Assessing the Performance of Data Fusion Algorithms Using Human Response Models

A Thesis

Submitted to the Faculty

of

Drexel University

by

Donald J. Bucci

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

February 2015

Dedications

To Korban, Grace, and my sweet Chelsea.

Acknowledgments

My time as a graduate student has been wonderful and thought-provoking. I have been privileged to work with many intelligent and insightful individuals who contributed to what I consider the best six years of my life.

I would first like to thank my advisor, Dr. Moshe Kam. Moshe has been an incredible mentor during my time in the Data Fusion Lab. I am very grateful for the time he has invested in helping me formulate and complete my thesis research. More importantly, Moshe has demonstrated to me what it means to be a working professional. Whenever I am solving problems, writing papers, or giving presentations I constantly find myself thinking, "How would Moshe approach this?" His wisdom is something that I will take with me through the rest of my career.

I am also very grateful for the counsel provided by Dr. Timothy Pleskac and his student Shuli Yu. Both Dr. Pleskac and Shuli made themselves available to me to answer questions regarding two-stage dynamic signal detection. All of the cognitive psychology models used in this thesis were graciously provided by them in the form of published and unpublished data. Without their contributions, this work would not have been possible.

There were many Drexel University faculty members that contributed to my professional development. I would like to thank the members of my dissertation committee: Drs. Thomas Chmielewski, Leonid Hrebien, Pramod Abichandani, Spiros Mancoridis, and Matthew Stamm. It has also been a privilege to work with Drs. Richard Primerano and Eli Fromm on the Drexel Freshman Engineering Design curriculum and with Dr. Kapil Dandekar on a wireless communications research project for the Office of Naval Research.

Moshe and I were able to use a subset of my thesis research to co-write a white paper for the Army Research Lab and a grant proposal for the National Science Foundation. On both of these efforts, we were very fortunate to collaborate with Dr. Pramod Varshney and his student Aditya Vempaty from Syracuse University. Working with Dr. Varshney and Aditya has been a wonderful

experience, and I am very thankful for their comments and insights.

My lab-mates in the Data Fusion Laboratory have been instrumental to my professional and personal growth. In particular, I would like to thank Ryan Measel, Christopher Lester, Sayandeep Acharya, Raymond Canzanese, Bradford Boyle, Marco Janko, George Sworo, Feiyu Xiong, Jeffrey Wildman, Gabriel Ford, and David Dorsey. I have been very fortunate to work with these individuals, and I consider them good friends.

Finally, I would not be here without the love and support of my family. This includes my wife Chelsea, my mother-in-law Sharon, my mother Dorothy, and my father Donald. All of you have invested so much in me during these past six years while I pursued my dreams. Thank you Chelsea, for loving and supporting me unconditionally. Thank you Dad, for all of the sacrifices you made for us. There are simply no words to fully express my gratitude.

Table of Contents

List of Tables

List of Figures

List of Symbols

Set Theoretic and Probabilistic Terminology

$\overline{A}$	The compliment (or negation) of a set, denoted here as A.
$A \cup B$	The union of two sets, denoted here as A and B.
$A \cap B$	The intersection of two sets, denoted here as A and B.
$\mathcal{F}_\Omega$	A set algebra over Ω, defining all possible unions and complements of the elements $\omega \in \Omega$.
$\emptyset$	The empty set.
2^Ω	The power set of Ω, defined as the set of all subsets of Ω including the empty set $\emptyset$.
$\|A\|$	The cardinality of a set, denoted here as A. Since the sets here are assumed finite, it denotes the number of elements contained by the set.
$\mathcal{F}_\Omega$	A set algebra over Ω, defining all possible unions and complements of the elements $\omega \in \Omega$.
$a \in A$	Read as "the element a is contained in the set A."
$A \subseteq B$	Read as "the set A is a subset or equal to B."
$A \subset B$	Read as "the set A is a subset of B."
$\mathbb{R}$	Set of all real numbers.
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2.
$\mathcal{U}(a, b)$	Uniform distribution defined on the interval $[a, b]$.

Data Fusion Terminology

Ω	A finite set of disjoint alternatives or states that a phenomenon can exist in.
M	The number of alternatives denoted by Ω (*i.e.*, $M = \|\Omega\|$).
N	The number of information sources being combined by a data fusion operator.
$P_\Omega(\cdot)$	A probability measure defined over the set Ω. This notation is used when describing statistical probabilities (*i.e.*, propensities of occurrence).
$\mathcal{P}_\Omega(\cdot)$	A subjective probability assignment defined over the set Ω. This notation is used when describing epistemic probabilities (*e.g.*, decision confidence assessments).
u_i	A decision generated by the i^{th} source amongst a set of sources.

u_0	A global decision, generated by a data fusion center.
u^*	The correct alternative amongst Ω.
p_{F_i}	Probability of false alarm for the i^{th} source amongst a set of sources.
p_{M_i}	Probability of missed detection for the i^{th} source amongst a set of sources.
p_{D_i}	Probability of correct detection for the i^{th} source amongst a set of sources.
p_{E_i}	Total classification error for the i^{th} source amongst a set of sources.
p_{F_0}	Probability of false alarm associated with the decision rule used at the data fusion center.
p_{M_0}	Probability of missed detection associated with the decision rule used at the data fusion center.
p_{D_0}	Probability of correct detection associated with the decision rule used at the data fusion center.
p_{E_0}	Total classification error associated with the decision rule used at the data fusion center.
$m_\Omega(\cdot)$	A belief mass assignment defined over the power set of Ω.
$\mathrm{Bel}(\cdot)$	A Dempster-Shafer theory belief function.
$\mathrm{Pl}(\cdot)$	A Dempster-Shafer theory plausibility function.
$\mathrm{Un}(\cdot)$	A Dempster-Shafer theory uncertainty function.
$\mathcal{C}$	Core of a belief function according to a belief mass assignment m over Ω.
$\mathcal{K}$	Measure of conflict after conjunctively combining a set of belief mass assignments.
$\xi(u, p)$	The evidence strength a subject's decision u and decision confidence assessment p.
$\overline{\xi}$	Sample average of a subject's observed evidence strengths.
$\xi'(A, p)$	The imprecise evidence strength of a subject's decision A and decision confidence assessment p_A.
$\overline{\xi'}$	Sample average of a subject's observed imprecise evidence strengths.
ξ_{Bel}	Lower bound of the evidence strength interval when using fusion operators from Dempster-Shafer theory.
ξ_{Pl}	Upper bound of the evidence strength interval when using fusion operators from Dempster-Shafer theory.
α_i	Discount rate for the i^{th} subject's belief mass or subjective probability assignment.

Two-Stage Dynamic Signal Detection Terminology

$\mathcal{A}$	The set of alternatives defining a two-alternative forced-choice task.
t	Elapsed time (in seconds).
$L(t)$	A stochastic process representing an accumulation of evidence that occurs during human deliberation.
$\Delta L(t)$	Discrete increment of $L(t)$ when modeled using a random walk approximation of a Wiener drift diffusion process.
Δt	Discrete time step (in seconds) used in a random walk approximation of a Wiener drift diffusion process.
δ	Drift rate of a drift diffusion process.
L_0	The initial condition of the stochastic difference equation for computing $L(t)$ (*i.e.*, $L_0 = L(0)$).
θ_A	Decision threshold on $L(t)$ for choosing alternative $A \in \mathcal{A}$.
$\epsilon(\cdot)$	White noise process having variance σ^2.
σ	Drift coefficient of a drift diffusion process.
a	2DSD declared decision (simulated).
p	2DSD declared decision confidence (simulated).
t_d	Elapsed time between the start of deliberation and decision declaration.
t_c	Elapsed time between the start of deliberation and decision confidence declaration.
τ	Time between the point of decision and decision confidence declaration (*i.e.*, inter-judgment time).
$\mathbf{P}^{(a)}$	Set of K_a possible decision confidence assessment values associated with the simulated 2DSD decision $a \in \mathcal{A}$.
$\mathbf{C}^{(a)}$	Set of $K_a - 1$ bin values that are used to map the value of $L(t_c)$ into a decision confidence assessment when the decision $a \in \mathcal{A}$ is produced by a 2DSD simulation.
$\mathcal{S}_{OST}$	Base parameters of the 2DSD optional stopping model. These parameters are estimated per subject based on experimentally determined decision, confidence assessment, and response time statistics.
$\mathcal{S}_{INT}$	Base parameters of the 2DSD interrogation model. These parameters are estimated per subject based on experimentally determined decision, confidence assessment, and response time statistics.
t_{ED}	Mean non-decision time. Represents the average motor time exhibited by a subject when making a decision.

t_{EJ} Mean non-judgment time. Represents the average motor time exhibited by a subject when assessing their decision confidence.

ν Optional 2DSD parameter. Specifies the mean drift rate when assuming that the 2DSD model drift rate δ is normally distributed per simulation.

η Optional 2DSD parameter. Specifies the drift rate standard deviation when assuming that the 2DSD model drift rate δ is normally distributed for a given simulation.

z Optional 2DSD parameter. Specifies the center of an interval in which the 2DSD model initial condition L_0 may be uniformly distributed for a given simulation.

s_z Optional 2DSD parameter. Specifies the size of an interval in which the 2DSD model initial condition L_0 may be uniformly distributed for a given simulation.

γ Optional 2DSD parameter. Specifies the rate of evidence decay in the 2DSD evidence accumulation process.

ρ Optional 2DSD parameter. Specifies the fractional amount that the 2DSD drift rate should be attenuated to during the interjudgment period.

Pairwise-Successive Aggregation Terminology

$\mathcal{Z}$ A set of unique stimuli associated with a class of M-ary decision tasks that involve choosing the most apparent stimulus amongst a set of M stimuli.

z_i The i^{th} stimuli of $\mathcal{Z}$, where $i = \{1, \ldots, M\}$.

z_{i^*} The most apparent stimuli of those defined by $\mathcal{Z}$.

l The imprecision level of the decision generated by the pairwise-successive aggregation method (*i.e.*, the cardinality of the generated decision, where $l \in \{1, \ldots, M-1\}$).

Abstract

Assessing the Performance of Data Fusion Algorithms Using Human Response Models
Donald J. Bucci
Moshe Kam

There is ongoing interest in designing data fusion systems that make use of human opinions (*i.e.*, "soft" data) alongside readings from various sensors that use mechanical, electromagnetic, optical, and acoustic transducers (*i.e.*, "hard" data). One of the major challenges in the development of these hard/soft fusion systems is to determine accurately and flexibly the impact of human responses on the performance of the fusion operator. Examples, counterexamples, and thought experiments have been used to illustrate specific performance aspects of soft and hard/soft fusion operators. However, these methods do not provide a systematic means for calculating performance statistics. Data sets of human responses developed through human testing have also been used. However, results obtained in this manner were often hard to generalize and difficult to tune up due to the experimental and administrative limitations imposed on human testing. Models of human decision making and confidence assessment from cognitive psychology offer a unique opportunity to assess the performance of fusion systems which make use of human opinions. Such models, which can programmed and modified readily and inexpensively, can be used to assess the performance of hard/soft fusion systems and to make credit assignments to the multiple sources of data that lead to the final estimate/decision of the fusion architecture.

The main contribution of this thesis is a set of algorithms for determining the performance of various soft and hard/soft fusion operators using existing models of human decision-making. In the first part of the thesis, we discuss the current state of the art and introduce applicable techniques for hard/soft fusion. In the second part of the thesis, we present techniques and examples for determining the performance of a family of soft and hard/soft fusion operators on a set of binary and multihypothesis tasks. We also present a method for simulating "vague" human responses, and correspondingly assess the statistical performance of soft fusion systems that have been proposed for combining sources of such imprecise information.

Chapter 1: Introduction

1.1 Context and Motivation

Data fusion is a broad field of research on combining data from multiple, often heterogeneous, information sources which report on the same underlying phenomenon. Often the purpose of data fusion is to develop improved decisions and/or parameter estimates compared to what could be achieved by the use of a single information source.

Data fusion has been used in a wide variety of fields and applications. For example, the defense research community has employed data fusion to improve the detection and tracking of objects in oceanic and battlefield surveillance, air-to-air defense, and strategic warning systems [1]. Non-defense related applications were demonstrated in air traffic control [2], forensics [3], medical imaging and diagnostics [4], robotics [5], and automotive navigation systems [6].

Most traditional data fusion applications make use of data from electronic, optical, or mechanical sensors. Such data tend to be relatively well structured and exhibit fixed and readily modeled sensor-error characteristics. These are often called "hard sources." There is an increasing interest in designing fusion systems that also make use of opinions from human decision makers (*i.e.*, "soft sources") [7]. These human opinions can be combined by themselves (*i.e.*, "soft fusion") [8–16]) or integrated with "hard" sensors. In the latter case we speak of "hard/soft fusion" [17–22]. Such "soft sources" can consist of the direct opinions of experts, the opinion of a crowd, or information obtained from the internet or social media. They are characterized as having highly unstructured data representations, and their governing statistics may be very complex and affected by many environmental factors [20]. Figure 1.1 is an attempt to depict hard and soft data fusion in a block diagram.

