for $\mathcal{T}_0(\chi_0)$:

$$
\begin{aligned}
\mathcal{T}_0(\chi_0) \ &= \ \frac{1}{\Pi_0(\chi_0)}\int_0^\infty \sum_k c_{0k}\gamma_k(\chi_0)\phi_k(t)dt, \\
&= \ \frac{1}{\Pi_0(\chi_0)}\sum_k c_{0k}\gamma_k(\chi_0)\int_0^\infty \phi_k(t)dt, \\
&= \ \frac{1}{\Pi_0(\chi_0)}\sum_k c_{0k}\gamma_k(\chi_0)\phi_k(t)\left(\frac{-8\vartheta\zeta^2}{k^2\pi^2 + 4\vartheta^2\zeta^2}\right).
\end{aligned}
\tag{2.115}
$$

2.5.2 Distribution of Decision Times for Passage Through B_1

We now solve for the coefficients c_{1k} for the distribution of DTs for passage through B_1.

$$
\begin{aligned}
c_{1k} &= \int_{-\zeta}^{\zeta} \Pi_1(\chi_0)\gamma_k(\chi_0)\varpi(\chi_0)d\chi_0, \\
&= \frac{2e^{\vartheta\zeta}k\pi\sqrt{\zeta}}{k^2\pi^2 + 4\vartheta^2\zeta^2}.
\end{aligned}
\tag{2.116}
$$

General Solution for Passage Through B_1

We now have

$$
\begin{aligned}
g_1(\chi_0, t) &= \sum_k c_{1k}\gamma_k(\chi_0)\phi_k(t), \\
&= \sum_k \left((i)^{-k+1}\frac{e^{\vartheta\zeta}k\pi}{k^2\pi^2 + 4\vartheta^2\zeta^2} \left[e^{(-\vartheta + \frac{i\pi k}{2\zeta})\chi_0} + (-1)^{k+1}e^{(-\vartheta - \frac{i\pi k}{2\zeta})\chi_0} \right] \right. \\
&\quad \left. \cdot e^{-\frac{1}{2}(\vartheta + \frac{\pi^2 k^2}{4\vartheta\zeta^2})t} \right).
\end{aligned}
\tag{2.117}
$$

The cdf, $q_1(t)$ is then given by

$$
q_1(t) = \Pi_1(\chi_0) - g_1(\chi_0, t),
\tag{2.118}
$$

and the corresponding pdf is

$$p_1(t) = \sum_k \left(\frac{i^{-k+1}k\pi}{8\vartheta\zeta^2} \exp\left[\frac{1}{8}\left(-4\vartheta(t+2\chi_0-2\zeta) - \frac{k^2\pi^2 t}{\vartheta\zeta^2} - \frac{4ik\pi\chi_0}{\zeta} \right) \right] \right.$$
$$\left. \cdot \left((-1)^k - \exp\left[\frac{i\pi k\chi_0}{\zeta} \right] \right) \right). \tag{2.119}$$

Mean Decision Time for Choosing H_1

As before, we can now find the mean DT for choosing B_1:

$$\begin{aligned}
\mathcal{T}_1(\chi_0) &= \frac{1}{\Pi_1(\chi_0)} \int_1^\infty \sum_k c_{1k}\gamma_k(\chi_0)\phi_k(t)dt, \\
&= \frac{1}{\Pi_1(\chi_0)} \sum_k c_{1k}\gamma_k(\chi_0) \int_0^\infty \phi_k(t)dt, \\
&= \frac{1}{\Pi_1(\chi_0)} \sum_k c_{1k}\gamma_k(\chi_0)\phi_k(t) \left(\frac{-8\vartheta\zeta^2}{k^2\pi^2 + 4\vartheta^2\zeta^2} \right). \tag{2.120}
\end{aligned}$$

2.6   Simulations

Since the SPRT operates in discrete time, we expected that the results from simulating the SPRT in a 2AFC task would approach the values calculated above as the step size decreased towards zero. Since we are interested in tasks that are fairly difficult, it is reasonable to expect that the decision variable for the SPRT will move with a small step size. In this section, we verify this intuition as well as our analytical solutions via simulation.

2.6.1 A Simple Test Case

Here, we present an example to make our earlier analysis concrete. We take

$$\mu_0 = 0.9,$$

$$\mu_1 = 1,$$

$$\sigma = 1, \qquad\qquad (2.121)$$

$$x_0 = 0,$$

$$\alpha_0 = \alpha_1 \equiv \varepsilon = 0.01.$$

We assume H_1 is true, so $\text{ER} = \Pi_0(\chi_0)$.

Analytical Solution

From Equations (2.31), (2.34), (2.78) - (2.80) and the above values, we have

$$m = 0.005,$$

$$D^2 = 0.01,$$

$$\zeta = 919.02, \qquad\qquad (2.122)$$

$$\vartheta = 0.25,$$

$$\chi_0 = 0.$$

Plugging this into our solutions, we find that for the above values, $\Pi_0(\chi_0) = 0.01$ and the mean DT is $\mathcal{T}(\chi_0) = 900.643$ seconds.

Figures 2.2(a) and 2.2(b) show the constants c_{ik} for $i = 0, 1$ respectively, for the values given above. As expected, the coefficients for the correct hypothesis (c_{1k}) are

100

significantly larger in magnitude than the coefficients for the incorrect hypothesis (c_{0k}).

We can now use these values to calculate the theoretical cdfs $q_i(t)$ and pdfs $p_i(t)$ for $i = 0, 1$. We show $q_0(t)$ in Figure 2.3(a), $p_0(t)$ in Figure 2.3(b), $q_1(t)$ in Figure 2.3(c), and $p_1(t)$ in Figure 2.3(d). As expected, the cdfs asymptote to the set ER $q_0(t) \to 0.01$ and PC $q_1(t) \to 0.99$.

Since our solution includes an infinite sum, we also investigated the effect of truncating terms. We found that for our test case, the even-k terms in the series were equal to zero. In Figures 2.4(a), 2.4(b), 2.4(c), and 2.4(d), we show $q_0(t)$, $p_0(t)$, $q_1(t)$, and $p_1(t)$, respectively, plotted using an increasing number of odd k terms in the sum. As k increases, the left tail of the plot more approaches zero more quickly, and the areas under the pdfs approach the correct value (0.01 for p_0, and 0.99 for p_1). The spike near zero is an artifact from truncating the sum, which is consistent with our expectation that the probability that the decision variable exits at time zero be zero. The distributions reach their limiting shapes very quickly: the cdfs achieve the final shape essentially with only three terms, and the pdfs reach their approximate final shape with only four terms. This indicates that we can simplify the infinite sum considerably, and achieve similar results with a very low number of terms.

(a) c_{0k} for the Simple Test Case

(b) c_{1k} for the Simple Test Case
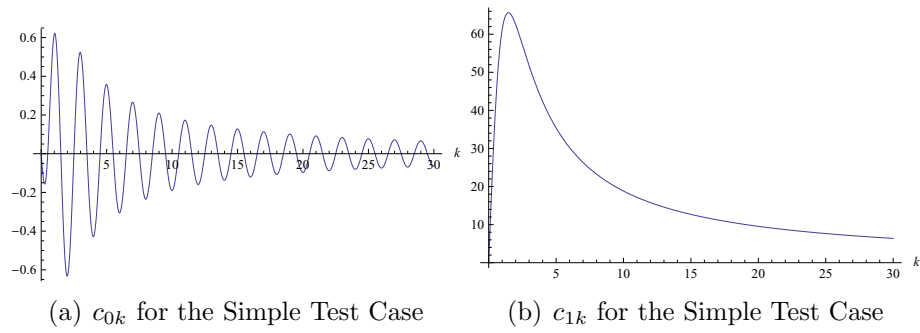
Figure 2.2: (a) Plot of c_{0k} versus k, for the simple test case in Section 2.6.1. (b) Plot of c_{1k} versus k. The parameters used in the simulation are given in Equation (2.121). In the test case, H_1 is correct; thus, the coefficients in (b) are much larger than those in (a).

(a) $q_0(t)$ for the Simple Test Case

(b) $p_0(t)$ for the Simple Test Case

(c) $q_1(t)$ for the Simple Test Case

(d) $p_1(t)$ for the Simple Test Case
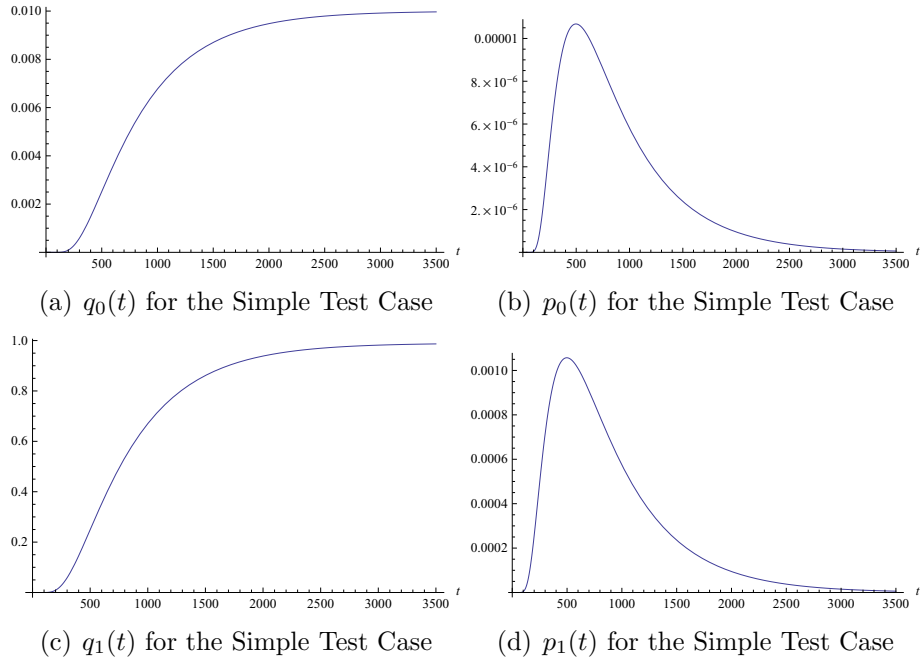
Figure 2.3: Graphs of the cdfs and pdfs of DTs for passage through each boundary. Our graphs show the first 20 odd-k terms of the sum. The cdf graphs' first 10 seconds and the pdf graphs' first 35 seconds were set to zero to remove the spurious tail. (a) Plot of the cdf of DTs for passage through B_0 for the simple test case in Section 2.6.1. The cdf levels off at 0.01, as expected. (b) Plot of the pdf of DTs for passage through B_0. The area under the curve approaches 0.01 as the number of terms increases, as expected. (c) Plot of the cdf of DTs for passage through B_1. The cdf levels off at 0.99, as expected. (d) Plot of the pdf of DTs for passage through B_1. The area under the curve approaches 0.01 as the number of terms increases, as expected.

102

(a) $q_0(t)$, Varying Number of Terms

(b) $p_0(t)$, Varying Number of Terms

(c) $q_1(t)$, Varying Number of Terms

(d) $p_1(t)$, Varying Number of Terms
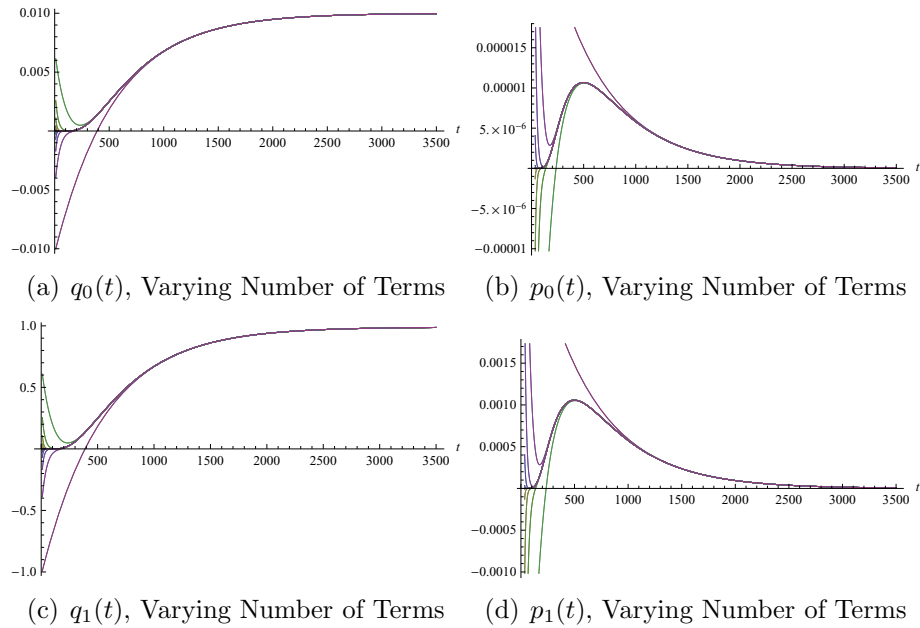
Figure 2.4: Plots of the pdfs and cdfs for passage through each boundary for the simple test case in Section 2.6.1, using increasing numbers of terms k in the sum. As discussed in the text, only terms for which k is odd contribute, since the even-k terms are zero. In all of the graphs, increasing the number of k terms pushes the spurious spike near zero further to the left. (a) $q_0(t)$, the cdf of exit times through B_0. (b) $p_0(t)$, the pdf of exit times through B_0. (c) $q_1(t)$, the cdf of exit times through B_1. (d) $p_1(t)$, the pdf of exit times through B_1.

Simulation Results

The solutions found above for a DM's pdf of DTs can now be verified using simulation: in theory, running a simulation of a DM using the SPRT to perform a 2AFC task is equivalent to randomly selecting a DT from $p(t) = p_1(t) + p_0(t)$, the pdf of DTs (for either hypothesis), or flipping an appropriately biased coin to choose between H_0 and H_1 and then randomly selecting a DT from the appropriate pdf. Simulation of a single DM employing the SPRT in a 2AFC test with the parameter values in Equations (2.121) yielded an actual ER $= 0.009587$ and DT $= 913.1323$, averaged over 1 million trials. The simulation took 937.675 seconds to run. This is very close to the predicted values: there is only a 2% difference in time and 4% difference in ER from the analytical results. This is verified in Figure 2.5, where the histogram of DTs over 10,000 trials for an individual using the SPRT have been compared with the analytical result $p(t)$ derived in this section.

2.6.2   Investigating the Small-Overshoot Assumption

Our test case in Section 2.6.1 verified our analytical results. Next, we explore the effect of changing $\Delta\mu = \mu_1 - \mu_0$. We found that the results were relevant to the small-overshoot assumption taken in Equations (2.6) and (2.7).

In our simulation, the location of the boundaries B_0 and B_1 depends only on α_0 and α_1. The amount by which the decision variable moves after receiving each observation depends only on the two distributions generating the data. Since we

Individual DM
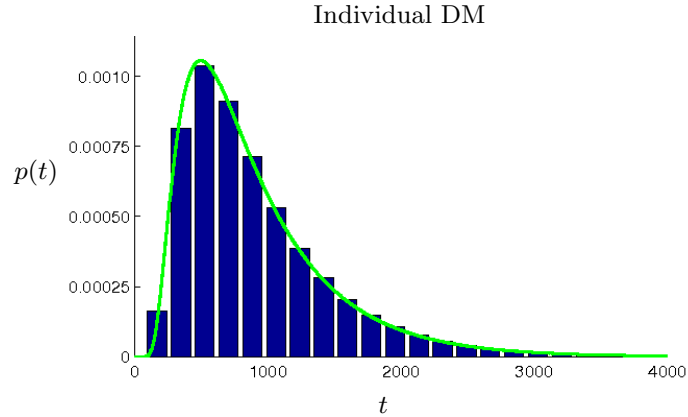
Figure 2.5: Histogram of DTs for a DM using the SPRT over 10,000 trials, normalized by the number of trials and bin width, and overlaid with the analytically calculated pdf of DTs, $p(t) = p_1(t) + p_0(t)$. This verifies that our DDM solutions accurately model an individual DM using the SPRT. In this simulation, we set $\mu_0 = 0.9$, $\mu_1 = 1$, $\sigma = 1$, $x_0 = 0$, and $\alpha_0 = \alpha_1 = 0.01$.

took the simplifying assumption above that the two distributions differ only in mean, the distance the decision variable moves in each step depends only on the difference of the means of the distributions ($\Delta\mu = \mu_1 - \mu_0 = d'$ for our example). Thus, the locations of the boundaries are independent of the difficulty of the task, even though the decision variable's step size (and therefore overshoot) increases with increasing d'. (A larger value of d' indicates a simpler task.) Therefore, we were interested in seeing how a simpler or harder task compared with our test case. We highlight the effect of overshoot from selecting a large value for d' with an example below.

A Too-Simple Test Case

We ran a simulation identical to the test case in Section 2.6.1 above, except that we set $\mu_0 = 0$, which makes $d' = 1$.

Translating this into the variables used in our analysis, we have that $m = \frac{1}{2}$, $D^2 = 1$, $\zeta \approx 9.19024$, and $\vartheta = \frac{1}{4}$. Plugging these values into our analytical solutions, we find that $\Pi_0(\chi_0) = 0.01$ and $\mathcal{T} = 9.00643$.

Simulations of a 2AFC system with these parameters had a mean ER of 0.005696 and mean DT of 9.5024, averaged over 1 million trials. The simulation took 26.98 seconds to run. This is a 5% difference in DT and 43% difference in a ER. Clearly the analytical results found for the DDM are no longer accurate for the SPRT simulation. The simulation is clearly being too conservative – the ER is too low, which indicates that the task was much simpler than the DDM model predicted. If we relaxed the allowed error rate ε to 0.019, then the simulation had an average DT of 8.1237 and ER of 0.010628. This provides an average ER closer to what we expected (0.01), but the simulation still finishes significantly more quickly than expected. This indicates that the boundaries we calculated using Equations (2.6) and (2.7) become too conservative when d' increases. We first sought to characterize the overshoot and its effect on the ER, then found a way one could compensate for the overshoot to recover the expected ER.

Error and Overshoot

Since changing d' affected the mean ER more than the mean DT in our first simulation, we focused on the effect of d' on the mean ER. We set ER = 0.01, and found the average ER from simulation as a function of d' for a range of d'
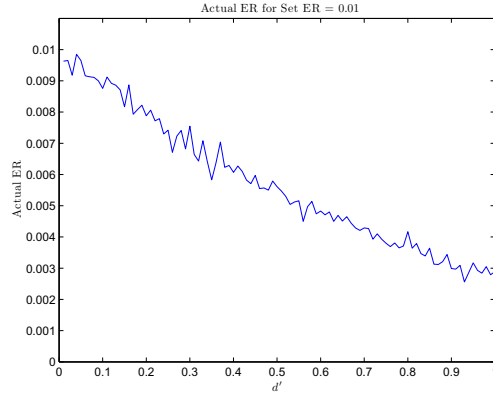
Figure 2.6: Actual ER as a function of d', for the expected ER set to 0.01, averaged over 1 million trials. The noise in the data is due to the finite number of samples used in the simulation. The plot indicates that there is an approximately linear relationship between d' and the actual ER for a set value of ER.

values. The results are shown in Figure 2.6. The plot indicates that there is an approximately linear relationship between d' and the actual ER from simulation.

We then found the average overshoot and step size as a function of d', shown in Figure 2.7(a). As we expect, the average step size is a function of d', the difficulty of the task. We also found that the average overshoot scaled with d', at approximately half the step size. The corresponding DTs for this simulation is found in Figure 2.7(b). As d' increases, the DT exponentially decreases.

Adjusting for the Overshoot

Using the characterization of the error due to overshoot found in Figure 2.7, we can introduce a simple fix to compensate for the effect of overshoot on ER in our simulations.

The error rates being too low for large d' indicates that the boundaries we used

107
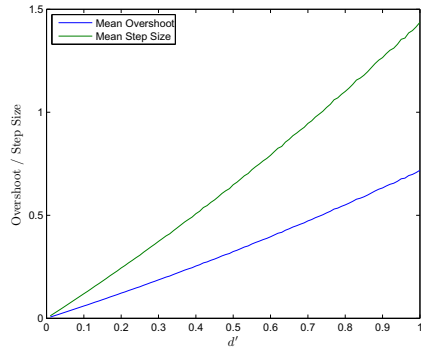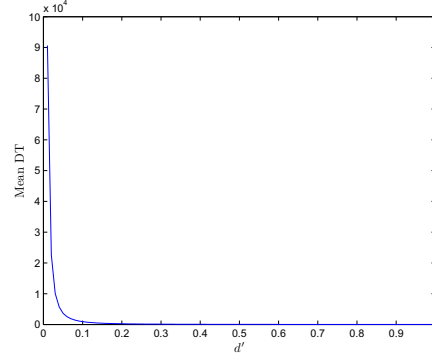
(a) Mean Overshoot and Step Size   (b) Mean DT

Figure 2.7: Overshoot, Step Size, and DT as a function of d'. (a) As we expect, the mean overshoot and step size are a function of d'. The overshoot appears to consistently scale at approximately half the step size. Using these results, we calculated a fix that corrected for the overshoot, to provide the expected ER (though the simplicity of the task was then expressed through a lower DT). (b) The corresponding mean DTs in the simulation. Our results show that the mean DT decreases exponentially with increasing d'.

were too conservative, or too far away from the decision variable's initial condition. Therefore, we ran a simulation in which we moved the boundaries inwards by the average amount of overshoot we found in our previous simulations, for each value of d'. The results are shown in Figure 2.8. The general trend of the error in the ER seems to be reversed for larger d'. This indicates that our adjustment may be too large for those values of d'. However, when compared to the ER in Figure 2.6, we see that the adjustment corrects for the overshoot well. It is possible that the trend for larger d' could be partly due to noise in the overshoot adjustment numbers, which would be noise from the finite number of trials over which we averaged the overshoot.
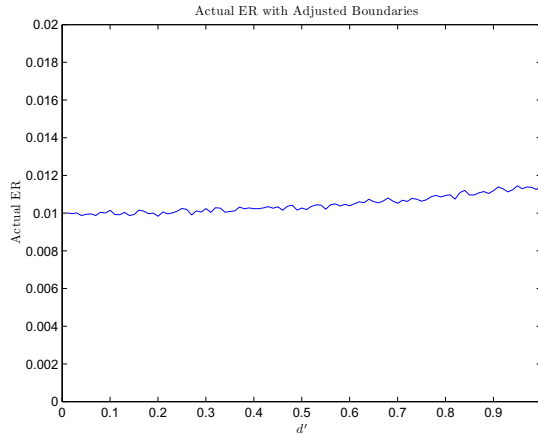
Figure 2.8: Average ER with boundaries adjusted inwards by the average overshoot. Our adjustment moves the boundaries slightly too far inwards, since the experiments with higher d' have a slightly higher error rate than desired; however, it is a very good adjustment, since the error rates for all values of d' are quite close to the desired ER of 0.01.

Actual and Set Error

Since we found that our analytical results could be too conservative for simulations with too large a d', we also investigated the relation between the actual and set error rates. Since we had reasonably good results in our simple test case, we chose $d' = 0.1$ for this comparison.

The results are shown in Figure 2.9, which shows the relationship between the set ER and the actual ER from simulation in blue, compared to a dashed line of slope 1 in green. This again shows that for smaller set ERs, the simulation will return what we expect. However, the difference is not large, which indicates that the biggest factor influencing how close a simulation's results will be to the analytical results is the choice of d'. Our results also verify that $d' = 0.1$ is a good parameter value to use, since we can expect our simulation results to match our
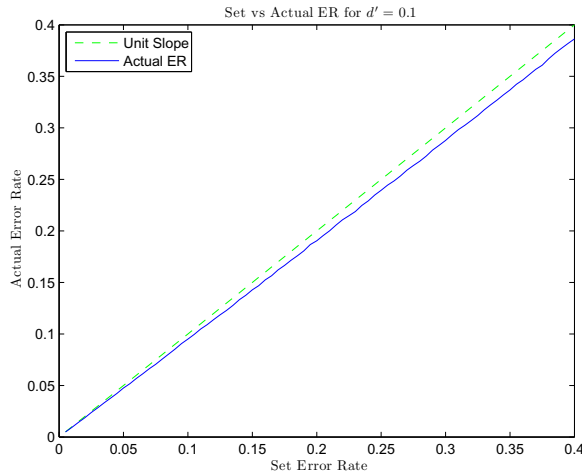
Figure 2.9: Actual versus set ERs for $d' = 0.1$ (shown in blue). The dashed green line shows a 1:1 relationship between the actual and set ERs. Each point on the plot is averaged over 1 million trials. Since the plot is approximately linear with slope close to one, the simulation confirms that there are few effects from overshoot at this value of d', even if the boundaries change.

analytical results well. We will be using these values in further examples in the next chapter.

2.7 Conclusion

We derived an explicit solution to the SPRT-based DDM, and thereby characterized the ER and pdf of DTs of a DM using the SPRT while under the understanding that the DM could possibly be a human observer or a detector (device).

Using these results as a building block for the performance of an individual, we now show how the performance of a group of DMs can be characterized in the next chapter.

Chapter 3

Group Decision-Making Models

3.1 Introduction

Our society has a long tradition of valuing the wisdom of groups. Groups offer the potential for redundancy and robustness, and are generally thought of as being more cautious, more creative, more informed, and more accurate than individuals. Many studies in social psychology have been dedicated to supporting or refuting these beliefs from a relatively qualitative social standpoint [35, 90, 102, 103]; see [26, 28] for reviews. Systems with human operators are qualitatively considered in [89]. A more quantitative approach to how human groups make decisions in a restricted task has been pursued in cognitive psychology [53, 58, 87, 88]. These studies generally focus on group accuracy and the weight placed on each individual's opinion as the main measures of performance. The group performance is typically

compared to that of an "Ideal Group", which represents the best that the group can do given the members in it, in order to find the group's efficiency [86]. The experiments cited above use fixed-sample procedures, are typically data-driven, and are generally based on Signal Detection Theory [40]; also see [59]. We will explore this general area and philosophy further in Chapter 4. A separate group of work in economics and political science [6, 7, 11, 46, 47] focuses heavily on various aspects of Condorcet's (Jury) Theorem [21]. These works use more theoretical modeling than the psychology literature, and take an approach to a group's error rate similar to the one we present later. These human-group-based studies inspire our current work as well as future directions and considerations in designing a cybernetic (including both humans and devices) group-based decision-making system.

A more mathematics- and engineering-based approach to group decision-making is taken in the design of distributed decision-making systems that utilize multiple sensors. This has been an intensive area of study, particularly in the past thirty years. These works generally focus on a collection of devices minimizing a cost or risk function in choosing between two hypotheses, and thereby assert the resulting decision-making scheme as "optimal". A very good introduction to and overview of decentralized detection is presented in [94], and a mathematical approach to sequential decision theory can be found in [8, 82]. Centralized systems, in which measurements from peripheral sensors are sent to a fusion center for processing, are largely considered a finished problem; thus, most literature focuses on decentralized

systems. Despite the name, in most decentralized systems, a minimal amount of processing is done at the peripheral sensors: a compressed "decision" based on a single observation is sent to the fusion center for processing at each time step. Decentralized decision-making systems, particularly ones which consider sequential processes, are a very large area of research [12, 51, 64, 67, 91, 92, 96, 97, 105], with many variations, including the following: general hypotheses [32, 95], multiple hypotheses [31, 60], quickest detection problems [25], sequential test truncation [52, 79], and different group communication topologies [65, 93]. Various applications are also considered, including networks with constrained communication [19, 33, 50, 56], networks with power constraints [57], vehicle classifiers [49], and probabilistic search [20]. Though our particular models are different from the ones cited here, these studies illustrate a very different philosophy on how a collective can make a decision, and show the methods and applications of interest in strictly device-based engineering setups.

In the previous chapter, we derived an explicit solution to a SPRT-based DDM, while focusing on the probability distribution function (pdf) of decision times (DTs) and average error rate (ER) as the primary measures of performance. Here, we begin by briefly discussing notation and some useful identities. In Part I, we derive the group error rate (GER) and pdf of group decision times (GDTs) for N independent decision-makers (DMs) using one of three simple group decision rules: Race, Majority Total, and Majority First. We illustrate each general solution with an

example using N independent and identically distributed (iid) DMs characterized by our individual solutions, and verify our analytical results via simulation. We also mention some similar group decision-making rules from the literature. In Part II, we compare the iid performance of the different group decision rules under different constraints, and discuss the relative merits of each scheme. In Part III, we show an example of a general (non-identical) group. We demonstrate the accuracy of our general solutions through simulation and discuss the effect of having non-identical group members. Finally, in Part IV, we generalize the Majority Total scheme to the η-Total rule, and the Majority First scheme to the η-First rule to further demonstrate the power of our approach.

The methods used to calculate the group's performance and general design philosophy can be extended to other group decision rules and more complicated hierarchical group topologies. We do not specify if the DMs are (human) observers, or detectors (devices), since our group models can accommodate a group of either type as well as a general group whose members' individual characteristics differ, as long as the individuals are independent. For simplicity, our simulations assume that each individual DM has quantifiable abilities and adequate motivation, and for now, we assume the DMs can be modeled using the analysis we presented in the previous chapter. Our goal is to establish a way to objectively and quantitatively compare different group decision rules and model group performance in an intuitive manner that is accessible to a wide range of communities.

3.1.1 Notation

We denote a pdf generically as $p(t)$, but will occasionally distinguish between the pdf of a general individual DM, an iid individual DM, and the group through the subscript. In the general individual DM case, we assume that the ith DM, denoted DM_i, has a defined individual cdf $q_i(t)$ and pdf $p_i(t)$ of DTs, with the single constraint that the DMs are independent of each other. In other words, the individual DMs in the group can have very different cdfs and pdfs of DTs. In the iid case, as the name suggests, we assume that the DMs are iid, so each DM is described by the generic cdf $q_\iota(t)$ and pdf $p_\iota(t)$. In addition, we may also use subscripts to indicate the decision made. To do this, we will use notation inspired by Signal Detection Theory [40, 59]. Without loss of generality, let N ("[N]oise only") represent H_0, and S ("[S]ignal present") represent H_1. Then, $p_N(t) = p(t|H_0)$ and $p_S(t) = p(t|H_1)$. Similarly, $p_{iN}(t) = p_i(t|H_0)$ and $p_{iS}(t) = p_i(t|H_1)$. Note that N, representing H_0, differs from N, the total number of DMs. We will denote the pdf of GDTs by $p_g(t)$. For the group pdf only, we will indicate the group decision rule, type of individual DMs in the group, and the number of DMs in the group in a superscript, which will be shown later for each scheme.

Let t_g be the time at which the group makes a decision. Let the shorthand notation "L" denote that the DM reached a decision at some time $t \leq t_g$, and "G" denote that the DM decides at some time $t \geq t_g$.

We previously referred to the error rate of an individual with "ER". Now that

errors can also occur at the group level, we distinguish between the Group Error Rate (GER) and the Local Error Rate (LER). The GER represents the error rate of the fusion center's decision, and The LER refers to the error rate of an individual DM. If the DMs are general, we specify the LER of DM_i by LER_i when necessary.

3.1.2 Some Useful Identities

Here we explicitly state some identities which will be highly useful in this chapter. By definition, the integral of a pdf $p(t)$ over all time must sum to unity. We can split this integral into two parts, defining the cutoff point as t_g. Then,

$$\int_0^\infty p(\tau)d\tau = \int_0^{t_g} p(\tau)d\tau + \int_{t_g}^\infty p(\tau)d\tau = 1. \tag{3.1}$$

Equivalently, the probability that a DM makes a decision before time t_g plus the probability that it makes a decision after time t_g is equal to unity, because our setup requires that the DM make a decision in a finite amount of time.

Also by definition, the integral of the pdf from time 0 to time t_g is the cumulative probability distribution function (cdf) at time t_g, $q(t_g)$:

$$\int_0^{t_g} p(\tau)d\tau \equiv q(t_g). \tag{3.2}$$

116

The first two identities can be combined to yield

$$\int_{t_g}^{\infty} p(\tau)d\tau = 1 - q(t_g), \tag{3.3}$$

the cumulative probability that a decision is made after time t_g.

By definition, the derivative of the cdf is the pdf. We can also use the Fundamental Theorem of Calculus to show that

$$\frac{d}{dt_g}[q(t_g)] = \frac{d}{dt_g}\left[\int_0^{t_g} p(\tau)d\tau\right] = p(t_g). \tag{3.4}$$

Equations (3.1) and (3.4) can be combined to yield

$$\frac{d}{dt_g}[1 - q(t_g)] = \frac{d}{dt_g}\left[\int_{t_g}^{\infty} p(\tau)d\tau\right] = -p(t_g). \tag{3.5}$$

Lastly, given that H_1 and H_0 are the only two possible hypotheses in the experiment, we have that

$$q(t) = q_S(t)P(S) + q_N(t)P(N), \tag{3.6}$$

$$p(t) = p_S(t)P(S) + p_N(t)P(N), \tag{3.7}$$

where $P(S)$ (resp., $P(N)$) is the prior probability that S (resp., N) is correct.

3.2   Part I: Simple Group Decision Rules

We present the analytical and simulation-based results for a group of N independent DMs using one of three simple group decision rules: Race, Majority Total, and Majority First. In these schemes, the individual DMs can only communicate with the fusion center, as shown in Figure 3.1. We will also briefly discuss related group decision making models from the literature.

In our analytical results, we derive the solution for both N general DMs and N iid DMs, then use the iid case as an example to show how group size affects the pdf of GDTs.

In our simulations, each DM takes and processes one sample of data at each time step. As soon as a DM chooses a hypothesis, it sends that decision to the fusion center: if the DM chooses H_1 (resp., H_0), it submits $+1$ (resp., -1). In each time step, the fusion center analyzes the decisions that have arrived in that time step, and checks to see if the group decision rule has been satisfied. If the group decision rule is not yet satisfied, the fusion center does nothing, and the process repeats in the next time step. While the DMs and fusion center in our simulations are synchronized for convenience, we stress that our analysis and solutions are directly applicable to completely asynchronous systems.

We will generally assume that after a DM has sent a decision to the fusion center, it shuts itself down to save power, and that the process ends once the fusion center returns a decision. All of the group decision rules presented here work in
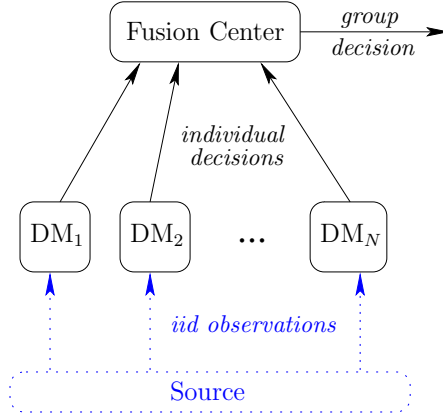
Figure 3.1: Illustration of how our group models are organized. Each DM takes and processes iid observations of the source (which represents the correct hypothesis). The observations are iid both within and across DMs. Once a DM makes a decision, it sends that decision to the fusion center, which then applies the group decision rule and issues the group's decision once the group rule has been satisfied.

cases where the fusion center can send a message to the individual DMs so they shut down after the fusion center's process ends as well as cases where the fusion center cannot send out messages.

In our analytical and simulated numerical examples below, we continue the Simple Test case from Section 2.6.1, and take $\mu_0 = 0.9$, $\mu_1 = 1$, $\sigma = 1$, $x_0 = 0$, and $\alpha_0 = \alpha_1 = 0.01$.

3.2.1 Race Scheme

In the Race scheme, the fusion center simply accepts the decision of the fastest DM. It is a race in the sense that only the first individual decision made counts towards the group's decision. We expect that the Race scheme will have the fastest GDT of all of our schemes, including a single DM. Thus, we are primarily interested

in this scheme as a benchmark for speed. An equivalent group decision scheme was proposed in [1, 2]; however, that treatment only finds the cdf of GDTs and qualitatively discusses the scheme, dismissing a full analysis of the pdf of GDTs as too complex.

Below, we first discuss the GER under the Race scheme, then solve for the pdf of GDTs for both the general and iid cases.

Race Scheme GER

For the Race scheme with general DMs, we expect the GER to be equal to the LER of the fastest DM. On average, we expect the GER to be approximately equal to the LER of the DM with the lowest mean DT. In the case where all of the DMs use the SPRT-based DDM (but are not necessarily identically distributed), the GER is given by $\max_i(\mathrm{LER}_i)$, $i = 1, ...N$, where LER_i is the LER of the ith DM, which we will denote DM_i. If the DMs are iid, then GER = LER.

Race Scheme GDT for General DMs

Consider the case where there is a group of general DMs using the Race group decision rule. We first solve the $N = 2$ case, then the $N = 3$ case, and finally extend the results to the arbitrary N case for general independent DMs.

Race GDT for $N = 2$ General DMs   Here we derive an expression for the cdf and pdf of GDTs. For the group to make a decision by time t_g, one of the following

120

three mutually exclusive cases must occur: (i) DM_1 makes a decision before t_g and DM_2 makes a decision after t_g; (ii) DM_1 makes a decision after t_g and DM_2 makes a decision before t_g; or (iii) both DM_1 and DM_2 make a decision before time t_g. Using these cases, we can define the cdf of GDTs, $q_g^{\text{rg2}}(t_g)$, where the superscript specifies the decision rule, type of individual DMs, and number of DMs ([r]ace scheme, [g]eneral individual DMs, $N = [2]$ DMs):

$$q_g^{\text{rg2}}(t_g) = q_1(t_g)\left[1 - q_2(t_g)\right] + \left[1 - q_1(t_g)\right]q_2(t_g) + q_1(t_g)q_2(t_g). \qquad (3.8)$$

To find an expression for $p_g^{\text{rg2}}(t_g)$, we take $\frac{d}{dt_g}$ of both sides to get:

$$p_g^{\text{rg2}}(t_g) = p_1(t_g)[1 - q_1(t_g)] + p_2(t_g)[1 - q_2(t_g)], \qquad (3.9)$$

which is intuitive: for the group to reach a decision at time t_g, either DM_1 reaches a decision at time t_g and DM_2 reaches a decision at some time after that, or vice versa, and taking the sum over all possible combinations of i and j then gives the probability that the group decides at time t_g.

Race GDT for $N = 3$ General DMs   There are multiple strategies one can use to calculate $p_g^{\text{rg3}}(t)$. Below, we show two equivalent ways of deriving the cdf of GDTs.

The first way follows the strategy used for the $N = 2$ case; however, adding another DM to the group increases the number of conditions under which the

DM$_1$	DM$_2$	DM$_3$
L	L	L
G	L	L
L	G	L
L	L	G
L	G	G
G	L	G
G	G	L

Table 3.1: All possible conditions under which a group of 3 DMs using the Race scheme reaches a decision before time t_g. "L" indicates that the DM reached a decision at some time $t \leq t_g$ ([L]ess than t_g), and "G" indicates that the DM reaches a decision at some time $t \geq t_g$ ([G]reater than t_g). Note that the combination "GGG" is not included, because if all 3 DMs take more than time t_g to reach a decision, the group will not reach a decision by time t_g.

group can reach a decision. See Table 3.1 for all possible combinations of individual decision times for this group which can lead to a group decision before or at time t_g. Accounting for these cases leads to the following cdf of GDTs:

$$
\begin{aligned}
q_g^{\text{rg3}}(t_g) = \ & q_1(t_g)q_2(t_g)q_3(t_g) + [1 - q_1(t_g)]q_2(t_g)q_3(t_g) + q_1(t_g)[1 - q_2(t_g)]q_3(t_g) \\
& + q_1(t_g)q_2(t_g)[1 - q_3(t_g)] + q_1(t_g)[1 - q_2(t_g)][1 - q_3(t_g)] \\
& + [1 - q_1(t_g)]q_2(t_g)[[1 - q_3(t_g)] + [1 - q_1(t_g)][1 - q_2(t_g)]q_3(t_g).
\end{aligned}
$$

(3.10)

The second way one can find the cdf of GDTs is to note that the first strategy covers all possible combinations except "GGG", the case where all three DMs decide after time t_g. Thus, subtracting the probability for the single case where no DM reaches a decision by time t_g from unity, we get:

122

$$q_g^{\mathrm{rg}3}(t_g) = 1 - [1 - q_1(t_g)][1 - q_2(t_g)][1 - q_3(t_g)]. \tag{3.11}$$

We note that Equations (3.10) and (3.11) are equivalent, which can be shown by using the identities outlined in Section 3.1.2.

Now that we have the cdf of GDTs, we can find the pdf by taking the derivative with respect to t_g and reducing the result by using the identities in Section 3.1.2 to get:

$$
\begin{aligned}
p_g^{\mathrm{rg}3}(t_g) = \;\; & p_1(t_g)[1 - q_2(t_g)][1 - q_3(t_g)] + [1 - q_1(t_g)]p_2(t_g)[1 - q_3(t_g)] \\
& + [1 - q_1(t_g)][1 - q_2(t_g)]p_3(t_g).
\end{aligned} \tag{3.12}
$$

Race GDT for N General DMs Following the second strategy outlined for the $N = 3$ case, we subtract the probability for the single case where no DM reaches a decision by time t_g from unity to get an expression for the cdf of group DTs, $q_g^{\mathrm{rg}N}(t_g)$ ([r]ace scheme, [g]eneral individual DMs, [N] DMs):

$$q_g^{\mathrm{rg}N}(t_g) = 1 - \prod_{i=1}^{N} [1 - q_i(t_g)]. \tag{3.13}$$

The corresponding pdf of GDTs is

$$p_g^{\mathrm{rg}N}(t_g) = \frac{d}{dt_g}\left[q_g^{\mathrm{rg}N}(t_g)\right] = \sum_{i=1}^{N}\left(p_i(t_g)\prod_{\substack{j=1,\\j\neq i}}^{N}[1-q_j(t_g)]\right). \tag{3.14}$$

Race Scheme pdf of GDTs for iid DMs

A simple example for a group decision-making problem is one in which the DMs are iid. Below, we specialize the general results to the iid case (denoted by ι). We note that one can equivalently derive these iid formulas from first principles by accounting for the number of possible combinations for each case.

Race GDT for $N = 2$ iid DMs Following the naming convention used previously, the group's cdf of GDTs is $q_g^{ri2}(t_g)$ ([r]ace scheme, [i]id individual DMs, $N = $ [2] DMs), given by

$$q_g^{ri2}(t_g) = [q_\iota(t_g)]^2 + 2q_\iota(t_g)[1 - q_\iota(t_g)]. \tag{3.15}$$

The corresponding pdf is

$$p_g^{\mathrm{ri2}}(t_g) = 2p_\iota(t_g)[1 - q_\iota(t_g)]. \tag{3.16}$$

Race GDT for $N = 3$ iid DMs When we specialize the general solution for 3 DMs to the case where the individuals are all iid, we find that the cdf of GDTs is

$$q_g^{\mathrm{ri3}} = 1 - [1 - q_\iota(t_g)]^3, \tag{3.17}$$

124

and the pdf of GDTs is

$$p_g^{\text{ri3}}(t_g) = 3p_\iota(t_g) \left[1 - q_\iota(t_g)\right]^2 . \tag{3.18}$$

Race Scheme GDT for N iid DMs As before, we can simplify $q_g^{\text{rg}N}(t_g)$ to $q_g^{\text{ri}N}(t_g)$, ([r]ace scheme, [i]id individual DMs, [N] DMs):

$$q_g^{\text{ri}N}(t_g) = 1 - \left[1 - q_\iota(t_g)\right]^N . \tag{3.19}$$

The corresponding pdf of GDTs for N iid DM is

$$p_g^{\text{ri}N}(t_g) = \frac{d}{dt_g}\left[q_g^{\text{ri}N}(t_g)\right] = N p_\iota(t_g) \left[1 - q_\iota(t_g)\right]^{N-1} . \tag{3.20}$$

Using the parameters given in Section 3.2, we can now plot the pdf of GDTs for a group of N iid DMs using the Race scheme. This is shown for $N = 1$ to 41 in Figure 3.2. As N increases, the group's pdf moves to the left, and the distribution becomes more peaked. This is consistent with what we would intuitively expect: as N increases, the probability that one of the DMs in the group draws samples that make it finish more quickly increases, so the minimum of the N individual DTs decreases.