The inclusion of human responses in a data fusion system is becoming a popular research topic across a variety of applications. The initial need for structured research into the problem of hard and soft fusion has risen over the last ten years out of the U.S. Army's needs in asymmetric urban warfare

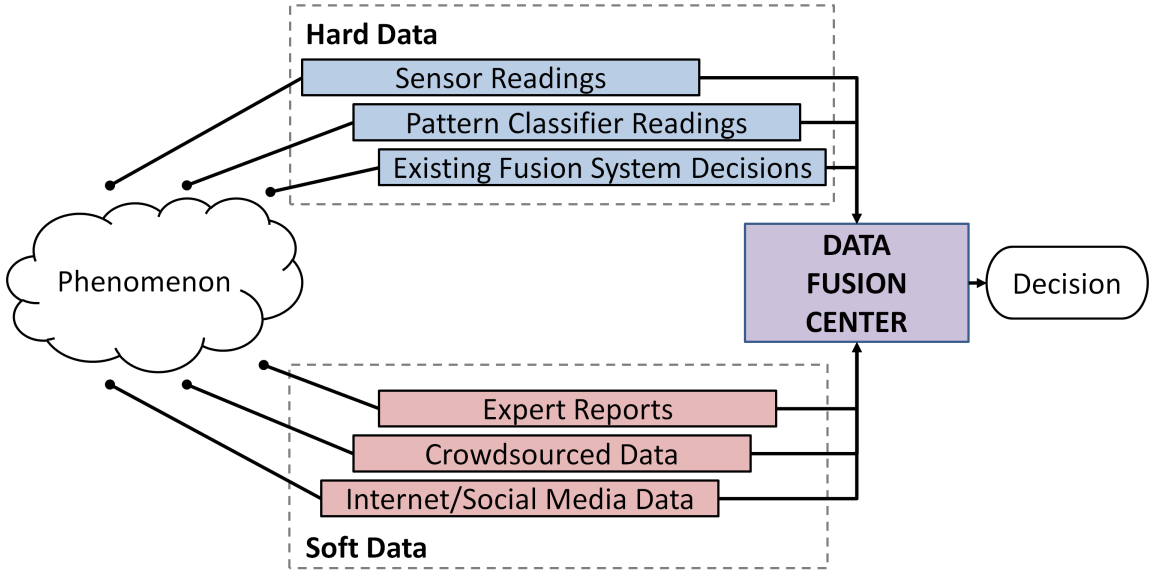


Figure 1.1: Hard and soft data fusion, conceptual block diagram.

[7]. For example, counter-insurgency (COIN) operations have been known to rely on a wide spectrum of information support sources including radio frequency sensors and imaging systems, along with reports from soldiers and intelligence analysts. In this setting, a hard and soft fusion system could be used to consolidate and combine relevant data in order to facilitate easier and improved decision-making by analysts and warfighters [23]. Deeper research into hard and soft data fusion however, has the potential for much broader impact [24]. The area of medical diagnostics, for example, makes use of information from lab tests and medical imagery in addition to patient reports and the opinions of doctors. These sources are integrated to arrive at an evaluation of some underlying pathology that, which it is hoped, is accurate and actionable. Emergency response applications are marked by a large influx of both hard and soft information (*e.g.*, geospatial data acquired from satellite imagery; human reports through social media). There are also many applications of hard and soft fusion in criminal justice systems, where the end goal may be to determine the identity of a person of interest given surveillance footage and fixed evidence from a crime scene, in addition to eyewitness accounts and expert testimonies.

1.2 Performance of Hard/Soft Fusion Systems

One of the challenges facing the development of state-of-the-art soft and hard/soft fusion systems is how one should to accurately and systematically evaluate their performance. A recent state-of-the-art survey of multisensor data fusion suggested that the role of soft data has not yet been extensively researched by the fusion community [25]. This apparent lack of research may be related to the difficulties in representing and characterizing human opinions statistically in practical soft and hard/soft fusion system applications [7]. In most studies, the performance of a given soft or hard/soft fusion system is evaluated through the use of examples, counterexamples, and thought experiments (*e.g.*, [26–35]), or through direct human testing on standardized data sets (*e.g.*, [13–15, 21, 36]). Although the use of examples and counterexamples can illustrate specific performance aspects of a data fusion operator, this approach does not lend itself towards wider applicability, nor does it allow the prediction of performance in hitherto-untested circumstances. Direct human testing and the use of predetermined data sets can provide estimates of average performance for specific fusion problems. However this approach often depends on the specific experiment (*i.e.*, human subjects; data exposition schemes). Furthermore, the repeatability of the experiments may prove difficult to achieve. This limitation is significant when a fusion system designer is interested in repeating an experiment with the same humans many times, with only a small modification of parameters between experiments; this is how sensitivity and record trends are often studied. Employing a large number of humans for testing also tends to be logistically cumbersome, making opportunities to re-test the same humans on modified data presentations for extensive periods of times impractical.

1.3 Probabilistic Human Response Models and Data Fusion

Probabilistic models of human decision making and confidence assessment from cognitive psychology offer an opportunity to assess the performance of fusion operators which make use of human opinions. However, little work has been conducted on using models of human decision-making and confidence assessment in this manner. For example, human decision-making probability models have been used to improve the performance of human-in-the-loop systems, where the human acts as the "data fusion

center" and controls system-wide parameters in response to evolving performance metrics. Models of human decision-making have been used in such systems to estimate strategies employed by the human element for the purposes of optimizing human-system interactions [37–40]. After performing a literature survey, we did not discover any soft or hard/soft fusion studies that made specific use of probabilistic models of human decision-making from cognitive psychology. Such models may need further validation with human decision-makers, but a lot of design, parameter selection, tuning, as well as stability and convergence analyses can be significantly accelerated and made more efficient by using them. They appear more promising than human testing or, as often is the case at the present time, reliance on researcher intuition.

1.4 Main Contributions

This dissertation presents methods for assessing the statistical performance of several soft and hard/soft fusion operators using models of human responses from cognitive psychology. Using data available in the cognitive psychology literature, we showcase how one can estimate the statistical performance (*e.g.*, error rates and confidence assessment accuracies) of an arbitrary collection fusion operators frequently used in soft or hard/soft data fusion. These operators come from detection theory [41], Bayesian Epistemology [42], and Dempster-Shafer Theory [43]. We also propose methods for extending human response simulation techniques in order to assess the performance of multihypothesis and/or vague[1] fusion operators. The main contributions of this dissertation are as follows.

- This work is one of the first to use probabilistic models of human decision-making, confidence assessment, and response time from cognitive psychology in assessing the performance of automated soft and hard/soft data fusion systems.

- This work is one of the first to associate meaningful statistical performance with data fusion operators from Dempster-Shafer theory using human decision-makers.

- This work explains the roles of decision confidence assessments and reliabilities of those as-

[1] For more information on how we define "vague" decisions, see Section 2.1.

sessments in the context of soft and hard/soft fusion systems.

- This work addresses the utility of the probability information content (PIC) metric in assessing the performance of a probability transformation[2].
- This work has used models of human responses to showcase the effects of vague evidence in terms of realistic fusion system performance metrics.
- This work is one of the first to implement a fully simulated hard and soft fusion system using models of human-decision making and confidence assessment from the cognitive psychology literature.

1.5 Organization of the Dissertation

This dissertation is organized into two parts and an epilogue. The first part consists of Chapters 2 and 3. The second part consists of Chapters 4 through 6. The final two chapters encompass the epilogue.

- **Part I: An Introduction to Data Fusion Tools and Human Response Simulation**. This part contains background material for this thesis. **Chapter 2** reviews of a selection of information fusion techniques used in soft and hard/soft fusion systems. **Chapter 3** presents a brief history and background on human response probability models, with a focus on *two-stage dynamic signal detection (2DSD)* [44, 45]. We also discuss a selection of decision tasks from [44, 45] that were used in this thesis to study the performance of several data fusion operators.
- **Part II: Performance Assessment Techniques for Hard/Soft Fusion Systems**. This part uses 2DSD models from [44, 45] to simulate the statistical performance of several fusion operators. **Chapter 4** studies soft fusion operator performance on binary decision tasks. **Chapter 5** discusses methods for generating multihypothesis and vague decisions and confidence assessments using models of binary human decision-making and confidence assessment.

[2] A probability transformation is a mathematical construct used in Dempster-Shafer theory. It is explained later in Section 2.5.6.

These methods are then used to simulate the statistical performance of several soft fusion operators which combine multihypothesis and vague human decisions and confidence assessments. **Chapter 6** studies hard/soft fusion operator performance on binary decision tasks.

- The **Epilogue** of the dissertation is contained in Chapters 7 and 8. **Chapter 7** discusses future work and areas of potential collaboration between researchers in data fusion and cognitive psychology. A summary of the dissertation and concluding remarks are presented in **Chapter 8**.

Part I

An Introduction to Data Fusion Tools and Human Response Simulation

Chapter 2: Data Fusion Terminology

We present the necessary terminology and techniques required to implement the data fusion operators discussed in Chapters 4 through 6. We begin by outlining the general hard/soft fusion problem as it pertains to the case studies of this thesis. Although all of the operators discussed in this thesis make use of the rules of probability, their interpretations may not necessarily have the same underlying meaning. We first discuss the definition of probability as a measure, and then its statistical and epistemological interpretations. We then use these definitions to discuss an arbitrary selection of fusion operators from detection theory [41], Bayesian Epistemology [42], and Dempster-Shafer Theory [43] that could potentially be used in the hard/soft fusion problem presented here.

2.1 Problem Overview

Figure 2.1 displays the general data fusion problem considered by this thesis. A fixed group of machines and humans make observations on the state of some phenomenon. The machines and humans each provide a decision towards the true state and a corresponding decision confidence level. Each decision and confidence assessment is communicated to a *data fusion center*, which is designed to aggregate the responses of all the sources into a single representation. This representation may be, for example, an aggregated set of beliefs towards each of the possible outcomes or a single decision that best captures the decisions and confidence assessments of the sources.

The decisions provided by some of the sources can be vague in the sense that they can consist of the disjunction of two or more states of the phenomenon. To illustrate this, consider a simple game between an illusionist and a volunteer where the volunteer chooses a card from a standard deck of playing cards and the illusionist must make a decision as to which card the volunteer chose. A precise decision would be one in which the illusionist decides that a specific card out of the 52 possible cards was chosen (*e.g.*, the 7 of Clubs). A vague decision would be one in which the illusionist decides that the volunteer has selected a specific type of card, but makes no distinction

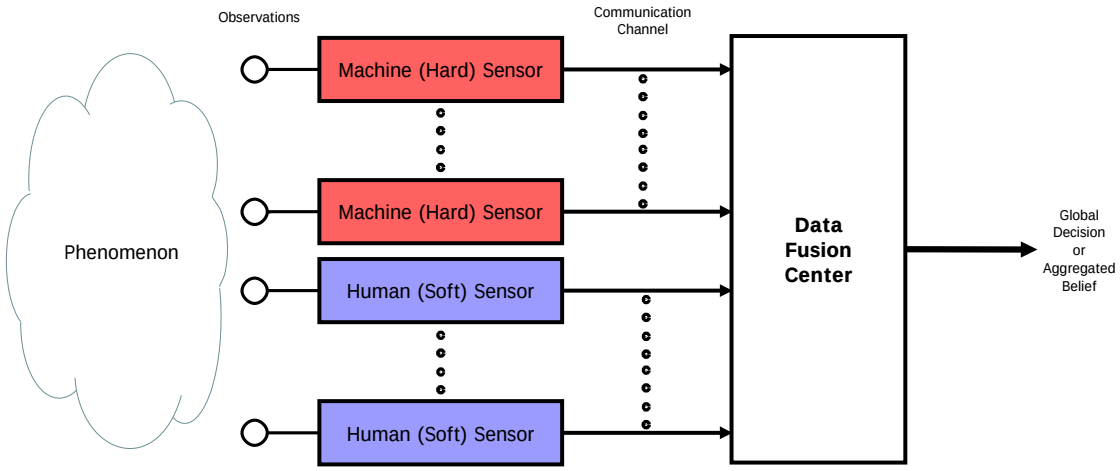


Figure 2.1: Parallel hard and soft fusion problem considered in this thesis.

as to the exact identity of the card (*e.g.*, a black face card, which is the disjunction of the Jack of Clubs, Jack of Spades, Queen of Clubs, and so on).

When fusing hard sensors, a corresponding set of conditional probability models can usually be estimated for each of the possible alternatives, and hence used in the construction of an optimal fusion rule. With soft sensors however, such a characterization is more difficult to achieve since the underlying probabilistic mechanisms may be very complicated and a closed form equation may not necessarily exist. An example of the relationships between the cognitive processes involved in a soft source is shown in Figure 2.2. In this abstraction, the soft source observes the phenomenon and translates his/her observations into evidence for the each of the possible alternatives. The source continues to collect observations and deliberate between alternatives before formulating a response. In some cases, this response may be freeform text or speech, and an information extraction technique may be needed in order to associate a decision and confidence rating. There are many external factors that can potentially lead to errors in the decision/confidence values provided by a soft source. For example, environmental factors such as stress, anxiety, and time constraints all would change the statistical characterization of the source. Internally, the ability for a source to adequately sense and deliberate between hypotheses may also be affected by their own experience or context.

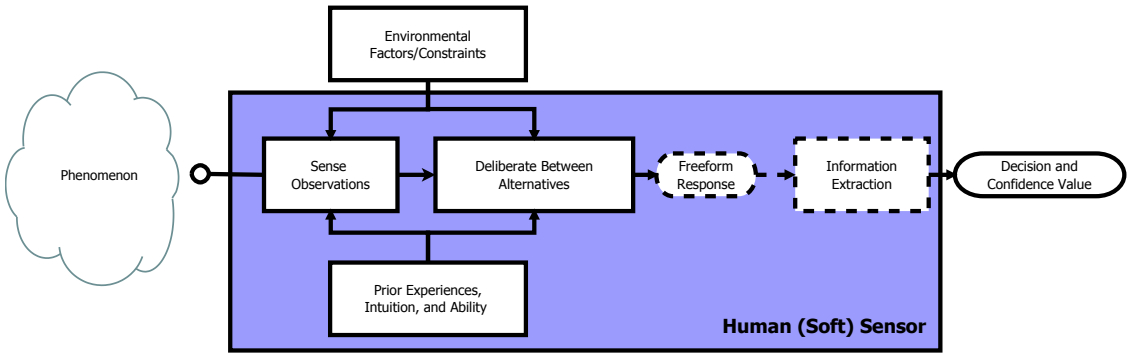


Figure 2.2: Abstraction of a soft source and the relevant cognitive processes involved in the formulation of a decision. A decision and confidence assessment may not be immediately available in the source's response, in which case some sort of information extraction technique may need to be applied (depicted here as dashed lines).