Race Scheme for Varying N

$N = 41$

$N = 1$

$p_g^{\mathrm{ri}N}(t_g)$

t_g
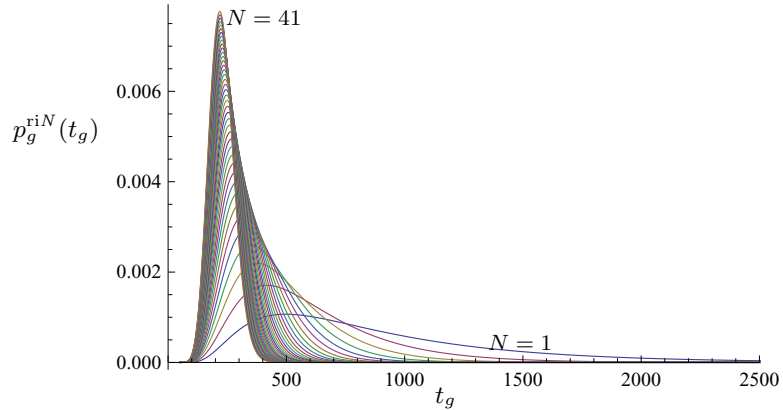
Figure 3.2: Plots of $p_g^{\mathrm{ri}N}(t_g)$, the pdf of GDTs for a group of N iid DMs using the Race scheme, where t_g denotes the GDT. In this plot, N varies from 1 to 41. As the number of DMs increases, the mean GDT decreases and the group pdf becomes more peaked. This is consistent with the intuition that a larger group of DMs using the Race scheme has a lower average GDT, and provides more information than a simple mean-only-based analysis.

Race Scheme Simulation Results

As noted in the beginning of Section 3.2, the fusion center and individual DMs operated synchronously in our simulations for convenience only, and the Race scheme does not require synchronous operation. In the unlikely event that an equal number of DMs reached opposing decisions in the same time step, we set the fusion center to cancel out the two decisions. We could have also chosen one at random or always gone with the DM_i with the lower index i; for our synchronous simulation it did not have a large impact, since the situation arose very rarely. In a more general asynchronous simulation, we expect that it would matter even less.

In Figure 3.3, we show a histogram of GDTs from simulating a group of $N = 5$ iid DMs, scaled by the number of trials and bin width and overlaid with the pdf
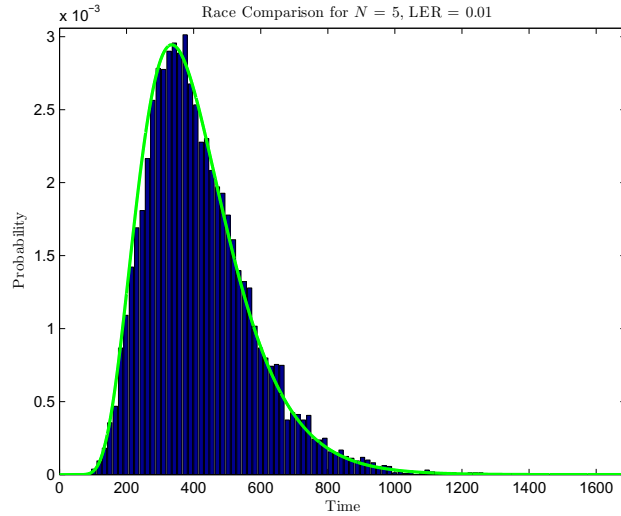
Figure 3.3: Comparison of our analytically derived DDM-based pdf of GDTs with a histogram of GDTs from a simulation of the same group of 5 iid DMs using the Race group decision rule over 10,000 trials. The histogram was scaled by the number of trials and bin width. The agreement between the pdf of GDTs and the histogram validates our analytical results for the Race scheme.

of GDTs calculated for the same group. Despite the fact that we truncated the analytical results (see Section 2.6.1), and used only 10,000 trials in the simulation, our analytical results match the simulation results well. This verifies that our analytical continuous-time solutions predict the actual discrete-time decisions made by the group whose members use the SPRT.

3.2.2 Majority Total Scheme

Here we analyze the first of two different types of Majority rule schemes, the Majority Total group decision rule. In this scheme, the fusion center waits until all N DMs have submitted a decision before declaring the group decision, which is chosen using a majority rule. To avoid ties in the fusion center decision, we consider

only N odd, though it would be simple to include additional constraints such as ignoring one DM at random or ignoring the slowest DM to account for N even. The Majority Total scheme has been mentioned in [94] as a possible group decision rule; however, to the best of our knowledge, this rule has not been analyzed in detail.

Below, we first discuss the GER of a group using either of our Majority schemes, then solve for the pdf of GDTs for the Majority Total scheme.

Majority Scheme GER

We first derive an expression for the GER of N iid DMs, then derive the GER for N general DMs. We indicate that the results below are relevant to either of our majority-based group voting schemes by stating that the following is the GER for a Majority scheme (rather than the Majority Total scheme, specifically).

Majority Scheme GER for iid DMs   Let $\Theta = \frac{N-1}{2}$ be the number of DMs in the largest minority possible ("maximal minority"), and let $\Upsilon = \frac{N+1}{2}$ be the number of DMs in the smallest majority possible ("minimal majority"). Note that by our problem definition, the probability that a DM makes an error is given by LER, and the probability that it makes a correct decision is given by (1-LER). Then, the GER for a group of N independent DMs using a Majority scheme is:

$$
\begin{aligned}
P(\text{group errs}) \quad &= P(\text{majority errs}) \equiv \text{GER} \\
&= P(N \text{ err}) + C(N,1) \cdot P(1 \text{ correct}, [N-1] \text{ err}) \\
&\quad + C(N,2) \cdot P(2 \text{ correct}, [N-2] \text{ err}) + \cdots \\
&\quad + C(N,\Theta) \cdot P(\Theta \text{ correct}, \Upsilon \text{ err}),
\end{aligned}
\tag{3.21}
$$

where $C(N,k) \in \mathbb{R}$ are scalar constants representing the number of possible ways N DMs can collectively make k correct decisions. This translates into the following equation for the overall predicted GER:

$$
\text{GER} = \sum_{k=0}^{\Theta} \binom{N}{k} \text{LER}^{N-k} \cdot (1 - \text{LER})^k.
\tag{3.22}
$$

Majority Scheme GER for General DMs Following the general form of Equation (3.21) for the GER, we can now derive the GER for a group of N general DMs using a Majority scheme, by accounting for each possible combination of DM responses instead of only the number of dissenting DMs. We can write this out for each term:

- $P(0 \text{ correct}, N \text{ err}) = \displaystyle\prod_{m=1}^{N} \text{LER}_m$

- $P(1 \text{ correct}, [N-1] \text{ err}) = \displaystyle\sum_{j=1}^{N} \left([1 - \text{LER}_j] \prod_{\substack{m=1, \\ m \neq j}}^{N} \text{LER}_m \right)$

129

- $P(2 \text{ correct}, [N-2] \text{ err}) = \sum\limits_{j_1=1}^{N-1} \sum\limits_{j_2=j_1+1}^{N} \left(\prod\limits_{\theta=1}^{2} [1 - \text{LER}_{j_\theta}] \prod\limits_{\substack{m=1, \\ m \neq \{j_1, j_2\}}}^{N} \text{LER}_m \right)$

  Note that we now refer to each DM that answers correctly using a sub-index. Also note that the limit on the first summation is $N-1$ rather than N. This prevents overcounting.

  ...

- $P(\Theta \text{ correct}, \Upsilon \text{ err}) = \sum\limits_{j_1=1}^{N-\Theta+1} \sum\limits_{j_2=j_1+1}^{N-\Theta+2} \cdots \sum\limits_{j_\Theta=j_{\Theta-1}+1}^{N} \left(\prod\limits_{\theta=1}^{\Theta} [1 - \text{LER}_{j_\theta}] \prod\limits_{\substack{m=1, \\ m \notin \mathcal{J}}}^{N} \text{LER}_m \right)$

where $\mathcal{J} = \{j_1, \ldots j_\theta\}$, the subset of DMs who answered correctly in the combination being considered.

We then combine the above equations to find the general equation for the mean GER of a group of N general DMs using a Majority scheme:

$$\text{GER} = \prod\limits_{i=1}^{N} \text{LER}_i + \sum\limits_{\theta=1}^{\Theta} \left[\sum\limits_{j_1=1}^{N-\theta+1} \sum\limits_{j_2=j_1+1}^{N-\theta+2} \cdots \sum\limits_{j_\theta=j_{\theta-1}+1}^{N} \left(\prod\limits_{k=1}^{\theta} [1 - \text{LER}_{j_k}] \prod\limits_{\substack{m=1, \\ m \notin \mathcal{J}}}^{N} \text{LER}_m \right) \right],$$

(3.23)

It has been shown that when the LER_i are not identical, the optimal procedure is to assign a weight to each DM's decision that is proportional to its accuracy [63]. Here we consider only the simple case where each DM is given only one vote. This is realistic for cases in which one cannot satisfactorily characterize or know the average performance of each DM as a constant number, or is otherwise unable

to assign non-uniform weights for other reasons. For example, in a committee of professionals, the members generally do not have the same expertise on any single matter that the committee must decide on, yet each member is typically given a single vote, which results in equal weighting. The GER for DMs with unequal weighting is discussed in the economics and political science literatures [6, 7, 47].

Majority Total Scheme pdf of GDTs for General DMs

Given the individuals' cdfs and pdfs of DTs, we can now calculate the cdf and pdf of GDTs. In the Majority Total scheme, the GDT is effectively equal to the DT of the slowest group member: the probability that a group of N general DMs using the Majority Total rule makes a decision by time t_g is the same as the probability that all N DMs make a decision by time t_g. Thus, even though we consider only N odd for the Majority schemes to avoid ties in the group decision rule, we can solve for the GDT for any N, independent of the fusion center's decision. We first solve the $N = 2$ and $N = 3$ cases as simple examples of how to calculate the pdf of GDTs, then extend the results to the case of N general DMs.

Majority Total GDT for $N = 2$ General DMs The probability that a group of 2 general DMs using the Majority Total rule makes a decision by time t_g is equal to the probability that both DMs make a decision by time t_g. We denote the cdf of GDTs for a group of 2 general DMs using the Majority Total decision rule as $q_g^{\text{mtg2}}(t_g)$ ([m]ajority [t]otal scheme, [g]eneral individual DM, $N = [2]$ DMs),

$$q_g^{\text{mtg2}}(t_g) = q_1(t_g)q_2(t_g). \tag{3.24}$$

We can take $\frac{d}{dt_g}$ of both sides to solve for the pdf of GDTs:

$$p_g^{\text{mtg2}}(t_g) = p_1(t_g)q_1(t_g) + p_2(t_g)q_1(t_g). \tag{3.25}$$

This intuitively makes sense - the probability that the group finishes at time t_g is equal to the sum of the probability that DM_1 makes a decision at time t_g and DM_2 made a decision some time before then, and the probability that DM_2 makes a decision at time t_g and DM_1 made a decision some time before then.

Majority Total GDT for $N = 3$ General DMs   Again, the probability that a group of 3 DMs makes a decision using the Majority Total rule by time t_g is again equivalent to the probability that all 3 DMs reach a decision by time t_g. We use this to write out the group cdf as a function of the cdfs of the individual DMs:

$$q_g^{\text{mtg3}} = q_1(t_g)q_2(t_g)q_3(t_g). \tag{3.26}$$

As before, we take $\frac{d}{dt_g}$ of both sides to solve for the group pdf of GDTs,

$$p_g^{\text{mtg3}}(t_g) = p_1(t_g)q_2(t_g)q_3(t_g) + q_1(t_g)p_2(t_g)q_3(t_g) + q_1(t_g)q_2(t_g)p_3(t_g). \tag{3.27}$$

We see a pattern similar to the $N = 2$ case, where the pdf of GDTs is a sum of all possible combinations in which one DM makes a decision at time t_g and the others in the group have already made a decision by time t_g.

Majority Total GDT for N General DMs Extending the patterns found in the results above, we can write out the cdf of GDTs, $q_g^{\mathrm{mtg}N}(t_g)$ ([m]ajority [t]otal scheme, [g]eneral individual DMs, [N] DMs), as

$$q_g^{\mathrm{mtg}N}(t_g) = \prod_{i=1}^{N} q_i(t_g).$$
(3.28)

The pdf of GDTs is then

$$p_g^{\mathrm{mtg}N}(t_g) = \sum_{i=1}^{N} \left(p_i(t_g) \prod_{\substack{j=1, \\ j \neq i}}^{N} q_j(t_g) \right).$$
(3.29)

Majority Total GDT for iid DMs

We can now solve for the pdf of GDTs for iid DMs. We first solve the $N = 2$ case, the $N = 3$ case, and then extend the results to the arbitrary N case. We also show a plot of the analytical results for the iid case, for $N = 1$ to 41.

Majority Total GDT for $N = 2$ iid DMs For iid individual DMs, the cdf of GDTs is given by

$$q_g^{\mathrm{mti}2}(t_g) = [q_\iota(t_g)]^2.$$
(3.30)

The corresponding pdf of GDTs is given by

$$p_g^{\mathrm{mti2}}(t_g) = 2p_\iota(t_g)q_\iota(t_g). \tag{3.31}$$

In other words, the probability that the group finishes at a given time t_g is determined by the probability that one DM makes a decision at time t_g and the other DM has already made a decision by that time.

Majority Total GDT for $N = 3$ iid DMs Similarly, for the $N = 3$ case, the cdf of GDTs is given by

$$q_g^{\mathrm{mti3}}(t_g) = [q_\iota(t_g)]^3, \tag{3.32}$$

and the pdf of GDTs is given by

$$p_g^{\mathrm{mti3}}(t_g) = 3p_\iota(t_g)[q_\iota(t_g)]^2. \tag{3.33}$$

Majority Total GDT for N iid DMs Following the pattern laid out above, the cdf of GDTs for N iid DMs is

$$q_g^{\mathrm{mti}N}(t_g) = [q_\iota(t_g)]^N, \tag{3.34}$$

and the pdf of GDTs is

134
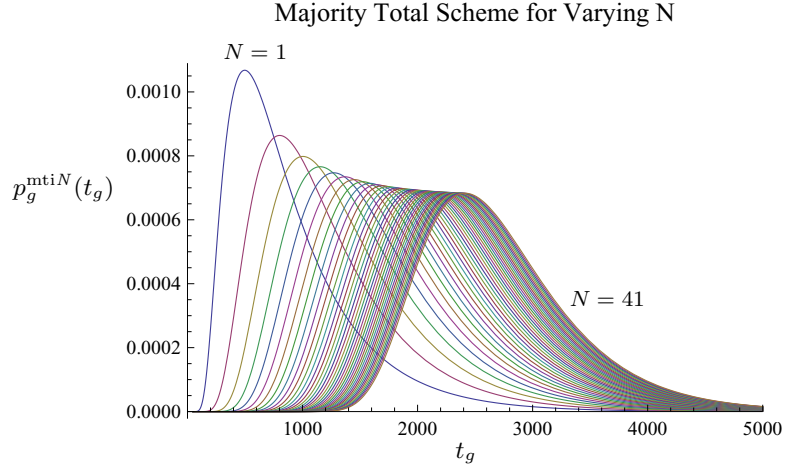
Figure 3.4: Plot of $p_g^{\mathrm{mti}N}(t_g)$, the pdf of GDTs for N iid DMs using the Majority Total decision rule, with N varying from 1 to 41. As the number of DMs increases, the mean GDT also increases, and the distribution spreads out. This is expected, since the fusion center declares the group's decision only after the slowest DM has responded, and as N increases, the slowest DM's DT tends to increase and tends to take on a wider range of values.

$$p_g^{\mathrm{mti}N}(t_g) = N p_\iota(t_g) \left[q_\iota(t_g) \right]^{N-1}. \tag{3.35}$$

These results are plotted for $N = 1$ to 41 in Figure 3.4. As N increases, the group pdf spreads out and drifts to the right, as we would expect - a larger group should take longer to decide, since the group must wait for the slowest DM and a larger N increases the chances of having a DM whose individual DT is from farther along the right tail of the individual's DT distribution.

Majority Total Simulation Results

In Figure 3.5, we show a histogram of GDTs from simulation, overlaid with the pdf calculated for the same group, and scaled by the number of trials and bin

135
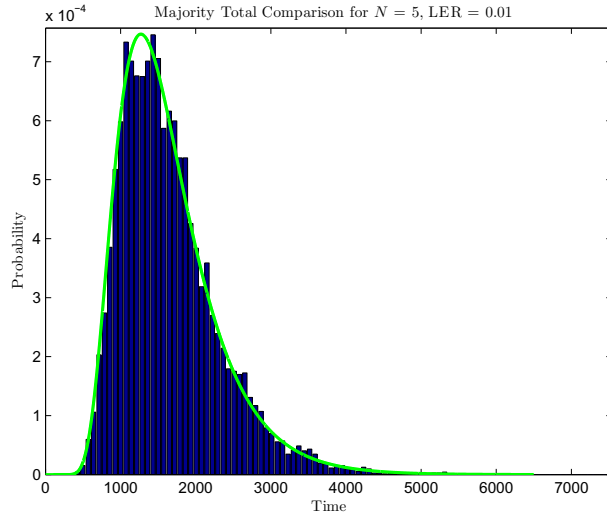
Figure 3.5: Comparison of the analytically-derived pdf of GDTs for a group of $N = 5$ iid DMs using the Majority Total scheme with a histogram of GDTs from simulating the same group over 10,000 trials. The histogram was scaled by the number of trials and bin width. The agreement between the pdf of GDTs and the histogram validates our results for the Majority Total scheme.

width. We took $N = 5$ and used the parameter values given in the introduction to Section 3.2. The agreement between the two verifies our results.

3.2.3 Majority First Scheme

The Majority First scheme is a slight modification of the Majority Total scheme. Given that our two major measures of performance are GDT and GER, it naturally follows that one can speed up the Majority Total scheme by ignoring the DMs whose decisions will not contribute to the fusion center's decision. Thus, in the Majority First scheme, the fusion center makes a decision as soon as the minimal majority Υ of DMs reach the same decision. For example, for $N = 3$, if the two fastest DMs both choose H_1, a fusion center using the Majority First scheme will immediately

136

choose H_1, whereas a fusion center using the Majority Total scheme will wait until the slowest DM responds before declaring the group's decision. Since the GDT is now inextricably linked to the individual DMs' decisions, we consider only N odd in calculating the cdf and pdf of GDTs for a group of DMs using the Majority First scheme. We have conducted an extensive search of past literature, and to the best of our knowledge, this particular group decision-making rule is novel.

Majority First Scheme GER

For a given group of DMs, the Majority First and Majority Total schemes have identical mean GERs, because they use the same decision rule to choose a hypothesis, and differ only in their stopping conditions. Thus, the GER for N iid DMs using the Majority First scheme is also given by Equation (3.22) and the GER for N general DMs by Equation (3.23).

Majority First Scheme pdf of GDTs for General DMs

The calculations used to find the pdf of GDTs of a group using the Majority First scheme differ from the previous two schemes in that they must track the decision made in addition to the decision time. We start with $N = 3$, since for $N = 2$, the group effectively uses either a Race or Majority Total scheme, depending on whether one considers 1 to be a minimal majority (Υ) or maximal minority (Θ). For $N = 3$, the minimal majority is $\Upsilon = 2$ and the maximal minority is $\Theta = 1$. We will then show the results for the $N = 5$ case before presenting the general N case,

$\mathbb{T}$			$\mathbb{D}\vert$S			$\mathbb{D}\vert$N		
1	2	3	1	2	3	1	2	3
L	L	L	S	S	S	N	N	N
G	L	L	N	S	S	S	N	N
L	G	L	S	N	S	N	S	N
L	L	G	S	S	N	N	N	S

Table 3.2: Decision Times ($\mathbb{T}$) and Decisions ($\mathbb{D}$) for $N = 3$ DMs. The numbers at the top of the columns identify the individual DMs in the group. The left-most column shows the possible DTs which possibly lead to a group decision by time t_g. The center and right-most columns show the possible decisions which lead to a group decision by time t_g, ending in choosing H_1 and H_0, respectively. A list of all possible $\mathbb{TD}$ combinations that result in the group deciding by time t_g can be easily constructed from this table. The full list corresponding to this table is given in Table 3.3.

for both general and iid DMs.

Majority First GDT for $N = 3$ General DMs   For $N = 3$, the possible DM decision times ($\mathbb{T}$) and decisions ($\mathbb{D}$) that can possibly lead to a group response by time t_g are shown in Table 3.2. However, not all combinations of the decision time and decision columns lead to a group decision. The specific combinations which lead to a group decision by time t_g are shown in Table 3.3, where entries on the left lead to the group choosing S (H_1) and the entries on the right lead to the group choosing N (H_0). Because the number of combinations are relatively small for $N = 3$, we specifically show both Table 3.2 and Table 3.3 as an example.

One can write a cdf showing all 20 entries from Table 3.2 as separate terms, but we are interested in developing a compact general formula, so we now introduce some simplifications. Note that the entries in Table 3.3 have been divided into eight partitions, four rows and two columns. The first row partition represents a

	S			N		
1	2	3	1	2	3
LS	LS	LS	LN	LN	LN
LN	LS	LS	LS	LN	LN
GS	LS	LS	GN	LN	LN
GN	LS	LS	GS	LN	LN
LS	LN	LS	LN	LS	LN
LS	GS	LS	LN	GN	LN
LS	GN	LS	LN	GS	LN
LS	LS	LN	LN	LN	LS
LS	LS	GS	LN	LN	GN
LS	LS	GN	LN	LN	GS

Table 3.3: All possible combinations of decisions and decision times which a group of 3 DMs can make and still have the fusion center make a decision by time t_g under the Majority First group decision rule. "S" indicates that the DM chooses H_1 ("Signal Present"), and "N" indicates that the DM chooses H_0 ("Signal Absent"). "L" indicates that the DM decided at some time before t_g and "G" indicates that the DM decides after time t_g. As a note, some useful identities with this additional notation can be found in Equation (3.36).

unanimous decision by time t_g. The second through fourth row partitions represent decisions in which there is one dissenting DM, which either reaches a decision by time t_g but disagrees with the majority or reaches a decision after time t_g. Under this expanded notation, we have additional identities, including

$$
\begin{aligned}
q_{\mathbb{D}}(t_g) &= \mathrm{L}\mathbb{D}, \\
q(t_g) &= \mathrm{L} = \mathrm{LS} + \mathrm{LN}, \\
1 - q(t_g) &= \mathrm{G} = \mathrm{GS} + \mathrm{GN}, \\
P(\mathbb{D}) &= \mathrm{L}\mathbb{D} + \mathrm{G}\mathbb{D}, \\
1 &= \mathrm{LS} + \mathrm{LN} + \mathrm{GS} + \mathrm{GN}, \\
1 - q_{\mathbb{D}}(t) &= \mathrm{L}\hat{\mathbb{D}} + \mathrm{GS} + \mathrm{GN},
\end{aligned}
\tag{3.36}
$$

where $\hat{\mathbb{D}}$ is the opposite decision of $\mathbb{D}$. For example, $[1 - q_{1S}(t_g)]$ is the probability that DM_1 does not contribute towards a group decision for S (i.e., it either selects N or answers after time t_g). Now each partition can be collapsed into a single term, which reduces the 20 terms to 8:

$$
\begin{aligned}
q_g^{\mathrm{mfg3}}(t_g) = \ & q_{1S}(t_g)q_{2S}(t_g)q_{3S}(t_g) + q_{1N}(t_g)q_{2N}(t_g)q_{3N}(t_g) \\
& + [1 - q_{1S}(t_g)]q_{2S}(t_g)q_{3S}(t_g) + q_{1S}(t_g)[1 - q_{2S}(t_g)]q_{3S}(t_g) \\
& + q_{1S}(t_g)q_{2S}(t_g)[1 - q_{3S}(t_g)] + [1 - q_{1N}(t_g)]q_{2N}(t_g)q_{3N}(t_g) \\
& + q_{1N}(t_g)[1 - q_{2N}(t_g)]q_{3N}(t_g) + q_{1N}(t_g)q_{2N}(t_g)[1 - q_{3N}(t_g)].
\end{aligned}
\tag{3.37}
$$

This is expression can be simplified further:

$$
\begin{aligned}
q_g^{\mathrm{mfg3}}(t_g) = \ & \prod_{i=1}^{3} q_{iS}(t_g) + \prod_{i=1}^{3} q_{iN}(t_g) + \sum_{i=1}^{3}\left([1 - q_{iS}(t_g)] \prod_{\substack{j=1, \\ j \neq i}}^{3} q_{jS}(t_g) \right) \\
& + \sum_{i=1}^{3}\left([1 - q_{iN}(t_g)] \prod_{\substack{j=1, \\ j \neq i}}^{3} q_{jN}(t_g) \right).
\end{aligned}
\tag{3.38}
$$

We can now differentiate Equation (3.38) to get the pdf of GDTs:

$$
p_g^{\mathrm{mfg3}}(t_g) = \sum_{i=1}^{3} \left[p_{i\mathrm{S}}(t_g) \prod_{\substack{j=1, \\ j \neq i}}^{3} q_{j\mathrm{S}}(t_g) + p_{i\mathrm{N}}(t_g) \prod_{\substack{j=1, \\ j \neq i}}^{3} q_{j\mathrm{N}}(t_g) \right]
$$

$$
- \sum_{i=1}^{3} \left[p_{i\mathrm{S}}(t_g) \prod_{\substack{j=1, \\ j \neq i}}^{3} q_{j\mathrm{S}}(t_g) + p_{i\mathrm{N}}(t_g) \prod_{\substack{j=1, \\ j \neq i}}^{3} q_{j\mathrm{N}}(t_g) \right] \tag{3.39}
$$

$$
+ \sum_{i=1}^{3} \sum_{\substack{j=1, \\ j \neq i}}^{3} \sum_{\substack{k=1, \\ k \neq i,j}}^{3} \left([1 - q_{i\mathrm{S}}(t_g)] p_{j\mathrm{S}} q_{k\mathrm{S}} + [1 - q_{i\mathrm{N}}(t_g)] p_{j\mathrm{N}} q_{k\mathrm{N}} \right),
$$

which simplifies to

$$
p_g^{\mathrm{mfg3}}(t_g) = \sum_{i=1}^{3} \sum_{\substack{j=1, \\ j \neq i}}^{3} \sum_{\substack{k=1, \\ k \neq i,j}}^{3} \left([1 - q_{i\mathrm{S}}(t_g)] p_{j\mathrm{S}}(t_g) q_{k\mathrm{S}}(t_g) + [1 - q_{i\mathrm{N}}(t_g)] p_{j\mathrm{N}}(t_g) q_{k\mathrm{N}}(t_g) \right).
$$

$$
\tag{3.40}
$$

This makes intuitive sense: the fusion center makes a decision when one DM makes a decision at time t_g and chooses the decision that Θ DMs have already chosen. This gives the Υ agreeing DMs required for the group to make a decision. We now consider the next simplest case, $N = 5$, below.

Majority First GDT for $N = 5$ General DMs For $N = 5$, the minimal majority is $\Upsilon = 3$ and the maximal minority is $\Theta = 2$. Adding two DMs vastly increases the number of possible combinations. In Table 3.4, we list out the $\mathbb{T}$ and $\mathbb{D}$ and combinations which may lead to a group decision by time t_g. Though

we could again write out a list of the specific combinations of decision times and decisions (with somewhat more effort), the list would be extremely long.

Note that θ is equivalently the number of G terms in the $\mathbb{T}$ column, the number of N terms in the $\mathbb{D}|$S column, and the number of S terms in the $\mathbb{D}|$N column. In other words, it represents the number of dissenting (minority) members. Using the same combination patterns as was done for the $N = 3$ case, we can find the cdf of GDTs for 5 general DMs using the Majority First scheme (dropping the explicit (t_g) notation on the right-hand side for brevity):

$$
\begin{aligned}
q_g^{\mathrm{mfg5}}(t_g) = \ & \prod_{i=1}^{5} q_{i\mathrm{S}} + \prod_{i=1}^{5} q_{i\mathrm{N}} + \sum_{i=1}^{5} \left([1 - q_{i\mathrm{S}}] \prod_{\substack{j=1, \\ j \neq i}}^{5} q_{j\mathrm{S}} + [1 - q_{i\mathrm{N}}] \prod_{\substack{j=1, \\ j \neq i}}^{5} q_{j\mathrm{N}} \right) \\
& + \sum_{i_1=1}^{4} \sum_{i_2=i_1+1}^{5} \left([1 - q_{i_1\mathrm{S}}][1 - q_{i_2\mathrm{S}}] \prod_{\substack{j=1, \\ j \neq i_1, i_2}}^{5} q_{j\mathrm{S}} \right. \\
& \left. + [1 - q_{i_1\mathrm{N}}][1 - q_{i_2\mathrm{N}}] \prod_{\substack{j=1, \\ j \neq i_1, i_2}}^{5} q_{j\mathrm{N}} \right).
\end{aligned}
\tag{3.41}
$$

Again, we differentiate and simplify to get the pdf of GDTs,

DM:	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
			$\mathbb{T}$					$\mathbb{D}|S$					$\mathbb{D}|N$		
$\theta = 0$	L	L	L	L	L	S	S	S	S	S	N	N	N	N	N
	G	L	L	L	L	N	S	S	S	S	S	N	N	N	N
	L	G	L	L	L	S	N	S	S	S	N	S	N	N	N
$\theta = 1$	L	L	G	L	L	S	S	N	S	S	N	N	S	N	N
	L	L	L	G	L	S	S	S	N	S	N	N	N	S	N
	L	L	L	L	G	S	S	S	S	N	N	N	N	N	S
	G	G	L	L	L	N	N	S	S	S	S	S	N	N	N
	G	L	G	L	L	N	S	N	S	S	S	N	S	N	N
	G	L	L	G	L	N	S	S	N	S	S	N	N	S	N
	G	L	L	L	G	N	S	S	S	N	S	N	N	N	S
$\theta = 2$	L	G	G	L	L	S	N	N	S	S	N	S	S	N	N
	L	G	L	G	L	S	N	S	N	S	N	S	N	S	N
	L	G	L	L	G	S	N	S	S	N	N	S	N	N	S
	L	L	G	G	L	S	S	N	N	S	N	N	S	S	N
	L	L	G	L	G	S	S	N	S	N	N	N	S	N	S
	L	L	L	G	G	S	S	S	N	N	N	N	N	S	S

Table 3.4: Combinations of Decisions ($\mathbb{D}$) and Decision Times ($\mathbb{T}$) which produce a group decision by time t_g for $N = 5$ DMs. Within each column, the row of numbers at the top identifies the individual DMs, numbered 1 through 5. The columns are partitioned into three parts, denoted by double vertical lines. The first partition on the left shows the value of θ, which denotes the number of dissenting DMs in each row and ranges from $\theta = 0$ to $\theta = \Theta$. The Majority First for a group of 5 DMs requires that at least 3 DMs must reach the same decision by time t_g for the fusion center to make a decision by time t_g. The second column partition shows the combinations of individual DTs ($\mathbb{T}$) for which the group reaches a decision by time t_g under the Majority First scheme. This is decoupled from the actual decision each DM makes. As before, "L" indicates that the DM made a decision before time t_g (i.e., the DT is [L]ess than t_g), and "G" indicates that the DM makes a decision after time t_g (i.e., the DT is [G]reater than t_g). The right-most column shows the combinations of individual decisions, decoupled from DT, required for the fusion center to choose S or N. The $\mathbb{D}|S$ (resp., $\mathbb{D}|N$) subcolumn shows the individual decision combinations that lead to the fusion center selecting S (resp., N). In the first row partition, $\theta = 0$, so there are no dissenting DMs: the $\mathbb{T}$ column is LLLLL, the $\mathbb{D}|S$ column is SSSSS and the $\mathbb{D}|N$ column is NNNNN. In the second row partition, $\theta = 1$, so there is one dissenting DM per column, and the rows in this partition show all unique combinations of one dissenting DM and four agreeing DMs. Now that we have all combinations of individual DTs and individual decisions that could give rise to a fusion center decision by time t_g, we can easily construct the full explicit list of all $\mathbb{D}\mathbb{T}$ pairs that result in a fusion center decision by time t_g; however, due to the large number of combinations, we will not explicitly state every $\mathbb{D}\mathbb{T}$ pair as we did for $N = 3$ in Table 3.3. We note that not all combinations of the center and right column partitions result in the group making a decision by time t_g, so one should be careful in listing out all possibilities.

$$
\begin{aligned}
p_g^{\mathrm{mfg5}}(t_g) = {} & \sum_{i=1}^{4} \sum_{j=i+1}^{5} \left([1 - q_{i\mathrm{S}}(t_g)][1 - q_{j\mathrm{S}}(t_g)] \sum_{\substack{k=1, \\ k \neq i,j}}^{5} \left[p_{k\mathrm{S}}(t_g) \prod_{\substack{m=1, \\ m \neq i,j,k}}^{5} q_{m\mathrm{S}}(t_g) \right] \right) \\
& + \sum_{i=1}^{4} \sum_{j=i+1}^{5} \left([1 - q_{i\mathrm{N}}(t_g)][1 - q_{j\mathrm{N}}(t_g)] \right. \\
& \left. \sum_{\substack{k=1, \\ k \neq i,j}}^{5} \left[p_{k\mathrm{N}}(t_g) \prod_{\substack{m=1, \\ m \neq i,j,k}}^{5} q_{m\mathrm{N}}(t_g) \right] \right).
\end{aligned}
$$

$$(3.42)$$

Majority First GDT for N General DMs The cdf of GDTs for N general

DMs, where we drop the explicit (t_g) notation on the right-hand side for brevity, is

given by

$$
q_g^{\mathrm{mfg}N}(t_g) = \prod_{i=1}^{N} q_{i\mathrm{S}} + \prod_{i=1}^{N} q_{i\mathrm{N}} + \sum_{\theta=1}^{\Theta} \left(\Gamma_{\mathcal{J}\mathrm{S}}^{\theta\mathrm{g}} \prod_{\substack{k=1, \\ k \notin \mathcal{J}}}^{N} q_{k\mathrm{S}} + \Gamma_{\mathcal{J}\mathrm{N}}^{\theta\mathrm{g}} \prod_{\substack{k=1, \\ k \notin \mathcal{J}}}^{N} q_{k\mathrm{N}} \right). \qquad (3.43)
$$

The first two terms express the situation where all N DMs agree on the group

decision and answer before time t_g. The last term denotes all combinations $[\mathcal{J}]$

where up to $[\theta]$ [g]eneral DMs either answer by time t_g but disagree with the (final)

group decision $[\mathbb{D}]$ or answer after time t_g. The subfunction $\Gamma_{\mathcal{J}\mathbb{D}}^{\theta\mathrm{g}}$ accounts for the

cases with dissenting DMs, and is defined as

$$\Gamma_{\mathcal{J}\mathbb{D}}^{\theta\mathrm{g}} = \sum_{j_1=1}^{(N-\theta+1)} \sum_{j_2=j_1+1}^{(N-\theta+2)} \cdots \sum_{j_\theta=j_{\theta-1}+1}^{N} \left(\prod_{m=1}^{\theta} [1 - q_{j_m\mathbb{D}}] \right), \qquad (3.44)$$

where $\mathbb{D} \in \{\mathrm{S}, \mathrm{N}\}$, $\mathcal{J} = \{j_1, \ldots, j_\theta\}$ and $q_i(t_g) = q_{i\mathrm{S}}(t_g) + q_{i\mathrm{N}}(t_g)$. The summation indices indicate which DMs are in each unique combination ($\mathcal{J}$) of DMs, and the product term uses the sub-index m to iterate through each DM in the combination.

Taking the derivative with respect to t_g and simplifying, we get the pdf of GDTs for N general DMs using the Majority First scheme:

$$p_g^{\mathrm{mfg}N}(t_g) = \sum_{j_1=1}^{N-\Theta+1} \sum_{j_2=j_1+1}^{N-\Theta+2} \cdots \sum_{j_\Theta=j_{\Theta-1}+1}^{N} \left(\prod_{i=1}^{\Theta}[1 - q_{j_i\mathrm{S}}] \sum_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} \left[p_{k\mathrm{S}} \prod_{\substack{m=1, \\ m\notin\mathcal{J}, \\ m\neq k}}^{N} q_{m\mathrm{S}} \right] \right)$$

$$+ \sum_{j_1=1}^{N-\Theta+1} \sum_{j_2=j_1+1}^{N-\Theta+2} \cdots \sum_{j_\Theta=j_{\Theta-1}+1}^{N} \left(\prod_{i=1}^{\Theta}[1 - q_{j_i\mathrm{N}}] \sum_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} \left[p_{k\mathrm{N}} \prod_{\substack{m=1, \\ m\notin\mathcal{J}, \\ m\neq k}}^{N} q_{m\mathrm{N}} \right] \right).$$
$$(3.45)$$

This follows the general form shown above for the $N = 3$ and $N = 5$ cases: there are Θ terms of the form $q_{\mathbb{D}}$, which means that Θ DMs made the same decision before time t_g; one term of the form $p_{\mathbb{D}}$, which means that one DM made the same decision at time t_g; and Θ terms of the form $[1 - q_{\mathbb{D}}]$, which means that Θ DMs did not contribute to the final group decision at time t_g: those DMs either chose the other hypothesis or did not make a decision by time t_g. See Section C in the Appendix for a proof of this formula.

Majority First Scheme GDT for iid DMs

We can simplify the above cdfs and pdfs for the case where the DMs are iid. We present these below.

Majority First GDT for $N = 3$ iid DMs The cdf of GDTs for 3 iid DMs using the Majority First decision rule is

$$q_g^{\mathrm{mfi3}}(t_g) = [q_{\iota\mathrm{S}}(t_g)]^3 + [q_{\iota\mathrm{N}}(t_g)]^3 + 3[1 - q_{\iota\mathrm{S}}(t_g)][q_{\iota\mathrm{S}}(t_g)]^2 + 3[1 - q_{\iota\mathrm{N}}(t_g)][q_{\iota\mathrm{N}}(t_g)]^2,$$

(3.46)

and the pdf of GDTs is

$$p_g^{\mathrm{mfi3}}(t_g) = 6[1 - q_{\iota\mathrm{S}}(t_g)]p_{\iota\mathrm{S}}(t_g)q_{\iota\mathrm{S}}(t_g) + 6[1 - q_{\iota\mathrm{N}}(t_g)]p_{\iota\mathrm{N}}(t_g)q_{\iota\mathrm{N}}(t_g).$$

(3.47)

Majority First GDT for $N = 5$ iid DMs The cdf of GDTs for 5 iid DMs using the Majority First decision rule is

$$
\begin{aligned}
q_g^{\mathrm{mfi5}}(t_g) = \ & [q_{\iota\mathrm{S}}(t_g)]^5 + [q_{\iota\mathrm{N}}(t_g)]^5 + 5 \left([1 - q_{\iota\mathrm{S}}(t_g)][q_{\iota\mathrm{S}}(t_g)]^4 \right. \\
& \left. + [1 - q_{\iota\mathrm{N}}(t_g)][q_{\iota\mathrm{N}}(t_g)]^4 \right) + 10 \left([1 - q_{\iota\mathrm{S}}(t_g)]^2 [q_{\iota\mathrm{S}}(t_g)]^3 \right. \\
& \left. + [1 - q_{\iota\mathrm{N}}(t_g)]^2 [q_{\iota\mathrm{N}}(t_g)]^3 \right),
\end{aligned}
$$

(3.48)

and the pdf of GDTs is

$$p_g^{\mathrm{mfi5}}(t_g) = 30 \left([1 - q_{\iota \mathrm{S}}(t_g)]^2 p_{\iota \mathrm{S}}(t_g) [q_{\iota \mathrm{S}}(t_g)]^2 + [1 - q_{\iota \mathrm{N}}(t_g)]^2 p_{\iota \mathrm{N}}(t_g) [q_{\iota \mathrm{N}}(t_g)]^2 \right).$$

(3.49)

Majority First GDT for N iid DMs As before, this can be simplified for N iid DMs. The cdf of GDTs is now

$$q_g^{\mathrm{mfi}N}(t_g) = [q_{\iota \mathrm{S}}]^N + [q_{\iota \mathrm{N}}]^N + \sum_{\theta=1}^{\Theta} \binom{N}{\theta} \left([1 - q_{\iota \mathrm{S}}]^\theta [q_{\iota \mathrm{S}}]^{N-\theta} + [1 - q_{\iota \mathrm{N}}]^\theta [q_{\iota \mathrm{N}}]^{N-\theta} \right),$$

(3.50)

and the group pdf for N iid DMs is

$$
\begin{aligned}
p_g^{\mathrm{mfi}N}(t_g) &= \binom{N}{\Theta} \Upsilon \left([1 - q_{\iota \mathrm{S}}(t_g)]^\Theta p_{\iota \mathrm{S}}(t_g) [q_{\iota \mathrm{S}}(t_g)]^\Theta \right. \\
&\quad \left. + [1 - q_{\iota \mathrm{N}}(t_g)]^\Theta p_{\iota \mathrm{N}}(t_g) [q_{\iota \mathrm{N}}(t_g)]^\Theta \right).
\end{aligned}
$$

(3.51)

We plot the pdf of group DTs for N iid DMs using the Majority First scheme in Figure 3.6, for N odd, from $N = 1$ to $N = 41$. In the Majority First scheme, the slowest member is between the Υth and Nth to respond, and as N increases, the number of DMs that can potentially be ignored also increases. The increase

147

Majority First Scheme for Varying N
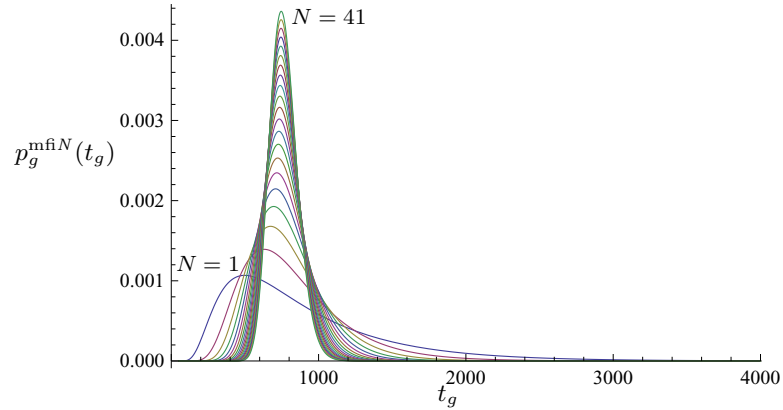
Figure 3.6: Plot of $p_g^{\mathrm{mfi}N}(t_g)$, the pdf of GDTs N iid DMs using the Majority First decision rule, with N varying from 1 to 41, for N odd. As N increases, the distribution appears to move slightly to the right, while the pdf becomes more peaked; however, the mean GDT actually decreases slightly.

in the number of ignorable DMs seems to balance out the increase in Υ, so the distribution moves to the right and becomes more peaked. Due to the shape of the curves, the mean GDT actually decreases slightly. It is very evident for the Majority First scheme that a full pdf of GDTs provides more useful information than a simple mean-GDT-based analysis.

Majority First Simulation Results

In Figure 3.7, we show a histogram of GDTs from simulating $N = 5$ iid DMs, scaled by the number of trials and bin width and overlaid with the pdf calculated for the same group. The agreement between the two verifies our results.
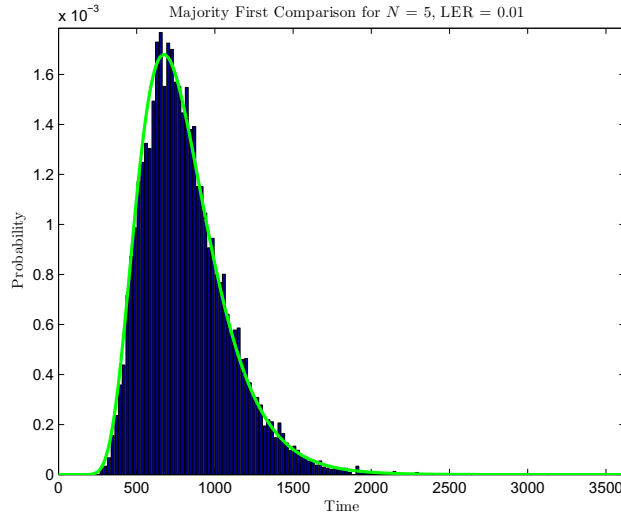
Figure 3.7: Comparison of the analytically derived group pdf and a histogram of GDTs from a simulation of the same group of 5 iid DMs using the Majority First group decision rule. The histogram was scaled by the number of trials and bin width. This simulation was run over 10,000 trials. The agreement between the pdf and the histogram validates our results for the Majority First scheme.

3.2.4 Decentralized SPRT (D-SPRT) Scheme

The Decentralized SPRT (D-SPRT) was proposed by A. Hussain in [45]. In this scheme, iid DMs use the SPRT to reach a decision, and then transmit their decisions to a fusion center. Here, unlike the schemes above, after an individual DM has reached a decision and transmitted it to the fusion center, it resets its decision variable and begins a new SPRT on new incoming data. The fusion center has its own decision variable, and performs a fixed step-size SPRT on the decisions from the local detectors, effectively aggregating the signals from the individual DMs. Once the fusion center's decision variable reaches a given threshold, determined by the maximum allowable error rate, the fusion center declares the group decision.