2.2 Interpretations of Probability

Consider the countable set Ω and a corresponding set algebra[1] $\mathcal{F}_\Omega$. By *probability* we mean a function which is a *normalized* and *additive* measure of Ω over the algebra $\mathcal{F}_\Omega$. These notions are summarized by the well known Kolmogorov axioms.

Definition 2.1: Kolmogorov Axioms. *Consider some function $P_\Omega : \Omega \to \mathbb{R}$. For any triple $(\Omega, \mathcal{F}_\Omega, P_\Omega)$, P_Ω is a probability measure on Ω over the set algebra $\mathcal{F}_\Omega$ if and only if*

(i) ***Non-Negativity:*** *$P_\Omega(A) \geq 0$ for all $A \subseteq \Omega$*

(ii) ***Unitarity:*** *$P_\Omega(\Omega) = 1$*

(iii) ***Additivity:*** *Any $A \subseteq \Omega$ with pairwise disjoint elements satisfies $P_\Omega(A) = \sum_{\omega \in A} P_\Omega(\omega)$.*

In an *aleatory* (*i.e.*, statistical) interpretation of probability, the elements of Ω represent testable hypotheses with observable outcomes. To avoid confusion, we denote a statistical probability assignment associated with each of the hypotheses $\omega \in \Omega$ as $P_\Omega(\omega)$. The values of each $P_\Omega(\omega)$ represent *propensities of occurrence*, and can be described as follows [47, Chapter 1].

Definition 2.2: Aleatory (Statistical) Probabilities. *Let the possible outcomes of a phenomenon be represented by each $\omega \in \Omega$. An aleatory interpretation of probability states that there*

[1] For a comprehensive overview of set theoretic concepts as they apply to this thesis, see [46].

exists a probability triple for this space, $(\Omega, \mathcal{F}, P_\Omega)$*, such that the probability measure* $P_\Omega(\omega)$ *for some* $\omega \in \Omega$ *represents the chance that* ω *will occur. Specifically, if one observes* N *total outcomes and finds that* $\omega^* \in \Omega$ *occurs* N_{ω^*} *times, then*

$$\lim_{N\to\infty} \frac{N_{\omega^*}}{N} = P_\Omega(\omega) \tag{2.1}$$

All statistical tools are constructed with an aleatory usage of probability in mind [47, Chapter 8]. This includes elements of detection and estimation theory [41] and portions of centralized/decentralized data fusion [48].

In an *epistemic* (*i.e.*, subjective) interpretation of probability, the elements of Ω represent choice alternatives that may not be statistically testable. Here, the Kolmogorov axioms are used in a literal sense to represent degrees of belief held towards each alternative $\omega \in \Omega$. We represent these probabilities as $\mathcal{P}_\Omega(\omega)$ for each alternative $\omega \in \Omega$. Evidence towards the alternatives ω can be cast objectively or subjectively. In the objective case, degrees of belief can be regarded as propensities of occurrence (*e.g.*, detection/error rates, system reliabilities, etc.). Hence the functions $\mathcal{P}_\Omega(\cdot)$ and $P_\Omega(\cdot)$ are the same. In the subjective case, degrees of belief are taken as a measure of one's confidences [43, Chapter 1], and hence $\mathcal{P}_\Omega(\cdot)$ may have no relation to the probabilities that describe the statistics of Ω. Work motivating the representation of subjective beliefs in a totally epistemic manner was first formalized by Frank Ramsey and De Finetti in [49] and [50]. Ramsey and De Finetti explained that epistemic probabilities could be thought of as betting rates an individual holds towards the outcomes of a phenomenon, and if one were to enter a betting situation which disobeyed the rules of probability then he or she would be considered irrational. With this in mind, the use of probabilities to represent subjective beliefs implies the following interpretation of the Kolmogorov Axioms.

Definition 2.3: Epistemic (Subjective) Probabilities. *Let* Ω *denote all possible alternatives (not necessarily testable) concerning a certain phenomenon. An epistemic interpretation of probability states that for a given observer, there exists a probability triple,* $(\Omega, \mathcal{F}, \mathcal{P}_\Omega)$*, such that the*

probability measure $\mathcal{P}_\Omega(\omega)$ represents the observer's beliefs for each $\omega \in \Omega$. When viewed in light of the Kolmogorov Axioms, this implies the following.

(i) ***Non-Negativity:*** *Beliefs are represented by non-negative numbers.*

(ii) ***Unitarity:*** *Full belief (tautology) must be attributed to the possibility that any occurrence in Ω can be true at a given time.*

(iii) ***Additivity:*** *The belief held for any collection of alternatives in Ω must be the same as the sum of the beliefs held for each one of those alternatives.*

Definition 2.3 describes a Bayesian interpretation of probability (*i.e.*, Bayesian Epistemology) [51–53]. The additivity requirement of the Kolmogorov axioms has been referred to as *Bayes' Rule of Additivity* [43, Chapter 1]. The empty set $\emptyset$ represents an impossible event (*i.e.*, a contradiction). As a consequence of the Kolmogorov axioms, it can be easily shown that the Bayesian theory of beliefs will always assign such contradictions zero belief (*i.e.*, $\mathcal{P}_\Omega(\emptyset) = 0$).

2.3 Detection Theory

When a fusion system designer has foreknowledge regarding the statistics of the phenomenon or the sensors (*i.e.*, decision makers), elements from classical detection theory [41, Chapter 1] can be applied to select the best performing fusion operator. Such a fusion operator is chosen in order to meet an optimization criteria (*e.g.*, the Bayes' criterion or the Neyman-Pearson criterion). When using the Bayes' criterion, additional statistical foreknowledge is required in the form of *a priori* probabilities and the decision costs associated with each possible course of action. When using the Neyman-Pearson criterion, the *a priori* probabilities and decision costs are not required. For a binary set of alternatives, both optimization problems reduce to a likelihood ratio test. For multi-hypothesis (M-ary) alternatives, the optimal decision rule for the Bayesian criterion is generalized to the *maximum a posteriori probability* (MAP) rule. The optimal decision rule for the Neyman-Pearson generalizes to a collection of likelihood ratio tests [41, Chapter 1]. The design and solution of such optimization problems may not be feasible in a soft or hard/soft fusion system since the required likelihood ratios may be too complicated to accurately compute in real time [27]. Further-

more, a representation of the statistics governing soft source responses may not exist in closed form. Nevertheless, we briefly mention two simple fusion operators from detection theory that could be applied *ad hoc* to the hard/soft fusion problem depicted in Figure 2.1, when restricted to binary hypotheses. The first is the *k-out-of-N* majority rule [54], which chooses an alternative when k or more decision makers declare it as the true outcome.

Definition 2.4: k-out-of-N Majority Rule. *Consider the binary hypothesis problem given by the set of alternatives $\Omega = \{H_0, H_1\}$. Suppose a set of N detectors employ a decision rule to make a decision u_i, $i = 1, \ldots, N$ such that*

$$u_i = \begin{cases} -1 & \text{if } H_0 \text{ is declared} \\ \\ +1 & \text{if } H_1 \text{ is declared} \end{cases}. \tag{2.2}$$

Then, the k-out-of-N majority rule makes the global decision u_0 such that

$$u_0 = \begin{cases} -1 & \sum_{i=1}^{N} u_i < 2k - N \\ \\ +1 & \text{otherwise} \end{cases}. \tag{2.3}$$

The value k is chosen based on the required sensitivity of the fusion rule. Notice that when $k = 1$ the k-out-of-N majority rule reduces to the OR fusion rule. When $k = N$ it reduces to the AND fusion rule. Finally, when $k = N/2$ it reduces to a majority rule. The corresponding false alarm rate p_{F_0} and correct detection rate p_{D_0} that describe the performance of the k-out-of-N majority rule are defined as

$$p_{F_0} = P_\Omega(u_0 = +1|H_0) = P_\Omega(\sum_{i=1}^{N} u_i \geq 2k - N|H_0) \tag{2.4}$$

$$p_{D_0} = P_\Omega(u_0 = +1|H_1) = P_\Omega(\sum_{i=1}^{N} u_i \geq 2k - N|H_1) \tag{2.5}$$

Alternatively, the *Chair and Varshney fusion rule* uses a weighted sum of local decisions where the weights capture the reliability (*i.e.*, missed detection and false alarm rates) of each decision maker

[54]. Assuming statistical independence of the source distributions conditioned on the hypotheses, the fusion rule is a likelihood ratio test that can be implemented in the following manner.

Definition 2.5: Chair and Varshney Fusion Rule (weighted k-out-of-N). *Consider a binary hypothesis testing problem defined by the set of alternatives* $\Omega = \{H_0, H_1\}$. *Suppose a set of* N *detectors employ a decision rule to produce a decision variable,* u_i *with* $i = 1, \ldots, N$, *such that*

$$u_i = \begin{cases} -1 & \text{if } H_0 \text{ is declared} \\ +1 & \text{if } H_1 \text{ is declared} \end{cases}. \tag{2.6}$$

Then, the Chair and Varshney fusion rule makes the global decision u_0 *such that*

$$u_0 = \begin{cases} -1 & \sum_{i=1}^{N} w_i u_i < \lambda \\ +1 & \text{otherwise} \end{cases} \tag{2.7}$$

where the weights w_i *are determined based on the false alarm rates,* $p_{F_i} = P(u_i = +1|H_0)$, *and missed detection rates,* $(p_{M_i} = P(u_i = -1|H_1))$, *of the detectors. That is,*

$$w_i = \begin{cases} \log\left(\frac{1-p_{F_i}}{p_{M_i}}\right) & u_i = -1 \\ \log\left(\frac{1-p_{M_i}}{p_{F_i}}\right) & u_i = +1 \end{cases}. \tag{2.8}$$

The threshold λ *is chosen to satisfy either the Bayes' or Neyman-Pearson criterion.*

For the Chair and Varshney Fusion rule, the corresponding false alarm rate p_{F_0} and correct detection rate p_{D_0} that describe its performance are defined as

$$p_{F_0} = P_\Omega(u_0 = +1|H_0) = P_\Omega(\sum_{i=1}^{N} w_i u_i \geq \lambda|H_0) \tag{2.9}$$

$$p_{D_0} = P_\Omega(u_0 = +1|H_1) = P_\Omega(\sum_{i=1}^{N} w_i u_i \geq \lambda|H_1). \tag{2.10}$$

2.4 Bayesian Theory of Beliefs

When using subjective confidences, deductive reasoning can be performed using conditional probabilities. The use of epistemic probabilities and conditionalization as a means of deductive reasoning has been referred to as the *Bayesian Theory of Beliefs* [43, Chapter 1]. In [42], Talbott describes the epistemological significance of probabilistic conditioning as follows.

Definition 2.6: Simple Principle of Conditionalization. *If one begins with prior subjective probabilities $\mathcal{P}_{\text{init}}$, and acquires new evidence which can be represented as becoming certain of some additional information x having non-zero subjective probability, then deduction can be applied by transforming one's initial probabilities to generate posterior subjective probabilities $\mathcal{P}_{\text{final}}$ by conditionalizing on x. Specifically, for any alternative $\omega \in \Omega$ we have that*

$$\mathcal{P}_{\text{final}}(\omega) = \mathcal{P}_{\text{init}}(\omega|x) = \frac{\mathcal{P}_{\text{init}}(\omega \cap x)}{\mathcal{P}_{\text{init}}(x)}. \tag{2.11}$$

There are many way that Bayesian conditioning can be implemented, depending on the application. For example, the well known *Bayes' theorem* gives a way of updating evidence when when beliefs are given as *likelihoods* of the different alternatives $\omega \in \Omega$ producing some observation $x \in \mathcal{X}$ (i.e., $\mathcal{P}(x|\omega)$).

Theorem 2.1: Bayes' Theorem. *Suppose we observe a piece of evidence $x \in \mathcal{X}$ related to the alternatives described by all $\omega \in \Omega$. The original (*i.e., a priori*) beliefs, $\mathcal{P}_\Omega(\omega)$, can be updated via Bayes' Rule:*

$$\mathcal{P}_\Omega(\omega|x) = \frac{\mathcal{P}_\mathcal{X}(x|\omega)\mathcal{P}_\Omega(\omega)}{\sum_{\hat{\omega}\in\Omega} \mathcal{P}_\mathcal{X}(x|\hat{\omega})\mathcal{P}_\Omega(\hat{\omega})} \tag{2.12}$$

where $\mathcal{P}_\mathcal{X}(x|\omega)$ is known as the likelihood that the alternative $\omega \in \Omega$ would produce the evidence $x \in \mathcal{X}$. The quantity $\mathcal{P}_\Omega(\omega|x)$ is known as the a posteriori belief in the alternative ω, given the observation of the evidence x.

When the a priori probabilities and likelihoods are constructed from statistical information, the *a posteriori* probabilities computed by Bayes' theorem are also statistical and represent the chances of

each alternative occurring given the observed evidence x. In this case, the application of Bayes' rule becomes synonymous with many of the fusion operators proposed in classical detection theory [41]. When evidence is provided by the sources in the form of subjective probabilities however, *Bayes' rule of probability combination* has a more meaningful interpretation of Bayesian conditioning [55]. Specifically, Bayes' rule of probability combination is the result of determining the joint subjective probability mass function between the sources, conditioned on the fact that the true outcome has not been ruled out (*i.e.*, assigned probability zero) by any of the sources.

Theorem 2.2: ***Bayes' Rule of Probability Combination.*** *Consider two sources whose confidences on a set of finite, disjoint alternatives Ω are given via the subjective probability functions $\mathcal{P}_1$ and $\mathcal{P}_2$. Let $\Omega_i = \{\omega \in \Omega | \mathcal{P}_i(\omega) \neq 0\}$ for $i = 1, 2$ (*i.e., *the outcomes which have not been ruled out by the sources being combined)*[2]*. Given that the true outcome $\omega^* \in \Omega_{1,2} = \Omega_1 \cap \Omega_2$ and assuming that the sources assess their confidences independently of each other, the combined probability function $\mathcal{P}_{1,2}$ can be represented as*

$$\mathcal{P}_{1,2}(\omega) = \mathcal{P}_{1,2}(\omega = \omega^* | \omega^* \in \Omega_{1,2}) = \frac{\mathcal{P}_1(\omega)\mathcal{P}_2(\omega)}{\sum_{\hat{\omega} \in \Omega_{1,2}} \mathcal{P}_1(\hat{\omega})\mathcal{P}_2(\hat{\omega})} \tag{2.13}$$

*for any $\omega \in \Omega_{1,2}$, provided that the Ω_1 and Ω_2 are non-contradictory (*i.e., *$\Omega_1 \cap \Omega_2 \neq \emptyset$).*

Although Bayes' rule of probability combination and Bayes' theorem are mathematically analogous, their interpretations are fundamentally different.

2.5 Dempster-Shafer (DS) Theory

2.5.1 History

The notion of *Dempster-Shafer Theory* (DS Theory) is in most instances attributed to work in the 1970s by Glenn Shafer [43]. When introduced, DS Theory consisted of two major components: (1) beliefs are represented by independently assigning evidence to all elements of the power set of alternatives, and (2) belief combination is achieved through a generalization of Bayesian conditioning, known as *Dempster's Rule of Combination (DRC)*. DS Theory has been proposed for many

[2] We use the notation $\mathcal{P}_1$ and $\mathcal{P}_2$ as shorthand for $\mathcal{P}_{\Omega_1}$ and $\mathcal{P}_{\Omega_2}$.

applications where uncertainty due to vague evidence is a factor.

When first introduced, Shafer promoted DS Theory as a universal theory of beliefs for both subjective and objective sources of evidence. Later Shafer admits in [56] that he had taken a "youthfully amibtious" view of the situation. Arguments regarding the usage of DS Theory began to surface in the early-to-mid 1980's through work done by Lotfi Zadeh in [30, 57]. Specifically, Zadeh had provided examples which showed that under certain cases of highly conflicting evidence, the DRC seemed to produce illogical and counterintuitive results. Although now these viewpoints could be attributed to a misunderstanding of what DS Theory represented (*e.g.*, [58]), this nevertheless caused a multitude of belief function alternatives to surface, leaving many users in the field of expert reasoning systems confused[3].

The tensions and disagreements resulting from this confusion resulted in a debate in the early 1990's between Glenn Shafer and Judea Pearl in the International Journal of Approximate Reasoning volume 4, issues 5-6. The main points of this debate were to assess how belief functions should be interpreted and their merit when used in place of a probabilistic approach[4]. Shafer's major points were aimed at further clarifying the use of probability functions as tools that model a "special case" of reality in which aleatoric knowledge is known, and hence epistemic and aleatoric knowledge are identical. In such cases, Shafer reaffirmed that probabilistic reasoning proved to be more than sufficient, however he noted that expert reasoning systems usually go beyond such statistical problems. Hence, he advocated a constructive interpretation of probability in which one should be advised not to speak of probabilities in an application unless (1) his/her meaning was clarified and (2) he/she was assured to a reasonable extent that such probabilities actually existed [61]. Alternatively, Judea Pearl pointed out that the merit of the belief function representation was still yet to be verified. For example, in his first position paper Pearl showed that belief functions could be mapped into an equivalent likelihood function for use in probabilistic reasoning. Hence, the same results could theoretically be achieved on a lower level of computational complexity[5] [60]. Ten years later, Jerome

[3]Many of these methods are summarized in [59], using this viewpoint as the overarching theme.

[4]The position papers given by Shafer and Pearl on the usage of belief functions can be found in [56] and [60] respectively. All response papers can be found in the *International Journal of Approximate Reasoning* Volume 6, Issue 3, including Shafer's and Pearl's responses in [61] and [62] respectively.

[5]Later, this was emphasized through the results of papers similar to [63]

Braun showed through simulations in [64] that reasoning via belief functions under traditional DS Theory provided a higher level of accuracy than a traditional probabilistic approach in cases of high uncertainty and high conflict. In almost all other cases, the accuracies of both approaches were shown to be identical.

Despite the controversies, Dempster-Shafer theory continues to be a popular tool in the data fusion community. It has been used in pattern recognition applications in which the images under consideration are of sufficiently low resolution, preventing standard probability-based edge detection techniques [65–68]. Similarly, DS Theory has been adapted into some facial recognition systems [69, 70] and various other multimodal biometric recognition systems [71]. There also exist a wide variety of applications in biomedical engineering [72–76], navigation systems [77–79], risk and reliability assessment systems [80–84], and various other detection systems [85–89].

2.5.2 General Terminology

Again, we let Ω denote the set of possible alternatives for a certain phenomenon. In a DS theoretic setting, Ω is commonly referred to as the *frame of discernment* [43]. Because we assumed the existence of M disjoint alternatives[6] in the problem overview of Figure 2.1, the corresponding algebra on Ω reduces to the power set $\mathcal{F} = 2^{\Omega}$. For convenience, we will refer to the elements of the power set as *propositions*. The fundamental measure of belief in DS theory is the *belief mass assignment* (BMA), which in a manner similar to Bayesian epistemology assigns values between zero and one to the propositions defined by power set.

Definition 2.7: Belief Mass Assignment (BMA). *A function* $m : 2^{\Omega} \to [0, 1]$ *is a belief mass assignment for some frame* Ω *if and only if* $\sum_{X \subseteq \Omega} m(X) = 1$.

The *belief function* describes the minimum amount of evidence that is committed to a proposition, according to some BMA on Ω.

Definition 2.8: Belief Function[7]. *Consider the BMA* m *on some frame* Ω. *The function* Bel :

[6] The most common assumption in DS theory is that Ω is finite and thus is isomorphic to some disjoint, finite set of alternatives.

[7] The notion of $\text{Bel}(\emptyset) = 0$ is mentioned by Smets in [90]

$2^{\Omega} \to [0,1]$ *is called a belief function on* Ω*, and is given for any* $A \subseteq \Omega$ *as*

$$\mathrm{Bel}(A) = \begin{cases} 0 & A = \emptyset \\ \sum_{\substack{X \neq \emptyset \\ X \subseteq A}} m(X) & A \neq \emptyset \end{cases}. \tag{2.14}$$

The relationship between a BMA and a belief function is unique. The inverse mapping is defined using a Möbius transformation [43]. We call the quantities $m(A)$ and $\mathrm{Bel}(A)$ the *belief mass* and the *committed belief* to the proposition $A \subseteq \Omega$. The quantities $\mathrm{Bel}(\Omega)$ and $m(\emptyset)$ represent the belief in the sufficiency and insufficiency of Ω respectively. In the special case of a *closed world assumption*, Ω is considered completely sufficient (*i.e.*, $m(\emptyset) = 0$) and hence the function Bel becomes a normalized monotonic measure [46].

Since it is not necessary to assign a belief mass to every possible subset of Ω, it becomes valuable to denote those belief masses which are non-zero.

Definition 2.9: Focal Element. *Any proposition* $A \subseteq \Omega$ *such that* $m(A) > 0$ *is called a focal element of some belief function over* Ω *as given by* m*.*

Definition 2.10: Belief Core. *The set of sets* $\mathcal{C} \subseteq 2^{\Omega}$ *consisting of all focal elements of* Ω *as given by* m *is called the core of the belief function.*

It is important to note that belief in any $A \subseteq \Omega$ does not fully describe the extent to which one doubts A, or rather one's belief in its negation $\mathrm{Bel}(\overline{A})$. Together, $\mathrm{Bel}(A)$ and $\mathrm{Bel}(\overline{A})$ provide a complete description of a source's oppinion towards $A \subseteq \Omega$. Alternatively, we can consider the degree of belief in A and the degree at which one finds A *plausible*.

Definition 2.11: Plausibility Function. *Consider the BMA* m *on some frame* Ω*. The plausibility function* $\mathrm{Pl} : 2^{\Omega} \to [0,1]$ *is given as*

$$\mathrm{Pl}(A) = \begin{cases} 0 & A = \emptyset \\ \mathrm{Bel}(\Omega) - \mathrm{Bel}(\overline{A}) = \sum_{X \cap A \neq \emptyset} m(X) & A \neq \emptyset \end{cases}. \tag{2.15}$$

Uncertainty representation in a belief function setting comes from the assessment of evidence on disjunctions (*i.e.*, vague evidence). For example, belief assigned to $\{\omega_1\}$ and $\{\omega_2\}$ are the amounts of evidence that *directly imply* the propositions $\{\omega_1\}$ and $\{\omega_2\}$ respectively. Evidence assigned to the proposition $\{\omega_1, \omega_2\}$ represents evidence which *could* be assigned to either $\{\omega_1\}$ or $\{\omega_2\}$, however due to uncertainty it cannot be further divided amongst them. When calculating belief and plausibility functions from a BMA, one can think of them as minimum and maximum amounts of evidence which can be attributed to a given proposition. The uncertainty measure of a proposition in Ω can thus be viewed as the amount of evidence which *could possibly* be committed to a given proposition. Specifically, it is the amount of evidence which does not imply a given proposition, but also does not contradict it.

Definition 2.12: Uncertainty Function. *Consider the BMA m on some frame Ω. The uncertainty function* $\mathrm{Un} : 2^{\Omega} \to [0, 1]$ *is given as*

$$\mathrm{Un}(A) = \mathrm{Pl}(A) - \mathrm{Bel}(A) = \sum_{\substack{X \cap A \neq \emptyset \\ X \not\subset A}} m(X). \tag{2.16}$$

Finally, we introduce the concept of three specialized belief structures: a *vacuous support function*, a *simple support function*, and a *categorical belief function.*

Definition 2.13: Vacuous Support Function. *For some belief function whose frame of discernment is given by Ω, total ignorance to all propositions can be represented by*

$$\mathrm{Bel}(A) = \begin{cases} 1 & A = \Omega \\ \\ 0 & A \neq \Omega \end{cases}. \tag{2.17}$$

The corresponding BMA for a vacuous support function is given as

$$\mathrm{m}(A) = \begin{cases} 1 & A = \Omega \\ \\ 0 & A \neq \Omega \end{cases}. \tag{2.18}$$

Definition 2.14: Simple Support Function. *For some belief function whose frame of discernment is given by Ω, a simple support function is a belief function whose evidence points precisely and unambiguously to a single non-empty $A^* \subset \Omega$. Specifically, we say that a simple support function holds the degree of support $s \in [0,1]$ for A^* such that for any $A \subset \Omega$ we have that*

$$\mathrm{Bel}(A) = \begin{cases} 0 & A \not\supset A^* \\ s & A \supset A^*, A \neq \Omega \\ 1 & A = \Omega \end{cases} . \tag{2.19}$$

The corresponding BMA for a simple support function is given as

$$\mathrm{m}(A) = \begin{cases} s & A = A^* \\ 1-s & A = \Omega \\ 0 & otherwise \end{cases} . \tag{2.20}$$

Definition 2.15: Categorical Support Function. *For some belief function whose frame of discernment is given by Ω, a categorical belief function is a belief function whose evidence is completely committed to a single non-empty $A^* \subset \Omega$. Specifically, a categorical support function on A^* is given for any $A \subseteq \Omega$ such that*

$$\mathrm{Bel}(A) = \begin{cases} 0 & A \not\supset A^* \\ 1 & A \supseteq A^* \end{cases} . \tag{2.21}$$

The corresponding BMA for a categorical support function is given as

$$\mathrm{m}(A) = \begin{cases} 1 & A = A^* \\ 0 & A \neq A^* \end{cases} . \tag{2.22}$$

2.5.3 DS Theory and the Bayesian Theory of Beliefs

As originally introduced by Shafer in [43], DS Theory assumes that the set of alternatives Ω is sufficient (*i.e.*, $m(\emptyset) = 0$) [43, Chapters 2]. With this in mind, it is straightforward to prove the following theorem.

Theorem 2.1. *Consider the belief function* Bel *resulting from the BMA* m *over the set of alternatives* Ω*. If* $m(\emptyset) = 0$*. then the following must be true*

(i) **Non-Negativity:** $\text{Bel}(A) \geq 0$ *for all* $A \subseteq \Omega$*.*

(ii) **Unitarity:** $\text{Bel}(\Omega) = 1$*.*

(iii) **Superadditivity**[8]**:** *For any* $A \subseteq \Omega$*,* $\text{Bel}(A) \geq \sum_{\omega \in A} \text{Bel}(\omega)$*.*

Notice that Theorem 2.1 is almost identical to the Kolmogorov Axioms, with the only difference being that the additivity requirement has been relaxed to superadditivity. From a measure theoretic standpoint, belief functions can also be described as *totally monotonic measures* [46, Chapter 4], [91].