149

The fusion center in the D-SPRT approximates a single DM using the SPRT, but includes redundancy and on average effectively processes approximately N times as much data as a single DM in each time step. The D-SPRT is not completely efficient because there will always be observations taken by all but one DM which are not transmitted to the fusion center before the fusion center process finishes, but the scheme's potential for speed, ability to utilize lower accuracy local sensors to achieve a higher level of accuracy in the group's decision, and robustness to individual sensor failure (in the sense that the sensor either stops taking data or can no longer transmit decisions) make it a reasonable group decision-making scheme to consider.

We note that the D-SPRT is reasonable to consider for device-based systems, but is non-ideal for systems with human DMs, since humans will likely be too biased by their previous decision to effectively reset their decision variable. Though we have not found a study that tests this specifically, there is ample evidence that human observers tend to discount others' opinions and prefer their own. In [42, 104], it was shown that people tended to discount others' opinions, and place a majority weight on their own, with weights of approximately (70% / 30%) on (self / other). If one considers the "advisor's opinion" to be additional data, the studies indicate that once a human observer has made a decision, he is likely to set a very high prior to repeat that decision thereafter. It has also been found that people tend to ignore information that does not support their opinion and tend to pay more attention

to information that does [90, 102]. This implies that a human observer may not only assign a high prior to the previous opinion, but may also selectively process additional information. Additionally, it was concluded in [53] that even though human observers could adjust their criteria to change their DT and accuracy to adapt to different situations, they rarely revised their decisions after making them. Thus, group decision-making rules which require a decision-maker to submit more than one decision to the fusion center are likely to be very inappropriate for systems with human observers. The D-SPRT scheme is also vulnerable to attack: if one DM is hijacked, it can repeatedly send a response to the fusion center, and thereby control the group's decision; in our majority-based schemes, a hijacker would have to control a majority of the DMs to ensure a certain group response. These weaknesses aside, we analyze the D-SPRT along with our group decision rules, since the D-SPRT may be relevant to many 2AFC tasks.

D-SPRT Scheme LER, GER, and Boundaries

Hussain generally models the individual DMs as a bi-marked renewal process, denoted $w_k(t)$ for individual k, where each individual outputs $+1$ or -1 with probability δ or $(1-\delta)$, respectively, at some time whose inter-decision intervals are described by a probability distribution function. The fusion center process $w(t)$ is then modeled as a superposition of N of the above processes. An equivalent description more consistent with our other schemes is as follows: the local DMs collect data and

151

update their individual decision variables at each time step. Each individual has decision boundaries at B_0 and B_1, and declares a decision when its decision variable is absorbed into one of the boundaries. When the kth DM declares a decision at time t, it sends output $w_k(t)$ to the fusion center, where

$$
w_k(t) = \begin{cases} +1 & \text{if detector } k \text{ decides } H_1 \\ -1 & \text{if detector } k \text{ decides } H_0 \end{cases}.
$$

That DM then resets its decision variable and begins a new SPRT on new incoming data. At each time step, the fusion center uses the sum of these signals to update its decision variable, moving its decision variable by $B_1 \cdot \sum_k w_k(t)$ at each time step. Like all of the schemes presented here, the individual detectors do not need to be synchronized.

The expected GER is (theoretically) set by the experimenter, and is determined by the fusion center's error rate, ERfc. This value sets the fusion center's decision boundaries, $\mathfrak{B}_0$ and $\mathfrak{B}_1$, using Equations (2.16) and (2.17). The experimenter is also free to choose η, the number of steps which the fusion center must take to reach a decision. Under the no-overshoot assumption, the fusion center decision boundaries and η then set the decision boundaries for the (local) individual DMs: $B_0 = \frac{\mathfrak{B}_0}{\eta}$ and $B_1 = \frac{\mathfrak{B}_1}{\eta}$, from which one can calculate the LER:

$$
\alpha_0 = \frac{1 - e^{B_0}}{e^{B_1} - e^{B_0}} \tag{3.52}
$$

152

$$\alpha_1 = \frac{e^{B_0}\left(e^{B_1} - 1\right)}{e^{B_1} - e^{B_0}}, \tag{3.53}$$

since LER $= \alpha_0 = \alpha_1$. We explore the D-SPRT further through simulations to build on the results of [45] below.

Investigating the Small-Overshoot Assumption in the D-SPRT

In our simulations of the D-SPRT, each DM in the group runs a SPRT until it reaches a decision, at which point it sends a signal to the fusion center, then resets its decision variable back to zero and begins a new SPRT. At each time step, the fusion center checks to see if any of the local DMs have reached a decision. If any have, the fusion center updates its decision variable and checks to see if it has reached a decision. If the fusion center decision variable crosses a threshold, then the group decision is made and the entire group shuts down. If not, the process continues.

Our initial series of simulations investigates the best value for η, since this was not specified in [45]. In theory, one would expect that a large η value would be best, since a large η corresponds to individual DMs with higher LERs and faster individual DTs. Thus, a larger value of η would cause the D-SPRT scheme to become more similar to the simple centralized (in the engineering literature sense of the word) case, where the fusion center processes N observations at each time step, and the individual DMs are just sensors that send information to the fusion center. However, as we saw in Chapter 2, at large LERs, overshoot becomes

non-negligible, and the small-overshoot assumption used to calculate the decision variable boundaries no longer holds. Therefore, we expect to find an optimal point, at which there are enough DMs to gain an advantage in numbers, but not so many that the LER increases past the point where overshoot affects our results.

We simulated a group of 3 DMs utilizing the D-SPRT, with ERfc = 0.01, since we found this error rate to provide performance reasonably close to what we expected from analytical calculation for an individual DM. This gives the boundaries of the fusion center's decision variable. We then used η to determine the boundaries of the individual DM, and solved for the individual DMs' LERs. The ideal LER, and mean actual GDT and GER are shown for each value of η in Table 3.5, averaged over 10,000 trials. This initial study allowed us to shorten the range of η values of interest. We then focused on a promising range of values and ran an additional set of simulations to more definitively find the optimal η. Our results, averaged over 100,000 trials, are shown in Table 3.6.

If the individual DM adequately fulfilled the small-overshoot assumption for all values of LER, then we would expect that the D-SPRT's mean GER would remain steady, while its mean GDT would decline with increasing η before asymptoting at approximately one-Nth the expected GDT of an individual using the SPRT with the same d' and an ER equal to the group's ERfc. Thus, for $N = 3$, one would expect the GDT to asymptote at a value bounded below by 300.214 for large η, ERfc = 0.01, and $d' = 0.1$. However, as shown in Figure 3.8, this does not

154

occur in simulation. While the plots are not smooth due to the finite sample size (10,000 trials), there is a clear trend showing that the GDT decreases from $\eta = 1$ to $\eta = 4$, then slowly increases with η. Meanwhile, the GER generally decreases with increasing η. We believe both of these effects to be due to overshoot.

Though the values we chose were suitable for an individual DM, larger values of η can cause the D-SPRT to amplify the overshoot on two levels. First, the larger η results in the individual DM having a larger LER, which, as discussed above, results in more error than a smaller LER would have. Second, this slightly larger error is multiplied by at least η, since the error accumulates over at least η steps before the fusion center reaches a decision. With larger η, the error increases to the point where it is no longer considered small. This results in the boundaries for the individual DM being too conservative, which results in the group using the D-SPRT to have a longer GDT but lower GER, as is shown in Figure 3.8.

From our results, we selected $\eta = 5$, since this value had both a low GDT and low GER across multiple simulations.

Now that we have reasonable parameters to use for the D-SPRT, we can compare its performance to our main three group decision rules, which we will do in the next section; however, first, we will detail a few other related group decision rules from the literature for completeness.

η	LER	Mean GER	Mean GDT
1	0.0100	0.0100	510.8921
2	0.0913	0.0100	366.3545
3	0.1777	0.0060	341.6355
4	0.2407	0.0072	332.3445
5	0.2852	0.0056	335.5477
6	0.3174	0.0062	338.0423
7	0.3415	0.0080	338.8554
8	0.3602	0.0060	344.3209
9	0.3751	0.0052	342.0504
10	0.3871	0.0062	350.8608
11	0.3971	0.0046	351.7345
12	0.4054	0.0056	357.7687
13	0.4125	0.0030	385.8191
14	0.4187	0.0056	368.5452
15	0.4240	0.0052	388.6162
16	0.4287	0.0046	384.6044
17	0.4328	0.0036	380.1343
18	0.4365	0.0036	391.2518
19	0.4398	0.0020	398.3662
20	0.4428	0.0038	402.9937
21	0.4455	0.0036	400.3097
22	0.4480	0.0018	413.8168
23	0.4502	0.0032	418.0568
24	0.4523	0.0040	412.3200
25	0.4542	0.0020	436.9519

Table 3.5: Results of our first set of simulations to find a good value for η. We set ERfc = 0.01, and for $N = 3$, varied the value of η. The second column shows what LER each DM was set to, and the third and fourth columns show the actual mean GERs and GDTs from simulation. This table corresponds to Figure 3.8. The values above are averaged over 10,000 trials.
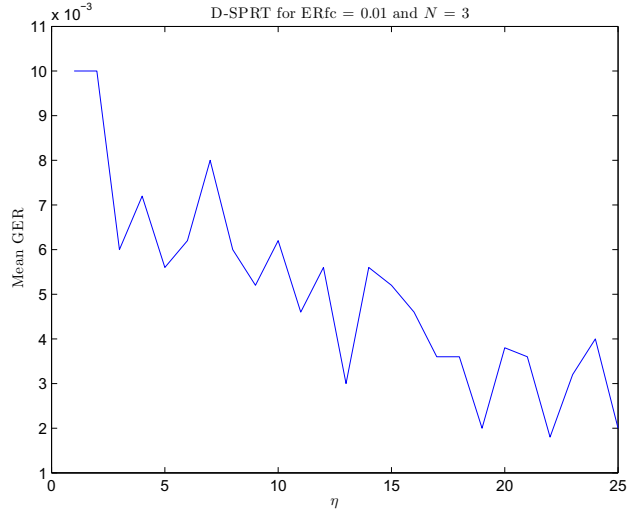
(a) Group Error Rate
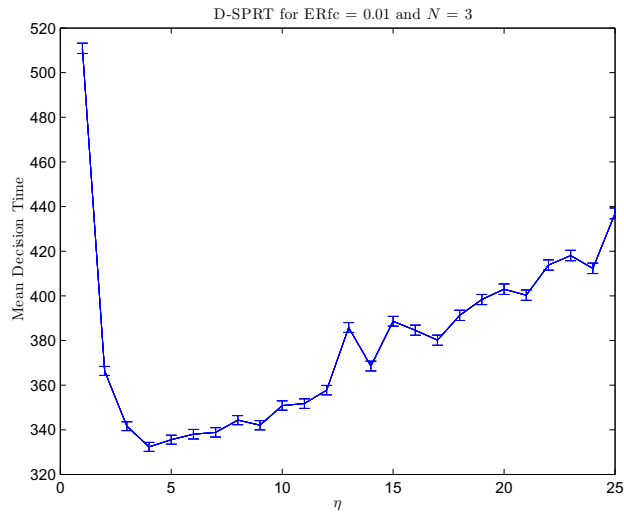
(b) Group Decision Time

Figure 3.8: Plots corresponding to the data in Table 3.5. Our goal is to select a value of η at which the D-SPRT functions well for a group of size $N = 3$ or larger, to compare to our other group decision rules. Figure (a) shows the actual GER. Figure (b) shows the mean GDT. Theoretically, we would have expected the GDT to continue decreasing with η, but we found that in simulation, overshoot becomes non-negligible for large η. Thus, there is a tradeoff, and a minimal value for the GDT appears. The error bars represent standard error. The values shown here are averaged over 10,000 trials. Based on these simulation results, we selected $\eta = 5$ to provide a good balance between low DT and low GER.

η	LER	Mean GER	Mean GDT
1	0.0100	0.0091	510.6508
2	0.0913	0.0086	366.3114
3	0.1777	0.0090	340.5606
4	0.2407	0.0073	334.4718
5	0.2852	0.0075	332.8404
6	0.3174	0.0091	338.6020
7	0.3415	0.0086	336.8180
8	0.3602	0.0090	341.6382
9	0.3751	0.0073	343.7072
10	0.3871	0.0075	350.8489

Table 3.6: Results of our second set of simulations to find a good value for η. We used the same parameter values as we did for Table 3.5, but averaged over 100,000 trials this time to better locate the best η value. Though it appears that $\eta = 4$ would also work well, we felt these results verified that $\eta = 5$ is a good choice for the D-SPRT applied to a group described by our parameter values.

3.2.5 Other Group Decision-Making Models

Here we present some brief notes on group decision-making models similar to those above that we have found in the literature: the AND, OR, and Audley models.

AND and OR Models

We note that the logical AND and OR group decision schemes from the engineering literature [1, 3] generate identical pdfs of GDTs to the Majority Total and Race schemes, respectively, because they share the same stopping rule. However, the schemes' decision rules are very different – the AND scheme requires a consensus to select H_0, otherwise, it selects H_1; in contrast, the Majority Total scheme selects the hypothesis with more votes once every DM has voted. The OR scheme requires at least one DM to choose H_0, otherwise, it selects H_1; in contrast, the

Race scheme selects the hypothesis that the fastest DM chooses. In a 2AFC task where there is little a priori knowledge about which hypothesis is correct, the AND and OR models may favor one hypothesis too heavily.

However, there are tasks in which the AND and OR scheme are appropriate. For example, in an item recognition task, an observer is first presented with a list of items to memorize (words, sounds, images, etc) for a finite period of time. After a delay, ranging from seconds to days, the observer is presented with a series of items which may or may not have been on the list, and asked if the item is "old" (i.e., was on the list), or "new", (i.e., was not on the list). It is proposed in [68] that during the task, the observer compares the item presented with all items evoked in memory, and a number of processes decide if the item presented matches one of the items on the list in memory. If there is a match, the process terminates when the fastest of the processes returns a "yes"; otherwise, the process does not end until the slowest process returns a "no". This is equivalent to using an AND (or OR) group rule to select the observer's decision, where the observer is the fusion center and the processes are the individual DMs. We note that one could then treat each observer as a DM in a higher-level group using a Majority Total or Race (or OR or AND) group decision rule to complete the task with even higher accuracy or speed, which demonstrates the flexibility of the group structures we consider.

We do not consider the AND and OR schemes further, since they are not as appropriate for the 2AFC tasks we consider.

Audley Model

The Audley Model [4] is similar to the D-SPRT in that at least some group members typically make more than one decision before the fusion center makes a decision, and the members reset their decision variables and begin a new decision process (SPRT) after sending a decision to the fusion center. However, in the Audley model, the fusion center selects the hypothesis that receives M consecutive decisions, where M is a parameter set by the experimenter. Though viable, we consider this rule to be less practical than the D-SPRT, and thus only mention it for completeness.

3.3 Part II: Group Decision Rule Comparison for iid DMs

In Part II, we compare our main three group decision rules and the D-SPRT under different constraints, while assuming that each group rule is applied to a group with iid DMs. We first verify that our simulations provide the GERs we expect, then compare the performance of the different group decision rules under the equal-LER and equal-GER conditions.

3.3.1   GER Verification

The following results verify that our simulations fulfill our expectations over a reasonable range of values, and indicate the extent to which overshoot may affect our group rules' simulation results.

Fixed $N = 3$, Different LER

This simulation shows how the different schemes are affected by changes in the DMs' LER. For each scheme, we held N fixed at 3, and varied the LER from 0.002 to 0.05, in steps of size 0.002.

We expect that the single DM and Race schemes should have GER $=$ LER, and that the Majority Total and Majority First schemes should have the same GER $<$ LER. Since the D-SPRT requires 5 net votes to make a decision, whereas the others require at most 2 total votes, we expect it to have a very low GER, but to be very slow. Figure 3.9(a) verifies this intuition for a group of 3 DMs, averaged over 10,000 trials. The fluctuation in the measurements is due to the finite number of trials that each point in the graph is averaged over.

Figure 3.9(b) shows that the average GDT from simulation of our main three group decision rules closely follows the mean GDT predicted by our analytical results, which indicates that the system is not affected appreciably by overshoot. For $N = 3$ and a given LER, as we expected, the D-SPRT is the slowest, the Majority Total scheme is the second slowest, the Majority First scheme is slightly

faster than a single DM, and the Race scheme is the fastest.

Fixed $N = 3$, Different GER

This simulation serves mainly to verify that the Group Error Rates are accurate at the parameter values we selected, and will not be noticeably affected by overshoot. We held $N = 3$ fixed and varied the GER for the different schemes varying the GER from 0.002 to 0.05, in steps of size 0.002, adjusting the LER as required by each group decision rule to obtain the desired GER.

We expect the simulation's GER to match the GER we programmed the simulation to have. Figure 3.10(a) verifies that the simulation's GER remains close to what we set it to be. The simulated and calculated GERs match best at smaller values, since lower GER values require lower LER values, which translates to less overshoot for each DM.

In Figure 3.10(b), the single DM is the slowest, followed by the Majority Total scheme. For lower set GER values, the Majority First scheme is the fastest of our main three schemes, since it utilizes higher LER values than the Race scheme; however, for higher GERs, the Race scheme becomes faster than the Majority First scheme. The D-SPRT is the fastest overall, since it uses individual DMs with the highest LERs.
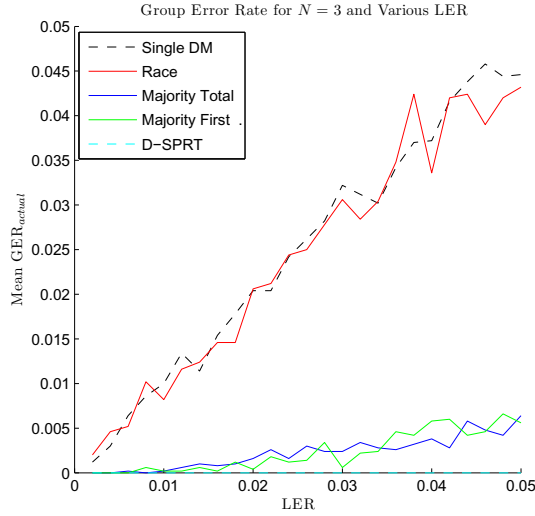
(a) Group Error Rate
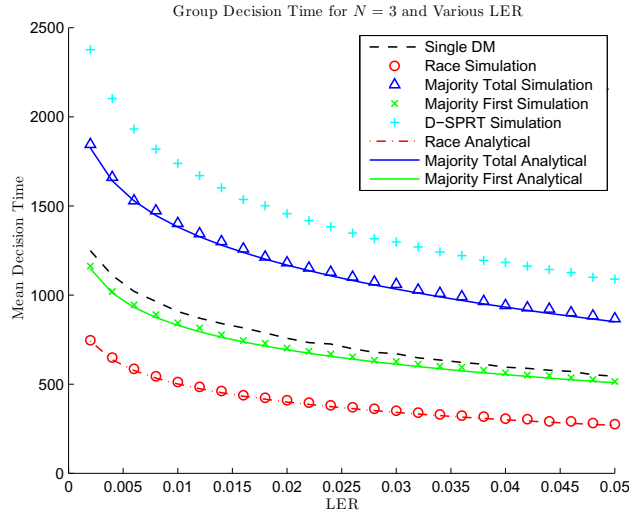
(b) Group Decision Time

Figure 3.9: Error rate verification for $N = 3$ and a given LER. In (a), the GER graph shows that overshoot is not significant for this simulation. The D-SPRT scheme has zero error because $\eta = 5$, and is thus more strict than even a consensus rule. The variation in the graph is due to the finite number of trials it is averaged over (10,000). In (b), the mean GDT is also as expected: since the D-SPRT must wait for at least 5 individual decisions before the fusion center chooses, it is by far the slowest. The Majority Total scheme is the next slowest, followed by the single DM, the Majority First scheme, and the fastest is the Race scheme. This ordering corresponds to the number of slowest DMs the group decision rules allow the fusion center to ignore: the D-SPRT must get 2 decisions from most of the DMs, the Majority Total scheme cannot ignore any DMs, the Majority First scheme can ignore up to Θ DMs, and the Race scheme can ignore $(N - 1)$ DMs.
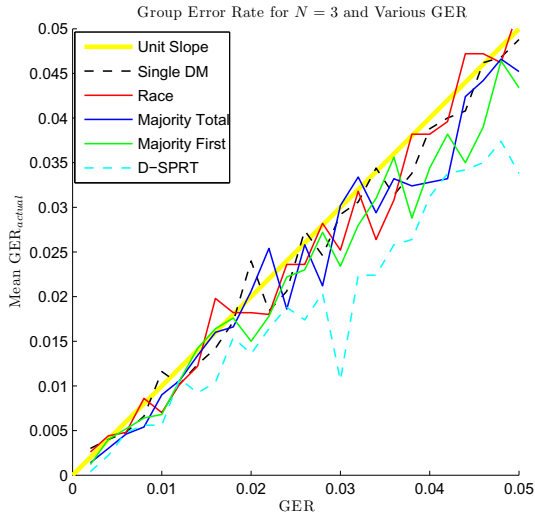
(a) Group Error Rate
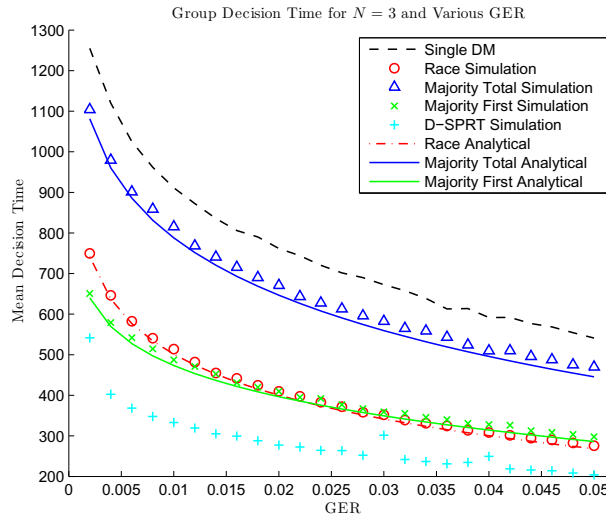
(b) Group Decision Time

Figure 3.10: GER verification for $N = 3$ DMs. The GERs shown in (a) show that all of the group decision rules, simulated over 10,000 trials, produce a mean GER close to what we set it to be. This confirms that our simulations and analytical results for our main three schemes match well. In (b), we show the corresponding mean GDTs. The most interesting point here is that the Majority First scheme is the fastest of our main three schemes for low GERs, but around GER $= 0.022$, the Race scheme becomes faster than the Majority First scheme. The D-SPRT is the fastest overall because its DMs have the highest overall LER.

164

3.3.2 Performance Comparison

The various fusion center decision rules each have different strengths and weaknesses, which we will discuss here in terms of performance. We use two different cases to explore the relative advantages of the schemes: the equal LER case for different N and the equal GER case for different N. The results shown in this section are averaged over 10,000 trials and compared to the results for a single DM for reference.

Equal LER, Different N

In many realistic applications, it is reasonable to assume that one is supplied with N iid DMs, and then must choose a group decision rule that will provide the best performance for one's goals. This situation is equivalent to providing all of the above schemes with N individual DMs with the same LER, and looking at the relative performance of each scheme.

The GERs for the various schemes are shown in Figure 3.11 for LER fixed at 0.01. The GER for the two Majority schemes drop off rapidly with N, while the GER for the Race scheme remains approximately steady and close to the LER. The D-SPRT's GER in this simulation was zero, since, under the parameters we used, it is using a significantly stricter rule than the other group decision rules, and thus had a GER closer to zero than $\frac{1}{10,000}$. The Race scheme's fluctuations are due to the finite number of trials that the values are averaged over.
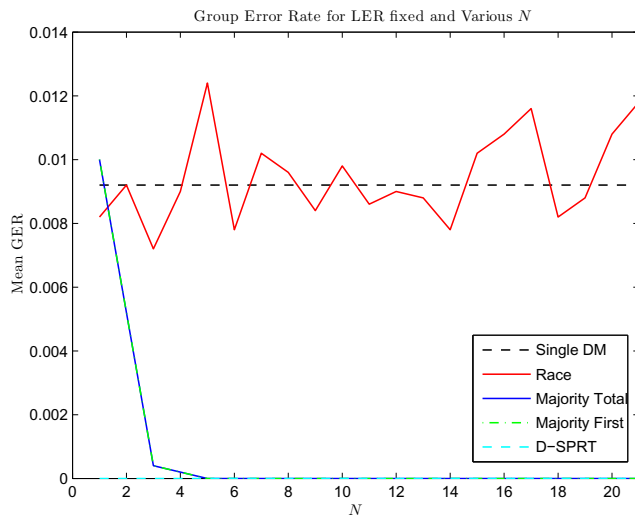
Figure 3.11: Simulation results for the mean GER of N iid DMs using the Race, Majority Total, and Majority First schemes for LER fixed at 0.01. All results are averaged over 10,000 trials. As predicted, the Majority First and Majority Total schemes have identical GERs that drop off rapidly with N, while the Race scheme's GER remains close to an individual's ER, and the D-SPRT's GER is extremely low; for this particular simulation, the D-SPRT has a GER of zero, due to the relatively low number of trials in the simulation. Similarly, the variation in the Race scheme's GER is due to the finite number of trials that the simulation is averaged over. We averaged over only 10,000 trials because it is a large number of trials but still within a reasonable order of magnitude for actual experiments to achieve.
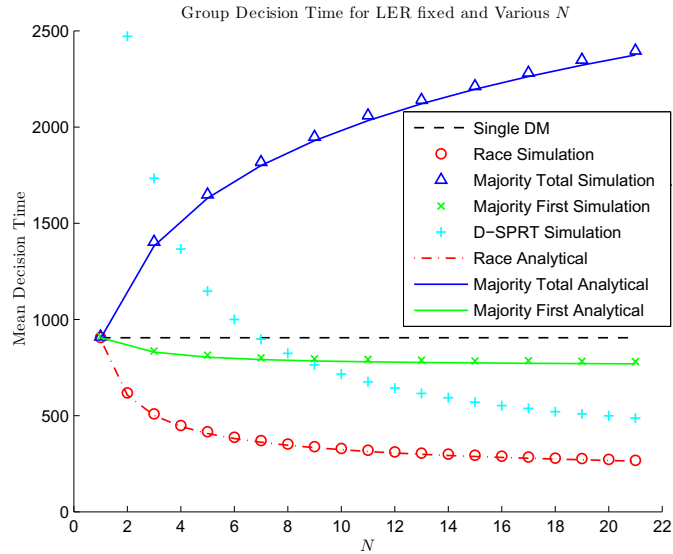
Figure 3.12: Simulation results for the mean GDT of N iid DMs using the Race, Majority Total, and Majority First schemes, overlaid on our analytical solution for the same group. The group schemes are also compared to an individual DM using the SPRT. In this case, the LER was fixed at 0.01, and N was varied from 1 to 21. The interesting trend in the D-SPRT is due to the fact that the number of agreeing cumulative decisions η remains fixed at $\eta = 5$, but the number of DMs available to contribute to those decisions increases with N.

The corresponding GDTs are shown in Figure 3.12. The GDTs for the Majority Total scheme rise with N, while it quickly levels off for the Majority First scheme. The Race scheme provides the fastest performance. Thus, in this case, if one must greatly prioritize decision time, the Race scheme is a reasonable option; otherwise, the Majority First scheme provides a good balance of low GER and low GDT. The D-SPRT shows an interesting behavior compared to the other group decision rules: it drops off approximately exponentially. This is due to the fact that as the number of group members increases, the number of net decisions that the D-SPRT requires to reach a decision does not change. Thus, the overall GDT decreases quickly as N increases.
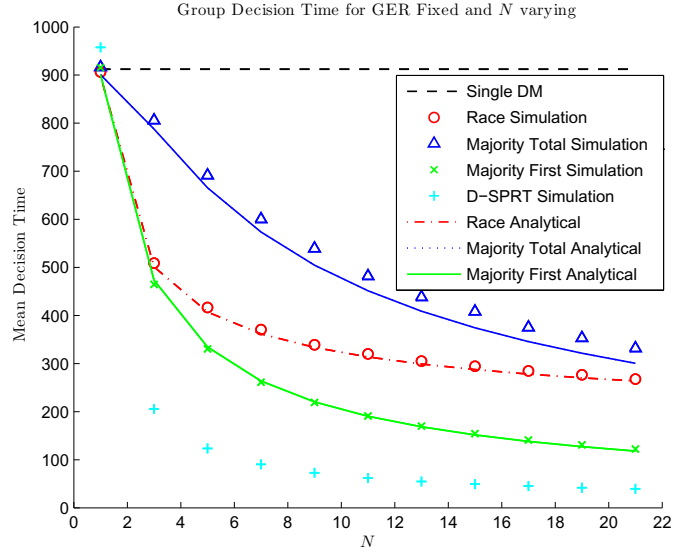
Figure 3.13: Simulation results for the mean GDT of N iid DMs using the Race, Majority Total, and Majority First schemes overlaid on our analytical results for the same groups, and the simulation results for the D-SPRT. The group schemes are compared to an individual DM using the SPRT. The GER was fixed at 0.01, and N was varied from 1 to 21. All results are averaged over 10,000 trials. Over the range of N shown, the Majority Total scheme remained the slowest, while the D-SPRT was the fastest by far, since it used DMs with a lower DT and required a decreasing fraction of the DMs to answer for the fusion center to finish as N increased.

Equal GER, Different N

As another way to compare the different group decision rules, suppose one has a desired GER in mind, and is interested in seeing which decision rule can achieve it with some other desirable properties (speed, cost, etc). This assumes that one can acquire individual DMs of any given LER to achieve the desired GER. As was shown in Figure 3.10, all of the schemes' simulations returned values close to what the simulation was set to provide. The mean GDT results are shown in Figure 3.13 for GER = 0.01 and different values of N.

The results are interesting in that they are not immediately intuitive. As N

increases, for a fixed GER, the LER for each scheme increases, which results in faster individual decisions. The Majority Total scheme rapidly becomes faster than an individual DM even though it must wait for its slowest member because that slowest member has a higher LER than the individual DM. However, this is not enough to make the Majority Total scheme faster than the Race scheme for the N values shown, even though the Race scheme samples from DMs each with LER = GER. We find that for GER fixed, the Majority First scheme is the fastest of our main three schemes for a given level of N, since it combines the benefits of a higher LER (like the Majority Total scheme) with being able to finish before every DM in the group has made a decision (like the Race scheme). In addition to providing faster results, the Majority First scheme may also be better than the Race scheme because it is generally more difficult and/or more expensive to obtain individual DMs with lower LERs. The fastest overall scheme in this situation was the D-SPRT, which had the advantage of using individual DMs with higher LERs (and therefore lower DTs) than the other group rules. The D-SPRT's mean GDT differs slightly from the other group rules at $N = 1$, probably due to overshoot. The best group decision rule for a particular experiment may depend on the experimenter's specific budget and cost function, which formally set the best trade-off between DM cost and GDT.

3.3.3   Group Decision Rule Comparison Discussion

The Race scheme provided consistently fast GDTs in both scenarios, and is the simplest to design, since the fusion center needs neither significant computational power, memory, nor the ability of the DMs to reset its decision variable. Therefore, for situations where one is given a set of iid DMs to work with, the Race scheme provides a simple and fast solution. However, we point out that in the case where one DM encounters a malfunction that causes it to very quickly return a decision that is not related to its observations (i.e., it is either faulty or is hijacked), the overall group decision is vulnerable: since the group scheme itself is very fast, it would be very difficult to differentiate between a group decision drawn from the left-hand tail of the pdf of GDTs and an erroneous or malicious response, especially for large N. The group decision scheme also does not provide a GER that is better than the individual DMs' LER. On the other hand, this scheme is robust to multiple individual DM failure, where the DM cannot communicate with the fusion center.

The Majority Total scheme was generally the slowest in the fixed-LER case, but had a GDT close to the other group schemes in the fixed-GER case and can use less accurate individual DMs to achieve a higher level of accuracy at the group level, which is desirable because lower-accuracy DMs are typically cheaper and easier to acquire. The scheme is also robust to a small number of DMs being hijacked or faulty in a way that has them respond without processing data. In addition, the Majority Total scheme has the advantage of allowing one to calculate a measure

of confidence in the group's decision, since one has responses from all N DMs: if all N agree, then the group is very confident in its decision, whereas if Θ disagree with the majority, then the probability that the group is incorrect is higher. On the other hand, the Majority Total scheme is vulnerable to sensor failure, since it must wait for the slowest DM – if even one sensor becomes unable to communicate with the fusion center, the group never reaches a decision.

The D-SPRT scheme was generally the fastest scheme at larger group sizes, largely because has access to more information than the other group rules. It is less prone to DM failure than the other group decision rules because it can still reach a group decision if up to $N - 1$ of the DMs cannot communicate with the fusion center. However, the D-SPRT is still very sensitive to its DMs being hijacked, since an attacker could easily repeatedly send out the desired decision and thereby hijack the overall group's decision. In addition, the D-SPRT is not a good choice for systems with human observers, as was discussed earlier, and it may over-weight information from certain sources if the rule is applied to a system whose information includes a spatial component or multiple sources of information. However, for a completely device-based system with only one source of information, it provides good performance.

The Majority First scheme represents a good middle-ground alternative to the other schemes: it provided relatively quick performance in the fixed-LER case and the second-quickest performance in the fixed-GER case. It also shares the same

LER-related advantages as the Majority Total scheme, and is the most robust to sensors responding without processing data (due to being faulty or hijacked) of all the group decision rules presented. At the same time, it is robust to a small number of sensors failing (due to either faulty communication equipment or being destroyed), since it only waits for the Υth slowest agreeing member. Thus, while the Majority First scheme may not be a time-optimal solution for all possible situations, its robustness and combination of being both accurate and relatively quick make it an attractive choice, especially when there is uncertainty in the situations the system may face.

3.4   Part III: An Example of A General Group

We stated earlier that our approach to setting up and analyzing our group decision rules can be generalized to groups with non-identical members. As a demonstration of this assertion, we present an example: we consider a general group with $N = 5$, since it is a relatively simple but non-trivial system. This example also serves to illustrate the effect of assuming that a group is iid when it is not.

In our earlier examples, we showed that the pdf of GDTs predicted by our analytical results matched well with a histogram of data from simulation for each of our main three group decision rules. For the Race scheme, this is shown in Figure 3.3; for the Majority Total scheme, this is shown in Figure 3.5; and for the Majority

First scheme this is shown in Figure 3.7. In each of these histograms, the DMs had identical sensitivity, $d' \equiv \mu_1 - \mu_0 = 0.1$, and were assumed to be sufficiently trained to be optimal. The DMs also were trained to have LER = 0.01. In other words, their decision variables took identically-sized steps for equal incoming information and they had identical boundaries, which lead to them having the same LER.

For our example, suppose that the DMs still utilize the SPRT, are still each trained to have the same LER = 0.01, and that the group still has an average d' equal to the previous examples, but that some of the individual DMs have slightly different d's (i.e., sensitivities or abilities). Specifically, we set our group to have Θ DMs with $d' = 0.05$, Θ DMs with $d' = 0.15$, and one DM with $d' = 0.1$. The individual DMs' cdfs and pdfs of DTs are shown in Figures 3.14(a) and 3.14(b), respectively. These will be combined using our three main group decision rules below. All simulations results were averaged over 10,000 trials.

3.4.1   General Group Example: Race Scheme

In Figure 3.15, we compare the output of a simulation of our general group with the analytical result for the pdf of GDTs from Equation (3.14) in red, and the analytical result for the pdf of GDTs for the iid case found in Equation (3.20) in green.

As shown, even though the DMs in both analytical solutions have the same average LER as a group, the individual differences were significant enough to alter

(a) Group Error Rate (b) Group Decision Time
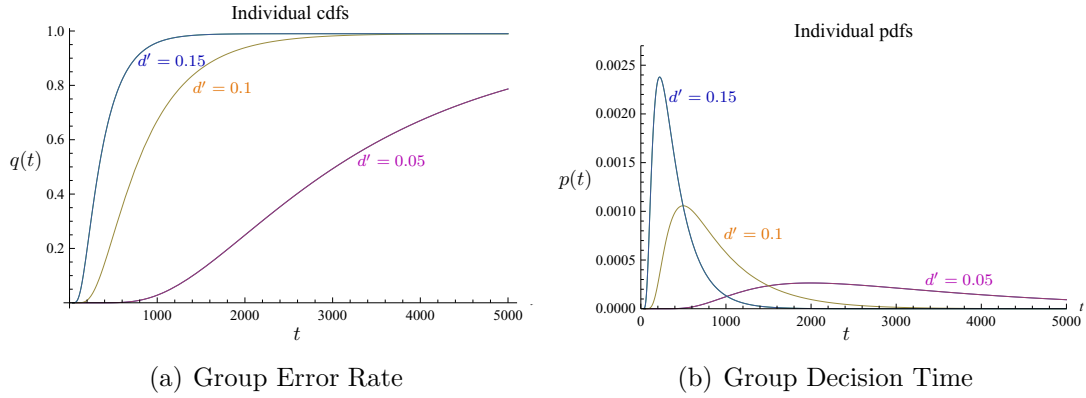
Figure 3.14: The cdfs and pdfs of the DMs in the general group example in Section 3.4. $\Theta=2$ DMs had $d' = 0.05$, Θ DMs had $d' = 0.15$, and one DM had $d' = 0.1$. Assuming that each member in the group utilized the SPRT and had been optimally trained to have a LER of 0.01, (a) shows each member's cdf of DTs (note that the top and bottom lines correspond to $d' = 0.15$ and $d' = 0.05$, respectively, and each of those lines represents two DMs, for a total of five), and (b) shows each member's corresponding pdf of DTs. We then combine these different distributions to find the group pdf of GDTs under our three main group decision rules.
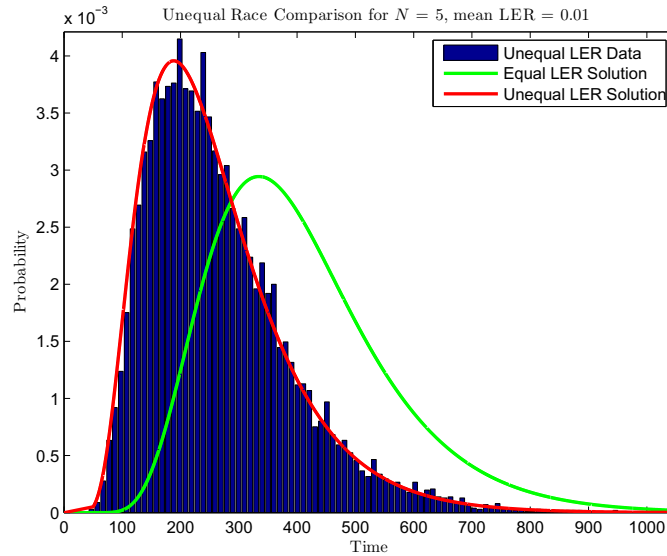
Figure 3.15: Comparison of our analytical results for general and iid DMs with a histogram of GDTs from a simulation of our example general group using the Race scheme. The analytical result for the general DMs is shown in red, and the analytical result for iid DMs with the same average d' as the general group is shown in green. It is evident that making the individual DMs non-identical has significantly shifted the group's pdf of GDTs to the left: using the approximation that the DMs are iid results in times that are too conservative. It is also evident from these results that our analytical formula matches the simulation well in only 10,000 trials.

174

the shape of the group pdf under the Race scheme. The group's pdf of GDTs shifted to the left because the DMs with higher d' values tend to be faster than the DMs with lower d's. This leads to lower group decision times under the Race decision rule. Thus, it is important to accurately represent the sensitivity of each DM: taking an average over all the DMs and then declaring the group to be iid will not provide sufficiently accurate results. We note that this demonstrates that the Race scheme's pdf of GDTs is sensitive to changes in the individuals' d's.

3.4.2 General Group Example: Majority Total Scheme

The comparison between simulation, our analytical results for the general group, and our analytical results for the iid group with the same average LER is shown in Figure 3.16 for the Majority Total group decision rule.

Like for the Race scheme, approximating that a group is iid and using the group's average LER does not accurately capture the performance of the general group. In the Majority Total scheme's case, the approximation leads to times that are significantly shorter than what simulation and our general analytical results return. This also demonstrates that the Majority Total scheme is quite sensitive to inaccuracies in each DM's value of d'.
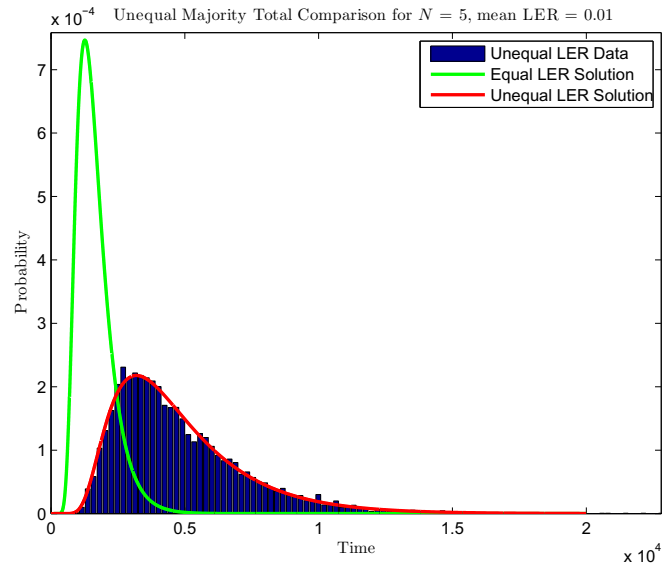
Figure 3.16: Comparison of our analytical results for general and iid DMs with a histogram of GDTs from a simulation of our example general group using the Majority Total scheme. The analytical results for the general DMs is shown in red, and the analytical result for iid DMs with the same average d' as the general group is shown in green. As one can see, the iid approximation does not work well, as it predicts GDTs that are far too optimistically short. It is also evident from these results that our analytical formula matches the simulation well in only 10,000 trials.
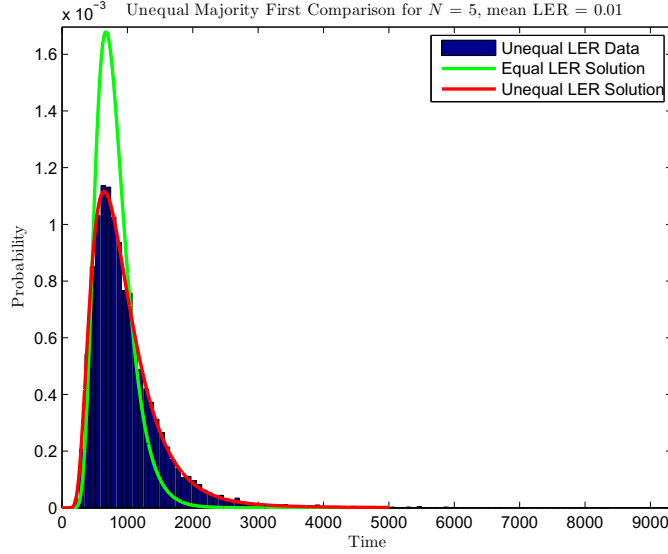
Figure 3.17: Comparison of our analytical results for general and iid DMs with a histogram of GDTs from simulation for the Majority First scheme. The pdf of GDTs predicted by our general analytical formula is shown in red, and the pdf of GDTs for an iid group with the same average LER is shown in green. For the Majority First rule, the general group's average GDT is approximately the same as the iid group's GDT. However, the distribution's overall shape did change as a result of the difference in the individuals' d' values, and only the general solution (red) remained close to the simulation results.

3.4.3 General Group Example: Majority First Scheme

In Figure 3.17, we compare the general analytical results found in Equation (3.45) in red, with a histogram from simulation of the general group in dark blue and our previously-found results for a group of iid DMs in green, for the Majority First scheme.

Since Θ members of the group are faster than the iid group, one member is the same speed, and Θ group members are slower than the iid group, we expect that the mean GDT should not change much if at all. This intuition is verified. However, the change in individual abilities does in fact change the shape of the pdf

of GDTs, and only the general analytical solution accurately represents this. In addition to verifying the accuracy of our general analytical solutions, this example also shows that the performance of the Majority First scheme is less sensitive to non-uniformities in the individual group members' abilities than either the Race or Majority Total schemes.

3.5 Part IV: General Group Rules

In the previous section we demonstrated that our approach could be used for general groups of DMs for our three rules; here, we demonstrate that our approach also can be used to generalize the group rules themselves. There are two obvious extensions to our group rules: the η-Total and η-First group schemes. We also show these rules in [48].

3.5.1 The η-Total Scheme

The general form of the Majority Total scheme is the η-Total rule [48], in which the fusion center returns a decision once η DMs have returned an individual decision. The fusion center then applies a majority rule to the individuals' responses to reach a group decision. For an actual decision rule, it is best to use odd values of η, since this guarantees that there will not be any ties; however, from a calculation-based standpoint, there is no reason why η must be odd, so we present formulas that also hold for η even. For this rule, N can easily be either even or odd.