Theorem 2.2. ***Monotonicity/Convexity of Belief Functions***[9]***.****Consider the belief function* Bel *resulting from the BMA* m *over the set of alternatives* Ω*. If* $m(\emptyset) = 0$*, then* Bel *is classified as a totally monotonic measure. That is, for all integers* $k \geq 2$ *and for all family of sets* $A_1, \ldots, A_k \subset \Omega$*, the following inequalities are satisfied.*

$$\text{Bel}\left(\bigcup_{j=1}^{k} A_j\right) \geq \sum_{\substack{K \subseteq \mathbb{N}_k \\ K \neq \emptyset}} (-1)^{|K|+1} \text{Bel}\left(\bigcap_{j \in K} A_j\right) \tag{2.23}$$

where $\mathbb{N}_k = \{1, 2, \ldots, k\}$*. Convexity is covered as the case in which* $k = 2$*,*

$$\text{Bel}(A_1 \cup A_2) \geq \text{Bel}(A_1) + \text{Bel}(A_2) - \text{Bel}(A_1 \cap A_2). \tag{2.24}$$

Unsurprisingly, probability functions can also be classified as totally monotonic measures. Using Theorems 2.1 and 2.2, we can now describe a *Bayesian belief function* (*i.e.*, subjective probability

[8] This follows immediately from Theorem 2.1 in [43, Chapter 2]
[9] The proof of this is given as a portion of Theorem 2.1 from [43, Chapter 2].

function) as a special type of belief function.

Theorem 2.3. Relation to Subjective Probabilities[10]**.** *The following statements are equivalent and define a special type of belief function over a frame* Ω *known as a Bayesian belief function (*i.e.*, subjective probability assignment).*

(i) *The core of the belief function consists only of singleton elements from* 2^{Ω}.

(ii) $\text{Bel}(A) = \text{Pl}(A)$ *for all* $A \subseteq \Omega$.

(iii) $\text{Un}(A) = 0$ *for all* $A \subseteq \Omega$.

(iv) $\text{Bel}(A) + \text{Bel}(\overline{A}) = 1$ *for all* $A \subset \Omega$.

(v) $\text{Bel}(A) = \sum_{\omega \in A} \text{Bel}(\omega)$ *for all* $A \subseteq \Omega$.

Subjective probabilities thus form a natural subset of belief functions. Relating Theorem 2.3 back to Definitions 2.8 and 2.11, it's easy to see that for any $A \subseteq \Omega$ that $\text{Bel}(A) \leq \text{Pl}(A)$, with equality only when $\text{Un}(A) = 0$. This relationship leads to a set of intervals $[\text{Bel}(A), \text{Pl}(A)]$ for every $A \subseteq \Omega$. The meaning of these belief and plausibility intervals has been debated at length. In general, one can think of them as merely minimum and maximum amounts of evidence which can be attributed to the proposition A. The intervals thus represent spectra of belief for evidence that *is* committed to evidence that *could be* committed to a given proposition [92, Chapter 7]. There is another interpretation however in which belief and plausibility intervals determine the envelope for a supporting class of subjective probability functions. This is not hard to see given the convexity of both belief and probability functions, and the results of Theorem 2.3. However, it has been discussed at length that these probabilities may not have any connection to any statistical probabilities, regardless of how their associated BMA is constructed. Instead, these probabilities should only be interpreted as possible *subjective probabilities* from which the evidence assessed by their associated BMA supports (*i.e.*, probabilities of provability, as discussed in [60]).

[10] This is a modified version of Theorem 2.8 from [43, Chapter 2]

2.5.4 Dempster's Rule of Combination

Shafer cites *Dempster's Rule of Combination (DRC)* in [43] as the primary method for belief aggregation and updating. It usage can have many interpretations, depending on the application. However, Shafer presents DRC as an axiom of DS Theory in [43], and motivates it as an orthogonal sum of two belief masses.

Definition 2.16: Dempster's Rule of Combination (DRC) [43]. *Consider the opinions of two sources, represented by the BMAs m_1 and m_2. Dempster's Rule of Combination is defined such that the resulting BMA after combination, $m_{1,2}(A)$ for any $A \subseteq \Omega$, is given by*

$$m_{1,2}(A) = \left(\frac{1}{1-\mathcal{K}}\right) \sum_{\substack{X_1, X_2 \subseteq \Omega \\ X_1 \cap X_2 = A}} m_1(X_1) m_2(X_2) \tag{2.25}$$

where $\mathcal{K} < 1$ is the measure of conflict after combination, given by

$$\mathcal{K} = \sum_{\substack{X_1, X_2 \subseteq \Omega \\ X_1 \cap X_2 = \emptyset}} m_1(X_1) m_2(X_2). \tag{2.26}$$

When $\mathcal{K} = 1$, Dempster's rule is undefined.

The DRC is both commutative and associative [93]. It can be viewed as a two step evidence combination process. In the first step, evidence is conjunctively combined through the sum of products expression in the numerator of Equation 2.25. Because Ω is assumed to consist of disjoint alternatives, a portion of the combined evidence is committed to the empty set. In the second step, the BMAs after conjunctive combination are normalized such that the belief mass assigned to the empty set is equal to zero.

2.5.5 Alternative Belief Combination Rules

Notice that Bayes' rule of probability combination (Definition 2.2) and DRC (Definition 2.16) perform source combination by conjunctively combining the evidence presented from the sources and renormalizing the combined evidence on the amount of non-conflicting evidence. In fact, it is not difficult to show that both of these combination rules are the same if the BMAs being combined are

Bayesian. If the combined sources present highly conflicting opinions however, evidence combination in this manner may not provide an acceptable result [28, 30]. For this reason, a plethora of BMA combination operators have been proposed as alternatives to DRC. We summarize a selection of them here, as they pertain to the soft and hard/soft fusion studies discussed later in the thesis.

One possible variation known as *Yager's Rule* calls for interpreting the evidence in conflict as total uncertainty.

Definition 2.17: ***Yager's Rule [94].*** *Consider two BMAs m_1 and m_2 over some set of alternatives Ω. The resulting BMA $m_{1,2}$ after combination via Yager's rule for any $A \subseteq \Omega$ is given as*

$$m_{1,2}(A) = \begin{cases} \sum_{\substack{X_1, X_2 \subseteq \Omega \\ X_1 \cap X_2 = A}} m_1(X_1) m_2(X_2) & A \neq \Omega \\ m_1(A) m_2(A) + \mathcal{K} & A = \Omega \end{cases} \tag{2.27}$$

where $\mathcal{K}$ is the degree of conflict between m_1 and m_2 as given in equation (2.26).

Yager's rule is commutative but in general not associative [95]. A generalization of Yager's rule is *Dubois and Prade's rule* [95, 96], which attributes conflict between two sources as uncertainty between the conflicting elements of the sources, rather than total uncertainty.

Definition 2.18: ***Dubois and Prade's Rule.*** *Consider two BMAs m_1 and m_2 over some set of alternatives Ω. The resulting BMA $m_{1,2}$ after combination via Dubois and Prade's rule for any $A \subseteq \Omega$, $A \neq \emptyset$ is given as*

$$m_{1,2}(A) = \sum_{\substack{X_1, X_2 \subseteq \Omega \\ X_1 \cap X_2 = A}} m_1(X_1) m_2(X_2) + \mathcal{K}(A) \tag{2.28}$$

where $m_{1,2}(\emptyset) = 0$ and

$$\mathcal{K}(A) = \sum_{\substack{X_1, X_2 \subseteq \Omega \\ X_1 \cap X_2 = \emptyset \\ X_1 \cup X_2 = A}} m_1(X_1) m_2(X_2). \tag{2.29}$$

Similar to Yager's rule, Dubois and Prade's rule is commutative but in general not associative [95]. It is also possible to subdivide evidence amongst conflicting propositions proportionate to the degrees

of belief held by the sources before combination. This concept has lead to the family of *Proportional Conflict Redistribution (PCR)* rules by Dezert and Smarandache. In this dissertation we only make use of *PCR5* for brevity.

Definition 2.19: ***Proportional Conflict Redistribution Rule #5 (PCR5) [97].*** *Consider two BMAs m_1 and m_2 over some set of alternatives Ω. The resulting BMA $m_{1,2}$ after combination via PCR5 for any $A \subseteq \Omega$ is given as*

$$m_{1,2}(A) = \sum_{\substack{A_1, A_2 \subseteq \Omega \\ A_1 \cap A_2 = A}} m_1(A_1) m_2(A_2) + \sum_{\substack{X \subseteq \Omega \\ X \cap A = \emptyset}} \left(\frac{m_1(X)^2 m_2(A)}{m_1(X) + m_2(A)} + \frac{m_2(X)^2 m_1(A)}{m_2(X) + m_1(A)} \right). \tag{2.30}$$

Similar to Yager's Rule, the PCR5 is commutative but in general not associative [97]. Next, the consensus operator [98] use the general terminology of BMAs, belief functions, and uncertainty while providing a way of representing any Ω as a binary frame focused on some $A \subseteq \Omega$ and its negation. The result is a four-valued vector known as an *opinion tuple*, which can be rephrased in terms of BMAs for a binary Ω as follows.

Definition 2.20: ***Consensus Operator (Binary Frames) [98].*** *Consider the binary frame $\Omega = \{A, \overline{A}\}$. The consensus operator $m_{1,2}$ can be represented completely in terms of the BMAs m_1 and m_2 as*

$$\begin{aligned} m_{1,2}(A) &= \frac{m_1(A) m_2(A \cup \overline{A}) + m_2(A) m_1(A \cup \overline{A})}{\mathcal{K}_C} \\ m_{1,2}(\overline{A}) &= \frac{m_1(\overline{A}) m_2(A \cup \overline{A}) + m_2(\overline{A}) m_1(A \cup \overline{A})}{\mathcal{K}_C} \\ m_{1,2}(A \cup \overline{A}) &= \frac{m_1(A \cup \overline{A}) m_2(A \cup \overline{A})}{\mathcal{K}_C} \end{aligned} \tag{2.31}$$

where $\mathcal{K}_C = m_1(A \cup \overline{A}) + m_2(A \cup \overline{A}) - m_1(A \cup \overline{A}) m_2(A \cup \overline{A})$. If $\mathcal{K}_C = 0$, then the consensus operator

becomes

$$\begin{aligned} m_{1,2}(A) &= \frac{m_1(A) + m_2(A)}{2}, \\ m_{1,2}(\overline{A}) &= \frac{m_1(\overline{A}) + m_2(\overline{A})}{2}, \\ m_{1,2}(A \cup \overline{A}) &= 0. \end{aligned} \tag{2.32}$$

The consensus operator is both commutative and associative [98]. Finally, BMAs can be combined by simply averaging them. In general, this averaging operation is known as the mixing combination rule [99].

Definition 2.21: ***Mixing Combination Rule [99].*** *Consider two BMAs m_1 and m_2 over some set of alternatives Ω. The resulting BMA $m_{1,2}$ after combination via the mixing combination rule for any $A \subseteq \Omega$ is given as*

$$m_{1,2}(A) = w_1 m_1(A) + w_2 m_2(A) \tag{2.33}$$

where the weights $w_1, w_2 \in [0, 1]$ and $w_1 + w_2 = 1$.

According to [59], the mixing combination rule is in general not commutative nor associative. When the weights are equal $w_1 = w_2 = 0.5$, the mixing rule reduces to Murphy's combination rule [100]. When the BMAs being combined are equivalent to subjective probability assignments, this combination rule is identical to linear opinion pooling [101].

2.5.6 Decision Making in Dempster-Shafer Theory

Although there have been methods proposed on relating belief and plausibility functions to subjective utility intervals (*e.g.*, [91]), decision making in DS theory is usually achieved by transforming a post-combination BMA into a subjective probability assignment [95]. Once a subjective probability assignment has been obtained from a BMA, it can then be used to implement a decision-making rule. There are a large number of these so called *probability transforms* proposed throughout the DS theory literature. We summarize a few of them here as they pertain to the fusion simulations described later in the thesis.

The pignistic probability transformation ($BetP$) was first proposed by Philippe Smets in [102]

and then included in [103] as a part of the Transferable Belief Model. The pignistic probability transform involves transferring the belief mass from each non-singleton element of a BMA to its respective singleton elements by dividing its mass equally (*i.e.*, according to its cardinality).

Definition 2.22: ***Pignistic Probability Transform [102].*** *The pignistic probability is defined for any* $A \subseteq \Omega$ *as*

$$BetP(A) = \sum_{\substack{X \subseteq \Omega \\ X \neq \emptyset}} \frac{|A \cap X|}{|X|} m(X), \tag{2.34}$$

where $|\cdot|$ *is the cardinality of a set. The pignistic probability transform satisfies all three Kolmogorov Axioms, and hence it only needs to be computed for the singleton elements* $\omega \in \Omega$*.*

Sudano has also proposed a suite of five probability transformations in [104]: $PrPl$, $PrNPl$, $PraPl$, $PrBel$, and $PrHyb$. In this dissertation, we will only focus on the $PrNPl$, $PrPl$, and $PrHyb$ transforms.