The general formula for the cdf of GDTs for the η-Total rule is $q_{g\eta}^{\mathrm{Tg}N}(t_g)$ (super-script: η-[T]otal group rule, [g]eneral individual DM, N DMs; subscript: [g]roup cdf, [η] individual DMs must finish for the fusion center to finish):

$$q_{g\eta}^{\mathrm{Tg}N}(t_g) = \sum_{\theta=\eta}^{N} \left[\sum_{i_1=1}^{(N-\theta-1)} \sum_{i_2=i_1+1}^{(N-\theta+2)} \cdots \sum_{i_\theta=i_{\theta-1}+1}^{N} \left(\prod_{m=1}^{\theta} q_{i_m} \right) \prod_{\substack{j=1, \\ j\notin\mathcal{I}}}^{N} [1-q_j] \right], \qquad (3.54)$$

where $\mathcal{I} = \{i_1, i_2, \ldots i_\theta\}$, which designates the set of DMs who have reached a decision by time t_g. To get the pdf of GDTs, we take the derivative of the cdf. After some simplification, we arrive at the following general formula:

$$p_{g\eta}^{\mathrm{Tg}N} = \sum_{i_1=1}^{(N-\eta+1)} \sum_{i_2=i_1+1}^{(N-\eta+2)} \cdots \sum_{i_\theta=i_{\theta-1}+1}^{N} \left(\sum_{m=1}^{\eta} p_{i_m} \prod_{\substack{k=1, \\ k\neq m}}^{\eta} q_{i_k} \right) \left(\prod_{\substack{j=1, \\ j\notin\mathcal{I}}}^{N} [1-q_j] \right). \qquad (3.55)$$

For iid DMs, the formula for the pdf of GDTs simplifies considerably: $p_{g\eta}^{\mathrm{Ti}N} = \binom{N}{\eta} \eta p_\iota q_\iota^{\eta-1} [1-q_\iota]^{N-\eta}$. For $N=11$, the effect of increasing η from 1 to N is shown in Figure 3.18. For $\eta=1$, the η-Total rule is identical to the Race scheme, and for $\eta=N$, the η-Total rule is identical to the Majority Total scheme. This can also be seen by comparing Equation (3.55) with the pdf of GDTs for the Majority Total scheme, given by Equation (3.29).

For iid DMs, we can also find the GER in a straightforward manner. Let $\Phi = \frac{\eta-1}{2}$ be the smallest possible minority for η total DMs, and let us consider only

η−Total Scheme for Varying η

Figure 3.18: Analytical results for the η-Total group rule for different values of η, with $N = 11$. For $\eta = 1$, this rule is identical to the Race scheme, and for $\eta = N$, this rule is identical to the Majority Total scheme.

odd values of η to avoid ties. Then we have

$$\text{GER}_\iota = \sum_{\phi=0}^{\Phi} \binom{\eta}{\phi} \text{LER}_\iota^{\eta-\phi} \left(1 - \text{LER}\right)_\iota^{\phi},\tag{3.56}$$

where ϕ represents the number of DMs in the minority.

3.5.2   The η-First Scheme

The general form of the Majority First scheme is the η-First rule, in which the fusion center returns a decision once η DMs have returned the same individual decision [48]. Realistically, any implementation of this scheme should require that $\eta \leq \Upsilon$, the minimal majority: if $\eta > \Upsilon$, there is no guarantee that the fusion center's rule will ever be satisfied. However, it is still possible to calculate the cdf

and pdf of GDTs given that the group does reach a decision in finite time, and we provide formulas for doing so below. When $\eta = \Upsilon$, the η-First rule is identical to the Majority First scheme, and when $\eta = 1$, the η-First rule is identical to the Race scheme. This rule can easily accommodate both even and odd N and η.

If η DMs finish by time t_g and agree, then up to $(N - \eta)$ DMs do not contribute. Using this, we can construct the cdf of GDTs, $q_{g\eta}^{\mathrm{Fg}N}$ (superscript: η-[F]irst scheme, [g]eneral individual DM, N DMs; subscript: [g]roup pdf of GDTs, [η] DMs must agree for the fusion center to finish). Because of the ordering that occurs under this scheme, our formula is split into two cases: $\eta \leq \Theta$ and $\eta \geq \Upsilon$. For simplicity, we define the subfunctions

$$
\Gamma_{\mathcal{I}}^{\gamma \mathrm{g}} = \sum_{i_1=1}^{(\eta+\gamma+1)} \sum_{i_2=i_1+1}^{(\eta+\gamma+2)} \cdots \sum_{i_{(N-\eta-\gamma)}=i_{(N-\eta-\gamma-1)}+1}^{N} \left(\prod_{m=1}^{N-\eta-\gamma} [1 - q_{i_m}] \right),
$$

$$
\Lambda_{\mathcal{J}\mathbb{D}}^{\xi \mathrm{g}} = \sum_{\substack{j_1=1, \\ j_1 \notin \mathcal{I}}}^{(N-\xi+1)} \sum_{\substack{j_2=j_1+1, \\ j_2 \notin \mathcal{I}}}^{(N-\xi+2)} \cdots \sum_{\substack{j_\xi=j_{\xi-1}+1, \\ j_\xi \notin \mathcal{I}}}^{N} \left(\prod_{k=1}^{\xi} q_{j_k \mathbb{D}} \right),
$$

where $\Gamma_{\mathcal{I}}^{\gamma \mathrm{g}}$ specifies all unique sets of $(N - \eta - \gamma)$ DMs who do not reach a decision by time t_g, $\Lambda_{\mathcal{J}\mathbb{D}}^{\xi \mathrm{g}}$ specifies all unique sets of ξ DMs who choose decision $\mathbb{D}$ by time t_g when the fusion center chooses $\hat{\mathbb{D}}$ (the other hypothesis: $\hat{\mathbb{D}} \in \{\mathrm{S}, \mathrm{N}\}, \hat{\mathbb{D}} \neq \mathbb{D}$), and

$\Gamma_{\mathcal{I}}^{(N-\eta)\mathrm{g}} = \Lambda_{\mathcal{JD}}^{0\mathrm{g}} = 1$. Then, for $\eta \leq \Theta$, we have

$$
\begin{aligned}
q_{g(\eta \leq \Theta)}^{\mathrm{Fg}N} = \quad & \sum_{\gamma=0}^{\eta-1} \left(\Gamma_{\mathcal{I}}^{\gamma\mathrm{g}} \sum_{\xi=0}^{\gamma} \left[\Lambda_{\mathcal{JS}}^{\xi\mathrm{g}} \prod_{\substack{\ell=1, \\ \ell \notin \mathcal{I},\mathcal{J}}}^{N} q_{\ell\mathrm{N}} + \Lambda_{\mathcal{JN}}^{\xi\mathrm{g}} \prod_{\substack{\ell=1, \\ \ell \notin \mathcal{I},\mathcal{J}}}^{N} q_{\ell\mathrm{S}} \right] \right) + \\
& \sum_{\gamma=\eta}^{N-\eta} \left(\Gamma_{\mathcal{I}}^{\gamma\mathrm{g}} \sum_{\xi=0}^{\eta+\gamma} \Lambda_{\mathcal{JS}}^{\xi\mathrm{g}} \prod_{\substack{\ell=1, \\ \ell \notin \mathcal{I},\mathcal{J}}}^{N} q_{\ell\mathrm{N}} \right).
\end{aligned}
\tag{3.57}
$$

The formula for $q_{g(\eta \geq \Upsilon)}^{\mathrm{Fg}N}$ is essentially the same, except that it only includes the first term of Equation (3.57) and the limit on the outermost summation in that term is $(N - \eta)$ instead of $(\eta - 1)$. Here, γ denotes the maximum number of excess members in the subgroup that sets the group's decision (i.e., for the $\gamma = 0$ case, exactly η DMs finish by time t_g and agree), and ξ denotes the number of DMs that finish by time t_g but do not agree with the fusion center's final decision. Since $p_{g\eta}^{\mathrm{Fg}}(t_g) = \frac{d}{dt_g}\left[q_{g\eta}^{\mathrm{Fg}}(t_g)\right]$ for both cases of η, it is straightforward to calculate the corresponding formulas for the pdf of GDTs.

Like before, the formula for the cdf of GDTs simplifies considerably for the iid case. Let $\Gamma^{\gamma\mathrm{i}} = \begin{pmatrix} N \\ N - \eta - \gamma \end{pmatrix} [1 - q_\iota]^{(N-\eta-\gamma)}$ and $\Lambda_{\mathbb{D}}^{\xi\mathrm{i}} = \begin{pmatrix} \eta + \gamma \\ \xi \end{pmatrix} q_{\iota\mathbb{D}}^{\xi}$. Then we have

$$
q_{g(\eta \leq \Theta)}^{\mathrm{Fi}N} = \sum_{\gamma=0}^{\eta-1} \Gamma^{\gamma\mathrm{i}} \left(\sum_{\xi=0}^{\gamma} \left[\Lambda_{\mathrm{S}}^{\xi\mathrm{i}} q_{\iota\mathrm{N}}^{\eta+\gamma-\xi} + \Lambda_{\mathrm{N}}^{\xi\mathrm{i}} q_{\iota\mathrm{S}}^{\eta+\gamma-\xi} \right] \right) + \sum_{\gamma=\eta}^{N-\eta} \Gamma^{\gamma\mathrm{i}} \sum_{\xi=0}^{\eta+\gamma} \Lambda_{S}^{\xi\mathrm{i}} q_{\iota\mathrm{N}}^{\eta+\gamma-\xi}, \tag{3.58}
$$

and again, the formula for $q_{g(\eta \geq \Upsilon)}^{\mathrm{Fi}N}$ contains only the first term, with the limit on
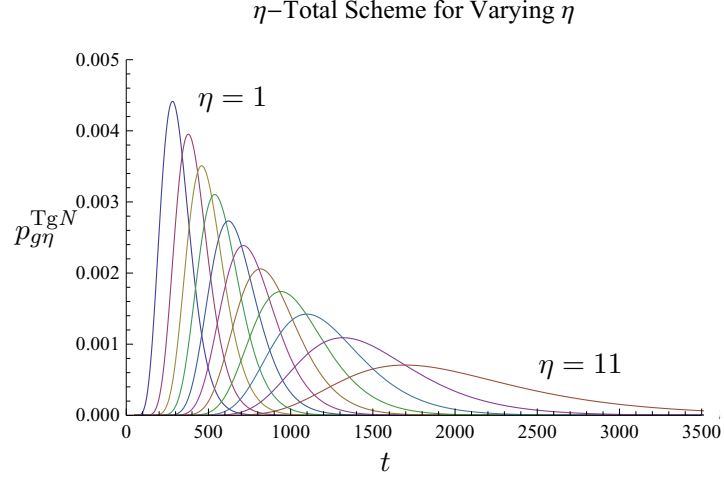
η−First Scheme for Varying η
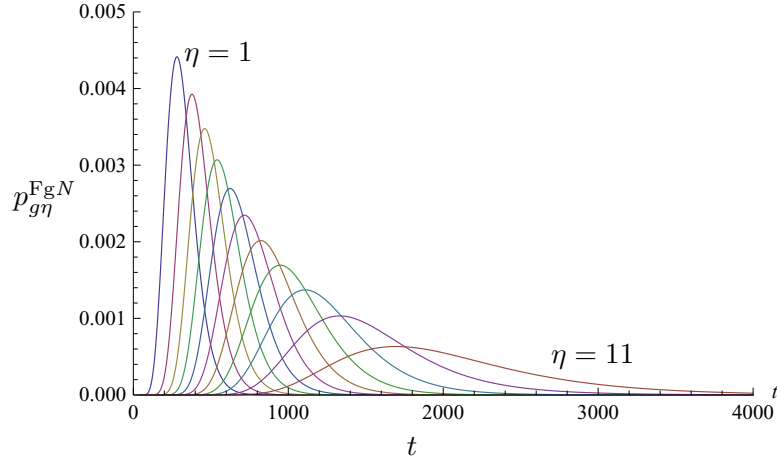
Figure 3.19: Analytical results for the η-First group rule for different values of η, with $N = 11$. For $\eta = 1$, this rule is identical to the Race scheme, and for $\eta = \Upsilon$, this rule is identical to the Majority First scheme. This rule is not guaranteed to finish for $\eta > \Upsilon$; however, we can still calculate the pdf of GDTs given that the group does finish for those values of η, which is shown in this figure.

the first summation sign modified to $(N - \eta)$. For $N = 11$, the effect of increasing η from 1 to N is shown in Figure 3.19. For $\eta = 1$, the η-First scheme is equivalent to the Race scheme, and for $\eta = \Upsilon$, it is equivalent to the Majority First scheme. We have verified these analytical results with simulation.

For the iid case, we can also calculate the GER for the η-First scheme. Like before, the exact form of the equation depends on the relative values of η and N. If $\eta \leq \Upsilon$, then

$$\text{GER}_{\iota, \eta \leq \Upsilon} = \sum_{\phi=0}^{\eta-1} \begin{pmatrix} \eta + \phi - 1 \\ \phi \end{pmatrix} (1 - \text{LER}_{\iota})^{\phi} \, \text{LER}_{\iota}^{\eta}. \tag{3.59}$$

The formula for $\text{GER}_{\iota, \eta > \Upsilon}$ is the same except that the limit on the summation is

183

$(N - \eta)$ instead of $(\eta - 1)$.

3.5.3 Discussion

For the particular examples shown, the pdf of GDTs for the correct answer dominates the overall pdf of GDTs. Therefore, while the formulas are different for the η-Total and η-First schemes, Figures 3.18 and 3.19 appear very similar. However, we stress that the pdfs in each figure for each value of η are not identical, except for the case of $\eta = 1$, where both schemes are equivalent to the Race scheme. The difference between the curves of the η-Total and η-First schemes with iid DMs will be more pronounced for DMs whose pdf of GDTs is different for the correct hypothesis than for the incorrect hypothesis, or who have a higher LER.

3.6 General Discussion

The approach we take to designing our group decision-making model is intuitive, and is supported by previous work in various experiment-based literatures. Our model for the individual sensor is based on Wald's SPRT, which has been shown to be optimal for the task we have described. We base our claim that one can reasonably characterize a human observer performance with a pure DDM on experimental data [14], which found that when sufficient factors are taken into account in the more realistic "extended" DDM, the factors tend to minimize each others' effect, and one can again approximate the result with a "pure" DDM. The

approach we take to establishing a group rule is similar to that described in cognitive psychology, economics and political science. We mention the social psychology and information pooling literature because it explains several issues that will need to be accounted for when implementing the group decision schemes with human observers, especially if the observers are allowed to interact.

Our work also shares a number of similarities with research in decentralized detection, distributed detection, and decision aggregation, but should be more accessible to a wider range of communities, since our results and analysis for the group are expressed in terms of pdfs, cdfs, ERs, and DTs, all of which are very intuitive and used in a wide range of disciplines. We also show exact solutions to our individual model and group models, and exhibit a novel approach to modeling the performance of a general (i.e., non-identical) group, all without the explicit use of specialized subjects such as measure theory or dynamic programming.

We do not specifically mention a cost function in our work; nor do we claim that our group solutions are optimal. It is likely that one can post-facto construct a specific cost function for each scheme we have presented for which that scheme is optimal; therefore, we feel that only valuing group decision rules which can be claimed as optimal in some narrow setting is limiting. Also, systems that are optimized for a particular situation may contain fragilities when confronted with situations outside of the one it was optimized for. Thus, when designing a system that may face a wide range of situations, the best solution may not necessarily be

185
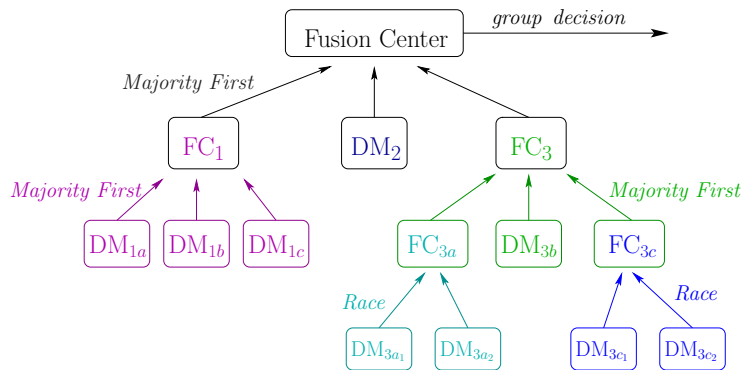
Figure 3.20: An example of a very general hierarchical group whose performance can be solved for by our models and methods. In this case, some of the individual "DMs" are actually fusion centers of subgroups of DMs, as denoted. In this figure, the font and line colors indicate the subgroups and their fusion centers, as well as the group decision rule used in each subgroup. The Race schemes shown here have an even number of DMs, since the Race scheme can easily accommodate any number of DMs with minimal modification – a tie in the Race schemes can be settled with a coin flip.

one that is "optimal", as we have noted in our discussion of the three group rules.

Our setup takes the performance of N independent members and finds how the group performs under each of three simple group decision rules and two generalized families of rules, the η-Total and η-First rules. However, we note that each "member" need not be an individual, as long as its performance can be characterized by an ER and a pdf of DTs. Therefore, it is straightforward to extend our models to more complicated hierarchical groups, like the one shown in Figure 3.20, where a "member" in some sub-groups is itself another group, characterized by one of the group models presented here or elsewhere in the literature.

3.7 Conclusion

In this chapter, we thoroughly investigated group decision rules for a 2AFC task.

In Part I, we derived the pdf of GDTs for N general DMs using a fusion center to apply the Race, Majority Total, or Majority First group decision rule in a 2AFC task, while under the understanding that each DM could be either a human observer or a detector (device). We also showed how the group decision rules' pdfs changed with increasing group size, for iid DMs. To verify these analytical results, we demonstrated that each solution matched the discrete-time results from simulating the group using the SPRT. We also analyzed the D-SPRT scheme via simulations to find a good value of η, and introduced other related group decision rules.

In Part II, we verified that simulation of our group decision rules returned the expected GER, and then compared the Race, Majority First, Majority Total, and D-SPRT schemes under fixed LER and GER conditions. In our simulations, the D-SPRT was shown to generally be the fastest, but the Majority First scheme promised to also be a good alternative. We also discussed the relative merits of each scheme with respect to robustness.

In Part III, we presented an example in which we used our general analytical results to predict the performance of a group of $N = 5$ independent, non-identical, SPRT-based DMs using the Race, Majority First, and Majority Total group decision rules. We found that our general formulas accurately predicted the group's

performance and that comparing a non-identical group to an iid group with the same average LER provided insight on the robustness of each group decision rule to variation in individual observer performance.

Finally, in Part IV, we generalized the Majority Total rule to the η-Total rule, and the Majority First rule to the η-First rule, further demonstrating the flexibility of our approach, and completely solving for the performance of a general group under one of these families of group decision rules.

The models presented here are relevant to many situations in which a group of DMs must reach a collective decision in a sequential task. Since our models only use each members' performance to find the group's performance, they can easily be used for cybernetic groups (including both human observers and detectors), and naturally extend to hierarchical and more complicated group topologies in which some "members" are groups. Our models are interesting because they present a novel and general way in which one can intuitively yet mathematically calculate a group's performance based on its members' statistics, establish a reasonable base model which can be extended to build up more complicated models for realistic groups, and provide a means by which one can compare different group decision rules.

Next, we change gears and show an analysis of data from a psychophysics experiment where a group of human observers collaboratively performed a simple Signal Detection task.

Chapter 4

Signal Detection-Based Group Decision-Making

4.1 Introduction

The question of how and how well groups of humans reach a decision has been under consideration for a long time. Early work by Condorcet [21] in 1785 established the idea that a group of observers will out-perform an individual decision-maker as long as each observer has a probability greater than 50% of being correct. This philosophy is prevalent in Western society - important decisions are generally made by boards, committees, or the general populace, rather than by a single individual, because it is widely believed that a group is more accurate and wiser than any individual.

However, there is a large body of experimental evidence showing that groups of human observers often do not out-perform the group's most capable member, and even when the group does, it generally does not perform as well as statistical analysis suggests it should. A comprehensive review by Davis [26] of works studying group decision making between 1950 and 1990 illustrates various studies that challenge commonly-held beliefs about groups. Research suggests that groups do not perform as well as statistical analysis predicts because of process losses like social loafing, the state in which some (possibly all) members of the group reduce their individual efforts because they believe that the other group members will make up the difference or that the overall group's ability will not be affected by their reduced contribution. Other sources of process losses include groupthink and intra-group competition [43]. Related areas of research in group decision-making include information pooling [28, 90, 102], optimal committee size [35], and the role of advice [42, 103, 104].

Work in this area by Sorkin et al. [86, 87, 88, 89] uses Signal Detection Theory (SDT) to analyze the performance of a group, and focuses mainly on majority-based rules, Condorcet-like schemes, and comparing a human group's performance to the performance of the "Ideal Group," which by definition represents the best long-term performance possible for a given group of individuals. In their experiments, each observer was given independent, non-identically-distributed samples of data, consisting of a series of measurements on vertical bars from a distribution centered

at one of two values, called "Noise" and "Signal" [87, 88]. The distributions that generated the observers' data samples had the same mean, but the experimenters manipulated the distributions' variances with the intention of setting the sensitivity of each observer. After viewing the data, the observers in the group were asked to confer and reach a binary Yes (Signal) or No (Noise) decision on which distribution everyone's samples were drawn from. The groups were given feedback on the accuracy of the group decision. In their first experiment [87], the observers did not provide a response in the individual task, so the experimenters could not accurately judge each observer's actual abilities during the experiment. The group's responder was selected at random, and an analysis of the data showed that the responder tended to downweight his groupmates' opinions and overweight his own. This result was based on correlations between the data shown to each observer and the group's responses in those trials. Also, the groups utilized a Condorcet-like voting strategy, rather than discussing their opinions (as the experimenters had hoped), because there was a slight premium on finishing the task quickly and the group's responder was tasked only with giving a binary response. It was argued that an observer's confidence was communicated with her binary response through verbal cues such as "I think ... " versus "It's definitely ...", but without a formal scale, it is not certain to what extent the other members of the group used or could interpret those cues. Many of these problems were addressed in the second round of experiments in [87]: the premium on time was eliminated, so the group was awarded only for accuracy,

191

and the individuals were required to input a rating that expressed a preference for one hypothesis as well as the level of confidence in that decision during the individual phase of the task. The individual ratings were then sorted by response and displayed for the group phase, which made it simpler for the group to see which alternative was generally favored. This also made it easier for the group to prevent the responder from overweighting his dissenting opinion by providing negative feedback when the responder's decision did not match the "obvious" choice, based on the individual ratings; thus, instances of a single observer hijacking the group's decision were supposedly minimized (but not eliminated). However, this setup did not identify which rating came from which observer, making it virtually impossible for the group members to accurately assess each other's ability: the observers typically self-identified only when their answer was a dissenting outlier.

Our study shares some similarities with the second experiment; however, in our experiment, the stimulus consisted of a Gaussian signal (if present) embedded in a white noise field, and in each trial, each observer was shown the same stimulus, so any individual noise in our model is internal. Our setup is more realistic for visual tasks: in applications involving a visual task such as cancer detection in mammograms, or threat detection in baggage screening, the same stimulus is available to all observers. Thus, our setup allows each observer to more accurately assess the abilities of the other observers: in the Sorkin et al. setup, a capable observer could be given a set of data with a larger amount of noise than a less capable observer.

This would make the two observers appear to have the same ability, or worse, make the less capable observer appear to be more capable, making it significantly more difficult for the observers to accurately estimate each others' abilities.

Our interests are based around five main subjects. First, we sought to determine the performance of three observers in a simple Signal Detection task and compare that with the performance of a group made up of those three observers. Is the group better than its members? How efficient are the observers and groups at the task? Our second aim was to explore different ways of visualizing the data, by analyzing it in different windows of time, to see what further information (if any) can be gleaned from the additional views. Current practice simply looks at the average over all trials, and often combines data across observers or groups, assuming that a "Law of Large Numbers"-type view of the data provides enough information: at most, a per-session view of the data is sometimes considered. Third, we wanted to determine how the observers arrived at a group decision. Did the observers use optimal weights? Did the weights change in time, and were they a function of observer ability? What strategies or rules were used to arrive at a group decision? What group rules are close to optimal and realistically implementable in a system? Fourth, we analyzed the results of a continuation study to determine if a group's performance and strategy evolved over a longer set of trials. Last, we explored the possibility that an observer's sensitivity may not be fixed.

We begin with a description of the experiment in Section 4.2, then discuss the

observers' performance. In Section 4.3, we discuss performance, including individual and group Percent Correct (PC) and sensitivity (d'), and the weights assigned to each member. We also analyze all of the above in different windows of time. In Section 4.4, we apply eleven different group rules in a manner similar to Chapter 3, to see if we can identify what strategy each group actually used to reach its group rating. We show the results of our continuation study in Section 4.5, demonstrating the extent to which the trends we found in the original study remained over a larger number of trials. We then present the results of simulating the groups under three different conditions in Section 4.6, to see the effect of different assumptions on the observers' d' values. We conclude with a discussion of our results in Section 4.7.

4.2 Experiment

4.2.1 Method

We thank Dr. Binh Pham, a former Vision & Image Understanding postdoctoral fellow and now a Los Alamos National Laboratory Senior Engineer, for sharing the raw data from this experiment with us.

Participants

Nine undergraduate UCSB students participated in the study. The nine students were divided into 3 groups of three members each for the experiment. All subjects

had normal or corrected vision. To provide relative anonymity, each observer will be referred to by his or her initials, as is customarily done in human studies. Group 1 was comprised of observers AC, KM, and KT; Group 2 consisted of observers JW, KV, and TS; and Group 3 had observers CG, MR, and SF.

All of the groups trained on 100 trials before participating in the experiment, then completed 100 trials (1 session) per day for 6 days, for a total of 600 trials. Group 3 returned for a continuation study to assess the long-term performance of a group of human observers, and completed an additional 54 sessions, for a total of 6,000 trials. All subjects were compensated for their time with course credit.

Apparatus and Stimuli

The stimulus was generated in Image Display Language (IDL) and presented on M17LMAX monochrome monitors (Image Systems, Minnetonka, MN) with a maximum resolution of 1664×1280 pixels. The luminance versus gray level relationship was set on the linear response function. The monitors were calibrated with the "black" luminance set to 0.00 cd/m^2, and the "white" luminance set to 40.00 cd/m^2. The observers maintained a constant distance of 50 cm from the display, which resulted in a subtended angle of 0.0376 degrees per pixel.

The signal (if present) was a positive (white) Gaussian, whose luminance at location (x, y) is given by

$$E(x, y) = E_0 e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}}, \tag{4.1}$$

where E_0 is the maximum luminance in the center of the Gaussian signal, $\sigma = 4$ pixels (0.1504 degrees) is the spatial standard deviation of the Gaussian signal's luminance distribution, and (x_0, y_0) is the location of the center of the Gaussian signal. In this experiment, $E_0 = 14.10$ gray levels; equivalently, 5.53% of the maximum luminance. A reference image of the signal without noise, labeled "Target", was displayed at the top of the screen throughout the experiment. The standard deviation of the white noise was equal to 25 gray levels, resulting in a Signal-to-Noise Ratio (equivalently, d') of 4.00001.

Procedure

The members of a group each sat in front of their own computer, separated by partitions, in a quiet dark room. Each observer first performed the individual signal detection task, then conferred with the other observers to arrive at a group response. All members within a group were shown the same image in each trial, and the same progression of images were used for all groups. Below, we describe the individual and group tasks in detail.

In the individual task, the observer was presented with a noisy stimulus for 1.5 seconds, and his task was to decide whether or not a target was present in the stimulus, and to rate the level of certainty with which he made that decision. The

stimulus consisted of a 180x180 pixel (6.768 degrees) white noise field with a 68x68 pixel (2.5568 degree) black box located in the center of the white noise field, to cue the observer to the location where the target would appear (if present). There was a 50% prior probability of the target being present in any given trial. After 1.5 seconds, the stimulus was removed, and the observer was prompted to select one response from a range of choices, by clicking on the appropriate box with a standard mouse. In addition to selecting either Yes (target present) or No (target absent) as a binary response, each individual chose one of four ratings to express his level of confidence: "Ex[tremely] cert[ain] NO", "Very cert[ain] NO", "Certain NO", "Some[what] cert[ain] NO", "Some[what] cert[ain] YES", "Certain YES", "Very cert[ain] YES", and "Ex[tremely] Cert[ain] YES". These options remained on the screen throughout the experiment. An example of the stimulus and response options is shown in Figure 4.1(a). To give a sense of the difficulty of the task, we note that Figure 4.1(a) shows a signal present trial.

Once the observer finished the individual task, the instructions "Discuss rating and record group answer" were placed on the screen between the cue and the response boxes, as shown in Figure 4.1(b), and he was required to wait for the rest of the observers in the group to finish the individual task before moving on to the group task. The group task was for the observers to confer and decide on a collective rating. The response options were the same for the individual and group tasks. The group members were not given explicit instructions on how to reach
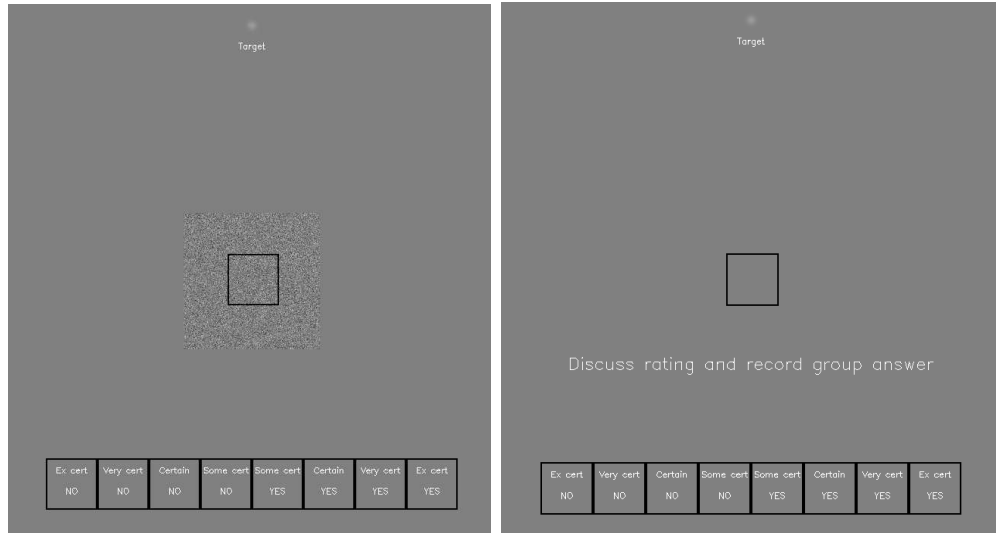
(a) Stimulus for the Individual Task   (b) Response Screen for the Group Task

Figure 4.1: Screenshots from the experiment: (a) a sample stimulus, which was displayed to each observer for 1.5 seconds during the individual task. After that time, the stimulus disappeared, and the observer was given an unlimited amount of time to decide on a response and confidence rating, which he submitted by clicking on the appropriate box on the bottom of the screen. Each observer had to wait for his groupmates to finish the task before moving on to the group task. (b) Once everyone in the group had responded, the message "Discuss rating and record group answer" was placed on the screen, prompting the group to confer and decide on a rating that expressed their collective decision. The response boxes at the bottom of the screen and the labeled reference image of the target at the top of the screen were available throughout the experiment.

a group rating. Once the group agreed on a rating for the trial, every member submitted the group's decision, again by clicking on the appropriate box. After all group members finished submitting the group response, each member's screen displayed feedback on the accuracy of the group response. The observers were given an unlimited amount of time to make their individual and group decisions.

4.2.2 Data Preprocessing

For our analysis, we converted the individual ratings to a numerical scale of 0-7, with 0 corresponding to "Extremely certain NO", and 7 corresponding to "Extremely certain YES". We will continue to use the convention introduced in previous chapters and generically refer to the "No" response as "Noise" (i.e., target absent), and the "Yes" response as "Signal" (i.e., target present).

Before analyzing our data, we screened the observers' group responses for inconsistencies. In trials where one group member's "group" response did not match the other two members' group responses and the other two members agreed, we ignored the dissenting group member and assumed that that group member had simply mis-clicked the group response. If all three group member's "group" scores disagreed, the trial was thrown out; however, while a number of trials had inconsistencies, none of the trials needed to be rejected. Thus, in our experiment setup, it was not possible for a single group member to alter the group's response post-facto, whether accidentally or intentionally. This is a key improvement on previous studies [87], in which one member was assigned at random to input the group's response and the experimenters relied on negative feedback from the other group members to discourage each trial's appointed responder from substituting his individual opinion for the agreed-upon group response.

4.3 Human Observer-Based Results

We begin with the observers' and groups' average statistics over all trials, then show the performance per session and in a moving window of 100 trials. We then used linear regression to estimate the weights that each group applied to its members, again considering different windows of time to visualize the data, and finish with some comments on the relative advantages of the different windows of time and some insights gained from that analysis.

4.3.1 Average Individual Statistics Over All Trials

Since our individual-based statistics are calculated using only the observers' ratings, we will treat the group as a fourth "individual" in this section. We begin by characterizing each observer's and group's observation spaces through the response statistics, response criteria, and Receiver Operator Characteristic (ROC) curves. We then show each observer's and group's average performance over all trials.

Response Statistics

The response statistics are relevant because they provide information about each observer's observation space and sensitivity. Intuitively, a higher area of overlap between the Signal and Noise curves corresponds to a lower observer d' because the area of the region of overlap determines the probability of an error (see Section 1.4.3 in Chapter 1); equivalently, the area of overlap defines the probability that

the less-likely response is correct.

Since it is customary to assume that the internal response distributions of the observer are Gaussians, we first verified that a Gaussian distribution could be used to reasonably fit the response statistics of the individuals and groups. For each group and individual, we fitted a Gaussian curve to the response data, normalized by the number of trials and the histogram's bin width. These are shown in Figures 4.2, 4.3, and 4.4 for Groups 1, 2, and 3, respectively. The red histogram is the distribution of ratings that the individual used when Noise was the correct response, and the green histogram is the distribution of ratings that the individual used when Signal was the correct response.

A t-test on the response statistics failed to reject the hypothesis that the responses were not from a Gaussian distribution at a 5% level of significance for all observers and groups except TS's Signal, JW's Signal and Noise, and Group 3's Signal response distributions.

Response Criteria

We calculated the response criteria used by each individual and group, and overlaid these criteria on the Gaussian response curves found above. Our results for Group 1 (resp., 2 and 3) are shown in Figure 4.5 (resp., 4.6 and 4.7). See [59] for a tutorial on calculating criteria for a Yes-No task with confidence ratings. The vertical red lines are rating criteria that result in a Noise response and the
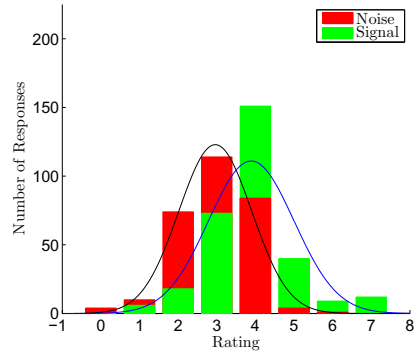
(a) Histogram of AC Responses
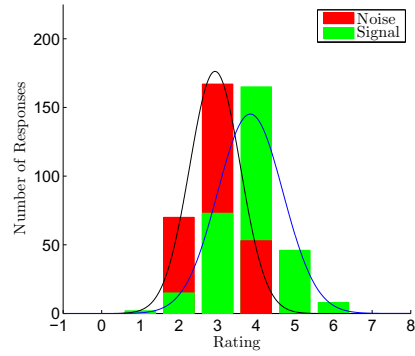
(b) Histogram of KM Responses
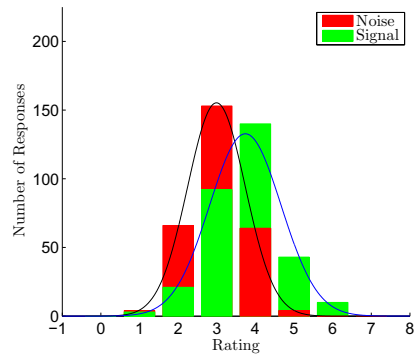
(c) Histogram of KT Responses
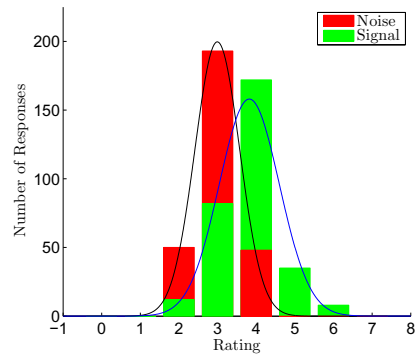
(d) Histogram of Group 1 Responses

Figure 4.2: Group 1's individual and group response histograms, overlaid with Gaussians fitted to the data. The red histogram shows the number of times each rating was used in trials where Noise was correct, and the green histogram shows the number of times each rating was used in trials where Signal was correct. In addition to being fitted to the data, the Gaussian curves are normalized to account for the unequal number of Noise and Signal trials. These diagrams provide a visual check on how reasonable it is to approximate the statistics as Gaussian as well as an approximation of each observer's observation space.
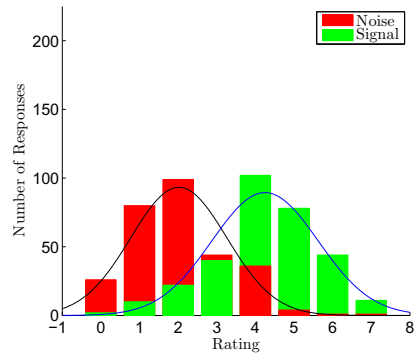
(a) Histogram of JW Responses
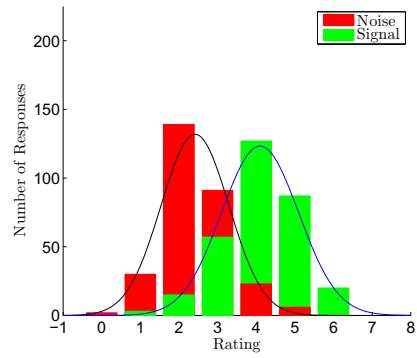
(b) Histogram of KV Responses
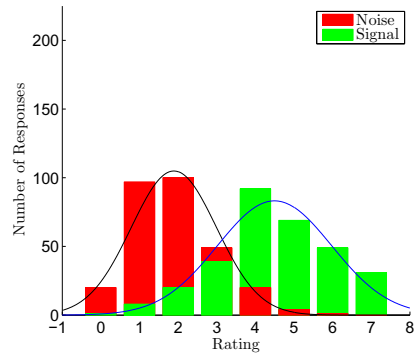
(c) Histogram of TS Responses
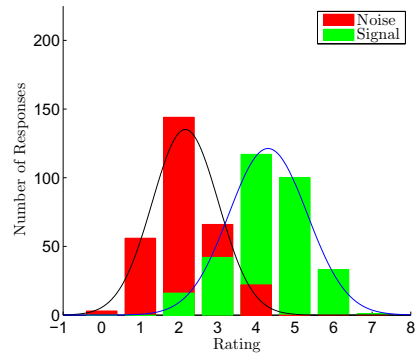
(d) Histogram of Group 2 Responses

Figure 4.3: Group 2's individual and group response histograms, overlaid with fitted Gaussians.
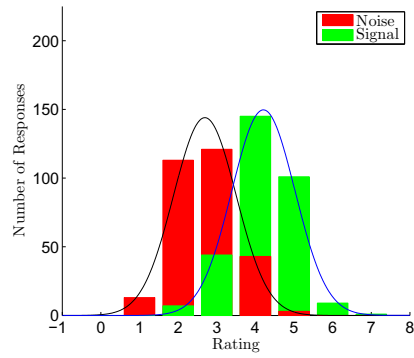
(a) Histogram of CG Responses
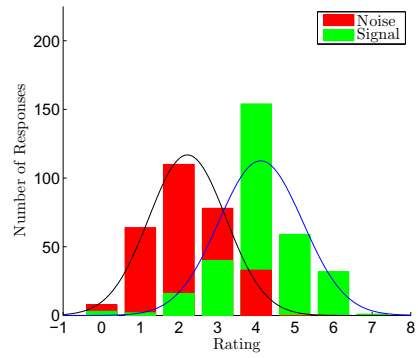
(b) Histogram of MR Responses
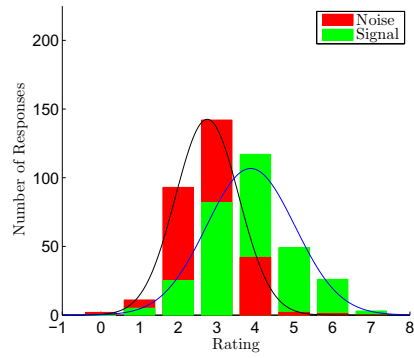
(c) Histogram of SF Responses
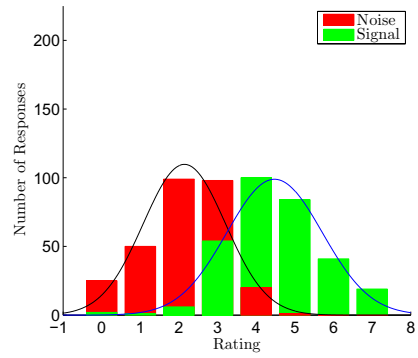
(d) Histogram of Group 3 Responses

Figure 4.4: Group 3's individual and group response histograms, overlaid with fitted Gaussians.

vertical green lines are rating criteria that result in a Signal response. By rating criteria, we mean the vertical lines that separate the regions that correspond to the observer responding with a particular rating. (Though the observers' ratings were integers, averaging over the different responses leads to continuous criteria.) The dark blue line is the response criterion used to decide between Signal and Noise responses, and will be generically referred to as "the criterion". The dashed cyan line shows the location of 3.5, the mean rating on our scale, and serves mainly to highlight any shifts in the system. The numbers on the plot denote what rating corresponds to which region of the observation space. For example, for the observer to respond "0", the observer's observation must have been drawn from the region of the observation space to the left of the first red criterion. Therefore, that region of the observation space is labeled "0". The region directly to the right of this is the region corresponding to the observer responding "1"; hence, the left-most red criterion is the 0-1 criterion. This pattern continues to the maximum rating of 7.

In Group 1, observer AC was unbiased and used an optimal criterion, but had poor performance due to a low d', as evident by the large area of overlap of AC's Signal and Noise distributions. The Group 1 plot does not have a 0-1 criterion, because the group never responded 0 or 1; thus, those two ratings are indistinguishable. It is interesting to note that the observers in Group 1 generally were not able to efficiently assign the numerical confidence ratings in 600 trials, which resulted in rating criteria that are not evenly distributed.

Group 2 used a wider range of ratings than Group 1, and in general were better at distributing the criteria for each rating evenly; however, they tended to be conservative (i.e., biased against selecting "Signal"). This is also clear in Table 4.1: the members of Group 2 tend to have a lower Hit Rate (HR) and lower False Alarm rate (FA) than other observers of similar ability. In general, however, Group 2 had the most evenly spaced rating criteria.

In Group 3, observer CG was the least biased, and used an optimal criterion. Observer SF was the most biased and used a very non-optimal criterion. SF's distributions also overlapped by a significantly larger amount than any other observer in Group 3, which is consistent with the fact that SF had the worst performance in the group. Observer MR's distributions are shifted slightly to the right, but MR's criterion is closer to optimal than the mean rating (3.5), indicating that he likely was quite capable at the task. Group 3's rating criteria are clustered towards the center of the observation space, and are slightly biased towards being conservative.

ROC Curves

ROC curves plot the FA versus the HR, and are used to characterize the abilities of each observer - a larger area under the ROC curve indicates a higher d'. The zROC curve is the z-transform of the ROC curve, which converts HR and FA to a z-score, or units of standard deviation. As defined in [59], this is equivalent to the inverse normal function. The shape of the zROC and ROC curves also indicate

(a) AC Criteria

(b) KM Criteria

(c) KT Criteria
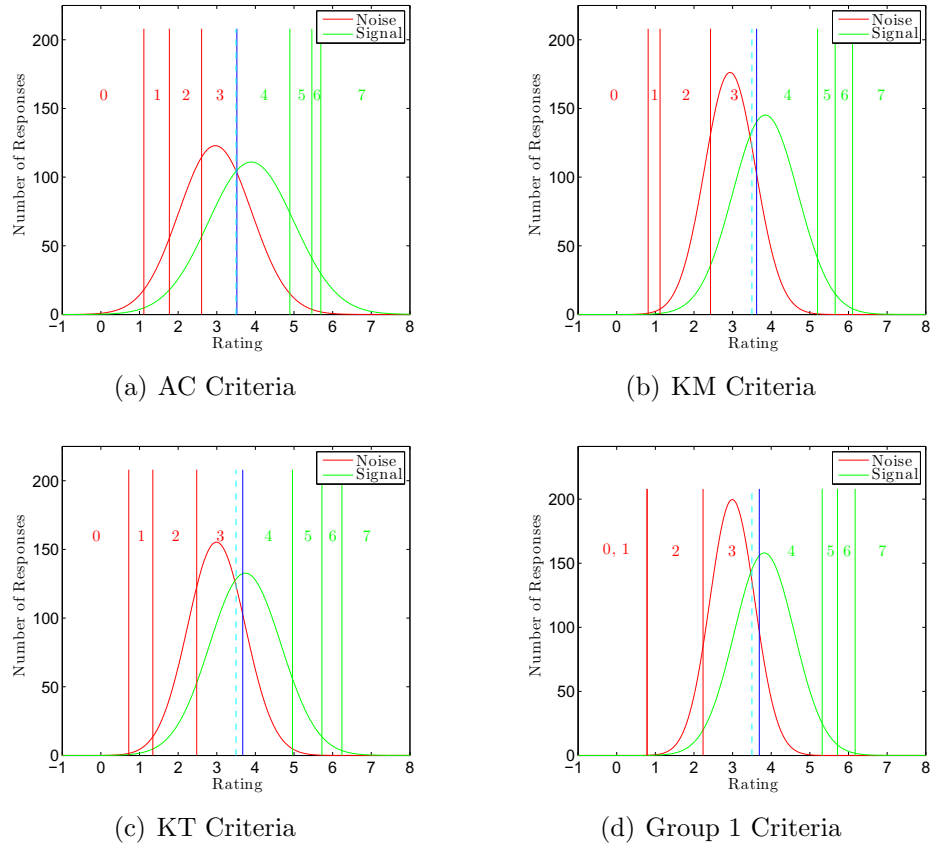
(d) Group 1 Criteria

Figure 4.5: Group 1 individual and group criteria overlaid on the fitted Gaussian response statistics from Figure 4.2. The dark blue line is the criterion used to decide between responding Signal or Noise, and the cyan dashed line marks the mean rating 3.5, which indicates if the distributions are symmetric around the mean rating. The vertical red lines indicate the regions of the response curves that correspond to a particular rating with a Noise response, and the vertical green lines indicate the regions of the response curves that correspond to a particular rating with a Signal response. The Group 1 plot (d) does not have a 0-1 criterion because the Group used neither the 0 nor 1 rating in the experiment; thus, the ratings 0 and 1 are indistinguishable. Group 1 was the least effective at evenly distributing its criteria over the observation space.

(a) JW Criteria

(b) KV Criteria

(c) TS Criteria
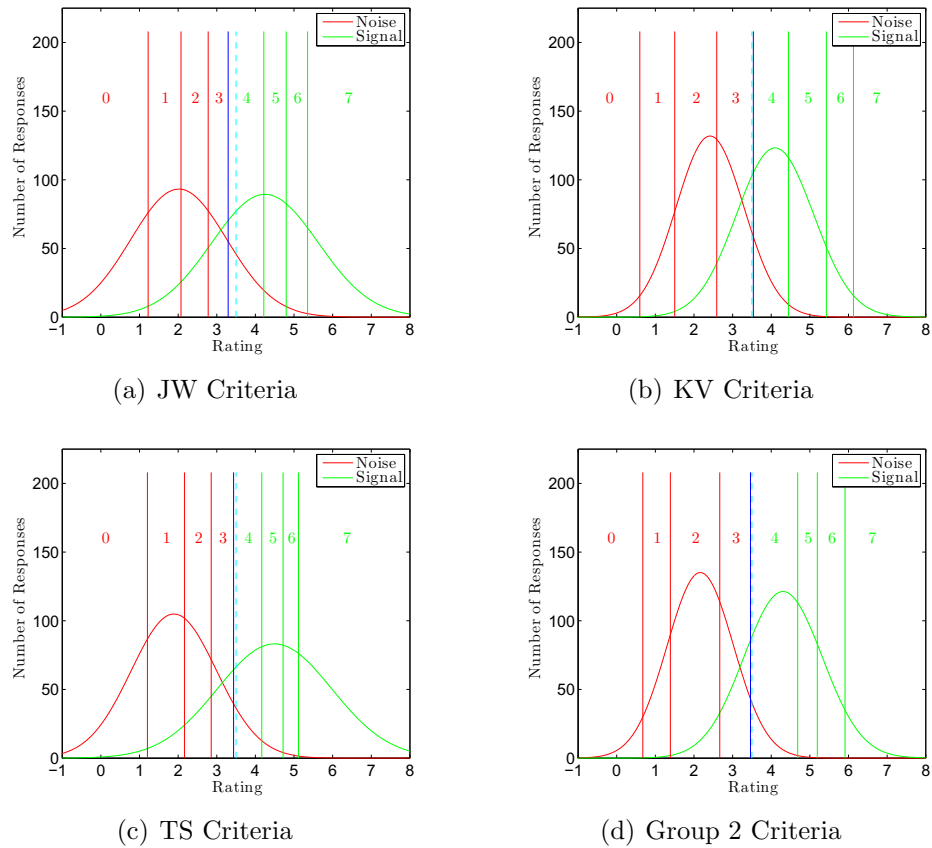
(d) Group 2 Criteria

Figure 4.6: Group 2 individual and group criteria overlaid on the fitted Gaussian response statistics from Figure 4.3. All of the members of Group 2 tended to be conservative. This is reflected in the group's and individuals' performances, since they tended to have lower HRs and FAs than other observers of similar ability, as shown in Table 4.1.

(a) CG Criteria

(b) MR Criteria

(c) SF Criteria
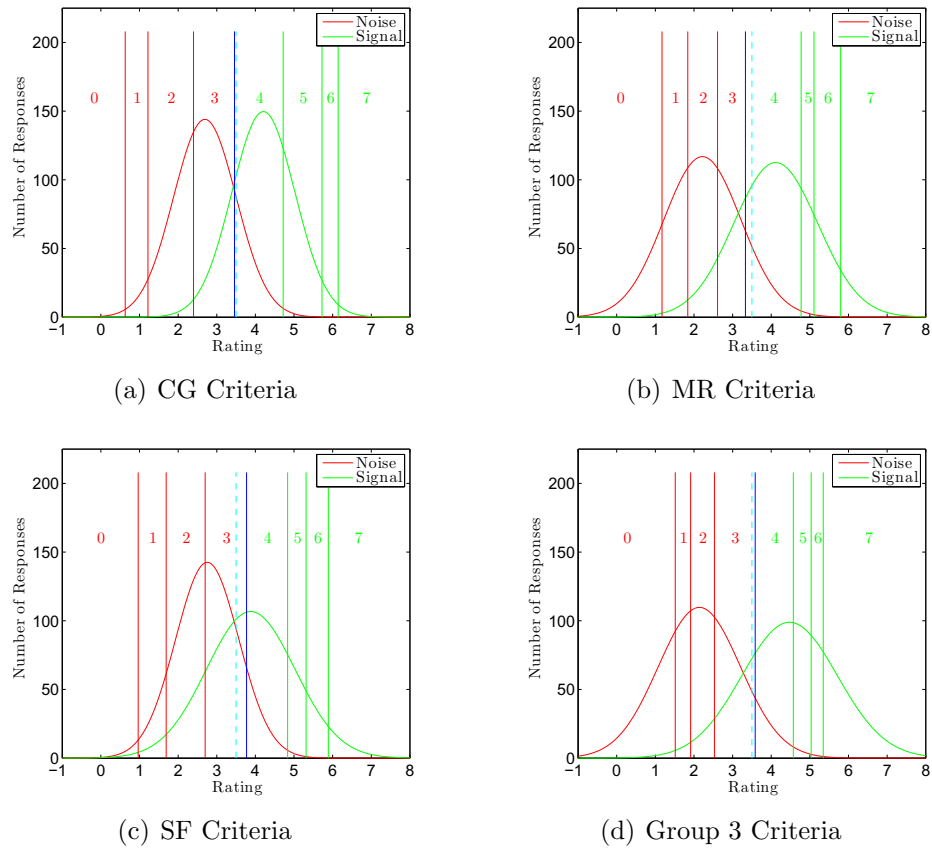
(d) Group 3 Criteria

Figure 4.7: Group 3 individual and group criteria overlaid on the fitted Gaussian response statistics from Figure 4.4. The individuals in Group 3 generally used the ratings efficiently, as evidenced by their reasonably evenly-distributed criteria. It is interesting that the group's responses, in contrast, did not use the ratings very efficiently.

the likelihood that the underlying distribution is Gaussian. A very good tutorial on calculating ROC and zROC curves can be found in [59]; also see [40].

The fitted ROC curves for the individuals and groups are shown in Figure 4.9. The parameters for the empirical ROC curves were found using the zROC fits, given in Figure 4.8. The distributions of points and the non-unit slope of most of the fitted lines in the zROC plots indicate that there are nonlinearities in the data.

Sensitivity, Revisited

As stated in the introduction, one important measure of performance is d', or sensitivity. The simple formula for d' shown in Equation (1.12) is appropriate for theoretical calculations; however, for data from an actual experiment, the assumptions implicit in the formula shown are not always appropriate. Specifically, we found in analyzing our data that much of it departs from the Gaussian assumption; therefore, before we discuss observer performance, we first define the Wilcoxon method, which provides a more fundamental method for calculating d'.

Wilcoxon Method   The Wilcoxon method is a non-parametric method for calculating the area under the ROC curves that is robust to distribution violations. In other words, it does not require that the behavioral data be Gaussian. We describe it in further detail below.

Suppose an observer's ratings are given in the vector X. Let X_S be the observer's ratings in trials where Signal was correct, and let X_N be the observer's ratings in
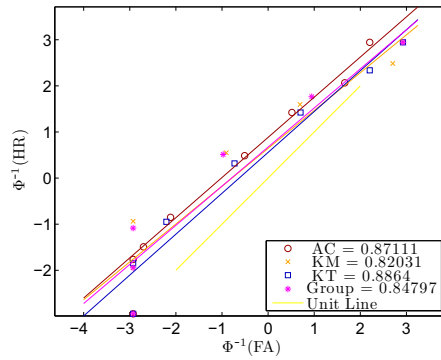
(a) Group 1 zROC Curves
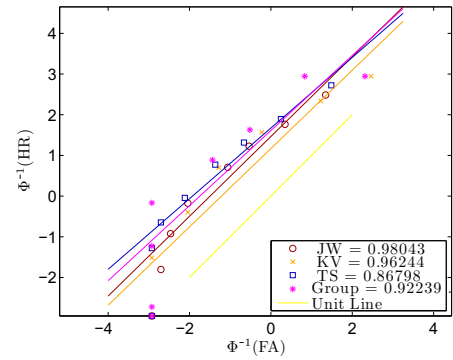
(b) Group 2 zROC Curves
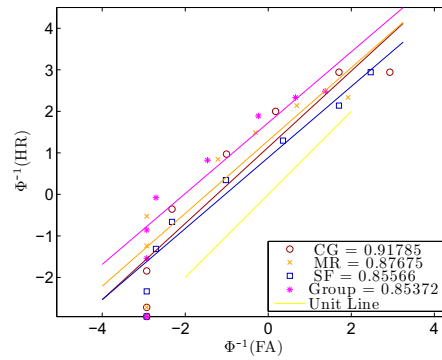
(c) Group 3 zROC Curves

Figure 4.8: Individual and group zROC curves, with linear fits. $\Phi^{-1}(\cdot)$ is the inverse normal function. The data is nonlinear, which is evident in the non-unit slope of the fitted lines and the general curved shape of the data points. The non-unit slope of the fitted lines indicates that the variances of the response distributions are unequal, and the curved shape of the data indicates that it departs from the Gaussian assumption. The parameters found by fitting a line to these data points were used to calculate the fitted lines in Figure 4.9. The legend shows each observer and group's slope.

211
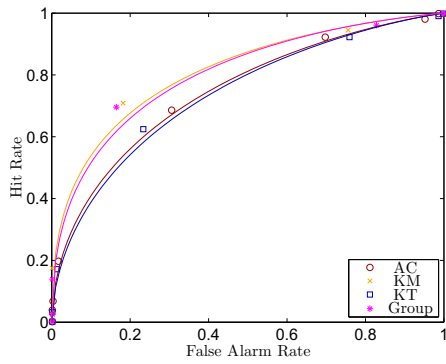
(a) Group 1 ROC Curves
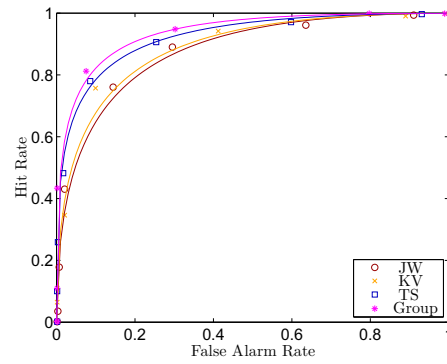
(b) Group 2 ROC Curves
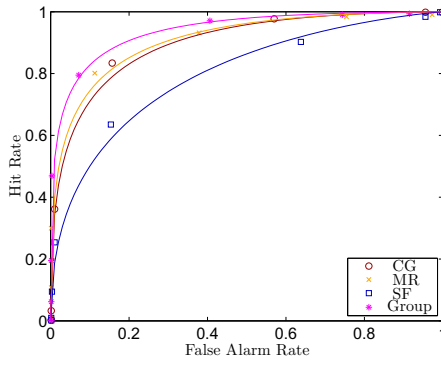
(c) Group 3 ROC Curves

Figure 4.9: Empirical ROC curves, fitted using least-squared error. The area under the ROC curve, A_Z, is directly related to an observer's d', so ROC curves are often used to characterize an observer's sensitivity.

trials where Noise was correct. For the Wilcoxon method, we first sort X_S (resp., X_N) into ascending order. We then initialize a counter, k, to zero. For each rating in X_S that is greater than a rating in X_N, we increment k by one. For each rating in X_S that is equal to a rating in X_N, we increment k by one-half. We then normalize k by the length of X_S multiplied by the length of X_N, to get a fraction less than or equal to one. This fraction is the area under the ROC curve A_Z, which can then be transformed to d' using the formula

$$d' = \sqrt{2}\Phi^{-1}(A_Z). \tag{4.2}$$

Performance

The d', PC, HR, and FA for each observer and group are shown in Table 4.1. In all three groups, the group performed better than its most capable individual member in PC, and two of the three groups had a higher d' than its most capable member.

The efficiency of each observer and group [87] is given by

$$\eta_{d'} = \left(\frac{d'_{\text{obs}}}{d'_{\text{IO}}}\right)^2, \tag{4.3}$$

where d'_{obs} is the d' of the observer, and d'_{IO} is the d' of the Ideal Observer. The Ideal Observer uses template matching to perform the same task as the observers, and represents the highest level of performance possible given the information available

Observer	d'	PC (%)	HR (%)	FA (%)	$\eta_{d'}$ (%)
AC	0.885	69.0	68.7	30.5	14.07
KM	1.15	76.2	70.8	18.2	23.67
KT	0.850	69.3	62.5	23.4	12.98
Group 1	1.12	76.3	69.5	16.5	22.72
JW	1.63	80.7	76.1	14.5	48.02
KV	1.69	82.7	76.0	9.90	51.48
TS	1.90	84.5	78.0	8.56	64.82
Group 2	2.10	86.7	81.2	7.65	79.43
CG	1.73	83.8	83.4	15.7	53.71
MR	1.78	84.3	80.1	11.3	56.85
SF	1.10	73.8	63.5	15.4	21.61
Group 3	2.04	86.0	79.5	7.17	74.65

Table 4.1: Average performance of each individual and group over all trials: sensitivity (d'), Percent Correct (PC), Hit Rate (HR), False Alarm rate (FA), and efficiency with respect to the Ideal Observer ($\eta_{d'}$). The d' values were found non-parametrically, using the Wilcoxon method to calculate A_Z, the area under the ROC curve, which was then transformed to d' using Equation (4.2).

in the visual task [66]. The average d' of the Ideal Observer in the task was 2.36. A graph of the average efficiency of each observer and group is shown in Figure 4.10. The error bars are standard error, calculated using the per-session d' values.

Now that we have characterized the observers' statistics over all trials, we consider their performance averaged over different windows of time, to check for trends in the data.

4.3.2 Performance in Different Windows of Time

Though the convention is to consider only the average performance over all trials, we present two additional views to provide a more complete characterization of the data.
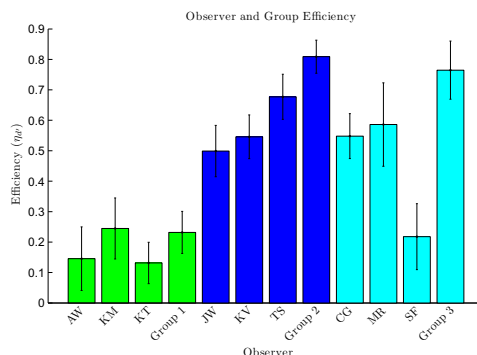
Figure 4.10: Individual and Group efficiencies with respect to the Ideal Observer, $\eta_{d'}$, defined by Equation (4.3).

Performance Per Session

The three groups' performance in each session is shown in Figure 4.11. The error bars were calculated using jackknifing [80]. As we expected from the mean performance, the group usually did better than the best of the observers, and in sessions where the best individual observer out-performed the group, the difference usually was not significant. Across all three groups, there is only one session in which an individual significantly out-performed the group: session 4 in Figure 4.11(b). The per-session view is informative because it shows that the amount by which the group out-performs the best observer is usually small but consistent. It also shows that the group's performance tends to be more robust than any individual observer's: in sessions where the group did not have the highest performance, it was not significantly different than the best performance (with the single exception mentioned above) and generally had the second highest performance (with

215

the single exception of Figure 4.11(e), Session 2, where the group is third best by a non-significant amount). In contrast, the best individual performer in each group has the third best performance at least once, and generally has more volatile performance than the overall group.

There are some interesting trends in the per-session graphs: Group 1's performance increases in the first three sessions, then levels off, Group 2's performance remains relatively flat, and Group 3's performance generally increases across all six trials. This indicates that Groups 1 and 3 exhibited learning during the course of this experiment.

Moving Window of 100 Trials

We used a window of 100 trials (1 session) because we wanted to have enough data within the window to perform the calculations in a meaningful manner, while keeping the window small enough to provide a large number of data points. Calculating the performance in a moving window of 100 trials is significantly more computationally expensive than finding the average overall performance or the performance per session, but it provides a continuous view of the dynamics of the data. The value at each point represents the performance calculated over that point, the preceding 50 trials, and the ensuing 49 trials. In other words, we plotted the value found in a 100-trial window at the midpoint (51st trial) of the window.

The individual and group PC scores for Groups 1, 2, and 3 are shown in Figures
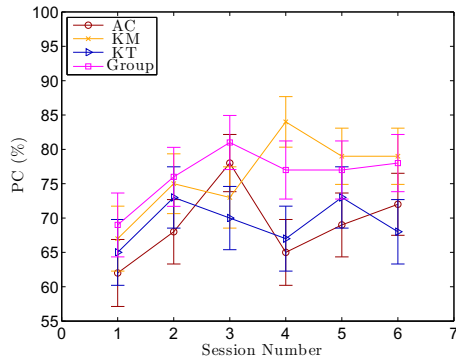
(a) Group 1 PC
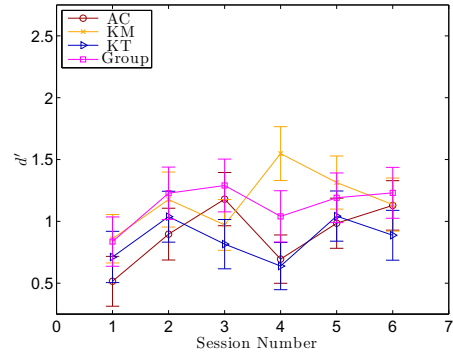
(b) Group 1 d'
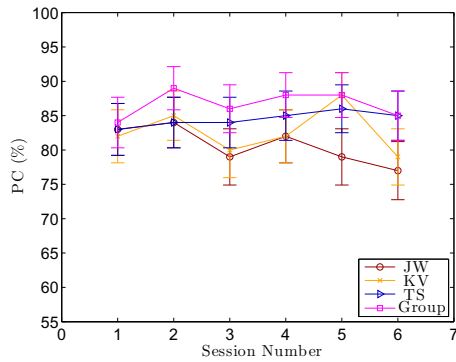
(c) Group 2 PC
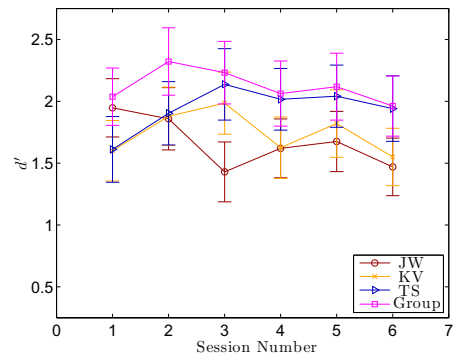
(d) Group 2 d'
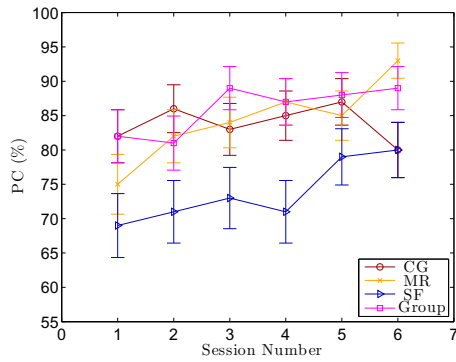
(e) Group 3 PC
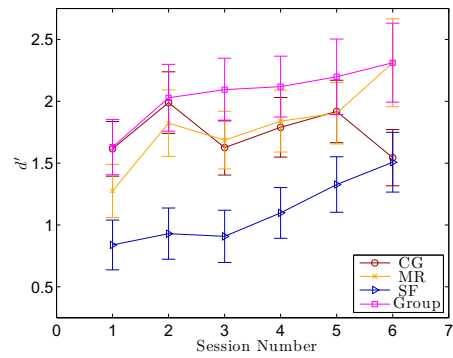
(f) Group 3 d'

Figure 4.11: Group and Individual PC and d' scores, calculated per session, for all three groups. The group had either the best performance or performance not significantly different from the best performance in all sessions except for session 4 in (b).

4.12(a), 4.13(a) and 4.14(a), respectively. In all of the graphs, the horizontal orange line is the performance averaged across all sessions, and the wide horizontal yellow line is the corresponding standard error. The step-like blue line is the average performance per session, and the red line is the average performance over a moving window of 100 trials. In general, Group 2 and its members had the steadiest performance, staying close to the overall average value, while the other groups and groups' members' PC values varied and differed significantly from the mean value in at least one session.

The corresponding d' values for Groups 1, 2, and 3 are shown in Figures 4.12(b), 4.13(b), and 4.14(b), respectively. In this case, the average d' over all trials encompasses the per-session values, but only because it has large error bars; thus, even though the results of the more detailed views are encompassed within the overall average, the average value is not very specific.

Generally, the per-session analysis provides a good balance of detail and simplicity: the overall average does not capture trends, and the moving window of 100 trials requires more computation. However, the overall average is simpler to compute, and in cases where there is little fluctuation from trial to trial, it provides a sufficient approximation: for example, except for KV's Session 5 in Figure 4.13(a), none of the PC values from Group 2 varied significantly from the mean. The average in a moving window of 100 trials provided the most detail on how performance changed in a continuous manner from trial to trial. Though it is the most compu-

218

tationally intensive analysis shown, it did not require significant computation time for our data.

Based on our results, we found that it is most worthwhile to view performance data in a per-session basis first. For some of the observers, the overall average provided approximately the same information as the moving window, while for others, the moving window analysis showed that the local performance changed noticeably even from the average per session. Since it is not always clear beforehand which view of the data will provide the most information, we conclude that it is worthwhile to use multiple windows of time to visualize the data.

4.3.3 Human Group Weights

Now that we have characterized the abilities of each observer and group in the signal detection task, we can establish how efficient each group was at assigning weights to its members. This indicates how well the group members were able to accurately assess each other's sensitivity. We begin with some detail on how we regressed the group weights.

Regressing the Actual Group Weights

We approximated the observers' weights with linear least-squares regression. In our analysis, the only constraint we imposed was that the weights must add to unity. In other words, we allowed negative weights and weights greater than 1.0,
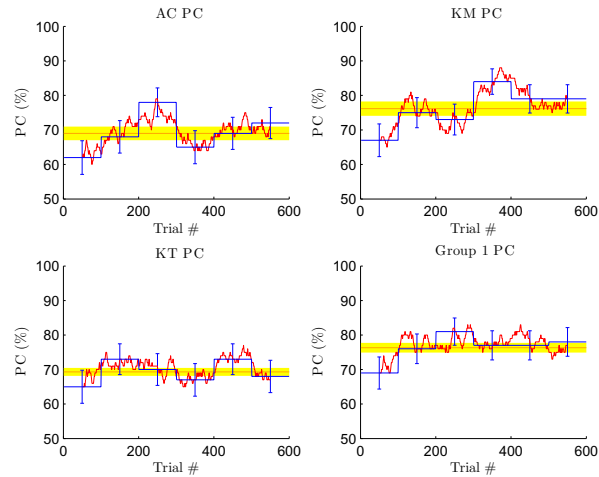
(a) Group 1 PC in different windows of time
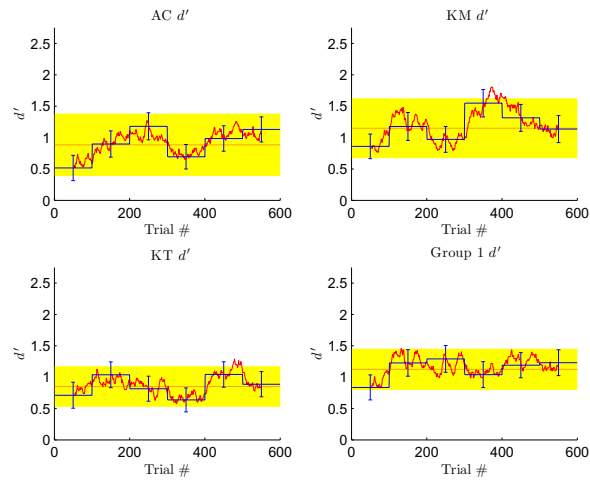
(b) Group 1 d' in different windows of time

Figure 4.12: Individual and group (a) PC scores and (b) d' values, calculated over all trials (orange), per session (blue), and in a moving window of 100 trials (red), for Group 1. The yellow region represents the error bars on the average over all trials (orange). This comparison visually shows how the different windows of time characterize Group 1's performance.
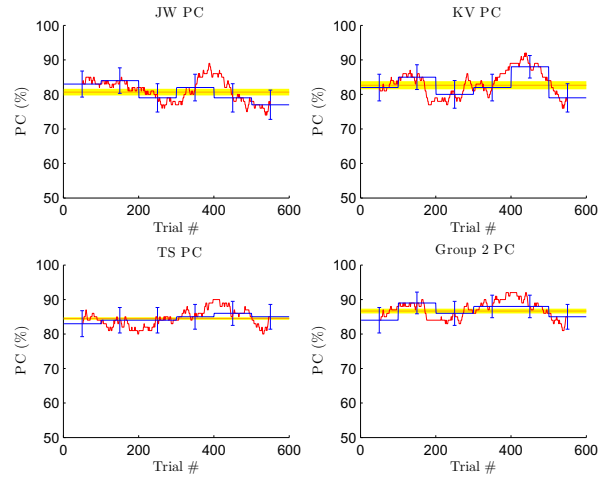
(a) Group 2 PC in different windows of time
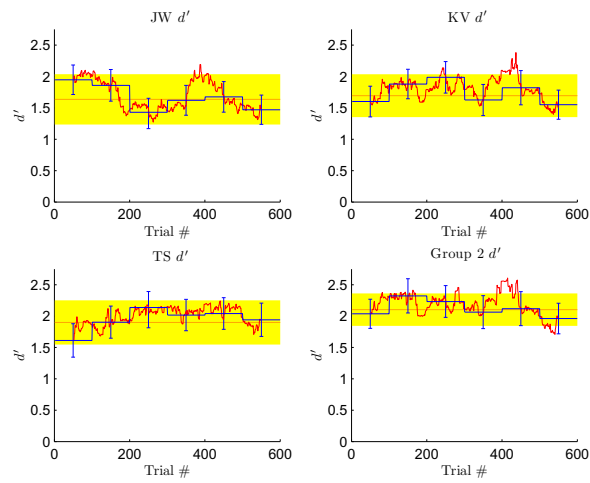
(b) Group 2 d' in different windows of time

Figure 4.13: Individual and group (a) PC scores and (b) d' values, calculated over all trials (orange, with error bars in yellow), per session (blue), and in a moving window of 100 trials (red), for Group 2.
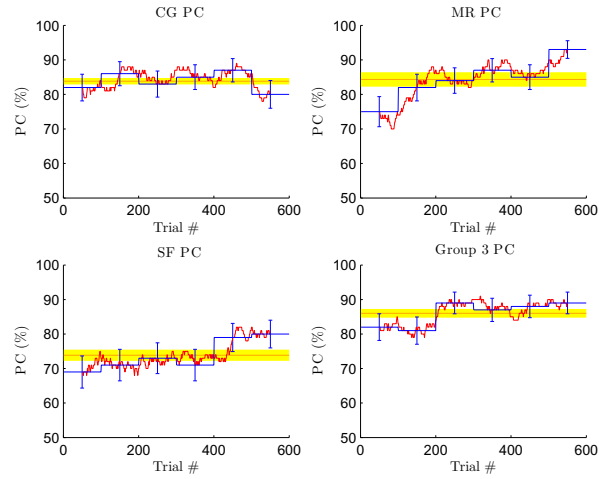
(a) Group 3 PC in different windows of time
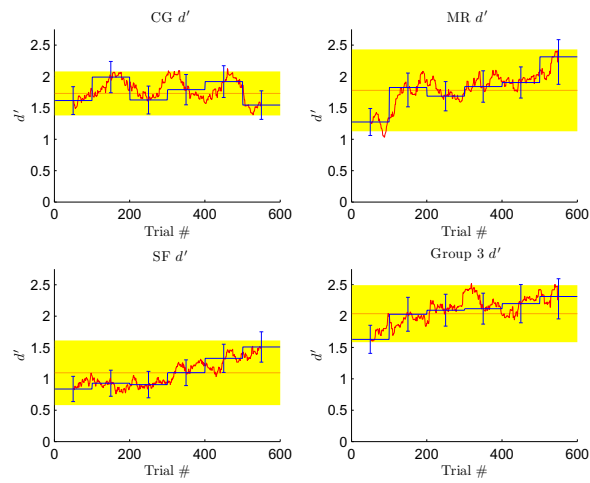
(b) Group 3 d' in different windows of time

Figure 4.14: Individual and group (a) PC scores and (b) d' values, calculated over all trials (orange, with error bars in yellow), per session (blue), and in a moving window of 100 trials (red), for Group 3.

should the analysis produce them. Though initially non-intuitive, we allowed this because it provides further information on how the members arrived at their group rating: an observer who is highly correlated with another observer would be down-weighted, while an observer who performed at chance (e.g., flipped a coin each round to determine his response) would receive a weight of approximately zero. An observer would not receive a negative weight unless he was frequently overruled. A negative weight with the constraint that the weights sum to unity could possibly induce a weight greater than unity for another observer in the group.

We used the following method to calculate the group weights over $\mathcal{T}$ trials: Let X be the $\mathcal{T} \times 3$ matrix of the individual observers' ratings. Let w be the 3×1 vector of weights for each observer, and let y be the $\mathcal{T} \times 1$ vector of group ratings. We then assumed that the observers had applied a linear weighting scheme,

$$y = Xw. \tag{4.4}$$

Then, given X and w, we can solve for the expected ratings, y. Conversely, given X and y, we can solve for the weights [41]:

$$w = \frac{(X^T X)^{-1} X^T y}{\|(X^T X)^{-1} X^T y\|_1}, \tag{4.5}$$

where X^T denotes the matrix transpose of X, $(X^T X)^{-1}$ is the inverse of $(X^T X)$, and $\| \cdot \|_1$ denotes the 1-norm. We verified that the inverse $(X^T X)^{-1}$ existed and

Group 1	AC	KM	KT
	0.2965	0.3164	0.3871
Group 2	JW	KV	TS
	0.3475	0.3579	0.2946
Group 3	CG	MR	SF
	0.2750	0.4119	0.3131

Table 4.2: Average weights used by the observers in the group signal detection task. The above weights were calculated using all of the trials in the least squares linear regression in Equation (4.5).

returned the expected result.

4.3.4 Group Weights in Different Windows of Time

Like performance, the group's weights are conventionally only considered over all trials. We also calculated the weights in different windows of time to better characterize the data and see how well common assumptions performed. Below, we provide greater detail on the windows considered for calculating the group weights, then finish with a discussion of the benefits of this analysis.

Average Group Weights Over All Trials

Using Equation (4.5), we solved for the observers' weights, averaged over all trials (i.e., $\mathcal{T} = 600$). Our results are shown in Table 4.2.

A comparison of the weights with the observers' performance (Table 4.1) shows that the weights are not optimal: in theory, the optimal weighting strategy is to assign each member a weight proportional to the member's d' [63]; however, in

Group 1, KT received the highest weight, despite having the lowest performance in the individual task. In Group 2, TS had the best performance but received the lowest weight. In Group 3, SF was given a higher weight than CG, despite that CG had much better performance than SF. Given these inefficiencies, it is quite impressive that in all three groups, the group generally out-performed its most capable individual observer.

Average Group Weights Per Session

As we stated earlier, assuming that the average weights calculated over all trials accurately represents the actual weights that the observers used is equivalent to assuming that there are no time dependencies in the data. Here, we investigate the possibility that the weights may change significantly in time on a per-session basis. For this analysis, we used Equation (4.5) on each session (i.e., $\mathcal{T} = 100$). The results are shown in Figure 4.15. This view shows that some of the weights varied significantly from session to session. For example, in Group 1, KT's per-session weight is very large in Sessions 4 and 5, which results in KT getting the largest overall weight; however, those two are the only sessions in which KT receives the highest weight - KT was even given the lowest weight in one session. In Group 2, KV was given the highest overall weight, but in half of the sessions, did not actually have the highest per-session weight.

As a note, during the experiment with Group 3, the proctor felt that observer

225
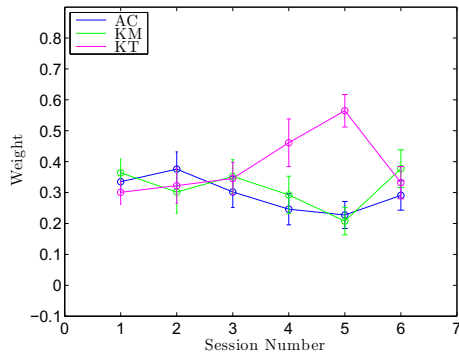
(a) Group 1 Weights per Session
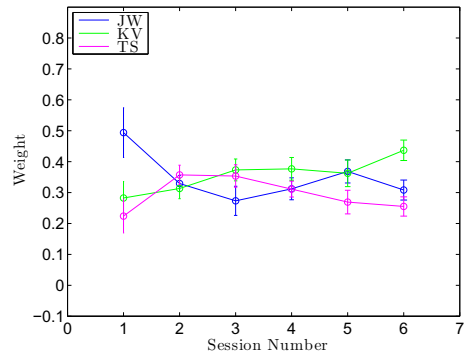
(b) Group 2 Weights per Session
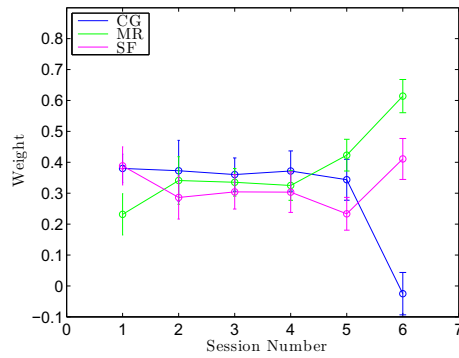
(c) Group 3 Weights per Session

Figure 4.15: Group weights per session, found using ordinary least squares linear regression. Some of the observers' weights varied significantly from session to session, indicating that an overall average may not be the most informative view.

CG was being to forceful in expressing his opinion during the group discussions, and therefore asked CG to be "less emphatic", partway through Session 5. It appears that the group took this suggestion to mean that CG should be ignored, since CG's weight went from being the highest in Sessions 1 through 4, down to a negative value in Session 6. This one session's weight was so low that CG's overall weight was affected: from the average over all trials, CG appears to have been given the smallest weight when in fact, CG was given the highest weight for most of the experiment. This underscores the importance of analyzing the dynamics of the data.

Group Weights in a Moving Window of 100 Trials

For each 100-trial window, we calculated the group weights, then plotted the results at the 51st trial in the window. Because this analysis assumes that there may be a time dependence in the data, we could not calculate weights for the first and last 50 trials.

Group Weights Averaged Over the Entire Past

These weights were calculated using all past trials. For example, the weights for trial 51 were calculated using trials 1 through 51. For simplicity, the rule sets the weights equal to a uniform 33.33% for the first 25 trials before using the past trials to calculate the weights to assure that there is enough data available to perform the calculation. This view shows how the weights would evolve if the observers

remembered the outcomes of all past trials and used that information to optimally select the group's weights for the present trial.

Comparison

The weights in the different windows of time for each observer in Group 1 (resp., Groups 2 and 3) are shown in Figure 4.16 (resp., Figures 4.17 and 4.18). In general, the weights from the moving window of 100 trials agree well with the per-session weights, but neither agree well with the weights calculated over the entire past.

In an ideal situation, we would expect that the observers remember the outcomes of all of the past trials and use that information to select a weight for each observer for the present trial. If this were the case, then the per-session weights and the weights from the moving window should converge to the mean weights calculated over all trials in a manner similar to the way that the weights calculated over the entire past converge to the overall mean. However, it is clear from the graphs that this is not true. This indicates that it is likely that a more local strategy was used. In other words, the observers did not use many (if any) past trials to calculate the weights to use for the present trial. This is reasonable because local strategies require a lower cognitive load and less (or no) memory to calculate the weights. This motivates the types of group rules we consider below.

(a) AC Weights

(b) KM Weights

(c) KT Weights
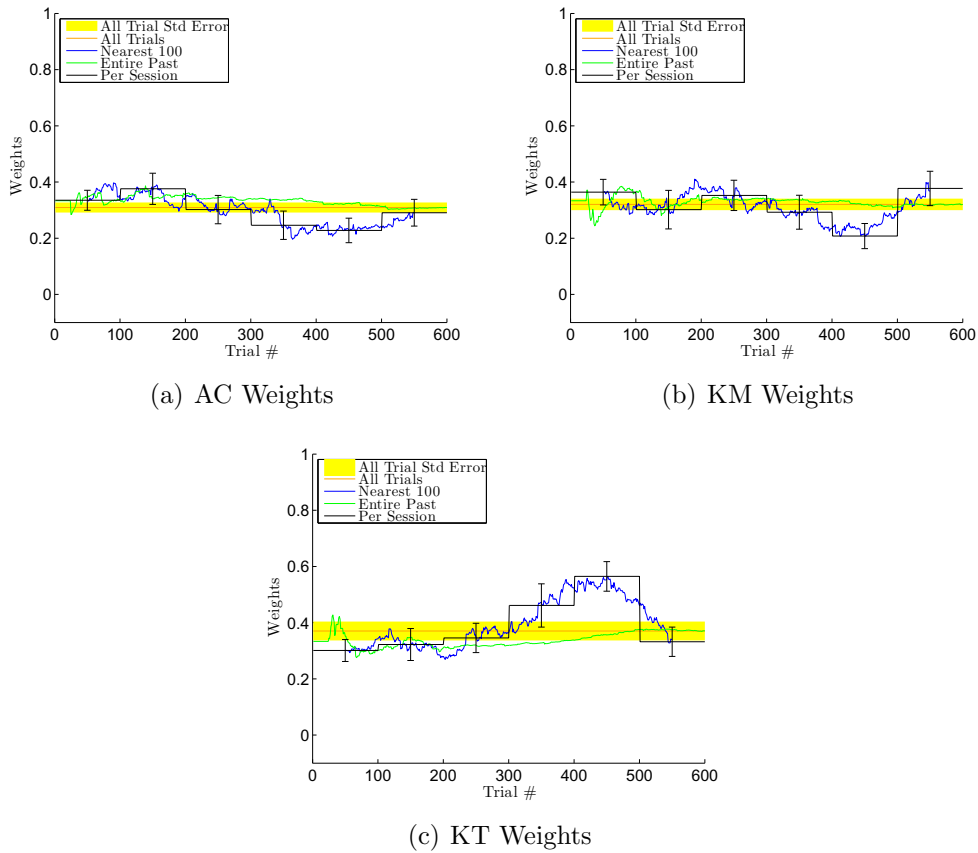
Figure 4.16: Group weights calculated over different windows of time, found using ordinary least squares linear regression, for Group 1. For the entire past calculation, the observers were given a uniform weight for the first 25 sessions to allow for a reasonable number of trials before beginning the regression. All of the observers' weights varied significantly from the mean in at least one session, indicating that the overall mean does not fully characterize the data.
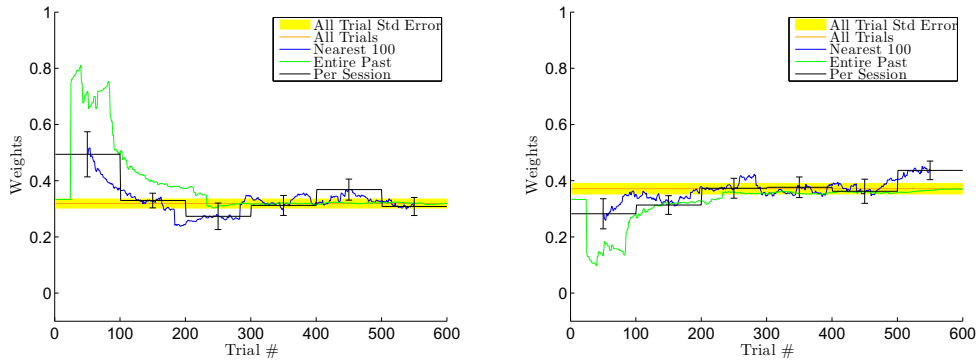
(a) JW Weights

(b) KV Weights
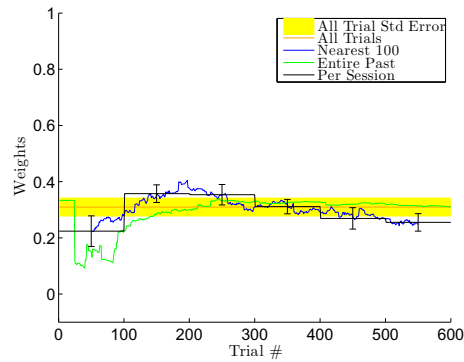
(c) TS Weights

Figure 4.17: Group weights calculated over different windows of time, found using ordinary least squares linear regression, for Group 2. The average over all sessions provides a good approximation of the actual weights over time: the per-session weights have relatively little fluctuation and small error bars, and they quickly settle to and remain near the overall mean.

(a) CG Weights

(b) MR Weights

(c) SF Weights
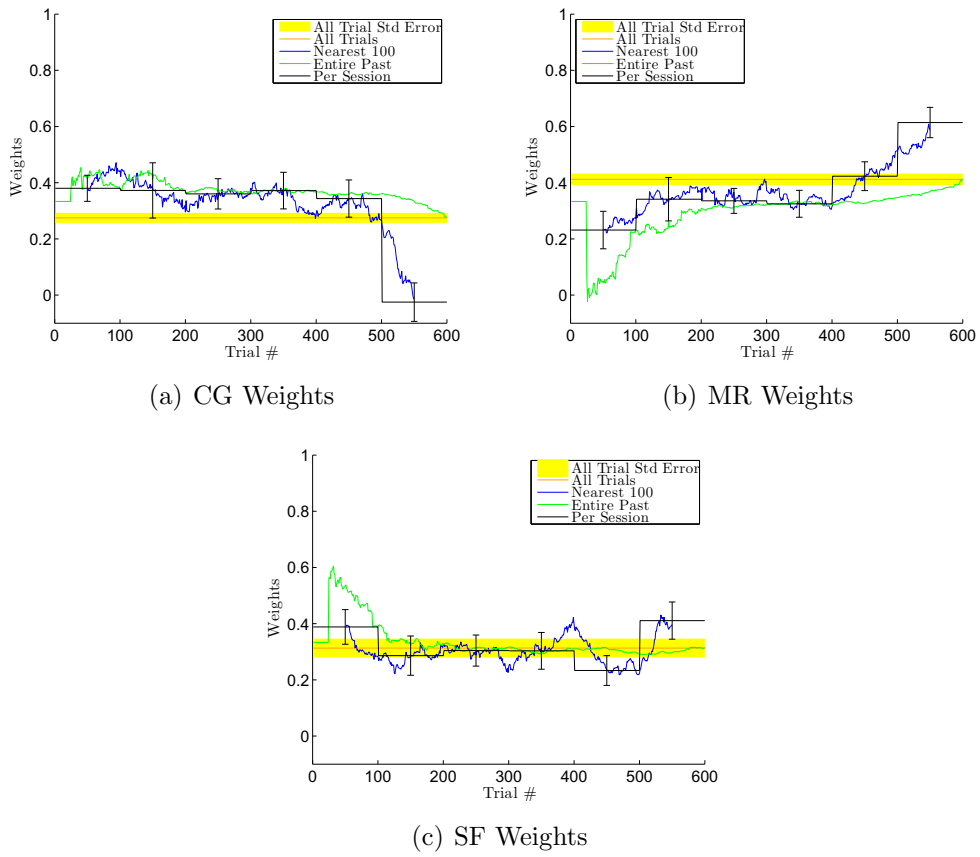
Figure 4.18: Group weights calculated over different windows of time, found using ordinary least squares linear regression, for Group 3. The mean weight over all trials does not do a very good job of describing the weights in the experiment as a function of time, since it does not capture CG's sharp downwards trend near Trial 500 or MR's general upward trend.