Definition 2.23: ***Sudano Probability Transforms [104].*** *Sudano's proportional plausibility (*$PrPl$*), normalized proportional plausibility (*$PrNPl$*), and hybrid (*$PrHyb$*) probability transforms are given for each* $\omega \in \Omega$ *and for the corresponding belief* Bel*, and plausibility* Pl*, and BMA* m *functions as*

$$PrPl(\omega) = \mathrm{Pl}(\omega) \sum_{\substack{Z \subseteq \Omega \\ \omega \in Z}} \frac{m(Z)}{\sum_{\hat{\omega} \in Z} \mathrm{Pl}(\hat{\omega})}, \tag{2.35}$$

$$PrNPl(\omega) = \frac{\mathrm{Pl}(\omega)}{\sum_{\hat{\omega} \in \Omega} \mathrm{Pl}(\hat{\omega})}, \tag{2.36}$$

and

$$PrHyb(\omega) = PraPl(\omega) \sum_{\substack{Z \subseteq \Omega \\ \omega \in Z}} \frac{m(Z)}{\sum_{\hat{\omega} \in Z} PraPl(\hat{\omega})}, \tag{2.37}$$

where

$$PraPl(\omega) = \mathrm{Bel}(\omega) + \left(\frac{1 - \sum_{X \subseteq \Omega} \mathrm{Bel}(X)}{\sum_{X \subseteq \Omega} \mathrm{Pl}(X)} \right) \mathrm{Pl}(\omega). \tag{2.38}$$

Finally, the recently proposed $DSmP$ probability transform is introduced by Dezert and Smarandache in [105]. $DSmP$ distributes belief masses assigned to the non-singleton elements of Ω pro-

portionally, according to the belief masses assigned to the singleton elements. The transformation is defined in [105] using Dedekind lattices (*i.e.*, the hyper powerset of alternatives). We define it here in terms of the power set.

Definition 2.24: ***Dezert-Smarandache Probability Transform [105].*** *The Dezert-Smarandache probability transform is defined for any subset $A \subseteq \Omega$ and a corresponding BMA m as*

$$DSmP_\epsilon(A) = \sum_{X \subseteq \Omega} \frac{\sum_{\hat{\omega} \in A \cap X} m(\hat{\omega}) + \epsilon |A \cap X|}{\sum_{\hat{\omega} \in X} m(\hat{\omega}) + \epsilon |X|} m(X), \tag{2.39}$$

where $\epsilon \in [0, \infty]$ is a tuning parameter.

As $\epsilon \to 0$, $DSmP$ approaches Sudano's $PrBel$ transform [104]; as $\epsilon \to \infty$, $DSmP$ approaches $BetP$ [105]. The authors of [105] suggest selecting a small value for ϵ in order to minimize the amount of entropy present in the resulting subjective probabilities. With this in mind, the studies in this dissertation use $\epsilon = 0.001$ as suggested in [105]. Similar to $BetP$, $DSmP$ satisfies all three Kolmogorov Axioms and only needs to be computed for the singleton elements $\omega \in \Omega$.

2.6 Other Notable Techniques

The fusion operators discussed in this section represent a small sample of those which could be applied to the hard/soft fusion problem. There are many other fusion rules which could be used; we mention a few of them here with references for further reading. Bayesian belief networks [106] are a useful tool in information fusion and are beginning to see applications in hard/soft fusion scenarios (*e.g.*, [107]). We have chosen not to implement any Bayesian network approaches here, since we will already be evaluating an instance of Bayes' rule (*i.e.*, Bayes' rule of probability combination). There have also been Dempster-Shafer operators proposed specifically for hard/soft fusion. For example, the authors of [19] propose the *conditional update equation* which makes use of Fagin-Halpern belief function conditioning [108]. We have chosen not to implement the conditional update equation in this dissertation because its usage is dependent on the selection of multiple tuning parameters. There are also many potential operators from probabilistic opinion pooling [101], fuzzy logic [46, Chapter

2], and random sets [109]. For a recent state-of-the-art review in fusion operators, we direct the reader to [25].

Chapter 3: Human Response Simulation

We present the relevant cognitive psychology background for understanding the human response simulation methods employed by this dissertation. We begin by discussing a sequential sampling theory of human decision-making that makes use of *drift diffusion* or *random walk* processes [110]. We then discuss the recently proposed *two-stage dynamic signal detection* (2DSD) model [44] as an extension of the drift diffusion process that captures the interrelationships between choice accuracy, confidence accuracy, and response time. Using 2DSD, we investigate the performance of fusion systems on three benchmark tasks which were studied by Pleskac et al. in [44] and Shuli et al. in [45]: a line length discrimination task, a city population size discrimination task, and a random dot motion discrimination task.

3.1 Random Walk/Diffusion Theory

Stochastic human decision-making models have been researched by psychologists since the early 1960s [111–115]. The majority of this work addressed human decision making in *two-alternative forced choice* (TAFC) tasks, where a subject is presented with a scenario and is forced to choose between two alternatives. Consider a TAFC task defined by the set of alternatives $\mathcal{A} = \{A, \overline{A}\}$. Sequential sampling models of human decision making based on TAFC tasks assume that (1) internal evidence favoring each alternative is accumulated over time; (2) the process of internal evidence accumulation is subject to random fluctuations; and (3) a decision is made when sufficient amount of internal evidence has been accumulated for one of the two alternatives [116]. These three assumptions have been frequently modeled in the cognitive psychology literature as a discrete (*i.e.*, directed random walk) or continuous (*i.e.*, Wiener drift diffusion) stochastic processes [110].

Let $L(t)$ be the amount of internal evidence a human decision-maker holds towards alternative A over alternative $\overline{A}$ after t seconds. For example, when $L(t) > 0$, alternative A is favored. When $L(t) < 0$, alternative $\overline{A}$ is favored. When $L(t) = 0$, both alternatives are equally preferred. In this

thesis, we make use of the directed random walk model, which simulates $L(t)$ by accumulating a white noise process, $\epsilon(t)$, with a *drift rate*, δ, over time in discrete time steps of Δt [44]. The variance of the white noise process is given as σ^2, where σ is referred to as the drift coefficient. The internal evidence $L(t)$ is determined using the stochastic difference equation

$$L(t + \Delta t) = L(t) + \Delta L(t + \Delta t), \tag{3.1}$$

with $L(0) = L_0$ and

$$\Delta L(t + \Delta t) = \delta \Delta t + \epsilon(t + \Delta t)\sqrt{\Delta t}. \tag{3.2}$$

The drift rate, δ, is a stimulus-dependent model parameter that affects the decision accuracy and the response time of the response. The initial condition, L_0, is also a model parameter representing the bias of the decision maker. The drift coefficient, σ, can be a model parameter, however it is set as a constant in many cases [44, 45]. By construction, the values of $\Delta L(t)$ are independent and identically distributed for all discrete values of t, with a mean of $\delta \Delta t$ and variance $\sigma^2 \Delta t$. The value of δ is either positive or negative, depending on which alternative is true. The time step Δt is chosen to be small enough such that the directed random walk model converges to a Wiener drift diffusion process [44,117]. As a consequence, the evidence accumulation process approaches a normal distribution (*i.e.*, $L(t) \sim \mathcal{N}(t\delta, t\sigma^2)$). To make a decision, $L(t)$ is accumulated until it passes one of two thresholds, θ_A or $\theta_{\overline{A}}$. These thresholds correspond to the two possible alternatives A and $\overline{A}$. Alternative A is chosen if $L(t) > \theta_A$; alternative $\overline{A}$ is chosen if $L(t) < -\theta_{\overline{A}}$. The time at which this decision occurs is denoted t_d.

3.2 Two-stage dynamic signal detection

3.2.1 Motivation and Definition

A model of human decision making based on random walk/diffusion theory does not directly account for subjective decision confidence assessments [118, 119]. The modeling and simulation of decision confidences, however, is still important to the types of soft and hard/soft fusion operators investigated by this thesis. The *two-stage dynamic signal detection* (2DSD) [44] is a recently developed

sequential sampling model of human decision making that accounts for a wide range of phenomena in human decision making, while also taking into account the modeling of confidence assessments. The rationale behind 2DSD is that decisions and confidence assessments are modeled as a two stage process. In the first stage, a decision and decision response time are modeled according to random walk/diffusion theory. In the second stage, the internal evidence $L(t)$, modeled by random walk/diffusion theory, continues to accumulate for a fixed *interjudgment time*, τ. The resulting accumulated evidence, $L(t_c = t_d + \tau)$ is then mapped into a subjective decision confidence via a binning operation.

We implement two forms of 2DSD: the optional stopping (*i.e.*, free response) model proposed in [44], and an interrogation (*i.e.*, cued response) model proposed in [45]. In the *optional stopping* model, a subject provides decisions and confidence assessments when he/she is ready. In the *interrogation* model, an external event (*e.g.*, an experimenter) interrupts the deliberation of a subject and asks him/her for a response. We present both implementations below, as well as the corresponding parameter sets which represent the modeled human decision-maker.

Definition 3.1: ***Two-stage Dynamic Signal Detection (2DSD), Optional Stopping [44].*** *Let $\mathcal{A} = \{A, \overline{A}\}$ denote the alternatives of a TAFC task. In the optional stopping model of 2DSD, the deliberation between alternatives is modeled via the stochastic difference equation*

$$L(t + \Delta t) = L(t) + \Delta L(t + \Delta t), \tag{3.3}$$

where

$$\Delta L(t) = \delta \Delta t + \sqrt{\Delta t}\, \epsilon(t + \Delta t), \quad L(0) = L_0, \tag{3.4}$$

and where $\epsilon(t)$ is a simulated white noise process with zero mean and variance σ^2. The drift rate δ is either positive or negative, depending on whether A or $\overline{A}$ is true. The evidence accumulation is simulated until a threshold, either $\theta_A, \theta_{\overline{A}}$, is crossed (where $-\theta_{\overline{A}} < L_0 < \theta_A$). A decision, $a \in \mathcal{A}$, is

determined such that

$$a = \begin{cases} A & L(t) > \theta_A \\ \overline{A} & L(t) < -\theta_{\overline{A}} \\ \text{wait} & \text{otherwise} \end{cases}. \tag{3.5}$$

Let $\mathbf{P}^{(a)} = [p_1^{(a)} \cdots p_{K_a}^{(a)}]$ *denote the* K_a *possible confidence values associated with choosing* $a \in \mathcal{A}$ *at time* t_d*. The assigned confidence level* $p \in \mathbf{P}^{(a)}$ *associated with deciding* a *after waiting* $t_c = t_d + \tau$ *is given as*

$$p = p_i^{(a)} \quad \text{when} \quad L(t_c) \in [c_{i-1}^{(a)}, c_i^{(a)}], \tag{3.6}$$

where $c_0^{(a)} = -\infty$ *and* $c_{K_a}^{(a)} = \infty$ *for each* $a \in \mathcal{A}$*. The value* τ *is known as the interjudgment time. The remaining confidence bin parameters* $\mathbf{C}^{(a)} = [c_1^{(a)} \cdots c_{K_a-1}^{(a)}]$ *are chosen such that* $c_{i-1} < c_i$ *for each* $i \in \{1, \ldots, K_a - 1\}$ *and each* $a \in \mathcal{A}$*. The time step* Δt *is chosen small enough such that the evidence accumulation approaches a continuous process. A modeled human decision-maker in this construction consists of a set of* $2(K_A + K_{\overline{A}}) + 4$ *parameters,* $\mathcal{S}_{OST}$*, where*

$$\mathcal{S}_{OST} = \{\delta, \sigma, L_0, \theta_A, \theta_{\overline{A}}, \tau, \mathbf{P}^{(A)}, \mathbf{P}^{(\overline{A})}, \mathbf{C}^{(A)}, \mathbf{C}^{(\overline{A})}\}. \tag{3.7}$$

The interrogation version of 2DSD is similar to the optional stopping version. The most notable difference is that the subject chooses alternative A when prompted at t_d if $L(t_d) > 0$. Otherwise, the subject chooses alternative $\overline{A}$. Another cue is presented to the subjects at $t_c > t_d$ to make a decision confidence assessment. The mapping between $L(t_c)$ and confidence values is again performed through a binning operation.

Definition 3.2: ***Two-stage Dynamic Signal Detection (2DSD), Interrogation [45].*** *Let* $\mathcal{A} = \{A, \overline{A}\}$ *denote the alternatives of a TAFC task. In the interrogation model of 2DSD, the deliberation between alternatives is again modeled via the stochastic difference equation*

$$L(t + \Delta t) = L(t) + \Delta L(t + \Delta t), \tag{3.8}$$

where

$$\Delta L(t) = \delta \Delta t + \sqrt{\Delta t}\, \epsilon(t + \Delta t), \quad L(0) = L_0, \tag{3.9}$$

and where $\epsilon(t)$ is a simulated white noise process with zero mean and variance σ^2. The drift rate δ is again either positive or negative, depending on whether A or $\overline{A}$ is true. The evidence accumulation occurs until a predetermined time t_d. A decision, $a \in \mathcal{A}$, is determined such that

$$a = \begin{cases} A & L(t_D) > 0 \\ & \\ \overline{A} & \text{otherwise} \end{cases}. \tag{3.10}$$

Let $\mathbf{P}^{(a)} = [p_1^{(a)} \cdots p_{K_a}^{(a)}]$ denote the K_a possible confidence values associated with choosing $a \in \mathcal{A}$ at time t_d. A response for a confidence assessment is cued at a predetermined time $t_c > t_d$. The assigned confidence level $p \in \mathbf{P}^{(a)}$ associated with deciding a at time t_c is given as

$$p = p_i^{(a)} \quad \text{when} \quad L(t_c) \in [c_{i-1}^{(a)}, c_i^{(a)}], \tag{3.11}$$

where $c_0^{(a)} = -\infty$ and $c_{K_a}^{(a)} = \infty$ for each $a \in \mathcal{A}$. The remaining confidence bin parameters $\mathbf{C}^{(a)} = [c_1^{(a)} \cdots c_{K_a-1}^{(a)}]$ are chosen such that $c_{i-1} < c_i$ for each $i \in \{1, \ldots, K_a - 1\}$ and each $a \in \mathcal{A}$. The time step Δt is chosen small enough such that the evidence accumulation approaches a continuous process. A modeled human decision-maker in this construction consists of a set of $2(K_A + K_{\overline{A}}) + 3$ parameters, $\mathcal{S}_{INT}$, where

$$\mathcal{S}_{INT} = \{\delta, \sigma, L_0, t_d, t_c, \mathbf{P}^{(A)}, \mathbf{P}^{(\overline{A})}, \mathbf{C}^{(A)}, \mathbf{C}^{(\overline{A})}\}. \tag{3.12}$$

Regardless of which implementation is used, the 2DSD parameter set for a given subject is estimated using decision accuracy, confidence accuracy, and response time statistics and a maximum likelihood estimation technique (*i.e.*, quantile maximum probability estimation [120]). For more information on the validation of the 2DSD model with respect to actual human responses, we direct the reader to [44] and [45].

3.2.2 Trial Variability of Parameters

Using the above formulations, 2DSD may not always be able to predict the relative response times associated with correct and incorrect responses. For example, the study in [121] demonstrates that incorrect responses can sometimes be slower than correct responses for difficult decision tasks. Sequential sampling models of human decision making may predict the opposite. To compensate, 2DSD can be amended by considering the drift rate δ and initial condition L_0 as random variables chosen at the start of each simulation, instead of model parameters[1]. Specifically, the authors of [44] suggest that δ be normally distributed with mean ν and variance η^2, and that L_0 be uniformly distributed in an interval of size s_z and centered at z. That is,

$$\delta \sim \mathcal{N}(\nu, \eta^2), \tag{3.13}$$

and

$$L_0 \sim \mathcal{U}(z - 0.5s_z, z + 0.5s_z). \tag{3.14}$$

In these cases, the parameter δ is replaced by ν and η, and L_0 is replaced by z and s_z. At the beginning of a simulation, a value of δ and L_0 would be randomly chosen based on their respective distributions, and then used to implement the evidence accumulation process, $L(t)$.

3.2.3 Non-Decision and Non-Interjudgment Time

As previously mentioned, the statistics of the subject's decision and confidence assessment response times, t_d and t_c are used in the estimation of a 2DSD parameter set. A non-negligible portion of t_d and t_c may be attributed to recognizing the stimulus and responding according to the constraints of the experiment (*i.e.*, motor time). Let the observed decision and interjudgment times for a subject be given as t'_d and τ'. Here, t_d is the decision time predicted by (or used in) the model, and τ is interjudgment time used in the model. To account for motor time in the estimation of the model parameters, two additional parameters known as the *mean nondecision time* t_{ED} and the *mean*

[1] This is also done for random walk/diffusion decision models, as described in [122].

nonjudgment time t_{EJ} are appended to the 2DSD parameter set in [44] and [45], such that

$$t'_d = t_{ED} + t_d, \tag{3.15}$$

and

$$\tau' = t_{EJ} + \tau. \tag{3.16}$$

In this formulation, t_{ED} and t_{EJ} are estimated per subject in place of t_d and τ. For information on how this change affects the parameter estimation of 2DSD subject model, we direct the readers to [44] and [45].

3.2.4 State-dependent decay

In [44], the authors state that the stochastic process $L(t)$ associated with the 2DSD model accumulates at a constant average rate (*i.e.*, with mean $\delta\Delta t$). A recent set of studies in [45] has suggested that the accumulation of $L(t)$ may tend to decay exponentially towards a steady state value. To account for this trend, the evidence accumulation increment $\Delta L(t)$ can be generalized to include a decay parameter $\gamma \geq 0$ in the diffusion process such that

$$\Delta L(t) = (\delta - \gamma L(t))\,\Delta t + \sqrt{\Delta t}\epsilon(t + \Delta t). \tag{3.17}$$

This decay parameter generalizes the drift diffusion model to an *Ornstein-Uhlenbeck* process with drift [45, 122]. In this construction, it is likely that evidence accumulated for correct choices will tend to decelerate whereas evidence accumulated for incorrect choices will accelerate to the correct state. As the interjudgment time is increased, this acceleration/deceleration implies that confidence assessments will remain relatively stable for correct choices, whereas confidence in incorrect choices falls [45]. Notice that for $\gamma = 0$, Equation (3.17) reduces to Equation (3.9).

3.2.5 Evidence (Drift Rate) Attenuation

For certain subjects and decision tasks, there may also be an attenuation in the rate of incoming evidence after a choice is made. This attenuation was observed in a set of decision tasks studied

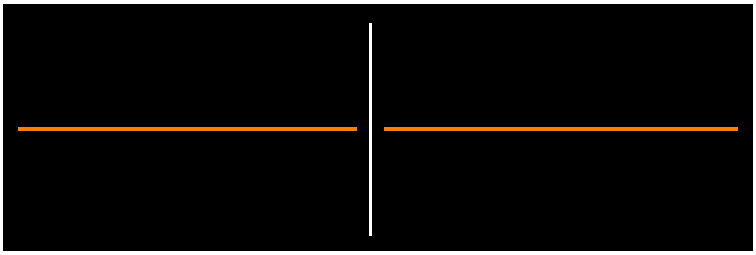


Figure 3.1: Stimulus example for the line length discrimination task. A 32.00 millimeter long line is shown on the left, and a 33.87 millimeter long line is shown on the right.

in [45], and attributed to reduced attention between the declaration of a decision and decision confidence assessment. To incorporate evidence attenuation, a parameter[2] ρ is appended to the 2DSD parameter set, to produce a new drift rate δ_{IJT} that replaces δ at the start of the interjudgment period, where

$$\delta_{IJT} = \rho\delta. \tag{3.18}$$

3.3 Human response tasks

We make use of three decision tasks studied in [44,45]. These are: a line length discrimination task, a city population size discrimination task, and a random dot motion discrimination task. We introduce these three tasks below, but direct the reader to [44] and [45] for more information regarding the experimental design and the discussion of fitting the human responses to the 2DSD model.

3.3.1 Line length discrimination task

In the *line length discrimination task*, subjects were positioned in front of a computer screen and asked to compare pairs of two horizontal orange lines with different lengths on a black background. The lines were separated by a vertical white line of fixed length as depicted in Figure 3.1. Subjects were asked to make a declaration on which of the two lines presented (*i.e.*, right or left) is longer, and then assess their own confidence in that decision on a subjective probability scale. The task becomes easier as the length difference between the two lines increases. Longer line length differences correspond to higher values of the estimated drift rates δ for each subject.

[2]This parameter in [45] is denoted θ. We use ρ in this thesis for consistency with other symbols.

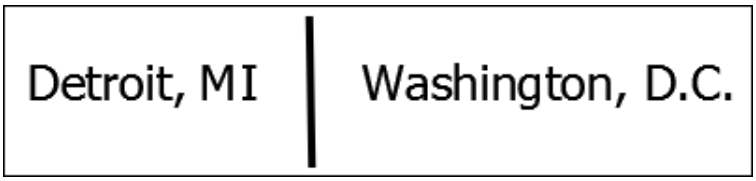


Figure 3.2: Stimulus example for the city population size discrimination task. According to the 2010 U.S. Census estimate, Detroit, MI has a population rank of 18 and Washington, D.C. has a population rank of 24.

3.3.2 City population size discrimination task

In the *city population size discrimination task*, subjects were positioned in front of a computer screen and asked to compare pairs of U.S. cities. Included with the cities were their corresponding state abbreviations. Similar to the line length task, the cities were separated by a vertical line of fixed length as in Figure 3.2. Subjects were asked to make a declaration towards which of the two cities (*i.e.*, left or right) has a higher population, and then assess their own confidence in that decision on a subjective probability scale. The difficulty of the city population size discrimination task decreases as the difference between the populations of the two cities increases. Equivalently, one could also consider U.S. city population ranks and rank differences, resulting from enumerating the cities from the most populated to the least populated. Higher city population differences (or population rank differences) correspond to higher values of the estimated drift rates δ for each subject.

3.3.3 Random dot motion discrimination task

In the *random dot motion discrimination task*, subjects were positioned in front of a computer screen and presented with a circular field of moving white dots on a black background, similar to Figure 3.3. In the center of the circle was a stationary red dot. A proportion of the dots were programmed to move either towards the left or right of the field, whereas the remainder of the dots were programmed to appear in random locations. The chance that a given dot remains moving in the "correct" direction is termed *motion coherence*. Subjects were asked to make a declaration as to which overall direction the dots were moving in (*i.e.*, left or right), and then assess their own confidence in that decision on a subjective probability scale. The task becomes easier as the

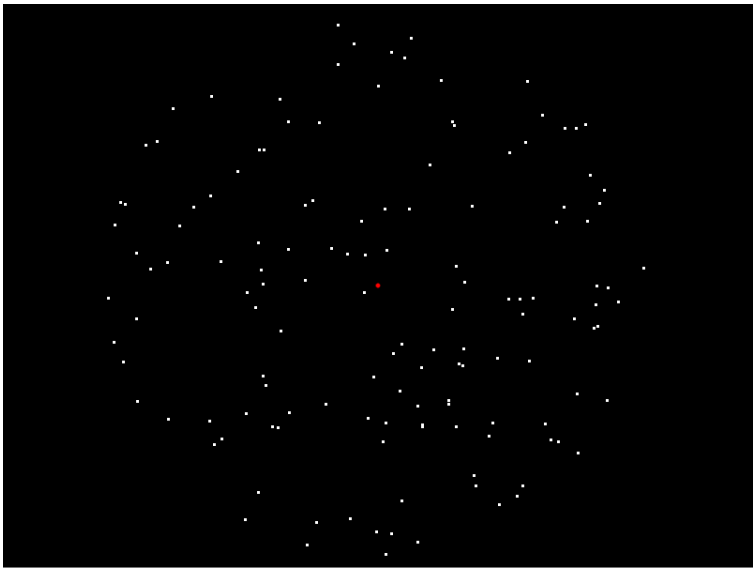


Figure 3.3: Stimulus example instance for the random dot movement task. Dots move either towards the right or left of the circle, masked by a subset of randomly moving dots.

motion coherence increases [45]. Higher motion coherence values correspond to higher values of the estimated drift rates δ for each subject.

3.4 2DSD Simulations

We present a few example 2DSD simulations to supplement the discussions presented in this chapter. The 2DSD parameter estimates for subjects #1, #2, and #3 in [44, Table 6] were simulated over a line length discrimination task. The 2DSD parameters for this particular task are fit to a 2DSD optional stopping model, employing trial variability on the drift rate, δ, and the initial condition, L_0. For this specific task, the authors of [44] constrain the 2DSD parameters such that the decision thresholds for each alternative are equal (*i.e.*, $\theta_A = \theta_{\overline{A}}$) and the confidence bin parameters for each alternative are negatives of each other (*i.e.*, $\mathbf{C}^{(A)} = -\mathbf{C}^{(\overline{A})}$). The simulated stimulus presented to the three subjects was a pair of lines which were 32.00 millimeters and a 32.27 millimeters in length. The 32.27 millimeter line was on the right side of the screen, whereas the the 32.00 millimeter line was on the left. Each subject was simulated to provide a decision as to which line was longer (*i.e.*, the right or left line) and a corresponding subjective probability to represent his/her confidence in

Table 3.1: Estimated correct detection probabilities under the speed focus and the accuracy focus for Subjects #1-3 when simulated on the line length discrimination task. Each simulation involved comparing the lengths of a 32.00 millimeter and a 32.27 millimeter long line placed on the left and right sides of the screen respectively. The estimation error is approximately ± 0.01 at the 95% confidence level.

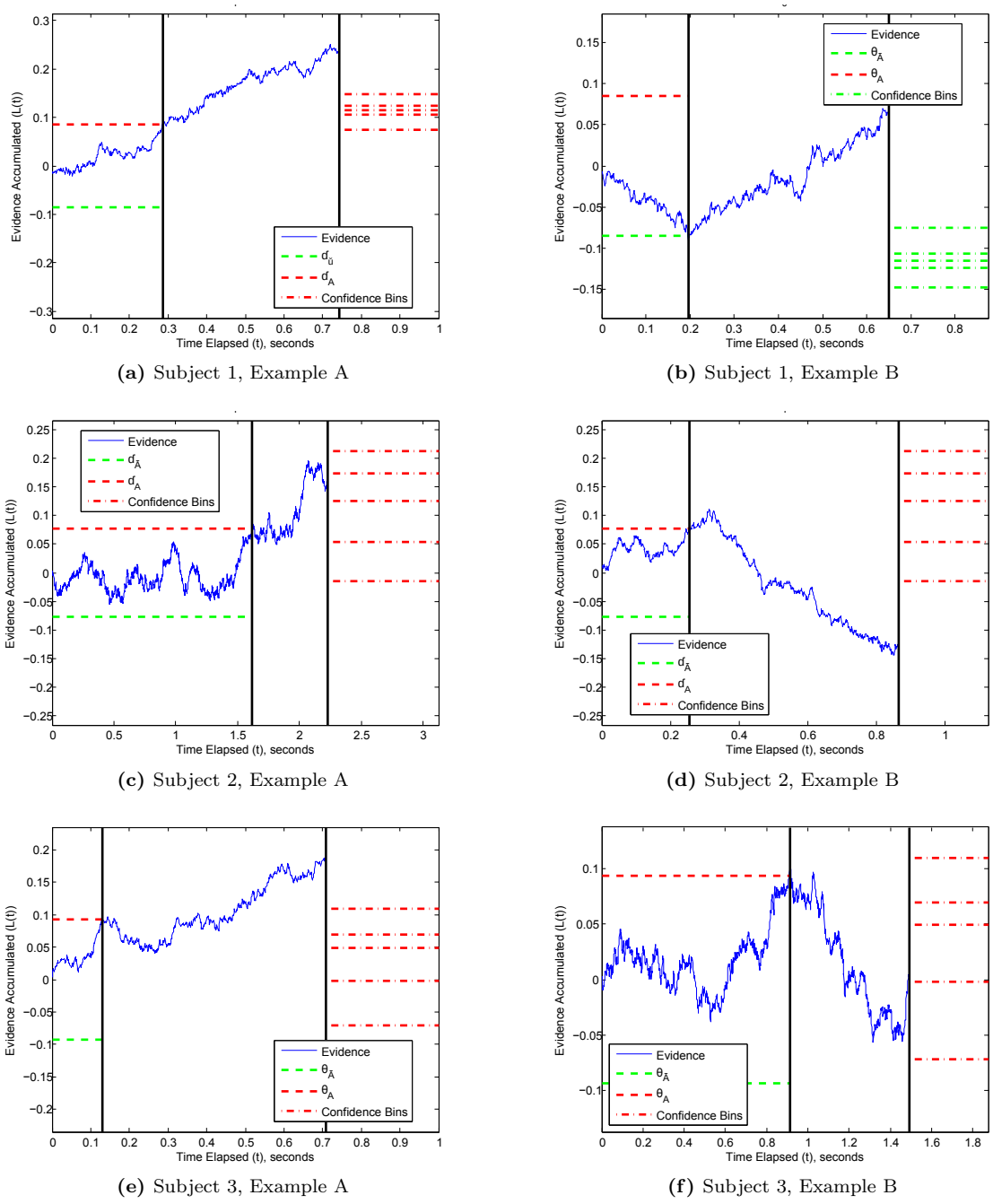
	Accuracy Focus	Speed Focus
Subject 1	0.61	0.60
Subject 2	0.57	0.56
Subject 3	0.58	0.55