4.4 Group Rule Performance

Here we blend psychophysics with the style of models we used in Chapter 3. Suppose that, instead of conferring, the observers submitted their individual ratings to a central fusion center that applies a group rule to the ratings to calculate a group rating. We first explain our eleven different group rules, then compare the group rules' performance, calculated observer weights, choice prediction (ability to predict which binary response the human group selected), and group rating correlation (the correlation between the fusion center's rating and the human group's rating) to determine which rule(s) provide the best performance and which best predict the actual group's ratings.

While we primarily evaluate each group rule based on its performance and ability to predict the observers' actual choices and ratings, we will also comment on realism (i.e., whether or not this group rule can be implemented in an experiment), with the intent of finding rules that could be added to a decision-making system with human observers.

4.4.1 The Group Rules

In this section, we introduce eleven group rules that a fusion center could use to reach a group decision in our simple 2AFC task. We use the term "fusion center" relatively loosely here, and apply it both to something that calculates the ideal weights and to something that applies a simple rule to the individuals' ratings.

The general setup is central, like the group shown in Figure 3.1. The first four rules are variants of the Ideal Group [87], which essentially is the group-based equivalent of the Ideal Observer: given a group of individuals, the Ideal Group (by definition) calculates and applies the optimal weights. We next consider two local rules, which train on the 100 trials adjacent to the test trial, using the nearest or past trials, respectively. We then consider five rules that do not require memory of the past. The first three are simple equality-based rules: the Mean rule, the Majority rule, and the Mean-Majority rule, which is a hybrid of the first two. We call them equality-based rules because they do not favor any observer in particular. We finish with two confidence-based group rules, the Exponential-Majority rule, which heavily favors more confident ratings, and the Max rule, which simply takes the most confident individual rating and uses it as the group rating. Since we are comparing so many rules, some of which have long names, we will use a shorthand name for each rule in our plots. The shorthand name for each rule is specified in parentheses in the rule's heading below.

Calculating the Ideal Weights For our group rules, we separated the individual observers' responses into training and testing trials. The ideal weights are based on a Frequentist approach to statistics; thus, it allows the data to be re-ordered, because it assumes no dependence in time. For each test trial, the Ideal Group calculates the ideal weights using the training data (typically all of the trials except the test trial), then applies those weights to the test trial's data to get the

group's rating for that trial. The Ideal Group then repeats this procedure for every trial in the data set; thus, it can get a group rating for every trial in the data set without ever using the test trial to generate the group weights, and all of the weights are calculated with the same number of trials.

According to Sorkin et al. [86], the ideal weights for a group of N observers is given by

$$w_{\text{ideal}} = \Sigma^{-1}\mu, \tag{4.6}$$

where Σ is the 3×3 covariance matrix calculated from the training data X_{tr}, and μ the 3×1 vector of the distances between each observer's Signal and Noise distributions, defined as follows: we divided our training data into two vectors, according to the correct response for each trial. Ratings from trials where Signal (resp., Noise) was the correct response were added to $X_{\text{tr}|\text{S}}$ (resp., $X_{\text{tr}|\text{N}}$). Then, we defined

$$\mu_{\text{S}} = \text{E}[X_{\text{tr}|\text{S}}],$$
$$\mu_{\text{N}} = \text{E}[X_{\text{tr}|\text{N}}],$$

and

$$\mu = \mu_{\text{S}} - \mu_{\text{N}}. \tag{4.7}$$

For our ideal group rules, we used Equation (4.6) on the training data to find the optimal weights, then applied those weights to the test trial X_{test} (a 1×3 vector)

to arrive at the fusion center's response for the test trial

$$y_{\text{IG}} = X_{\text{test}} w_{\text{ideal}}. \tag{4.8}$$

The fusion center's responses y_{IG} were then evaluated for performance.

Calculating the Optimal Criterion Z_C The ideal and local group rules used an optimal criterion. The optimal criterion Z_C is mentioned in [87], but no real information is provided on how to calculate it. We took the optimal Z_C to be "the criterion that maximizes PC in the training data". We found this by calculating PC for all possible criteria between 1.5 and 5.5, in steps of 0.01, and then selected the criteria corresponding to the highest PC as possible values of Z_C. This strategy usually produced multiple possible values for Z_C. After experimenting with taking the min, max, and median values, we settled on always selecting the median candidate value as the optimal criterion Z_C.

Ideal Group, Train on All Trials (IGAll)

This rule uses all of the data to calculate the weights that are used on each trial. This is not a reasonable approach for data analysis, since this strategy uses the testing data as training data, which results in accuracy that is biased upwards due to overfitting. We call it an Ideal Group because the results of analyzing the data in this manner should produce a theoretical upper bound on the performance

of the human group and "realistic" Ideal Group, given that the assumptions commonly made about observers in the literature are correct (i.e., that each observer's sensitivity is fixed over all trials).

Ideal Group, Train on All Trials Per Session (IGAllps)

This Ideal Group is similar to the IGAll rule, but it operates on only one session at a time. In other words, the IGAllps rule selects a session, then selects one trial from that session as the testing trial. It trains on the entire session and then tests on the test trial. It then repeats this procedure for all possible test trials in the session before moving on to the next session. Like the IGAll rule, the IGAllps rule cheats because it trains on the test trial. We investigated this rule partly to serve as a limit of performance, like the IGAll rule, but also to compare with the IGAll rule, in order to explore the effect of time dependence.

Ideal Group, One Testing Trial (IGm1)

A more reasonable Ideal Group has access to all of the trials except the test trial. We emphasize that this approach, though considered acceptable, is also unrealistic for a fusion center that can calculate the weights on the fly in an actual experiment, since this Ideal Group has access to more information than the observers in the experiment: it can access future as well as past trials. Even if the data's time dependence is discounted, as it would be in a Frequentist analysis, this Ideal Group still uses more training data than is available to the observers in every trial but

one.

Ideal Group Per Session (IGps)

This Ideal Group rule considers the data one session at a time, and does not train on the testing trial. In other words, the IGps rule selects a session, then designates one trial from that session as the test trial and trains on the other 99 trials. It then applies the weights found from the training data to the test trial and then repeats this procedure for every possible test trial in the session before moving on to the next session. The performance of this Ideal Group is based on the responses it generates on the testing trials. In theory, this Ideal Group should perform worse than the above Ideal Group rules, since it formulates its weights from a lower number of training trials – if the observers' abilities were truly constant, then the lower the number of training trials, the less accurate our approximation of the weights proportional to those abilities, and the lower the Ideal Group's performance. Like the IGm1 rule, the IGps rule cannot realistically be implemented.

Train on Nearest 100 Trials (Near100)

In this rule, the fusion center trains on only 100 trials, but it uses the 50 trials before and 50 trials after the test trial. If the observers' abilities vary smoothly and continuously in time, we expect this rule to perform best, since it would capture a change in ability in a way similar to how a central differencing method approximates a smooth solution well. The fusion center uses the same procedure as the

237

Ideal Groups to calculate the group weights and optimal criterion Z_c. Since this rule assumes that there is a time dependence in the data, it processes the data sequentially. Thus, for our task, this fusion center rule tests on trials 51 through 550 only, since it cannot calculate weights based on 100 trials for the first and last 50 trials. The performance and average weights for this rule are calculated on these 500 trials only.

Though this rule is more realistic than the Ideal Group rules in the sense that the fusion center has access to a limited amount of information to calculate the group weights and it does not train on the test trial, it still cannot be realistically implemented in an experiment because it requires knowledge of the future.

Train on Past 100 Trials (Past100)

In this group rule, we assume that the fusion center has access to the most recent past 100 trials only, which it uses to calculate the ideal weights and optimal criterion for the (present) test trial. Since the fusion center requires 100 trials to have already passed before it can calculate weights, it begins testing on the 101st trial. Like the Near100 rule, the performance of the Past100 rule is calculated over only 500 trials.

If the observers' sensitivities vary in a smooth and continuous manner in time, we expect this rule to produce weights that reflect that, but with a lag. We expect that this rule will not perform as well as any of the preceding rules; however, this

group rule has the advantage of being implementable in a real-time experiment because it does not rely on future trials, unlike all of the preceding rules.

Mean Rule (Mean)

As its name suggests, in the Mean rule, the fusion center averages the observers' ratings to arrive at a group rating. We expect this rule to perform worse than the preceding group rules because it gives each observer an equal weight, which is not optimal; however, this rule has the advantage of being implementable and extremely simple because it does not require any memory and needs very little computation. In this rule, the fusion center uses the mean rating (3.5) as the criterion.

Majority Rule (Majority)

In the Majority rule, the fusion center simply looks at the group members' binary decisions, then selects the response that the majority selected. The main part of this rule normally would not return weights using Equation (4.5), so we added a modification that allows us to regress weights: the fusion center averages the responses of the observers in the majority and uses that value as the group rating. This is equivalent to the members in the majority ignoring any observer in the minority. This rule also uses a criterion of 3.5.

We expect this rule to have performance somewhat similar to the Mean rule. The Majority rule is affected less by observers who are overconfident than the Mean rule, but it is affected more by observers who are unsure of their answer. Like the

Mean rule, the Majority rule is easily implemented, because it does not require memory of the past or significant computation power.

Mean-Majority Rule (MeanMaj)

As the name suggests, the Mean-Majority rule is a composite of the Mean and Majority rules. In this rule, the group uses a majority rule to decide on the binary group answer, but calculates the group's rating by finding the mean response. If the group's rating selects a different hypothesis than the binary group answer, then the fusion center uses the smallest rating possible that is consistent with the majority rule's decision (3.33 for Noise, 3.66 for Signal). These minimally-adjusted weights were selected to correspond to two observers responding 3 and one responding 4 for the Noise rating, and one observer responding 3 and two responding 4 for the Signal rating. We regressed the weights used under this rule with Equation (4.5). This rule uses a criterion of 3.5.

We predict that this rule will best match the observers' actual group ratings: it uses a Majority to reach a decision, which previous studies have suggested is frequently used as a default strategy in Yes-No tasks (see [43] for a review of relevant literature), and uses a Mean rule to decide on a rating, which would likely appease the dissenting member and maintain a good working relationship among the group's members.

A Note on the Equality-Based Rules: Mean vs. Majority vs. Mean-Majority We illustrate the differences among our three equality-based rules with an example: Suppose one observer responded 0 (Extremely certain NO) and the other two observers responded 4 (Somewhat certain YES). Under the Mean rule, the group rating would be $1\frac{1}{3}$, and the group would respond Noise. Under the Majority rule, the group rating would be 4, and the group would respond Signal. Under the Mean-Majority rule, the group return $3\frac{2}{3}$ and respond Signal. Thus, even though these three rules are similar, they do not always return the same ratings and responses, and are thus distinct. Because we used the Wilcoxon method to calculate d', these differences in ratings will result in different d' values even when the binary response is the same.

Exponential Mean Rule (ExpMean)

It was suggested in [87] that observers are able to accurately assess the probability that their decision was correct, and that an observer's level of confidence should be directly related to that probability as long as the observer is "honest". In our experiment, the observers were given an incentive to maximize the group's accuracy and had no incentive to lie; therefore, we would expect that if an observer's confidence accurately conveyed the probability that she was correct, then more confident responses should tend to be more accurate. We test this idea with the ExpMean rule.

In this rule, the fusion center gives a higher weight to more confident responses. A plot of the weights given to each rating are shown in Figure 4.19. These weights are symmetric around the 3-4 ratings, which represent a barely-certain response. The weights were calculated by first assigning a certainty value to each rating: for ratings [0 1 2 3 4 5 6 7], the corresponding certainty ratings were $c = [3\ 2\ 1\ 0\ 0\ 1\ 2\ 3]$. The certainty values c were then plugged into the formula

$$w = \exp(c/3) \tag{4.9}$$

to define the weights (w) that were applied to the individual ratings. The factor of $\frac{1}{3}$ was selected so that the more-certain responses did not completely dominate the less-certain responses: a response of 0 or 7 received a weight of 2.7183, while a 3 or 4 response received only a weight of 1.0. The weights in each trial were then normalized to sum to unity before applying them to the observers' responses, to generate the ExpMean group rule's ratings. This rule used a criterion of 3.5.

Like the equality based rules, this rule is very simple and easy to implement in an actual experiment. If more confident observers have a greater probability of being correct, then this group rule should perform better than the equality-based rules; however, if the more confident observers are not any more likely to be correct, then this rule should perform poorly.
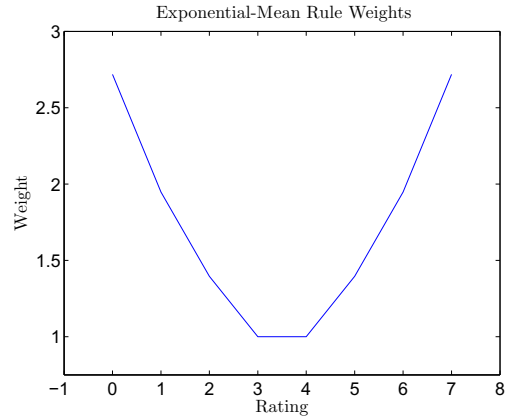
Figure 4.19: Relative exponential weights applied to ratings in the Exponential Mean rule. In each trial, each observer's rating was assigned a weight according to this function. The weights were then normalized to sum to unity and applied to the individual ratings to get the group's response. This rule emphasizes the opinions of the more confident observers.

Max Rule (Max)

The Max rule is a further exaggeration of the ExpMean rule: in the Max rule, the fusion center selects the most confident individual rating as the group rating. If two observers were equally confident in their answer but selected different hypotheses, the fusion center then selects the decision of the third observer. The Max rule is an extreme version of the ExpMean rule, so we expect to see it perform much better than the ExpMean rule if the observers are more confident when they have a greater probability of being correct, and significantly worse than any other group rule if an observer's confidence level is not related to the probability he is correct. Weights used under this rule were regressed using Equation (4.5), and the rule uses the mean rating (3.5) as the criterion.

4.4.2 Group Rule Comparison

We explored a wide variety of group decision rules with three aims in mind: first, we want to find the group rule with the best performance to verify that the Ideal Group rules are in fact ideal for our experiment. Second, we want to see which (if any) group rules perform nearly as well as the Ideal Group but can be realistically implemented. Lastly, we seek the actual group rule that the observers used in the experiment.

Below, we discuss the group rules' performance (PC and d'), weights, and relationship to the human groups' responses.

Overall Performance

The mean PC for each rule, applied to each group, is shown in Figure 4.20. The human group's mean PC is the thick horizontal black line behind the bar graph, and the dotted lines above and below that are the corresponding error bars. Even though some group rules used significantly more information than others, most of the group rules' PC scores were not significantly better than the human group's scores. The IGAllps rule had the highest PC of all rules, likely because it trained on the test trial, which represented a larger fraction of the training trials than it did in the IGAll rule. Though there may be a time dependence in the data, it is evident that the IGAllps rule's main advantage comes from the test trial making up 1/100th of the training data: if a per-session time dependence were the

main factor contributing to IGAllps's PC, then we would expect that the IGps rule would perform better than the IGm1 rule. The fact that the IGm1 rule always performed better than the IGps rule means that it is unlikely that the IGAllps's main advantage was its ability to track time dependence in the data.

It is interesting that the PC of the local rules (Near100 and Past100) varies widely from group to group. For Groups 1 and 3, the Past100 rule not only performs better than the Near100 rule but also competes with some of the less-optimal Ideal Group strategies; however, for Group 2, the Near100 rule has a higher PC than the Past 100 rule, which has the lowest PC among the rules. This is likely due to the individual characteristics of each group's members.

Among the realistically implementable strategies, it is clear that the best-performing rule varies from group to group. Though the difference is not significant, for Group 1, the best realistic rules were the Mean, ExpMean and Max Rules, which indicates that the observers who were more confident tended to have a higher probability of being correct. This is consistent with Group 1's observation spaces (Figure 4.5), where it was very rare for one of the observers to respond with an extreme rating and there were few if any extreme ratings that were incorrect. For Group 2, the Ideal Group strategies all performed significantly better than any other strategy, which is consistent with the idea that the Ideal Group strategies are ideal when the observers' d' values are approximately constant in time. It is clear from Group 2's performance in different windows of time (Figure 4.13) that

245

the observers in Group 2 had approximately constant performance, since the observers' local performance was generally very close to the mean. In Group 3, the best-performing rule was the Past100 rule, which out-performed even the IGAll rule, likely due to the fact that the group experienced volatility in member ability (i.e., MR was still learning the task) as well as group strategy for determining the weights (i.e., CG went from having the highest weight in the first four sessions to having a negative weight in the last session). The Majority and MeanMaj rules have the lowest PC, which indicates that the observers in Group 3 generally had a greater tendency to be correct when they were more confident.

The corresponding graphs for the d' values of the fusion center under each group rule are shown in Figure 4.21. Interestingly, the group rules with the highest PC did not necessarily have the highest d'. This is due to the way in which we calculated d': since we used the Wilcoxon method, group rules that provided less-certain responses on trials where the group was wrong and more-certain responses on trials where the group was correct have a higher d' than rules whose ratings were closer to being binary. In other words, PC is calculated using only one point on the ROC curve, whereas d' uses the entire ROC curve, so d' provides a more accurate measure of ability, and does not necessarily directly correlate with PC. For this reason, the Max rule has a low d' for all of the groups, as does the Majority scheme, which averages the ratings of the members in the majority. The Past100 rule has the highest d' for Groups 1 and 3, and the IGAllps rule has the highest d' for Group 2,
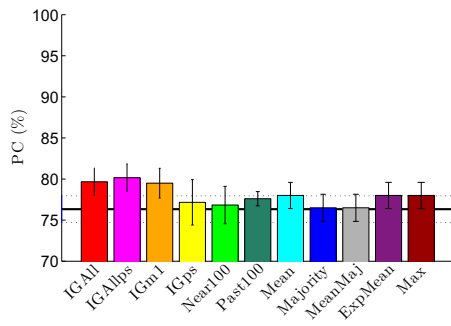
(a) Group 1 PC
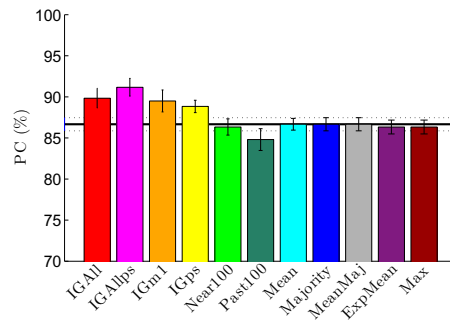
(b) Group 2 PC
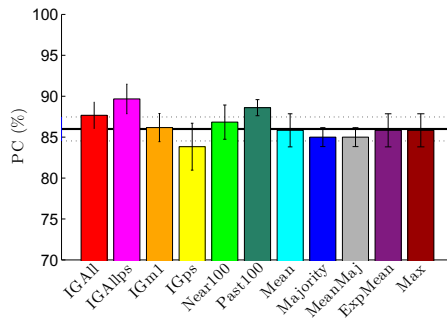
(c) Group 3 PC

Figure 4.20: Fusion center PC over all trials, under each of our eleven group rules. The heavy black horizontal line shows the human group's average performance, and the dotted lines show the error bars on that value.

likely for similar reasons as those given above for PC.

Observer Weights Per Session

The observer weights per session under each group rule for Groups 1, 2, and 3 are shown in Figures 4.22, 4.23, and 4.24, respectively. In this view, it is easier to see which group rules assigned weights similar to those that the human group actually used in each session, as well as the extent to which the different group
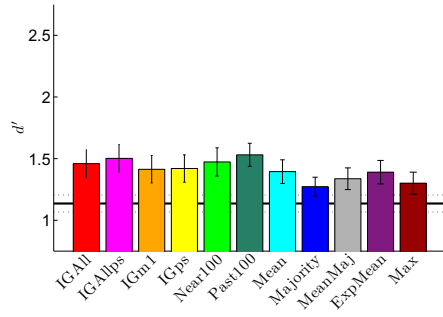
(a) Group 1 d'
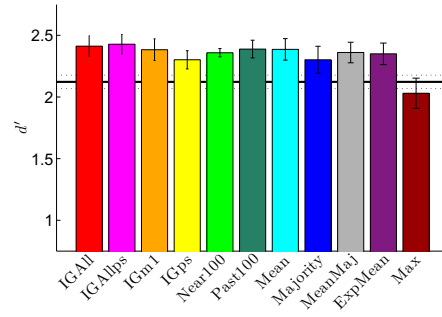
(b) Group 2 d'
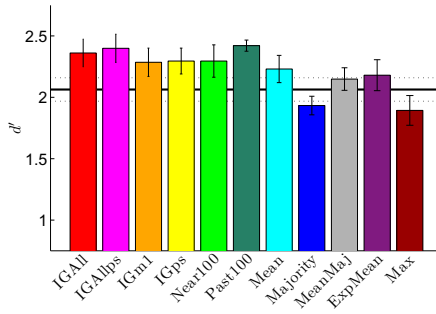
(c) Group 3 d'

Figure 4.21: Fusion center d' over all trials, for each group rule. The heavy black horizontal line shows the human group's average performance, with error bars given by the dotted lines above and below it.
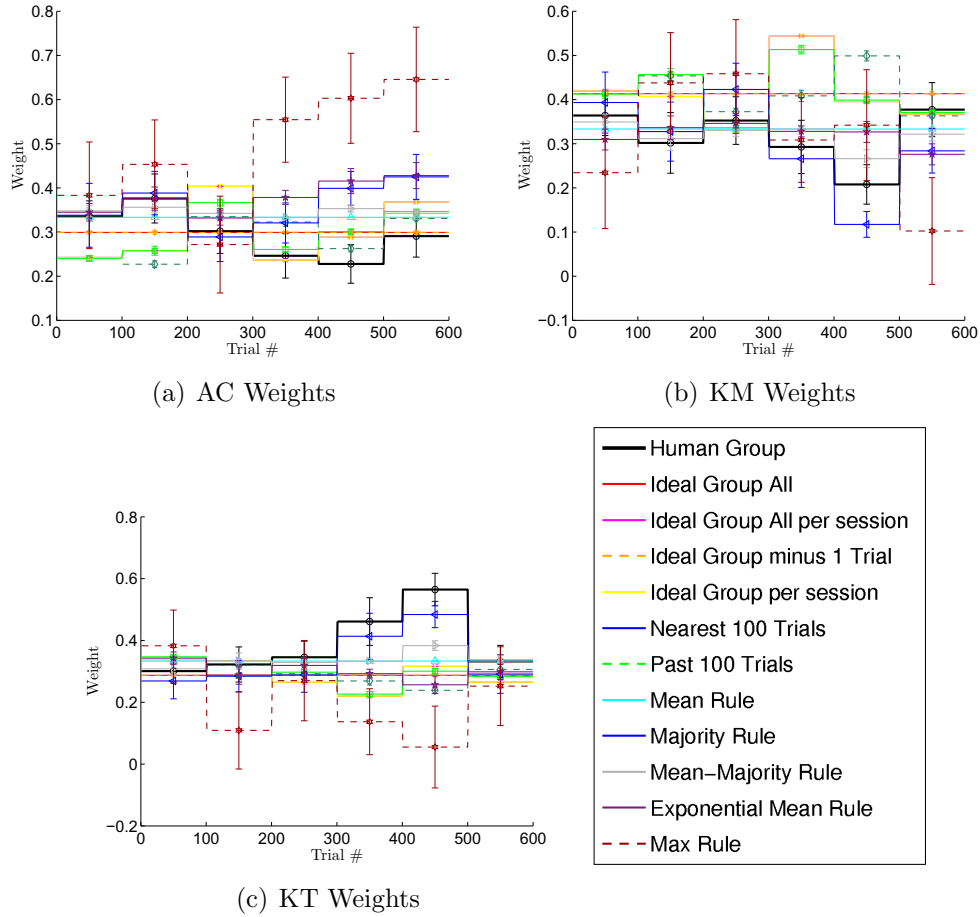
Figure 4.22: Average per-session weights assigned to each observer in Group 1 under each of the eleven group rules, compared to the weights used by the human observers.

rules' weights differed from the human group's weights.

Choice Prediction

Choice prediction is the percent of trials in which the group rule selected the same binary response as the human group, and a good measure for determining what strategy the group members likely used to arrive at a binary group response. The results are shown numerically in Table 4.3 and graphically in Figure 4.25.

(a) JW Weights

(b) KV Weights

(c) TS Weights
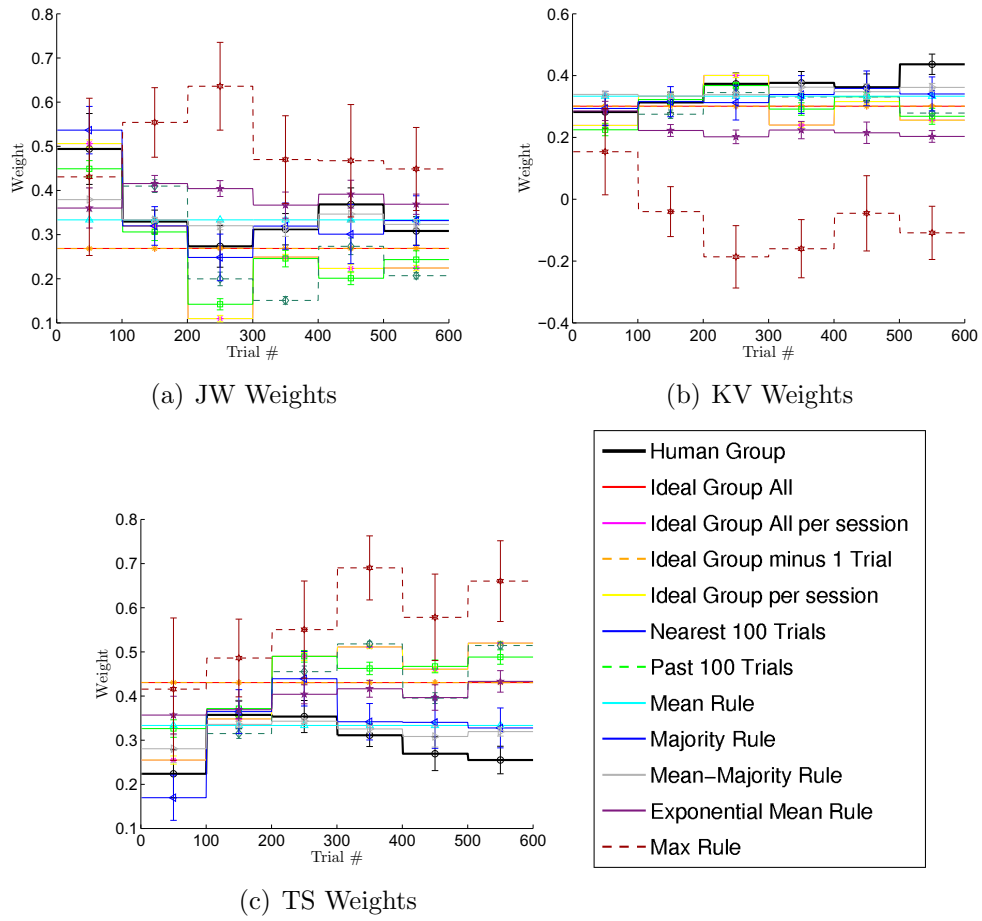
Figure 4.23: Average per-session weights assigned to each observer in Group 2 under each of the eleven group rules, compared to the weights used by the human observers.

(a) CG Weights

(b) MR Weights

(c) SF Weights
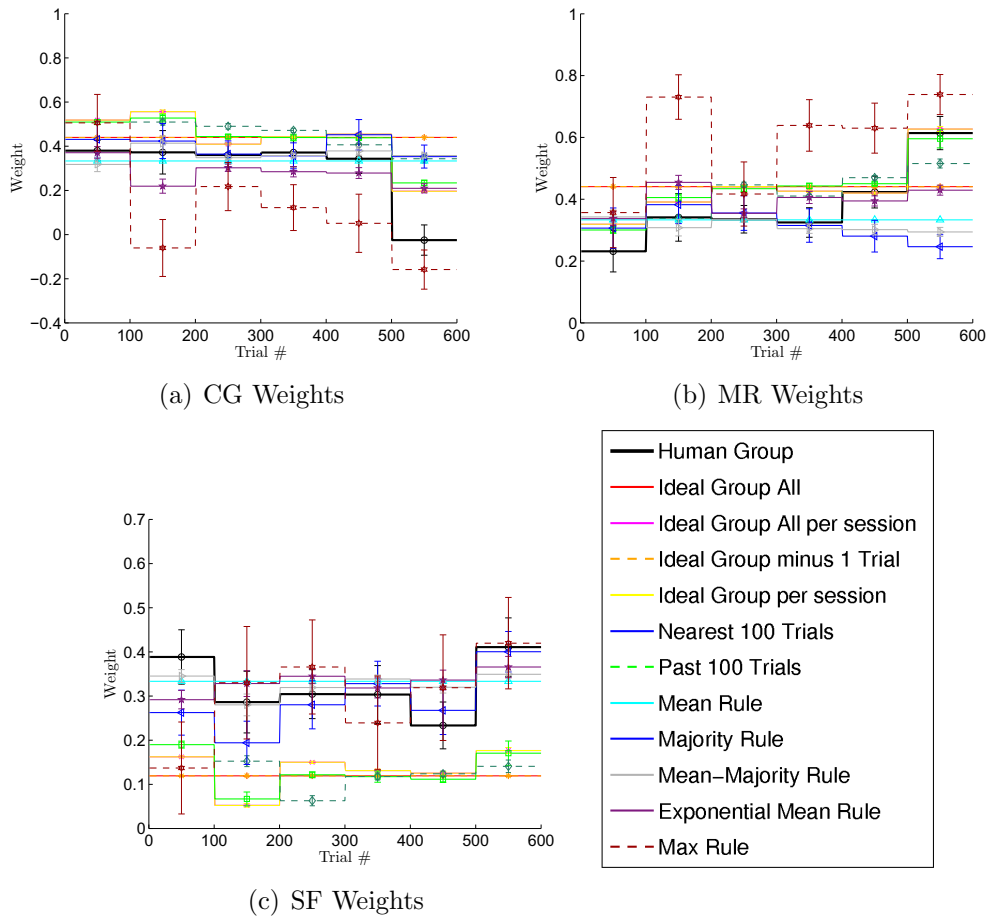
Figure 4.24: Average per-session weights assigned to each observer in Group 3 under each of the eleven group rules, compared to the weights used by the human observers.

For Groups 1 and 3, the Ideal Group rules had very low choice prediction scores, whereas the local rules (Near100 and Past100) had somewhat intermediate scores. This trend is reversed for Group 2. We attribute this to the fact that Groups 1 and 3 exhibited learning during the experiment.

It is clear that the human groups used a Majority-like rule to reach their binary decisions, since the Majority and MeanMaj rules have the highest choice prediction scores for all three groups. Those two rules have the same choice prediction score because they always select the same binary rule. In Figure 4.25(b), Group 2's Majority and MeanMaj rules do not have error bars because those rules had a choice prediction of 100% over the 600 trials. Though the data strongly suggests that Groups 1 and 3 primarily used a Majority-based rule, it is likely that they sometimes used other strategies that take an observer's confidence into account, since the Mean, ExpMean, and Max rules have the next highest choice prediction scores, which in Group 3's case, were not significantly lower than the Majority-based rules.

Group Rating Correlation

The correlation between the human group's ratings and the group rules' ratings are shown numerically in Table 4.4 and graphically in Figure 4.26. A higher level of correlation indicates that the group rule's ratings fluctuated from trial to trial in a manner similar to the human group's ratings.
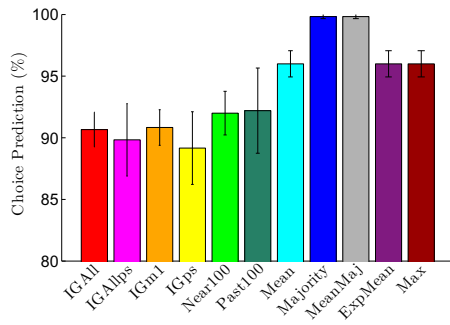
(a) Group 1 Choice Prediction
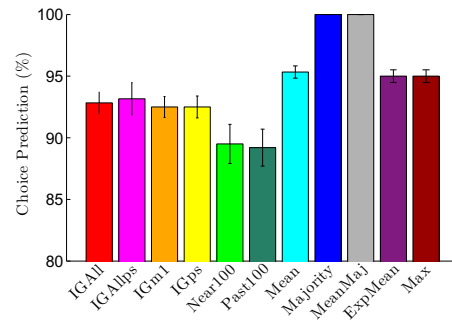
(b) Group 2 Choice Prediction
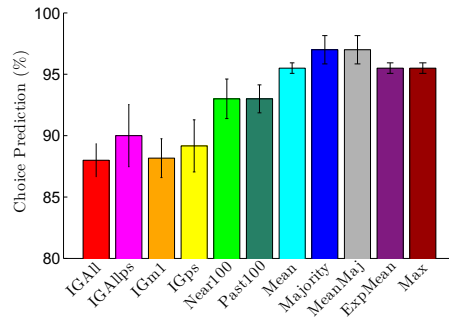
(c) Group 3 Choice Prediction

Figure 4.25: Choice Prediction of each group decision rule, for each group. The Majority rules clearly have the highest performance across all groups. In (b), there are no error bars for the Majority and Mean-Majority rules because both methods accurately predicted the group's choice in all trials.

Group Rule	Group 1	Group 2	Group 3
IGAll	90.67 ± 1.41	92.83 ± 0.87	88.00 ± 1.34
IGAllps	89.83 ± 2.94	93.17 ± 1.30	90.00 ± 2.53
IGm1	90.83 ± 1.45	92.50 ± 0.85	88.17 ± 1.58
IGps	89.17 ± 2.95	92.50 ± 0.89	89.17 ± 2.12
Near100	92.00 ± 1.77	89.50 ± 1.59	93.00 ± 1.61
Past100	92.20 ± 3.46	89.20 ± 1.50	93.00 ± 1.14
Mean	96.00 ± 1.06	95.33 ± 0.49	95.50 ± 0.43
Majority	99.83 ± 0.17	100.0 ± 0	97.00 ± 1.15
MeanMaj	99.83 ± 0.17	100.0 ± 0	97.00 ± 1.15
ExpMaj	96.00 ± 1.06	95.00 ± 0.52	95.50 ± 0.43
Max	96.00 ± 1.06	95.00 ± 0.52	95.50 ± 0.43

Table 4.3: Choice Prediction scores for our eleven different group rules, with standard error values. Choice Prediction is the percent of trials in which the group rule selected the same binary response as the human group. This information is graphically provided in Figure 4.25. The data strongly suggests that all of the groups used a Majority rule to select a binary response.

The group rule with the highest rating correlation varies from group to group: for Group 1, the Majority rule (which averages the ratings of the observers in the majority) had the highest score; for Groups 2 and 3, the MeanMaj rule had the highest score.

In general, the rules with the highest rating correlation are the equality-based rules. As we suggested earlier, this may be because the observers felt the need to make the group member in the minority feel that his opinion counted. A rule such as the Max rule, which gives one observer a weight of 1 and the other two observers a weight of 0 in each trial, would likely alienate members of the group and encourage social loafing or other negative behavior. Given this, it is not surprising that the Max rule had the lowest rating correlation for all three groups.
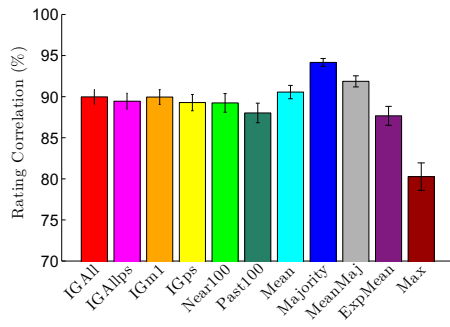
(a) Group 1 Rating Correlation
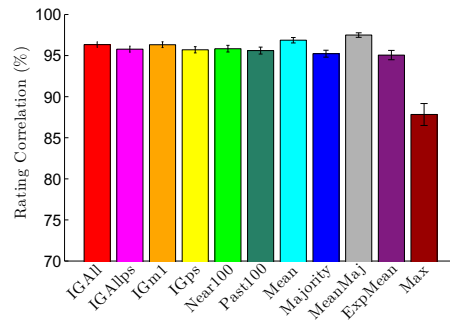
(b) Group 2 Rating Correlation
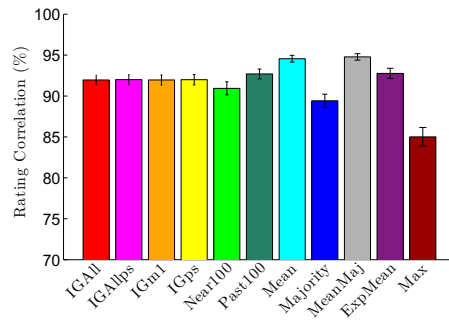
(c) Group 3 Rating Correlation

Figure 4.26: Rating Correlation of the fusion center under each group rule.

Group Rule	Group 1	Group 2	Group 3
IGAll	89.97 ± 0.89	96.33 ± 0.35	91.96 ± 0.60
IGAllps	89.44 ± 0.96	95.78 ± 0.37	92.00 ± 0.62
IGm1	89.96 ± 0.90	96.32 ± 0.35	91.96 ± 0.60
IGps	89.28 ± 0.99	95.70 ± 0.38	91.99 ± 0.64
Near100	89.24 ± 1.14	95.84 ± 0.40	90.94 ± 0.79
Past100	88.01 ± 1.19	95.60 ± 0.42	92.69 ± 0.60
Mean	90.55 ± 0.81	96.87 ± 0.33	94.56 ± 0.41
Majority	94.17 ± 0.48	95.22 ± 0.42	89.42 ± 0.80
MeanMaj	91.86 ± 0.68	97.49 ± 0.28	94.78 ± 0.39
ExpMaj	87.67 ± 1.15	95.06 ± 0.57	92.76 ± 0.62
Max	80.27 ± 0.67	87.83 ± 1.34	85.02 ± 1.14

Table 4.4: Rating correlation scores for the eleven group rules, with standard error values. Rating correlation is a measure of how similarly each rule's ratings fluctuated from trial to trial relative to the human group's ratings. While not all of the groups were most strongly correlated with the same group rule, all of the group rules with the highest rating correlation in each column are equality-based rules.

4.5 Continuation Study with Group 3

The observers of Group 3 were able to return for an additional 54 sessions, yielding a total of 6,000 trials (including the original study). We perform the same analysis as we did in Section 4.3 on this extended data set to be consistent with our previous results. Our primary interest is to see if the trends we saw in the first 6 sessions hold over the entire 60 sessions.

4.5.1 Individual Statistics

Below, we show the response statistics, response criteria, ROC curves, and performance of Group 3 in the continuation study.

Response Statistics

Group 3's response statistics are shown in Figure 4.27. The overall character-
istics are similar to those of the first 600 trials, but here, the histograms are more
evenly distributed, and the variance of each observer's Noise and Signal distribu-
tions are more similar than they were in the original study, likely an effect of the
larger number of trials. The distributions also cross closer to the mean rating (3.5)
in the continuation study than they did in the original study, indicating that the
observers learned to eliminate shifts in their observation spaces. A t-test failed to
reject the hypothesis that any of the distributions were not from a Gaussian at a
5% level of significance. Despite these improvements, some of the general trends
from the original study are still present: the group's distributions are the broadest,
yet have the smallest amount of overlap, and MR's distributions have the lowest
amount of overlap while SF's distributions have the highest, which corresponds
with the performance shown in Table 4.5.

Response Criteria

The observers' and group's response criteria are shown in Figure 4.28. While we
still see the interesting trend that the group's response criteria are more clustered
towards the center than any of the individual observers' criteria, it is also clear
that the criteria are more uniformly spaced than they were in the original study.
In addition, every criterion is now optimal. CG, SF, and the group also corrected

(a) CG Response Statistics

(b) MR Response Statistics

(c) SF Response Statistics

(d) Group 3 Response Statistics
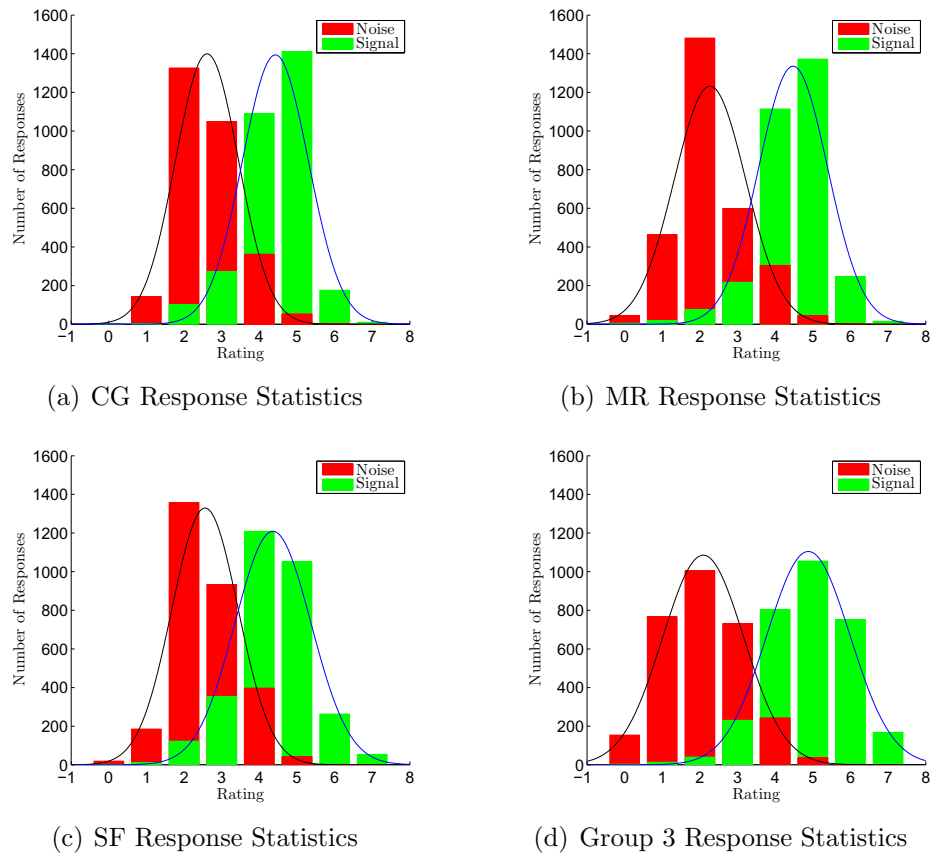
Figure 4.27: Response statistics for Group 3 in the continuation study. Though the distributions retain some of the characteristics found in the original study, overall, they are more regular, as one might expect from the larger number of trials.

(a) CG Criteria   (b) MR Criteria

(c) SF Criteria   (d) Group 3's Criteria
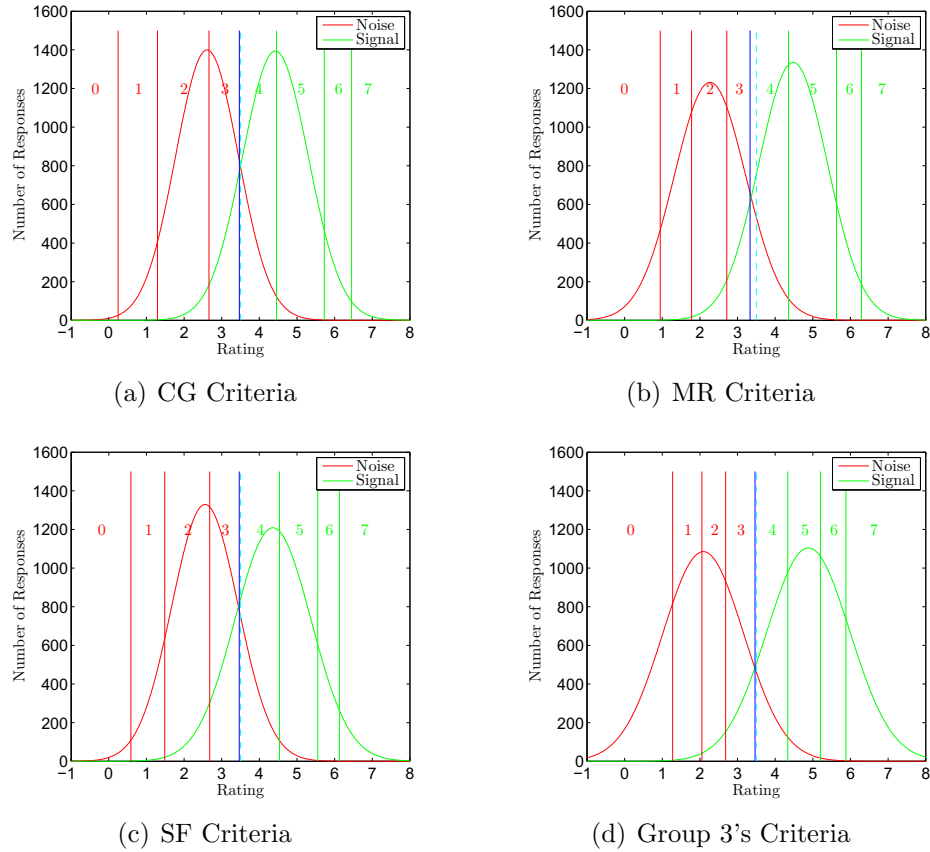
Figure 4.28: Rating criteria for Group 3 in the continuation study. The spacing between the criteria is significantly more regular than it was in the original study, which indicates that the observers learned to effectively use the ratings available and that the observers all learned the optimal criterion. MR's criterion is not at 3.5, but it is nonetheless optimal for his observation space.

the shifts in their distributions, but the shift in MR's distributions found in the original study has remained.

ROC Curves

The zROC and ROC curves for Group 3 in the continuation study are shown in Figures 4.29(a) and 4.29(b), respectively. Though the zROC plots still show nonlinearities, the extent of the nonlinearities has decreased, which results in a

259

(a) zROC Curves	(b) ROC Curves
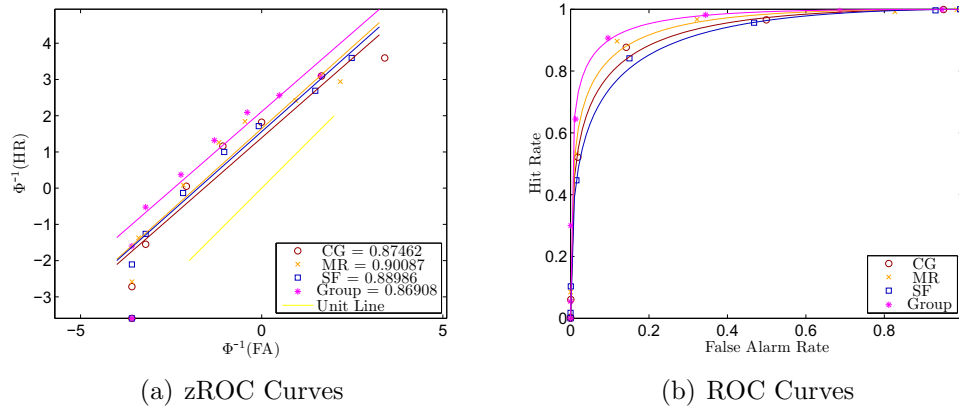
Figure 4.29: zROC and ROC curves for Group 3 in the continuation study.

better fit in the ROC plots. In addition, the area under the ROC curve has increased from the original study, because the observers improved at the task for reasons evident in their criteria and response statistics. Based on these results, we expect to see higher average performance.

Performance

Group 3's average performance over all trials is shown in Table 4.5, and the group's efficiency over all trials with respect to the Ideal Observer is shown in Figure 4.30. While all of the observers and the group improved, it is interesting to note that the overall group became significantly more efficient than the group's best individual observer (MR) over the course of the additional trials.

Due to the large number of trials, we present only the performance per session for a local view of the data (rather than using a moving window of 100 trials). Group 3's PC per session is shown in Figure 4.31(a), and the corresponding d' per
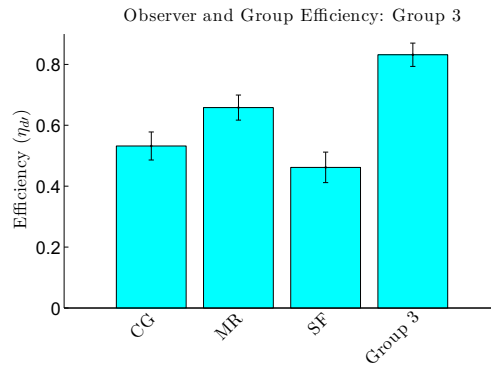
Figure 4.30: Group efficiency in the continuation study with Group 3.

session is shown in Figure 4.31(b). As we expected from the response statistics, criteria, and ROC curves, all of the observers and the group clearly improved from the original study. SF shows the most dramatic improvement in performance, but still remains the least capable member of the group. Since Group 3 favored equality-based group rules, it is likely that the main source of the group's higher performance in the continuation study is SF's improvement. It is very clear that a great deal of learning takes place during the first 9 sessions, but some additional learning takes place up to Session 18. This indicates the amount of training that observers may require to reach approximately steady-state performance.

4.5.2 Human Group Weights

Group 3's average weights in the continuation study are shown in Table 4.6. Compared to Table 4.5, it is clear that the weights still are not optimal, since CG is given a lower weight than SF despite the fact that CG has a higher d', PC, and
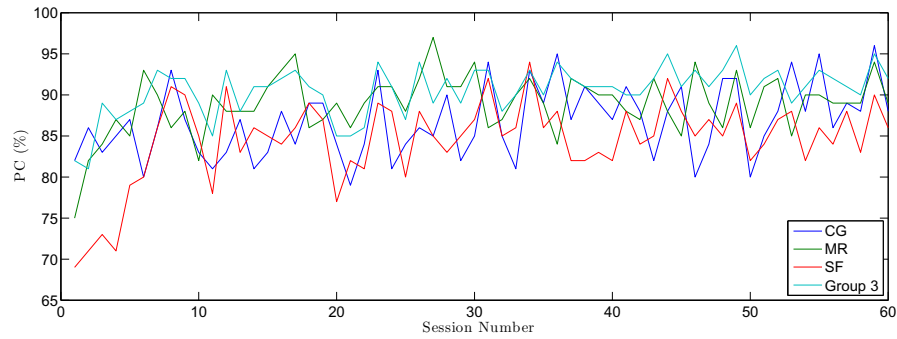
(a) PC per Session in the Continuation Study
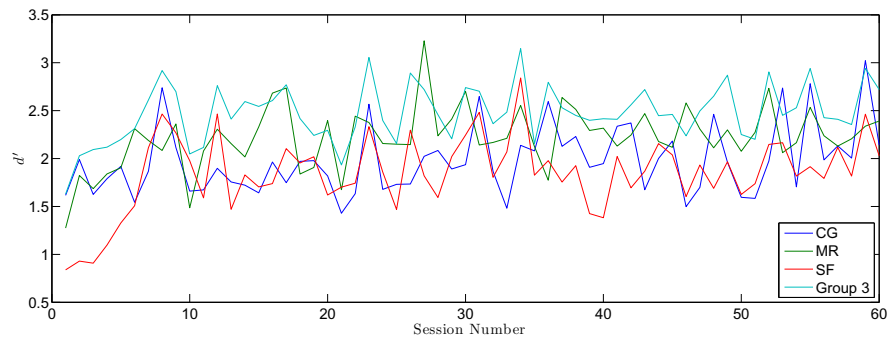
(b) d' per Session in the Continuation Study

Figure 4.31: Group 3's performance per session in the continuation study. It is clear that all of the observers and the group have improved at the task over the course of the additional 5,400 trials.

Observer	d'	PC (%)	HR (%)	FA (%)	$\eta_{d'}$ (%)
CG	1.93	86.8	87.7	14.2	53.2
MR	2.14	88.9	89.6	11.9	65.8
SF	1.79	84.5	84.1	15.0	46.2
Group 3	2.41	90.5	90.7	9.6	83.2

Table 4.5: Average performance of each individual and group over all trials in the continuation study: sensitivity (d'), Percent Correct (PC), Hit Rate (HR), False Alarm rate (FA), and efficiency with respect to the Ideal Observer ($\eta_{d'}$). The d' value was found non-parametrically, using the Wilcoxon method to calculate A_Z, the area under the ROC curve, which was then transformed to d' using Equation (4.2). The Ideal Observer's mean d' over all 60 sessions was 2.64.

	CG	MR	SF
Weight	0.2358	0.4014	0.3628

Table 4.6: Average Weights for Group 3 in the continuation study.
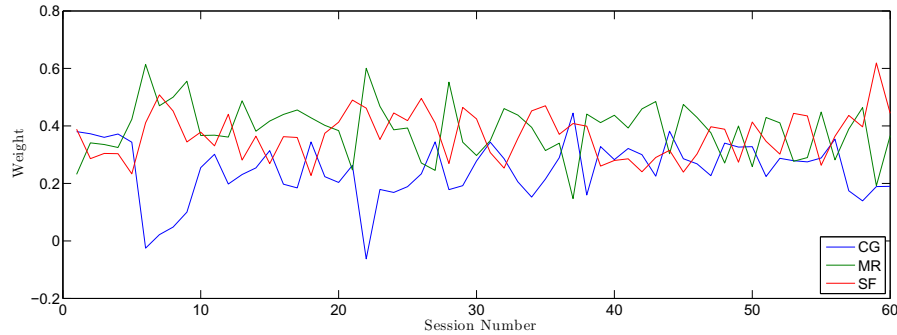
Figure 4.32: Weights assigned to each observer in Group 3 in the continuation study.

HR, and lower FA than SF.

The weights per session are shown in Figure 4.32. As was discussed in Section 4.3.4, CG's weights nose-dived in Session 6 after the proctor asked CG to be less emphatic during the group's discussions. While CG's weight did somewhat improve over the rest of the continuation study, it never recovered to the same level it was at during the first four sessions: aside from the first four sessions, there are only two sessions in which CG received the highest weight in the entire continuation study.

4.5.3 Group Rule Performance

Following our previous analysis, we applied the eleven group rules to the individual observers' ratings and found the performance of the fusion center under each group rule. We plotted the performance in 6-session increments in our detailed

view: the average PC per group rule is shown in Figure 4.33(a), and the PC per 6 sessions is shown in Figure 4.33(b). It is clear that the IGAllps group rule consistently has the highest PC of all of the group rules over the entire continuation study, while the rest of the group rules' PC values stay near the human group's PC values.

The corresponding d' over all trials and per 6-session window are shown in Figures 4.34(a) and 4.34(b), respectively. As we saw previously, the d' values do not necessarily show the same trends as the PC scores, and the Majority and Max rules have the lowest d'; however, the d' values for the other group rules appear to have all approached a similar value.

Choice Prediction

We show the choice prediction results over all trials in Figure 4.35(a), and per 6 sessions in Figure 4.35(b). As we expect from the original study, Group 3 primarily used a Majority-based rule to arrive at a binary decision in most of the trials in the continuation study. The IGAll and IGm1 rules' choice prediction scores have improved relative to the other group rules, likely because the observers' sensitivities eventually asymptote, and the IGAll and IGm1 rules perform best when there is no learning. However, other than that, the group rules have a similar level of choice prediction, indicating that Group 3 consistently used the same strategy to arrive at a binary decision throughout the experiment.
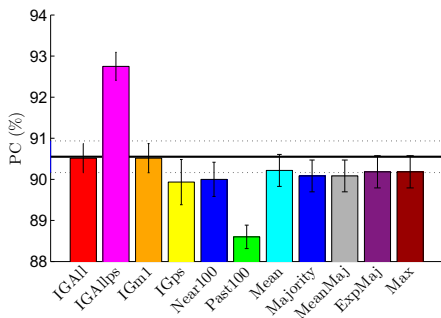
(a) PC Over All Trials in the Continuation Study
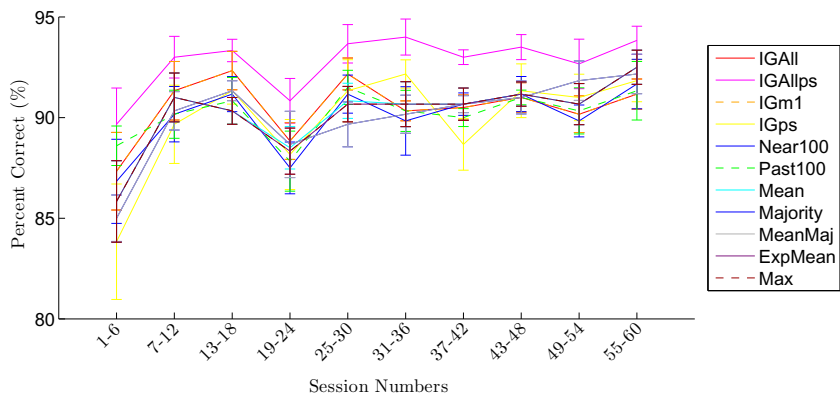
(b) PC per 6 Sessions in the Continuation Study

Figure 4.33: (a) Average PC over all trials for each group rule in the continuation study with Group 3. The horizontal black line behind the bar graph shows the PC of the human group, and the dotted lines show the error bars on that value. (b) PC per 6 sessions for each group rule.
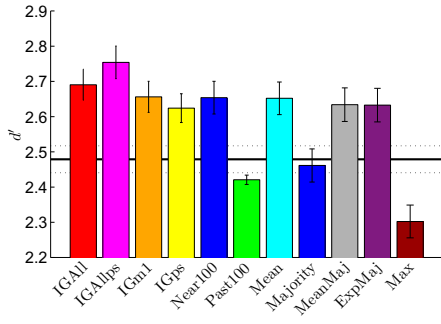
(a) d' Over All Trials in the Continuation Study
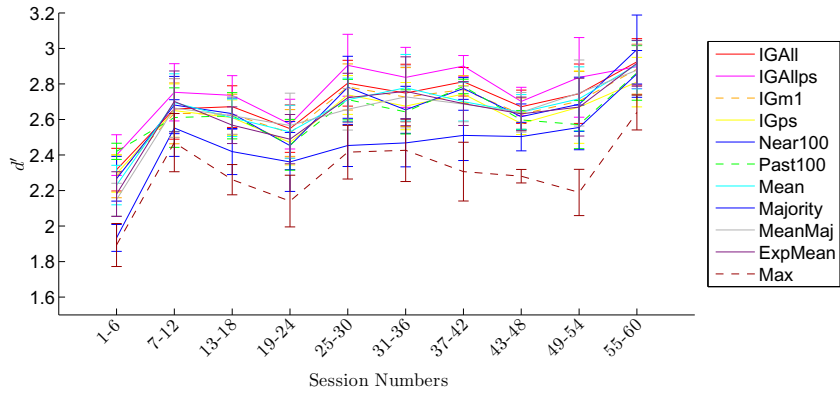
(b) d' per 6 Sessions in the Continuation Study

Figure 4.34: (a) Average d' over all trials for each group rule in the continuation study with Group 3. The horizontal black line behind the bar graph shows the d' of the human group, and the dotted lines show the error bars on that value. (b) d' averaged over 6-session intervals for each group rule.
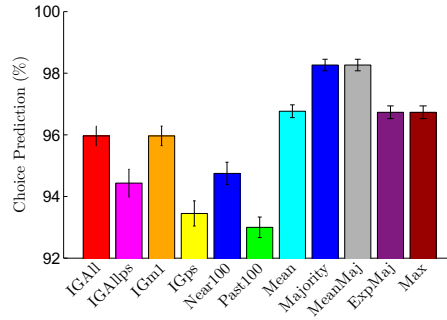
(a) Choice Prediction Over All Trials in the
Continuation Study
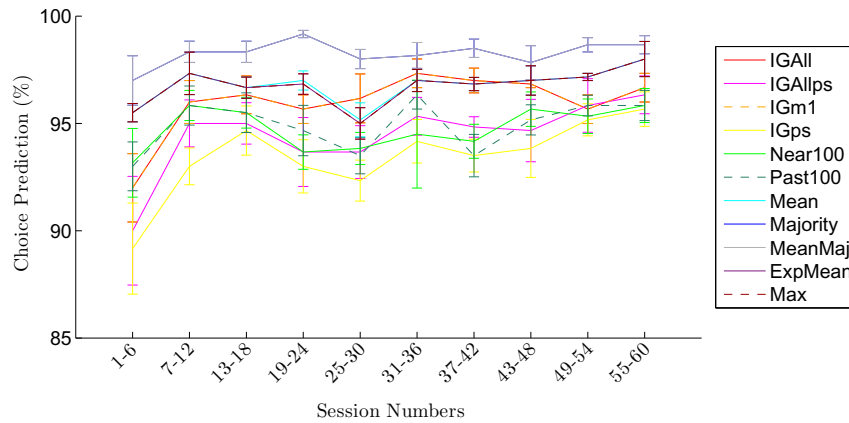
(b) Choice Prediction per 6 Sessions in the Continuation Study

Figure 4.35: Choice prediction of each group rule applied to Group 3 in the continuation study.
(a) Average choice prediction over all trials. It is clear that the Majority and MeanMaj rules have
the highest choice prediction. (b) Choice prediction per 6 sessions. From this view, we can see
that the Majority and MeanMaj rules have the highest choice prediction scores in each 6-session
period over the entire experiment. In the graph, the Majority line matches the MeanMaj line
exactly because they use the same rule to arrive at a binary decision, so the MeanMaj data points
obscure the Majority data points.

Rating Correlation

The rating correlation of each group rule, averaged over all trials, is shown in Figure 4.36(a), and the rating correlation per 6-session window is shown in Figure 4.36(b). The average rating correlation per group rule appears to be approximately unchanged from the original study in a relative sense, which further supports the idea that the strategy that Group 3 used did not change throughout the continuation study. Like in the original study, the Mean and MeanMaj rules have the highest rating correlation, indicating that even though the group used a Majority rule to reach a binary decision, the members in the majority factored in the opinion of the minority in selecting the group rating.

Between the choice prediction and rating correlation, it is clear that Group 3 continued to utilize a group rule very similar to the MeanMaj rule.

4.6   Simulation: Varying   the   Observation   Space in Time

It is a common assumption that an observer's sensitivity does not change in time. If observer sensitivity is fixed, then we would expect that the "Ideal Group" we defined earlier would indeed be ideal; however, if there is some variation in an observer's sensitivity from day to day or within a session, then the other group rules may be able to compete with the Ideal Group. We test this assumption via
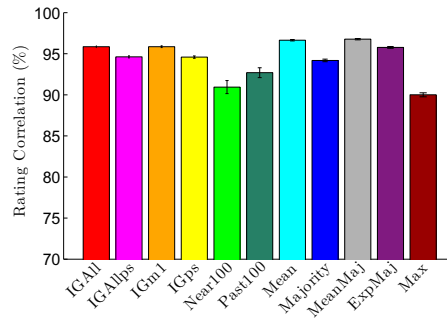
(a) Rating Correlation Over All Trials in the Continuation Study
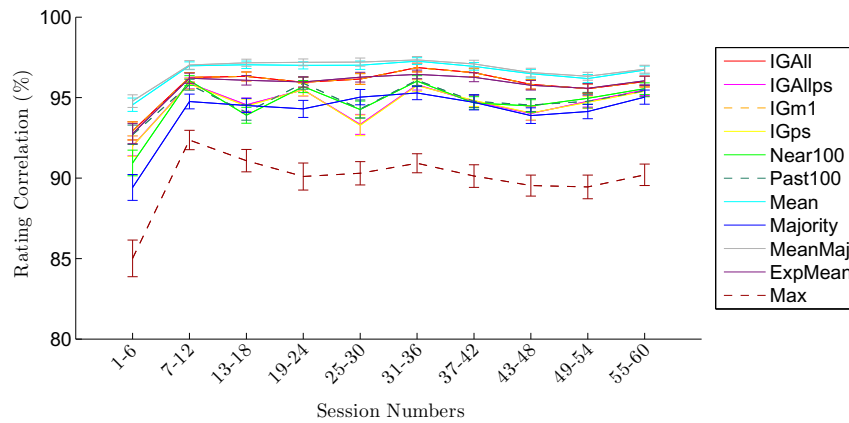
(b) Rating Correlation per 6 Sessions in the Continuation Study

Figure 4.36: (a) Average rating correlation of each group rule applied to Group 3 in the continuation study. It is clear that the group used a rule similar to the Mean or MajMean rule. (b) Rating correlation per 6 sessions. From this graph, we can see that the Mean and MajMean rules were heavily favored throughout the 6,000 trials.

simulation.

Our simulations follow the setup of our experiment, and we strongly tied our simulation's parameter values to the data, as we explain below. For each trial in the simulation, we used the same correct response (Signal or Noise) per trial as the actual experiment, which had a 50% prior probability of being Signal. For our stimulus type, even though each observer saw the same image, we expect each observer's internal, perceived observation to have only a 50% correlation with the perceived observations of the other members of the group. In other words, we expect half of the noise to be common to of the members of the group and the other half to be specific to each observer, due to internal noise, based on the results of [18]. Therefore, for each trial, we generated four normally-distributed random numbers: one served as the noise common to all members of the group, and the other three were designated as internal noise for each individual observer. Each observer's observation per trial was generated by averaging the common and individual noise values, then transforming that value to the correct distribution in the observer's observation space. We then applied each group member's rating criteria to arrive at a numerical rating.

Mirroring the analysis done in Section 4.3, we first show the results of a performance analysis of the simulated individuals and groups, then show the results of applying the eleven group rules to the simulated individuals.

4.6.1 Individual Performance

We simulated three different conditions, each representing a different assumption on the observers' d' values: (1) the observers had a fixed d' over all trials; (2) the observers' d' values differed from session to session, but were fixed within any given session; and (3) observers' d' values varied from trial to trial, and were drawn from a normal distribution. Below, we provide further detail on each condition.

Constant d' Over All Sessions   If an observer's Signal and Noise distributions are normally distributed, then the distance between the means of the distributions is equal to d'. We took this assumption to build the observation spaces of our simulated observers, and used each observer's mean d' over all trials to define the distance between his Signal and Noise distributions. The traditional formulation of this problem centers the Noise distribution at zero; however, we wanted the simulation's distributions to be similar to those calculated in Section 4.3.1, so we translated the distributions so that they were symmetric around the mean rating (3.5).

The PC scores of the simulated Groups 1, 2, and 3 are shown in Figures 4.37(a), 4.38(a) and 4.39(a), respectively, and the corresponding plots of d' are shown in Figures 4.37(b), 4.38(b) and 4.39(b), respectively.

Constant d' Per Session   For this case, we used the observers' actual per-session d' values to define their observation spaces for each session. We used the observers'
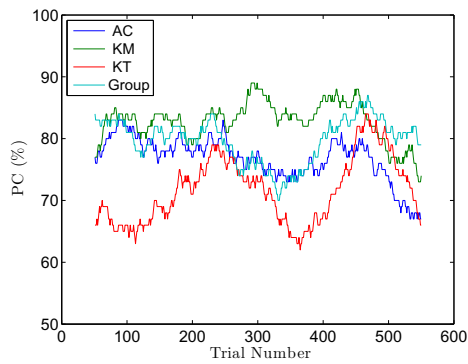
(a) Constant Group 1 PC
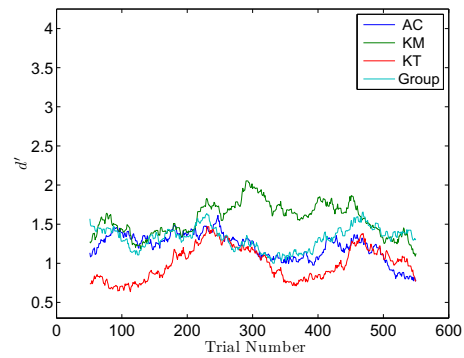
(b) Constant Group 1 d'
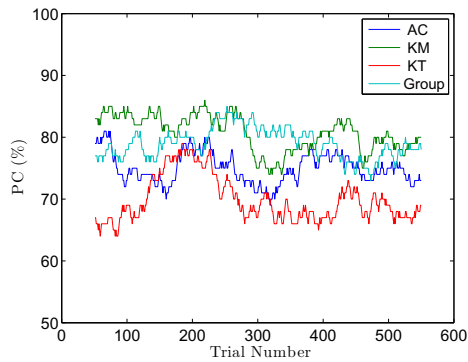
(c) Per Session Group 1 PC
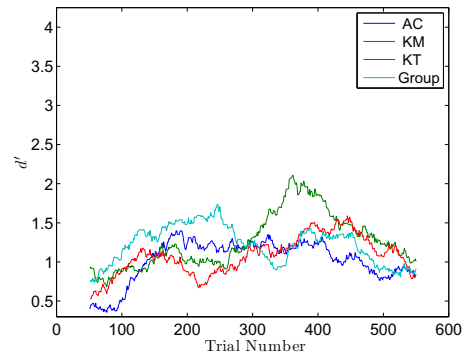
(d) Per Session Group 1 d'
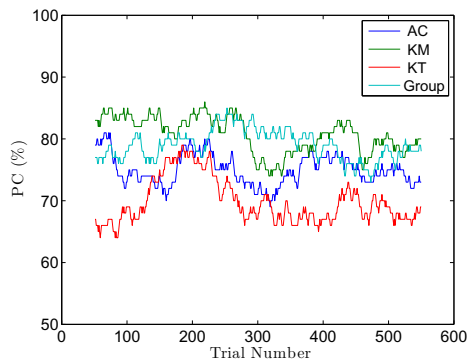
(e) Gaussian Group 1 PC
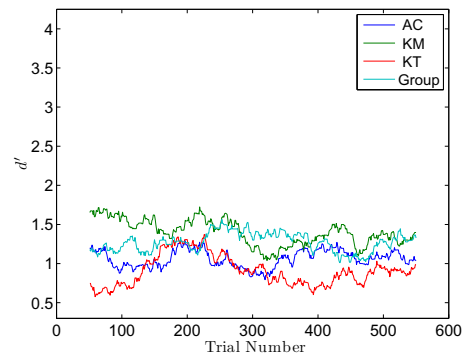
(f) Gaussian Group 1 d'

Figure 4.37: Simulated Group 1's performance for the three conditions - (a) & (b): constant d' over all trials, (c) & (d): constant d' per session, and (e) & (f): normal d' per trial.
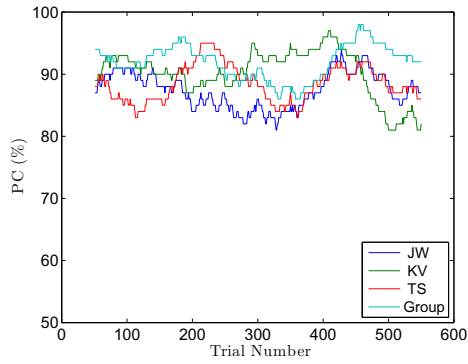
(a) Constant Group 2 PC
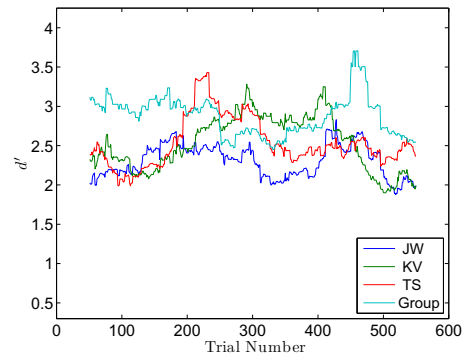
(b) Constant Group 2 d'
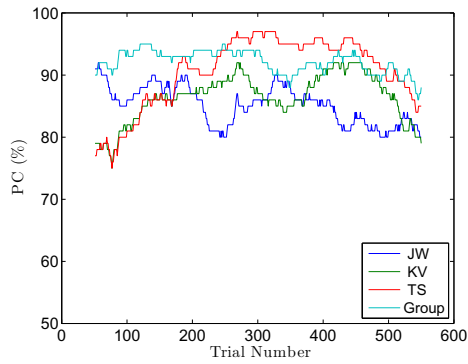
(c) Per Session Group 2 PC
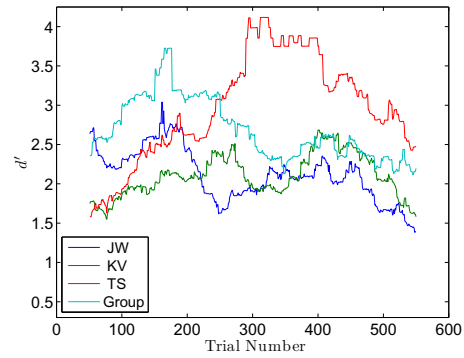
(d) Per Session Group 2 d'
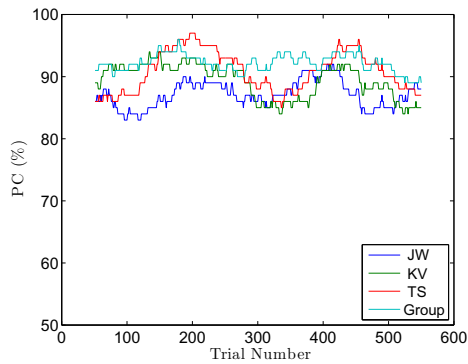
(e) Gaussian Group 2 PC
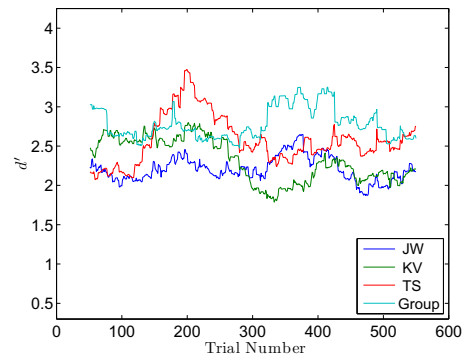
(f) Gaussian Group 2 d'

Figure 4.38: Simulated Group 2's performance for the three conditions - - (a) & (b): constant d' over all trials, (c) & (d): constant d' per session, and (e) & (f): normal d' per trial.

273
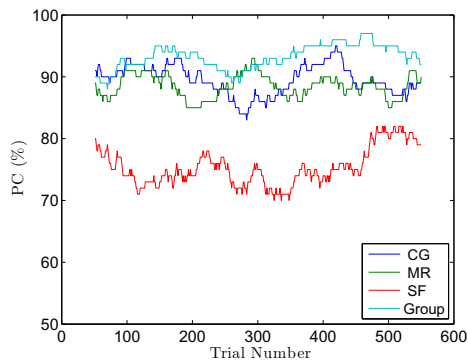
(a) Constant Group 3 PC
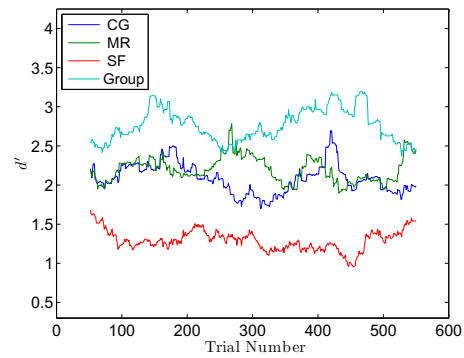
(b) Constant Group 3 d'
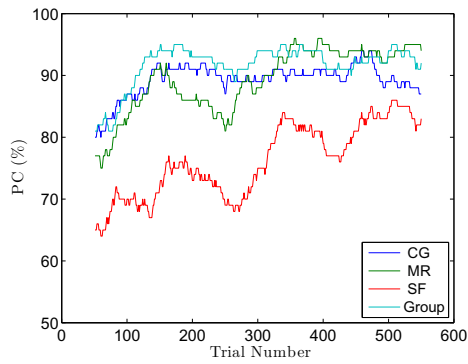
(c) Per Session Group 3 PC
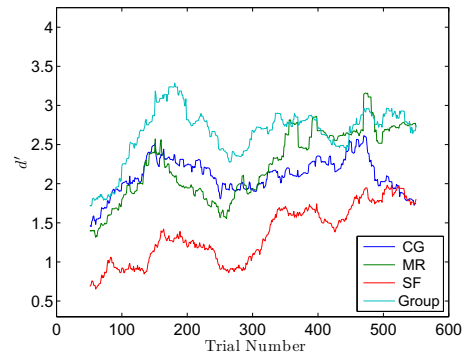
(d) Per Session Group 3 d'
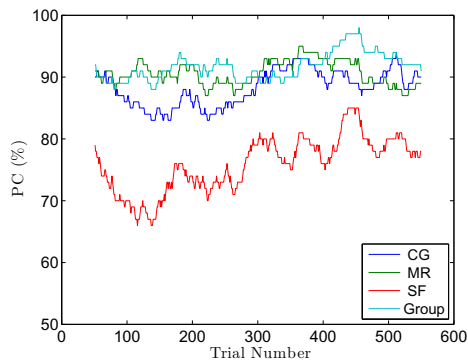
(e) Gaussian Group 3 PC
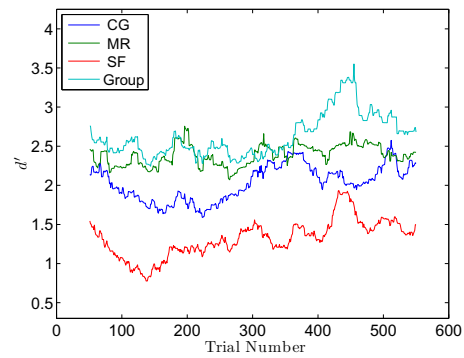
(f) Gaussian Group 3 d'

Figure 4.39: Simulated Group 3's performance for the three conditions - - (a) & (b): constant d' over all trials, (c) & (d): constant d' per session, and (e) & (f): normal d' per trial.

rating criteria over all trials to select the ratings per trial, to ensure that all of the rating criteria would be defined.

The PC scores of the simulated Groups 1, 2, and 3 are shown in Figures 4.37(c), 4.38(c) and 4.39(c), respectively, and the corresponding plots of d' are shown in Figures 4.37(d), 4.38(d) and 4.39(d), respectively.

Normal d' Per Trial   In this last variation, for each trial, we randomly pulled a d' value from a normal distribution centered at the observer's mean d', then used that value to define the observer's observation space for the trial. We again applied the criteria calculated over all trials to the observation to find the rating.

The PC scores of the simulated Groups 1, 2, and 3 are shown in Figures 4.37(e), 4.38(e) and 4.39(e), respectively, and the corresponding plots of d' are shown in Figures 4.37(f), 4.38(f) and 4.39(f), respectively.

4.6.2   Applying the Group Rules

After generating the observers' ratings, we applied our eleven group rules and calculated the performance of each rule. Our results are shown in Figure 4.40: (a), (c) and (e) show the PC for simulated Groups 1, 2 and 3, respectively, and (b), (d) and (f) show the corresponding d' values. In each plot, the blue circles show the performance of the constant d' data, the red x's show the performance of the constant d' per session data, and the green triangles show the performance of the normal d' per trial data. For comparison, we have also included information from

the human group. The horizontal orange line is the human group's actual average performance, and its error bars are shown in yellow. The gray-outlined white bars in the background are the performance of each group rule when applied to the human data, as shown in Figures 4.20 and 4.21.

For simplicity in the discussion, the human group's actual performance (shown in orange) will be referred to as the "human performance", the performance of the human group when its observers' ratings are given to a fusion center (shown in gray) will be referred to as the "human rule performance", the constant d' over all trials simulation condition (shown in blue) will be referred to as the "constant simulation", the constant d' per session simulation condition (shown in red) will be referred to as the "per-session simulation", and the normal d' per trial simulation condition will be referred to as the "normal simulation". We will continue to refer to the eleven group rules (labeled on the x-axis) collectively as the "group rules".

For Group 1, the per-session simulation consistently had the worst performance, but was most similar to the human and human rule performance. The constant d' over all trials condition consistently had the best performance, which generally was significantly better than the human and human rule performance, and frequently significantly better than the per-session performance. Overall, the simulation performance remained close to the human and human rule performance. This trend did not continue for Group 2. As shown in Figures 4.40(c) and 4.40(d), the normal simulation generally had the best performance for Group 2, but the simulations

did not have significantly different performance from each other. The simulations'
performance was also very significantly greater than the human rules' performance
in d', and in PC for the non-ideal group rules. For Group 3, the normal simulation
had the highest PC for the ideal and local group rules, and the constant simula-
tion had the best performance for the equality- and confidence-based rules. In d',
however, the normal simulation consistently had the best performance, though the
difference was not significant.

Overall, our results are fairly counterintuitive. We had expected the constant
simulation to perform the best among all group rules and simulations for the Ideal
Group rule, and for that performance to drop off for the other group rules, because
our Ideal Group rules were named "ideal" with the assumption that the observers'
d' values were fixed. We had also expected the normal simulation to perform best
for the rules that did not use any memory, and the per-session simulation to per-
form best in the IGps rule. Additionally, we had expected the performance of the
simulated groups to be similar to the performance of the human rules. The sim-
ulated groups' d' values appear to be correlated with the human rules' d' values,
which is reasonable because we used the values of d' found in the data in the simu-
lation; however, the simulations applied to the group rules performed significantly
better than the human rules. The simulated groups' PC values generally are neither
correlated with nor close to the human rules' PC values in only 600 trials.

This indicates that the human observers' observation spaces were not completely

277

described by any of our three conditions.

4.7 Discussion

We had three groups of 3 observers each perform a simple Signal Detection task individually and in a group. In addition to improving upon similar previous studies [87], we also explored five general areas: individual versus group performance, time-dependent data visualization, group decision strategies, persistence of our findings over a longer period of time, and the assumption that an observer's sensitivity is fixed.

From visualizing the data in different windows of time, we arrived at the conclusion that it was likely that the observers used a group rule that did not involve much memory; therefore, we considered five group rules that did not require memory to implement. We found that the group generally performed slightly better than its best individual observer despite the fact that the groups were not optimal at assigning weights. Since our experiment setup only gave the observers an incentive to maximize group accuracy and the group included only three members, it is unlikely that factors such as social loafing or intra-group competition affected our results.

From our group rule analysis, we found that it was most likely that Groups 2 and 3 used a Mean-Majority rule and Group 1 used our modified Majority rule, which explains the groups' inefficiencies in assigning weights: theory states that
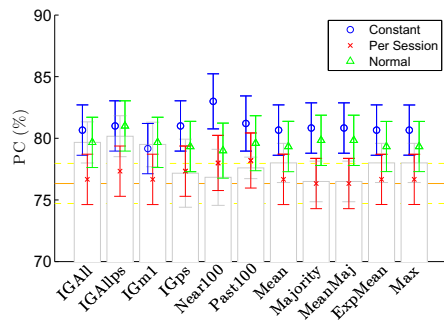
(a) Simulated Group 1 PC
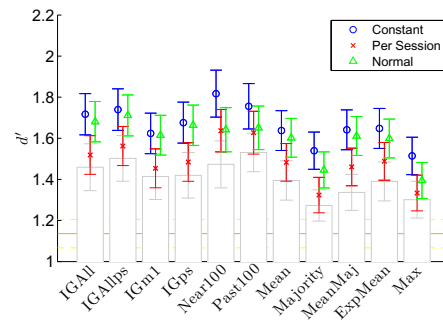
(b) Simulated Group 1 d'
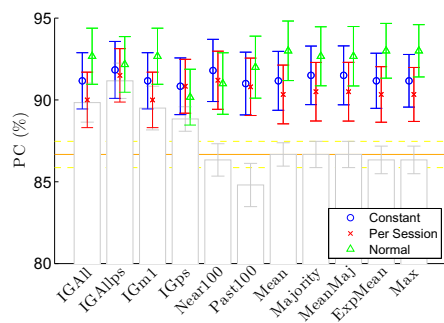
(c) Simulated Group 2 PC
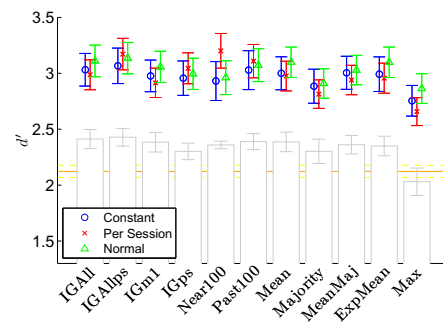
(d) Simulated Group 2 d'
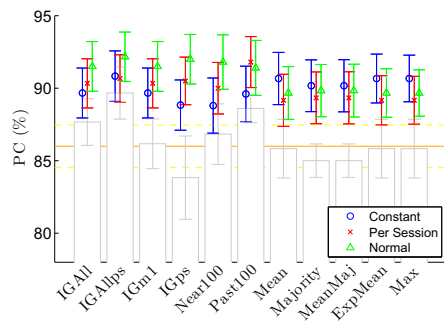
(e) Simulated Group 3 PC
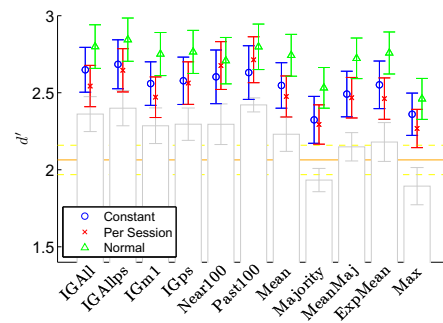
(f) Simulated Group 3 d'

Figure 4.40: The overall performance of each simulated group under the eleven group rules. The blue circles are the performance under the constant d' condition, the red x's show the performance under the constant d' per session condition, and the green triangles show the performance under the normal d' per trial condition. For comparison, we have also included the performance of the human group (shown in Figures 4.20 and 4.21): the horizontal orange line is the human group's average performance, with its error bars shown in yellow. The gray bars in the background show the performance of each group rule when applied to the human data.

the optimal strategy is to assign weights proportional to the observers' sensitivities [63]. This is not surprising: it is reasonable to default to a Majority rule to reach a binary response when no explicit instructions are given [43], and utilizing a Mean rule or some variant of it to reach a rating provides the majority with a way to appease the minority. Based on the results of our continuation study, we also found that observers generally do not change their weighting strategies in repeated tasks: the results related to Group 3's weighting strategy averaged over 6,000 trials in the continuation study showed the same general trends as the results from just the first 600 trials.

Though the IGAllps rule had the best performance, it is not a viable strategy, because it trains on test trials. It was surprising to find that many of the group rules that did not require memory had performance that was not significantly different than the Ideal Group rules, since these rules used significantly less information than the ideal or local rules. These rules also have the advantage of being realistically implementable and extremely simple to compute. It is likely that the Ideal Group rules would have been more ideal if we had applied them to a larger number of trials, due to the way they are defined; however, we felt it would not be a fair comparison if we used multiple realizations of the data to compute the performance of the Ideal Groups when the actual experiment used only 600 trials (in the original study, and 6,000 trials in the continuation study). Under this more realistic scenario, we found that the acceptable versions of the Ideal Group (IGm1 and IGps) generally

were not significantly better than the rules that used only local information, since those two rules only had significantly higher performance for Group 2's PC scores. Therefore, we conclude that the best rule for a particular group is related to the characteristics of the individuals in the group: for example, if even one group member's more confident ratings were not correlated with a higher probability of being correct, then the Max, ExpMax, and Mean rules would suffer noticeably.

Finally, we tested three hypotheses about the observers' internal d' values (or equivalently, observation spaces) by simulating the three cases: constant d' over all trials, constant d' per session, and normal d' per trial. Though we extracted our simulation parameters from the data, our simulation results frequently differed significantly from the human-based results. The simulation of Group 2 in Figure 4.40 (d) is the most extreme example of this. We believe that the difference is largely due to inefficiencies in the human observers that were not adequately captured in the simple model we used for the simulation. Our simulation assumed that the observers' observation spaces consisted of two normal distributions; however, our analysis of the observers' response statistics found that JW's distributions and TS's signal distributions were likely not Gaussian. Since both of those observers are from Group 2, this may explain the large difference between that group's actual and simulated performance. The (smaller) difference between the simulated and actual performance of Groups 1 and 3 may be due to differences in the variances of the individuals' Gaussian distributions in their observation spaces. Since the

shapes of the distributions differ, the area of overlap (that defines the error rates) differs, and thus, the performance differs. The group rules, when applied to the simulated data, performed in a manner similar to how they did for the human data, which suggests that the relative performance of the group rules depends more on the characteristics of the group it is applied to, rather than there being a fundamental hierarchy on which group rule is "better". Our Ideal Group rules were based on the "ideal weights" defined in [87]. We found that the performance of the Ideal Group rules when applied to the simulated data also did not out-perform our simpler group rules, which is consistent with what we found for the human data.

4.8   Conclusion

In our experiment, three observers in a group performed a simple signal detection task on the same image individually, then conferred to reach a group rating. Our results are relevant to visual Yes-No tasks in which observers in a team are given the same stimulus, such as determining whether or not a spot in a mammogram is likely cancerous and requires a biopsy, or deciding whether or not a piece of luggage at an airport security checkpoint contains something dangerous and is sufficiently suspicious to warrant a manual inspection. We found that a number of supposedly sub-optimal strategies; namely, the Past100, Mean, Majority, Mean-Maj, and ExpMean group rules; performed as well or occasionally better than the Ideal Group, as defined in [87], and that the human group generally performed as

well as or better than its best member. Through our analysis, we have provided further insight on how groups of humans reach decisions and how one might optimally organize a decision-making scheme for a particular group.

Chapter 5

Conclusion

The purpose of this dissertation is to explore the performance of groups in a two-alternative forced-choice (2AFC) task, with the intent of working towards developing a cybernetic group-based decision-making system.

We began with an applied mathematics-based approach to modeling the performance of a group of decision-makers (DMs) that is intuitive and supported by previous work in various experiment-based literatures. In Chapter 2, we derived explicit solutions for the performance of an individual DM in a 2AFC task using the Sequential Probability Ratio Test (SPRT), which has been shown to be the optimal test for the type of tasks we consider. We then demonstrated a novel way to combine the performance of N independent DMs to find the group's performance under each of three simple group decision rules in Chapter 3, and discussed the relative merits of each group rule. One of the main advantages of our models is

that they are flexible: each "member" in the group need not be an individual as long as its performance can be characterized, so it is straightforward to extend our models to more complicated hierarchical setups, where a "member" is itself another group, possibly using a different group rule. We also generalized our models to accommodate non-identical DMs, as long as they are independent, and generalized our group rules. We illustrated the accuracy of our DDM-based continuous-time results through a comparison of our solutions with the results of simulating of an identical system using the SPRT in discrete time.

In addition to developing general models, we also investigated various issues related to simulating the system. We found and characterized where and how Wald's approximation for the decision variable's boundary breaks down, and how that affected the individual and group models.

Our work shares a number of similarities with research in decentralized or distributed detection and decision aggregation, but is significantly more general, since our model can accommodate DMs that are humans as well as devices. Also, our models should be more accessible to a wider range of communities, since our results and analyses are expressed in terms of pdfs, cdfs, error rates, and decision times, all of which are very intuitive and used in a wide range of disciplines. We derived exact solutions to our individual and group models without the explicit use of specialized subjects such as measure theory or dynamic programming. Our models are interesting because they present a novel and general way in which one can intuitively

yet mathematically model a group's performance based on its members' statistics, establish a reasonable base model which can be extended to build up more complicated models for realistic groups, and provide a means by which one can compare different group decision rules without the need for a specific cost function.

Once we established a base model for cybernetic group decision-making in a potentially hierarchical scheme, we turned to investigating some of the issues specific to decision-making in a group of humans. Human observers are significantly more complicated than the simple devices often used in distributed decision-making because there are many factors outside of the task that can have a significant influence on the observer's performance: for example, devices never get bored with their task, or indulge in social loafing. To better learn and further the current understanding of group decision-making in humans, we analyzed the data from a simple experiment in which three groups of 3 observers each performed a visual signal detection task on the same image individually, then conferred to reach a group rating. In addition to improving upon similar previous studies [87], we also explored five general areas: individual versus group performance, data visualization, group decision strategies, persistence of findings in the first three areas over a longer period of time, and the assumption that an observer's sensitivity is fixed.

We found that the observers seemed to favor strategies that required little or no memory, and that the overall group generally performed better than the group's best individual, despite the fact that the groups did not use optimal weights. Our

data is fairly conclusive that the groups used a strategy similar to the Majority rule to reach a binary decision, which is consistent with previous studies [43]. We also found that the observers generally favored equality-based rules to decide the group's rating. We believe that the equality-based rules were attractive because they provide the majority (who determines the group's binary response) with a way to appease the minority, and thus maintain a good working relationship. Based on the results of our continuation study, we also found that observers generally do not change their weighting strategies in repeated tasks: the results related to Group 3's weighting strategy averaged over 6,000 trials in the continuation study showed the same general trends as the results from just the first 600 trials.

In addition to investigating which rule the groups actually used, we also searched for the best group rule among eleven different group rules. We found that the group rule with the best performance for a particular group of observers was largely dependent on the group members' characteristics, rather than some rules being intrinsically better than others. Notably, in a fair comparison, reasonable formulations of the Ideal Group [87] were not particularly ideal. Our results are directly relevant to visual Yes-No tasks where the stimulus is fixed, such as cancer detection in mammograms, and suggest that performing the task in a group may provide a sufficient increase in performance and robustness to warrant the added expense of the additional group members.

Further experimental data is required to confirm the results of our continuation

study, and to see to what extent our findings are applicable to sequential 2AFC tasks. Ultimately, we seek to experimentally confirm that our mathematical models can be applied to human observers and to find what experimental manipulations (if any) are required to do so.

Appendix A

Appendix for Chapter 1

A.1 Significance

In statistics, the term "significant" refers to the finding that a particular finding is likely to be true; in particular, it is unlikely to be a result of chance. It can also be defined as the probability that the test will not result in a false alarm (Type I error). The phrase "test of significance" was coined by R.A Fisher in 1925. Significance is quantified by the value $(1 - \alpha_0)$.

A.2 Power

In statistics, the "power" of a test is given by $(1 - \alpha_1)$, which is the probability that the test will not result in a miss (Type II error). In addition, there exists the notion a Uniformly Most Powerful (UMP) test. A UMP hypothesis test has the

greatest power $(1-\alpha_1)$ among all possible tests with a given α_0 (i.e., all equivalently

significant tests).

Appendix B

Appendix for Chapter 2

B.1 Wiener Process

A Wiener Process $W(t)$ on $[0, T]$ is a random variable depending continuously on $t \in [0, T]$ that satisfies

$$W(0) = 0, \tag{B.1}$$

$$\text{For } 0 \leq s < t \leq T, \quad W(t) - W(s) \sim \sqrt{t-s} N(0,1), \tag{B.2}$$

$$\text{For } 0 \leq s < t < u < v \leq T, \quad W(t) - W(s) \text{ is independent of } W(v) - W(u) \tag{B.3}$$

where $N(0,1)$ is a Normal (or Gaussian) distribution with zero mean and unit variance [36].

B.2   Ito Stochastic Differential Equation (SDE)

A stochastic quantity $x(t)$ obeys an Ito SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dW(t) \tag{B.4}$$

if for all t and t_0,

$$x(t) = x(t_0) + \int_{t_0}^{t} a[x(s), s]ds + \int_{t_0}^{t} b[x(s), s]dW(s). \tag{B.5}$$

Note that one can use the Cauchy-Euler procedure for constructing an approximate solution of the SDE. The solution is constructed by letting the mesh size approach zero.

There are two conditions required for the existence and uniqueness of solutions in the time interval $[t_0, T]$: the Lipschitz condition, which is a smoothness condition; and the growth condition, which ensures that the solution does not blow up too quickly [36]:

Lipschitz Condition: a K exists such that

$$|a(x, t) - a(\xi, t)| + |b(x, t) - b(\xi, t)| \leq K|x - \xi| \tag{B.6}$$

Growth Condition: a K exists such that for all $t \in t_0, T$,

$$|a(x, t)|^2 + |b(x, t)|^2 \leq K^2(1 + |x|^2).$$

B.3 Markov Property of the Solution to an Ito SDE

A Markov process is one in which knowledge of only the present determines the future. The Markov assumption is generally formulated in terms of conditional probabilities. For

$$t_1 \geq t_2 \geq \cdots \geq \tau_1 \geq \tau_2 \geq \cdots, \tag{B.7}$$

a Markov process's conditional probability is determined entirely by knowledge of the most recent condition, i.e., using the convention set in Equation (B.7),

$$p(\xi_1, t_1; \xi_2, t_2 \ldots | y_1, \tau_1; y_2, \tau_2; \ldots) = p(\xi_1, t_1; \xi_2, t_2; \ldots | y_1, \tau_1). \tag{B.8}$$

This is powerful and can simplify expressions. For example, by the definition of the conditional probability density,

$$p(\xi_1, t_1; \xi_2, t_2 | y_1, \tau_1) = p(\xi_1, t_1 | \xi_2, t_2; y_1, \tau_1) p(\xi_2, t_2 | y_1, \tau_1).$$

For a Markov process, we can simplify this to

$$p(\xi_1, t_1; \xi_2, t_2 | y_1, \tau_1) = p(\xi_1, t_1 | \xi_2, t_2) p(\xi_2, t_2 | y_1, \tau_1).$$

This generalizes to an ordered sequence of n times [36].

We now show that the continuous-time decision variable $x(t)$ is a Markov process. For a given initial condition $x(t_0)$, $x(t)$ is uniquely (stochastically) determined only by the particular sample path of $W(t)$ (for $t > t_0$) and the value of $x(t_0)$. Since $x(t)$ is a non-anticipating function of t, future values of $W(t)$ (for $t > t_0$) are independent of past values of $x(\tau)$, $(\tau < t_0)$. Thus, $x(t)$ for $t > t_0$ is independent of $x(\tau)$ for $\tau < t_0$ provided that $x(t_0)$ is known, and so $x(t)$ is a Markov process.

B.4 Ito's Formula

Ito's formula is useful because it provides a means to find the SDE for an arbitrary function of $x(t)$, $f[x(t)]$. To calculate this, we first expand $df[x(t)]$ to second order in $dW(t)$:

$$
\begin{aligned}
df[x(t)] &= f[x(t) + dx(t)] - f[x(t)], && \text{(B.9)} \\
&= f[x(t)] + f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)](dx(t))^2 + \cdots - f[x(t)], \\
&= f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)](dx(t))^2 + \cdots .
\end{aligned}
$$

Substituting in from Equation (B.4) and condensing notation,

$$
= f'[x(t)]\left\{adt + bdW\right\} + \frac{1}{2}f''[x(t)]\left\{a^2(dt)^2 + 2ab(dt)(dW) + b^2(dW)^2\right\}. \quad \text{(B.10)}
$$

Since $(dt)^2 << (dt)(dW) << (dW)^2 = dt$, the equation above simplifies to Ito's Formula [36]:

$$df[x(t)] = \left\{ a[x(t), t]f'[x(t)] + \frac{1}{2}(b[x(t), t])^2 f''[x(t)] \right\} dt + b[x(t), t]f'[x(t)]dW.$$

(B.11)

This demonstrates that for Ito SDEs, the change-of-variables formula is not given by ordinary calculus unless $f[x(t)]$ is linear in $x(t)$.

B.5   General Boundary Conditions

Suppose that at time $t = 0$, all of the probability of the location of the decision variable is concentrated at x_0, since we know that the decision variable is located there with probability 1. This is captured by the Dirac delta function initial condition. As time progresses, the location of the decision variable becomes increasingly uncertain due to noise. This corresponds to the probability distribution of its location spreading out in time, which is due to the diffusion term. Even though noise causes the probability distribution to spread out, the mean of the distribution should move towards the boundary corresponding to the correct decision. This corresponds to the drift term in the SDE. We assume that the decision is made in finite time, which means that all of the probability will eventually exit the system. Thus, from these assumptions, we choose absorbing boundary conditions for our
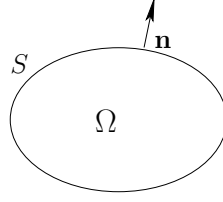
Figure B.1: The integral of the flux out of the system gives the change in probability in the system.

system.

Consider the general fFPE, where $\mathbf{x}$ is a vector and the coefficients a and b are generalized to the matrices $A(\mathbf{x}, t)$ and $B(\mathbf{x}, t)$:

$$\partial_t p(\mathbf{x}, t) = -\sum_i \frac{\partial}{\partial x_i} A_i(\mathbf{x}, t) p(\mathbf{x}, t) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} B_{ij}(\mathbf{x}, t) p(\mathbf{x}, t). \qquad \text{(B.12)}$$

We define the **probability current** as

$$J_i(\mathbf{x}, t) = A_i(\mathbf{x}, t) p(\mathbf{x}, t) - \frac{1}{2} \sum_j \frac{\partial}{\partial x_j} B_{ij}(\mathbf{x}, t) p(\mathbf{x}, t). \qquad \text{(B.13)}$$

Thus, we can write Equation (B.12) as

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} + \sum_i \frac{\partial}{\partial x_i} J_i(\mathbf{x}, t) = 0. \qquad \text{(B.14)}$$

This has the form of a local conservation equation [36]. The integral form of this equation can be used to express change in probability, as explained below.

B.5.1 Change in Probability

Consider a region Ω with boundary $\mathbf{S}$. We define the total probability in Ω, as shown in Figure B.1, to be

$$P(\Omega, t) = \int_\Omega p(\mathbf{x}, t) d\mathbf{x}. \tag{B.15}$$

Then we can substitute this into Equation (B.14) to get

$$\frac{\partial P(\Omega, t)}{\partial t} = -\int_{\mathbf{S}} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t) dS, \tag{B.16}$$

where $\mathbf{n}$ is the outward pointing normal to $\mathbf{S}$. Thus, the rate at which probability leaves Ω is given by the surface integral of $\mathbf{J}$ over the boundary $\mathbf{S}$ [36].

B.5.2 Flow of Probability Across a Surface

We also find that the current $\mathbf{J}$ has the stronger property that a surface integral over any surface $\mathbf{S}$ gives the net flow of probability across that surface.

Consider two adjacent regions Ω_1 and Ω_2, separated by surface S_{12}, as shown in Figure B.2. Region Ω_1 is enclosed by the surfaces S_1 and S_{12}, and region Ω_2 is enclosed by S_2 and S_{12}. We wish to know the net flow of probability from Ω_2 to Ω_1.

Since we are dealing with a process with continuous sample paths, the probability of crossing from Ω_2 to Ω_1 through S_{12} (in a sufficiently short period of time,

Δt) is given by the joint probability of being in Ω_2 at time t and in Ω_1 at time $t + \Delta t$,

$$\int_{\Omega_1} \int_{\Omega_2} p(\xi, t + \Delta t; \mathbf{x}, t) d\xi d\mathbf{x}. \tag{B.17}$$

The net flow of probability from Ω_2 to Ω_1 is then given by subtracting the flow in the opposite direction, dividing by Δt, and looking at the limit as $\Delta t \to 0$:

$$\Delta P = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\Omega_1} \int_{\Omega_2} \left(p(\mathbf{x}, t + \Delta t; \xi, t) - p(\xi, t + \Delta t; \mathbf{x}, t) \right) d\xi d\mathbf{x}. \tag{B.18}$$

We can Taylor expand this expression to get

$$\Delta P = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\Omega_1} \int_{\Omega_2} \left([p(\mathbf{x}, \tau; \xi, t) + \partial_\tau p(\mathbf{x}, \tau; \xi, t)\Delta t + \cdots] - \right.$$
$$\left. [p(\xi, \tau; \mathbf{x}, t) + \partial_\tau p(\xi, \tau; \mathbf{x}, t)\Delta t] \right)|_{\tau=t} d\xi d\mathbf{x}. \tag{B.19}$$

Since it is not possible to be in both Ω_1 and Ω_2 simultaneously,
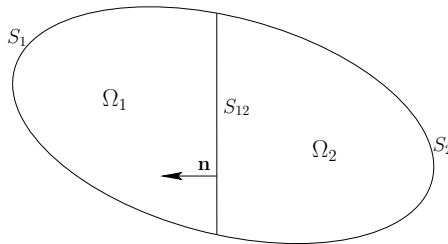
Figure B.2: This example system can be used to show that the integral of the flux over any surface gives the net flow of probability across that surface.

$$\int_{\Omega_1} \int_{\Omega_2} p(\mathbf{x}, t; \xi, t) d\xi d\mathbf{x} = 0,$$

$$\int_{\Omega_1} \int_{\Omega_2} p(\xi, t; \mathbf{x}, t) d\xi d\mathbf{x} = 0.$$

Then we have

$$\Delta P = \int_{\Omega_1} \int_{\Omega_2} \left[\partial_\tau p(\mathbf{x}, \tau; \xi, t) - \partial_\tau p(\xi, \tau; \mathbf{x}, t) \right]_{\tau=t} d\xi d\mathbf{x}. \tag{B.20}$$

We also note that

$$p(\mathbf{x}, t; \Omega_2, t) = \int_{\Omega_2} p(\mathbf{x}, t; \xi, t) d\xi \tag{B.21}$$

can be used to define a quantity $J_i(\mathbf{x}, t; \Omega_2, t)$, as it was done in Equation (B.13):

$$J_i(\mathbf{x}, t; \Omega_2, t) = A_i(x, t) p(\mathbf{x}, t; \Omega_2, t) - \frac{1}{2} \sum_j \frac{\partial}{\partial x_j} \left[B_{ij}(x, t) p(\mathbf{x}, t; \Omega_2, t) \right]. \tag{B.22}$$

A similar equation can be used to define $J_i(\xi, t; \Omega_1, t)$.

Now we substitute in the Fokker-Planck equation in the form in Equation (B.14),

$$\partial_t p(\mathbf{x}, t; \Omega_2, t) = - \sum_i \partial_{x_i} J_i(\mathbf{x}, t; \Omega_2, t),$$

to get

$$\Delta P = -\int_{\Omega_1} \sum_i \frac{\partial}{\partial x_i} J_i(\mathbf{x}, t; \Omega_2, t) d\mathbf{x} + \int_{\Omega_2} \sum_i \frac{\partial}{\partial \xi_i} J_i(\xi, t; \Omega_1, t) d\xi, \qquad \text{(B.23)}$$

and then convert the integrals to surface integrals:

$$
\begin{aligned}
\Delta P &= -\int_{S_1} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t; \Omega_1, t) d\mathbf{x} + \int_{S_{12}} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t; \Omega_1, t) d\mathbf{x} - \\
&\quad \int_{S_2} \mathbf{n} \cdot \mathbf{J}(\xi, t; \Omega_2, t) d\xi + \int_{S_{12}} \mathbf{n} \cdot \mathbf{J}(\xi, t; \Omega_2, t) d\xi. \qquad \text{(B.24)}
\end{aligned}
$$

The integrals over S_1 and S_2 vanish: the integral over S_1 involves $p(\mathbf{x}, t; \Omega_2)$ with $\mathbf{x} \notin \Omega_2$ or on its boundary except for a measure-zero set, and similarly for S_2.

Thus, we are left with the net flow from Ω_2 to Ω_1 as

$$\Delta P = \int_{S_{12}} \mathbf{n} \cdot [\mathbf{J}(\mathbf{x}, t; \Omega_1, t) + \mathbf{J}(\mathbf{x}, t; \Omega_2, t)] \, dS, \qquad \text{(B.25)}$$

and we can conclude that because $\mathbf{x} \in (\Omega_1 \cup \Omega_2)$, the net flow of probability per unit time from Ω_2 to Ω_1 is

$$
\begin{aligned}
\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\Omega_1} \int_{\Omega_2} &\left(p(\mathbf{x}, t + \Delta t; \xi, t) - p(\xi, t + \Delta t; \mathbf{x}, t) \right) d\xi d\mathbf{x} \\
&= \int_{S_{12}} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t) dS = \frac{\partial P}{\partial t}, \qquad \text{(B.26)}
\end{aligned}
$$

where $\mathbf{n}$ points from Ω_2 to Ω_1 [36].

B.5.3 Absorbing Boundary Conditions

Absorbing boundary conditions mean that when a particle reaches $\mathbf{S}$, it is removed from the system; hence the boundary absorbs. Therefore, the probability of being on the boundary is zero, and the boundary conditions are

$$p(\mathbf{x}, t) = 0 \quad \text{for } \mathbf{x} \in \mathbf{S}. \tag{B.27}$$

B.6 Sturm-Liouville Equation

The Sturm-Liouville Equation has the form

$$-\frac{d}{dx}\left[p(x)\frac{d}{dy}\right] + q(x)y = \lambda w(x)y \tag{B.28}$$

where y is a function of x, $p(x) > 0$, $q(x)$, and $w(x) > 0$ are specified at the outset. The function $w(x)$ is the weight function.

A general equation

$$P(x)y'' + Q(x)y' + R(x)y = 0 \tag{B.29}$$

can be converted into Sturm-Liouville form using the integrating factor

$$e^{\int \frac{Q(x)}{P(x)} dx}. \tag{B.30}$$

Appendix C

Appendix for Chapter 3

C.1   Proof of Equation (3.45)

We begin with the general cdf of GDTs for the Majority First scheme for N general DMs. Let $\Gamma^{0g}_{\mathcal{J}\mathbb{D}} \equiv 1$. Then Equation (3.43) becomes

$$q_g^{\mathrm{mfg}N}(t_g) = \sum_{\theta=0}^{\Theta} \left(\Gamma^{\theta g}_{\mathcal{J}\mathrm{S}} \prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathrm{S}} + \Gamma^{\theta g}_{\mathcal{J}\mathrm{N}} \prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathrm{N}} \right). \tag{C.1}$$

Taking the derivative with respect to t_g, we then get

$$p_g^{\mathrm{mfg}N}(t_g) = \sum_{\theta=0}^{\Theta} \left(\frac{d}{dt_g} \left[\Gamma^{\theta g}_{\mathcal{J}\mathrm{S}} \prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathrm{S}} \right] + \frac{d}{dt_g} \left[\Gamma^{\theta g}_{\mathcal{J}\mathrm{N}} \prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathrm{N}} \right] \right). \tag{C.2}$$

Since the S and N solutions are clearly related by symmetry and are non-interacting, for simplicity, consider the generic term

$$p_{g\mathbb{D}}^{\mathrm{mfg}N}(t_g) = \sum_{\theta=0}^{\Theta} \left(\frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{\theta\mathrm{g}} \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right] \right). \tag{C.3}$$

Clearly $p_{g\mathrm{S}}^{\mathrm{mfg}N}(t_g) + p_{g\mathrm{N}}^{\mathrm{mfg}N}(t_g) = p_{g}^{\mathrm{mfg}N}(t_g)$. If $p_{g\mathbb{D}}^{\mathrm{mfg}N}(t_g)$ can be simplified, it automatically follows that Equation (C.2) can be simplified. Now, note that by definition,

$$\frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{\theta g} \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right] = \frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{\theta g} \right] \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} + \Gamma_{\mathcal{J}\mathbb{D}}^{\theta g} \frac{d}{dt_g} \left[\prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right], \tag{C.4}$$

and that $(\Theta + 1)$ of these pairs appear during differentiation:

$$p_{g\mathbb{D}}^{\mathrm{mfg}N}(t_g) = \frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{0g} \right] \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} + \Gamma_{\mathcal{J}\mathbb{D}}^{0g} \frac{d}{dt_g} \left[\prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right] + \frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{1g} \right] \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} +$$
$$\Gamma_{\mathcal{J}\mathbb{D}}^{1g} \frac{d}{dt_g} \left[\prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right] + \cdots + \frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{\Theta g} \right] \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} + \Gamma_{\mathcal{J}\mathbb{D}}^{\Theta g} \frac{d}{dt_g} \left[\prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right]. \tag{C.5}$$

Now, the simplification step becomes significantly more straightforward if one can show that

$$-\frac{d}{dt_g} \left[\Gamma_{\mathcal{J}\mathbb{D}}^{\theta g} \right] \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} = \Gamma_{\mathcal{J}\mathbb{D}}^{(\theta-1)g} \frac{d}{dt_g} \left[\prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right], \tag{C.6}$$

since this leads to all of the terms canceling out except the first and last term. (Note that the second Gamma term has a different θ index from the first Gamma

term.)

On the left-hand side, we have

$$
- \sum_{j_1=1}^{N-\theta+1} \sum_{j_2=j_1+1}^{N-\theta+2} \cdots \sum_{j_\theta=j_{\theta-1}+1}^{N} \left(\frac{d}{dt_g} \left[\prod_{m=1}^{\theta} [1 - q_{j_m\mathbb{D}}] \right] \prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right). \qquad (C.7)
$$

There are θ terms in the square brackets, and $(N - \theta)$ terms in the right-hand product. After differentiating, the left-hand side becomes

$$
\sum_{j_1=1}^{N-\theta+1} \sum_{j_2=j_1+1}^{N-\theta+2} \cdots \sum_{j_\theta=j_{\theta-1}+1}^{N} \left[\sum_{m=1}^{\theta} \left(p_{j_m\mathbb{D}} \prod_{\substack{\ell=1, \\ \ell\neq m}}^{\theta} [1 - q_{j_\ell\mathbb{D}}] \right) \prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right]. \qquad (C.8)
$$

The summation terms outside of the square brackets define all unique subsets of size θ, and the terms in the parentheses iterate over all unique combinations of p and $[1 - q]$ within each subset. The product term on the right-hand side is made up of the remaining group members who were not in the subset.

Meanwhile, on the right-hand side, we have

$$
\sum_{j_1=1}^{N-\theta+2} \sum_{j_2=j_1+1}^{N-\theta+3} \cdots \sum_{j_{\theta-1}=j_{\theta-2}+1}^{N} \left(\prod_{m=1}^{\theta-1} [1 - q_{j_m\mathbb{D}}] \frac{d}{dt} \left[\prod_{\substack{k=1, \\ k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right] \right), \qquad (C.9)
$$

with $(\theta - 1)$ terms on the left, and $(N - \theta + 1)$ terms on the right. This expands to

$$\sum_{j_1=1}^{N-\theta+2} \sum_{j_2=j_1+1}^{N-\theta+3} \cdots \sum_{j_{\theta-1}=j_{\theta-2}+1}^{N} \left(\prod_{m=1}^{\theta-1}[1-q_{j_m\mathbb{D}}] \sum_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} p_{k\mathbb{D}} \prod_{\substack{\ell=1,\\\ell\notin\mathcal{J}\\\ell\neq k}}^{N} q_{\ell\mathbb{D}} \right). \qquad \text{(C.10)}$$

Note that there are $(\theta-1)$ terms in the left product and $(N-\theta)$ terms in the right product. We can rearrange this to

$$\sum_{j_1=1}^{N-\theta+2} \sum_{j_2=j_1+1}^{N-\theta+3} \cdots \sum_{j_{\theta-1}=j_{\theta-2}+1}^{N} \sum_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} \left(p_{k\mathbb{D}} \prod_{m=1}^{\theta-1}[1-q_{j_m\mathbb{D}}] \prod_{\substack{\ell=1,\\\ell\notin\mathcal{J}\\\ell\neq k}}^{N} q_{\ell\mathbb{D}} \right). \qquad \text{(C.11)}$$

This is close to the general form we desire. Next, note that the j subscripted summation signs on the left iterate over all unique subsets of size $(\theta-1)$, and the k subscripted summation sign iterates over the remaining choices. Together, the summations iterate over all unique subsets of size θ. Thus, we can rename the summation indices to

$$\sum_{j_1=1}^{N-\theta+1} \sum_{j_2=j_1+1}^{N-\theta+2} \cdots \sum_{j_\theta=j_{\theta-1}+1}^{N} \left[\sum_{m=1}^{\theta} \left(p_{j_m\mathbb{D}} \prod_{\substack{\ell=1,\\\ell\neq m}}^{\theta}[1-q_{j_\ell\mathbb{D}}] \right) \prod_{\substack{k=1,\\k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right], \qquad \text{(C.12)}$$

which is equal to Equation (C.8). It is now clear that Equation (C.6) holds. Thus, all of the terms in Equation (C.5) cancel, except for the first and last terms. However, since $\Gamma_{\mathcal{J}\mathbb{D}}^{0g} = 1$, the first term is zero. Thus, we have

$$p_{g\mathbb{D}}^{\mathrm{mfg}N}(t_g) = \Gamma_{\mathcal{J}\mathbb{D}}^{\Theta g} \frac{d}{dt_g} \left[\prod_{\substack{k=1,\\ k\notin\mathcal{J}}}^{N} q_{k\mathbb{D}} \right] = \Gamma_{\mathcal{J}\mathbb{D}}^{\Theta g} \sum_{\substack{k=1,\\ k\notin\mathcal{J}}}^{N} \left[p_{k\mathbb{D}} \prod_{\substack{m=1,\\ m\notin\mathcal{J},\\ m\neq k}}^{N} q_{m\mathbb{D}} \right], \qquad \text{(C.13)}$$

which expands to Equation (3.45) after adding the $\mathbb{D} = \mathrm{S}$ and $\mathbb{D} = \mathrm{N}$ cases. $\square$

Bibliography

[1] M. M. Al-Ibrahim. *On Distributed Sequential Hypothesis Testing.* PhD thesis, 1989.

[2] M. M. Al-Ibrahim and P. K. Varshney. A simple multi-sensor sequential detection procedure. In *Proc. 27th Conf. Decision and Control*, Austin, Texas, 1988.

[3] M. M. Al-Ibrahim and P. K. Varshney. A decentralized sequential test with data fusion. In *Proc. 1989 Amer. Control Conf.*, Pittsburgh, PA, 1989.

[4] R. J. Audley. A stochastic model for individual choice behavior. *Psych. Rev.*, 67(1):1–15, 1960.

[5] T. Bayes and R. Price. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Phil. Trans.*, 53:370–418, 1763.

[6] R. Ben-Yashar and J. Paroush. A nonasymptotic Condorcet jury theorem. *Soc. Choice Welf.*, 17:189–199, 2000.

[7] D. Berend and L. Sapir. Between the expert and majority rules. *Adv. in App. Probab.*, 35(4):941–960, 2003.

[8] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[9] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 1995.

[10] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1999.

[11] D. Black. On the rationale of group decision-making. *J. Political Economy*, 56(1):23–34, 1948.

[12] R. S. Blum, S. A. Kassam, and H. V. Poor. Distributed detection with multiple sensors: Part II - advanced topics. *Proc. IEEE*, 85(1):64–79, 1997.

[13] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. The physics of optimal decision-making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psych. Rev.*, 113:700–765, 2006.

[14] R. Bogacz, P. T. Hu, P. Holmes, and J. D. Cohen. Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Q. J. Exp. Psych.*, 63:863–891, 2010.

[15] W. M. Bolstad. *Introduction to Bayesian Statistics*. John Wiley & Sons, 2007.

[16] K. H. Britten, M. N. Shadlen, W. T. Newsome, and J. A. Movshon. Responses of neurons in macaque MT to stochastic motion signals. *Visual Neurosci.*, 10:1157–1169, 1993.

[17] E. Brown, J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat, and J. D. Cohen. Simple neural networks that optimize decisions. *Internat. J. Bifur. Chaos*, 15(3):803–826, 2005.

[18] A. E. Burgess and B. Colborne. Visual signal detection. IV. Observer inconsistency. *J. Opt. Soc. Am. A*, 5(4):617–627, 1988.

[19] Q. Cheng, P. K. Varshney, K. G. Mehrotra, and C. K. Mohan. Bandwidth management in distributed sequential detection. *IEEE Trans. Inform. Theory*, 51(8):2954–2961, 2005.

[20] T. H. Chung and J. W. Burdick. A decision-making framework for control strategies in probabilistic search. In *IEEE Intern. Conf. Robotics and Autom.*, Roma, Italy, 2007.

[21] N. C. Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix, Paris, 1785. (See McLean and Hewitt, translators, 1994).

[22] L. Conradt and T. J. Roper. Consensus decision making in animals. *TRENDS in Ecology and Evolution*, 20(8):449–456, 2005.

[23] B. J. Copeland. Colossus: Its origins and originators. *IEEE Ann. Hist. Comput.*, 26(4):38–45, 2004.

[24] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433:513–516, 2005.

[25] R. W. Crow and S. C. Schwartz. Quickest detection for sequential decentralized decision systems. *IEEE Trans. Aerosp. Electron. Sys.*, 32(1):267–283, 1996.

[26] J. H. Davis. Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950-1990. *Organ. Behav. and Human Decision Process.*, 52:3–38, 1992.

[27] J. L. Denebourg and S. Goss. Collective patterns and decision-making. *Ethology, Ecology & Evolution*, 1(295-311), 1989.

[28] A. R. Dennis. Information exchange and use in group decision making: You can lead a group to information, but you can't make it think. *Manag. Inform. Sys. Q.*, 20(4):433–457, 1996.

[29] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classfication.* John Wiley & Sons, New York, 2001.

[30] P. Eckhoff, P. Holmes, C. Law, P. M. Connolly, and J. I. Gold. On diffusion processes with variable drift rates as models for decision making during learning. *New J. Phys.*, 10(015006):1–27, 2008.

[31] B. Eisenberg. *Multihypothesis Problems*, chapter 9, pages 229–244. Marcel Dekker, Inc., New York, 1992.

[32] B. Eisenberg, B. K. Ghosh, and G. Simons. Properties of generalized sequential probability ratio tests. *Ann. Statist.*, 4(2):237–251, 1976.

[33] G. Ferrari and R. Pagliari. Decentralized binary detection with noisy communication links. *IEEE Trans. Aerosp. Electron. Sys.*, 42(4):1554–1563, 2006.

[34] V. P. Ferrera, M. Yanike, and C. Cassanello. Frontal Eye Field neurons signal changes in decision criteria. *Nat. Neurosci.*, 12(11):1454–1458, 2009.

[35] W. L. Francis. Legislative committee systems, optimal committee size, and the costs of decision making. *J. Politics*, 44:822–837, 1982.

[36] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer-Verlag, 2004.

[37] B. K. Ghosh. *Sequential Tests of Statistical Hypotheses*. Addison-Wesley, Reading, Massachusetts, 1970.

[38] B. K. Ghosh. *A Brief History of Sequential Analysis*, chapter 1, pages 1–19. Marcel Dekker, Inc., New York, 1992.

[39] J. I. Gold and M. N. Shadlen. Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36:299–308, 2002.

[40] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, Inc., New York, 1966.

[41] W. H. Greene. *Econometric Analysis*. Pearson Education, Inc., 2003.

[42] N. Harvey and I. Fischer. Taking advice: Accepting help, improving judgement, and sharing responsibility. *Organ. Behav. and Human Decision Proc.*, 70(2):117–133, 1997.

[43] R. Hastie and T. Kameda. The robust beauty of majority rules in group decisions. *Psych. Rev.*, 112(2):494–508, 2005.

[44] P. Holmes, E. Shea-Brown, J. Moehlis, R. Bogacz, J. Gao, G. Aston-Jones, E. Clayton, J. Rajkowski, and J. Cohen. Optimal decisions: From neural spikes, through stochastic differential equations, to behavior. *IEICE Trans. Fundam.*, E88-A(10):2496–2503, 2005.

[45] A. M. Hussain. Multisensor distributed sequential detection. *IEEE Trans. Aerosp. Electron. Sys.*, 30(3):698–708, 1994.

[46] D. Karotkin. The network of weighted majority rules and weighted majority games. *Games Econom. Behav.*, 22:299–315, 1998.

[47] D. Karotkin and J. Paroush. Optimum committee size: Quality-versus-quantity dilemma. *Soc. Choice Welf.*, 20:429–441, 2003.

[48] M. Kimura and J. Moehlis. Group decision-making models. Accepted to the SIAM Review.

[49] M. W. Koch and K. T. Malone. *Vehicle Classification in Infrared Video Using the Sequential Probability Ratio Test*, pages 63–84. Springer, London, 2009.

[50] O. P. Kriedl and A. S. Willsky. An efficient message-passing algorithm for optimizing decentralized detection networks. In *IEEE Conf. Decision and Control*, 2006.

[51] A. LaVigna, A. M. Makowski, and J. S. Baras. *A Continuous-Time Distributed Verson of Wald's Sequential Hypothesis Testing Problem*, pages 533–543. Springer-Verlag, 1986.

[52] C. C. Lee and J. B. Thomas. A modified sequential detection procedure. *IEEE Trans. Inform. Theory*, 30(1):16–23, 1984.

[53] M. D. Lee and M. J. Paradowski. Group decision-making on an optimal stopping problem. *J. Prob. Solv.*, 1, 2007.

[54] E. L. Lehmann. *Testing Statistical Hypotheses*. John Wiley & Sons, New York, 1959.

[55] F. P. Leite and R. Ratcliff. Modeling reaction time and accuracy of multiple-alternatives decisions. *Atten. Percept. Psych.*, 72(1):246–273, 2010.

[56] X. R. Li, Y. Zhu, J. Wang, and C. Han. Optimal linear estimation fusion - Part I: Unified fusion rules. *IEEE Trans. Inform. Theory*, 49(9):2192–2208, 2003.

[57] K. Liu and A. M. Sayeed. Type-based decentralized detection in wireless sensor networks. *IEEE Trans. Signal Processing*, 55(5):1899–1910, 2007.

[58] S. Luan, R. D. Sorkin, and J. Itzkowitz. Weighting information from outside sources: A biased process. *J. Behav. Decision Making*, 17:95–116, 2004.

[59] N. A Macmillan and C. D. Creelman. *Detection Theory: A User's Guide.* Cambridge University Press, New York, 1991.

[60] D. P. Malladi and J. L. Speyer. A generalized Shiryayev sequential probability ratio test for change detection and isolation. In *Proc. 35th Conf. Decision and Control*, pages 3115–3122, Kobe, Japan, 1996.

[61] J. A. R. Marshall, R. Bogacz, A. Dornhaus, R. Planque, T. Kovacs, and N. R. Franks. Optimal decision making in brains and social insect colonies. *J. R. Soc. Interface*, 6(40):1065–1074, 2009.

[62] J. Neyman and E. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. London*, A(231):289–337, 1933.

[63] S. Nitzan and J. Paroush. Optimal decision rules in uncertain dichotomous choice situations. *Internat. Econ. Rev.*, 23(2):289–297, 1982.

[64] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim. Sequential signal encoding from noisy measurements using quantizers with dynamic bias control. *IEEE Trans. Inform. Theory*, 47(3):978–982, 2001.

[65] J. D. Papastavrou. A distributed hypothesis-testing team decision problem with communications cost. In *Proc. 25th Conf. Decision and Control*, Athens, Greece, 1986.

[66] W. W. Peterson, T. G. Birdsall, and W. C. Fox. The theory of signal detectability. *Trans. IRE Profession Group on Info. Theory*, PGIT(4):171–212, 1954.

[67] R. Radner. Team decision problems. *Ann. Math. Stat.*, 33:857–881, 1982.

[68] R. Ratcliff. A theory of memory retrieval. *Psych. Rev.*, 85(2):59–108, 1978.

[69] R. Ratcliff. Group reaction time distributions and an analysis of distribution statistics. *Psych. Bull.*, 86(3):446–461, 1979.

[70] R. Ratcliff. Modeling aging effects on two-choice tasks: Response signal and response time data. *Psych. and Aging*, 23(4):900–916, 2008.

[71] R. Ratcliff, A. Cherian, and M. Segraves. A comparison of macaque behavior

and superior colliculus neuronal activity to predictions from models of two-choice decisions. *J. Neurophys.*, 90:1392–1407, 2003.

[72] R. Ratcliff and H. P. A. Van Dongen. Sleep deprivation affects multiple distinct cognitive processes. *Psych. Bull. Rev.*, 16(4):742–751, 2009.

[73] R. Ratcliff and M. J. Hacker. Speed and accuracy of same and different responses in perceptual matching. *Percep. Psych.*, 30(3):303–307, 1981.

[74] R. Ratcliff, Y. T. Hasegawa, R. P. Hasegawa, P. L. Smith, and M. A. Seagraves. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J. Neurophys.*, 97:1756–1774, 2007.

[75] R. Ratcliff and G. McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.*, 20:873–922, 2008.

[76] R. Ratcliff and J. N. Rouder. Modeling response times for two-choice decisions. *Psych. Sci.*, 9(5):347–356, 1998.

[77] R. Ratcliff, F. Schmiedek, and G. McKoon. A diffusion model explanation of the worse performance rule for reaction time and IQ. *Intelligence*, 36:10–17, 2008.

[78] R. Ratcliff and P. L. Smith. A comparison of sequential sampling models for two-choice reaction time. *Psych. Rev.*, 111(2):333–367, 2004.

[79] V. N. S. Samarasooriya and P. K. Varshney. Sequential approach to asynchronous decision fusion. *Opt. Eng.*, 35(3):625–633, 1996.

[80] S. Sawyer. Resampling data: Using a statistical jackknife. http://www.math.wustl.edu/~sawyer/handouts/jackknife.pdf, 2005.

[81] N. Schmitz. *Optimal Sequentially Planned Decision Procedures*. Springer-Verlag, New York, 1992.

[82] N. J. Schmitz. *Sequential Detection Theory*, chapter 17, pages 407–428. Marcel Dekker, Inc., New York, 1992.

[83] P. Simen, J. D. Cohen, and P. Holmes. Rapid decision threshold modulation by reward rate in a neural network. *Neural Netw.*, 19(8):1013–1026, 2006.

[84] P. Simen, D. Contreras, C. Buck, P. Hu, P. Holmes, and J. D. Cohen. Reward rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *J. Exp. Psych.: Human Percep. Perform.*, 35(6):1865–1897, 2009.

[85] P. L. Smith. Stochastic dynamic models of response time and accuracy: A foundational primer. *J. Math. Psych.*, 44:408–463, 2000.

[86] R. D. Sorkin and H. Dai. Signal detection analysis of the ideal group. *Org. Behav. Human Decision Process.*, 60:1–13, 1994.

[87] R. D. Sorkin, C. J. Hays, and R. West. Signal-detection analysis of group decision making. *Psych. Rev.*, 108(1):183–203, 2001.

[88] R. D. Sorkin, R. West, and D. E. Robinson. Group performance depends on the majority rule. *Psych. Sci.*, 9(6):456–463, 1998.

[89] R. D. Sorkin and D. D. Woods. Systems with human monitors: A signal detection analysis. *Human-Comput. Interact.*, 1:49–75, 1985.

[90] G. Stasser and W. Titus. Pooling of unshared information in group decision making: Biased information sampling during discussion. *J. Personality Soc. Psych.*, 48(6):1467–1478, 1985.

[91] D. Teneketzis and Y. C. Ho. The decentralized Wald problem. *Inform. and Comput.*, 73:23–44, 1987.

[92] J. N. Tsitsiklis. On threshold rules in decentralized detection. In *Proc. 25th IEEE Conf. Decision and Control*, Athens, Greece, 1986.

[93] J. N. Tsitsiklis. Decentralized detection. *Adv. Stat. Sig. Process.*, 2, 1989.

[94] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer-Verlag, New York, 1997.

[95] V. V. Veeravalli. Sequential decision fusion: theory and applications. *J. Franklin Inst.*, 336:301–322, 1999.

[96] V. V. Veeravalli, T. Basar, and H. V. Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Trans. Inform. Theory*, 39(2):433–442, 1993.

[97] R. Viswanthan and P. K. Varshney. Distributed detection with multiple sensors: Part I - fundamentals. *Proc. IEEE*, 85(1):54–63, 1997.

[98] A. Wald. *Sequential Analysis.* John Wiley & Sons, New York, 1947.

[99] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Ann. Math. Statist.*, 19:326–339, 1948.

[100] W. Allen Wallis. The statistical research group, 1942-1945. *J. Amer. Statist. Assoc.*, 75(370):320–330, 1980.

[101] C. White, R. Ratcliff, M. Vasey, and G. McKoon. Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition and Emotion*, 23(1):181–205, 2009.

[102] J. R. Winquist and J. R. Larson Jr. Information pooling: When it impacts group decision making. *J. Person. and Soc. Psych.*, 74(2):371–377, 1998.

[103] I. Yaniv. The benefits of additional opinions. *Curr. Direc. Psych. Sci.*, 13(2):75–78, 2004.

[104] I. Yaniv and E. Kleinberger. Advice taking in decision making: Egocentric

[96] V. V. Veeravalli, T. Basar, and H. V. Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Trans. Inform. Theory*, 39(2):433–442, 1993.

[97] R. Viswanthan and P. K. Varshney. Distributed detection with multiple sensors: Part I - fundamentals. *Proc. IEEE*, 85(1):54–63, 1997.

[98] A. Wald. *Sequential Analysis.* John Wiley & Sons, New York, 1947.

[99] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Ann. Math. Statist.*, 19:326–339, 1948.

[100] W. Allen Wallis. The statistical research group, 1942-1945. *J. Amer. Statist. Assoc.*, 75(370):320–330, 1980.

[101] C. White, R. Ratcliff, M. Vasey, and G. McKoon. Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition and Emotion*, 23(1):181–205, 2009.

[102] J. R. Winquist and J. R. Larson Jr. Information pooling: When it impacts group decision making. *J. Person. and Soc. Psych.*, 74(2):371–377, 1998.

[103] I. Yaniv. The benefits of additional opinions. *Curr. Direc. Psych. Sci.*, 13(2):75–78, 2004.

[104] I. Yaniv and E. Kleinberger. Advice taking in decision making: Egocentric

discounting and reputation formation. *Org. Behav. and Human Decision Proc.*, 83(2):260–281, 2000.

[105] L. J. Young. Computation of some exact properties of Wald's SPRT when sampling from a class of discrete distributions. *Biom. J.*, 36(5):627–637, 1994.