that decision. According to the 2DSD models in [44], possible subjective probabilities were simulated from $\{0.50, 0.60, \ldots, 1.00\}$. Additionally, two types of responses were simulated from the subjects: one in which subjects were focused on provided accurate responses, and another in which subjects were focused on providing fast responses.

Figure 3.4 shows two representations of the accumulated evidence $L(t)$ for each subject when focused on providing accurate responses. Also shown on each graph are the relevant thresholds, θ_A and $-\theta_{\overline{A}}$, and the confidence bins $c_1, \ldots, c_5$ parameters (depending on which response was declared). The comparison of a 32.00 and 32.27 millimeter long lines is difficult, hence there are times when the subjects will declare an alternative and reconsider during the interjudgment time (*e.g.*, Figure 3.4b, Figure 3.4d, and Figure 3.4f). Alternatively, the subject may waiver in their deliberation before settling on a decision (*e.g.*, Figure 3.4c).

Table 3.1 shows the estimated correct detection rates of Subjects #1-3 under the speed focus and the accuracy focus. The estimates were obtained by simulating the decisions of each subject over 10,000 trials. The corresponding estimation error was at most ± 0.01 at the 95% confidence level. For these three subjects on this task, time constraints do not seem to have a profound affect the statistics for each subject's correct detection rate. In fact, only Subject 3 exhibits a statistically significant change in the correct detection rate. These results are most likely attributed to the difficulty level simulated here for this task; the task is so difficult that allowing subjects more time to respond does not give them a significantly better chance at providing a better decision.

Figure 3.5 shows normalized histograms for the assessed confidence values of each subject under



(a) Subject 1, Example A

(b) Subject 1, Example B

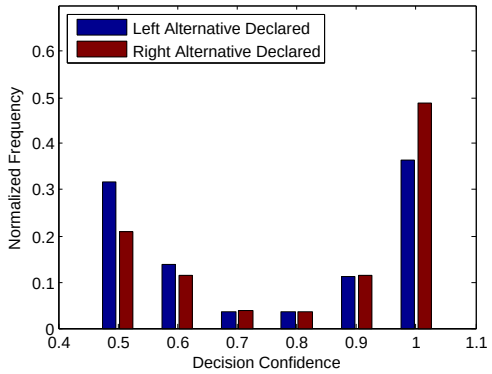
(c) Subject 2, Example A

(d) Subject 2, Example B

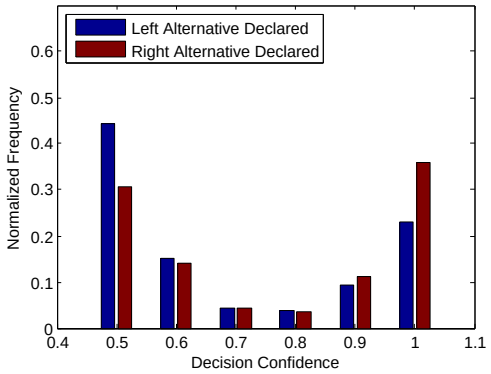
(e) Subject 3, Example A

(f) Subject 3, Example B

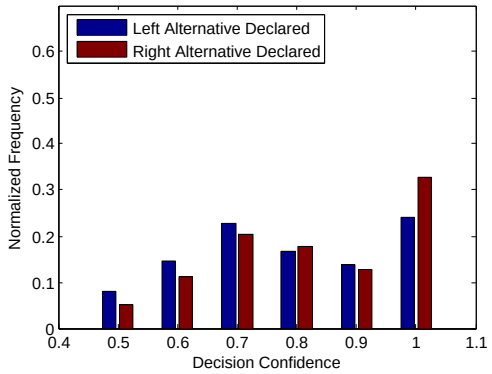
Figure 3.4: Four example 2DSD simulations, showing two simulations for the accumulated evidence $L(t)$ over time for three subjects on the line length discrimination task. Each simulation involved comparing the lengths of a 32.00 millimeter and a 32.27 millimeter long line. Parameters were obtained from [44] when subjects were asked to focus on making accurate responses. The time of decision declaration and confidence assessment are shown are vertical black lines. All thresholds and binning parameters are also shown.



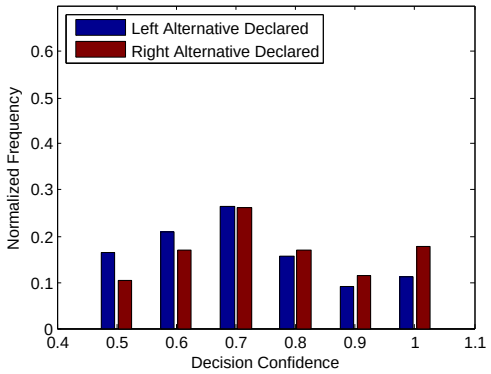
(a) Subject 1, Accuracy Focus



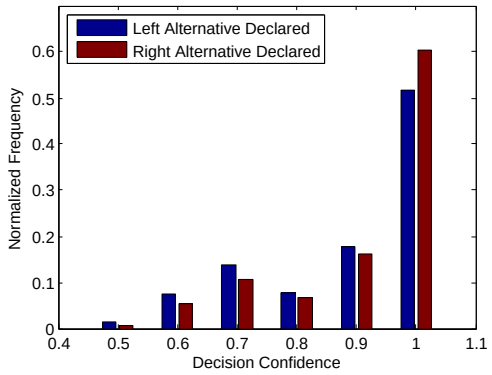
(b) Subject 1, Speed Focus



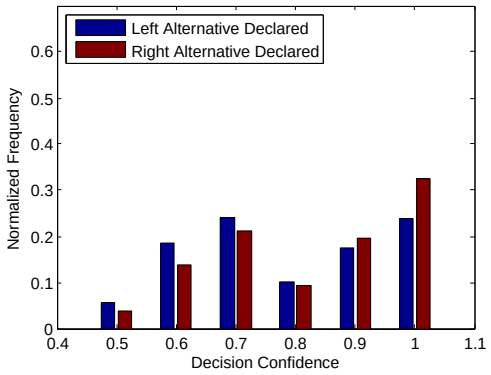
(c) Subject 2, Accuracy Focus



(d) Subject 2, Speed Focus



(e) Subject 3, Accuracy Focus



(f) Subject 3, Speed Focus

Figure 3.5: Normalized histograms for the decision confidence values simulated for three (3) subjects on the line length discrimination task of [44]. Each simulation involved comparing the lengths of a 32.00 millimeter and a 32.27 millimeter long line placed on the left and right sides of the screen respectively. For each subject, two types of responses were simulated: one which placed an emphasis on accurate responses, and another which placed an emphasis on fast responses.

the speed focus and the accuracy focus. The support of each histogram is the range of possible confidence values which could be simulated for this task (*i.e.*, $\{0.50, 0.60, \ldots, 1.00\}$). Each histogram was estimated over 10,000 simulation trials. The frequencies were normalized such that each bar in the histogram represents the conditional probability that one of the six possible confidence levels was selected, given that either the left or right line was declared as being longer. The three subjects were observed to declare full belief (*i.e.*, a subjective confidence of 1.00) more often towards the correct alternative than towards the incorrect alternative. The three subjects were also observed to declare indifference (*i.e.*, a subjective confidence of 0.50) more often towards the incorrect alternative than towards the correct alternative. When focusing on accurate responses, subject #2 appeared to make frequent use of all possible confidence assessment values (Figure 3.5c). Subject #1 appeared to declare mostly full belief or indifference (*i.e.*, a subjective confidence of 0.50) (Figure 3.5a). Subject #3 was observed to declare indifference for either alternative infrequently (Figure 3.5e). When focusing on fast responses, all subjects appeared to express indifference more often (*i.e.*, increased conditional probability towards selecting a subjective confidence of 0.50, given either alternative) and full belief less often (*i.e.*, decreased conditional probability towards selecting a subjective confidence of 1.00, given either alternative). Subject #2 (Figure 3.5d) appeared to also assess decision confidences more frequently on intermediate values (*i.e.*, 0.60 and 0.70) than what was observed when focusing on accurate responses. Otherwise, the overall shape of the normalized histograms for each subject when focusing on fast responses appeared similar to when they focused on accurate responses.

Part II

Performance Assessment Techniques for Hard/Soft Fusion Systems

Chapter 4: Soft Fusion Studies with Binary Alternatives

We simulate responses from a set of human decision-makers using two-stage dynamic signal detection (2DSD) (Chapter 3) in order to determine the performance of a selection of the soft fusion operators described in Chapter 2. Specifically, the studies presented in this chapter focus on the effects of time constraints, foreknowledge of subject reliabilities, subjective confidence, and superior/inferior performing sources on the performance of a fusion operator combining information on binary decision tasks. As a preliminary, we discuss the metrics used for assessing the decision-making performance of the subjects and of the fusion operators. Then, we present four case studies using 2DSD models from [44]. These case studies all focus on implementing several soft fusion operators on binary decision tasks.

4.1 Subject and Fusion Operator Performance Metrics

4.1.1 Decision Performance

Recall from Chapter 3 that the alternatives of a two-alternative forced choice task (TAFC) were denoted as $\mathcal{A} = \{A, \overline{A}\}$. To coincide with the notation used in classical detection theory [41], we can further specify these alternatives as the binary hypotheses $H_1 = A$ and $H_0 = \overline{A}$. Consider a pool of N subjects, enumerated according to $i \in \{1, 2, \ldots, N\}$. Recall from Chapter 2 that the decisions of each subject are denoted as u_i. If we do not consider subject-generated decision confidence assessments, we can define performance in terms of the false alarm p_{F_i} and correct detection p_{D_i} (or missed detection p_{M_i}) rates that were discussed in Chapter 2. For convenience, we repeat them below as

$$p_{F_i} = P(u_i = +1|H_0), \tag{4.1}$$

$$p_{D_i} = P(u_i = +1|H_1) \tag{4.2}$$

and

$$p_{M_i} = 1 - p_{D_i}. \tag{4.3}$$

Recall from Chapter 2 that $u_i = +1$ when the i^{th} source declares H_1, and $u_i = -1$ when it declares H_0. One could also look at the *total classification error rate* $p_{E_i} = p_{F_i}P(H_0) + p_{M_i}P(H_1)$ if the a priori probabilities $P(H_0)$ and $P(H_1)$ are available. When the fusion center output is a global decision $u_0 \in \{-1, +1\}$, the false alarm, correct detection, missed detection, and total classification error rates associated with it are analogously defined as p_{F_0}, p_{D_0}, p_{M_0}, and p_{E_0}. Since the studies presented in this dissertation are not concerned with utility and/or risk functions, we consider better performing sources (or fusion operators) as those which exhibit lower total classification error rates (*i.e.*, or similarly, lower false alarm and missed detection rates).

4.1.2 Decision Confidence (Belief) Performance

If decision confidence self-assessments are provided by the subjects, then subject performance can be assessed using a quadratic scoring rule. Suppose a subject provides a decision $u_i \in \{-1, +1\}$ and subjective confidence assessment $p_i \in [0, 1]$ on a TAFC task. This information can be summarized in the subjective probability assignment

$$\mathcal{P}_{\mathcal{A}}(u_i) = p_i \tag{4.4}$$

$$\mathcal{P}_{\mathcal{A}}(\overline{u_i}) = 1 - p_i \tag{4.5}$$

Let $u^* \in \mathcal{A}$ represent the correct choice. Better performing subjects should frequently assign high confidence to the correct choice. As a subject assigns less belief to u^*, the quality of the decision and confidence assessment decreases. Motivated by [44], we denoted this idea of subject opinion quality in [123] as *evidence strength*, ξ, where

$$\xi(u_i, p_i|u^*) = \begin{cases} 1 - (1 - p_i)^2 & u_i = u^* \\ 1 - p_i^2 & u_i \neq u^* \end{cases}. \tag{4.6}$$

Evidence strength is derived from the quadratic scoring rule known as *Brier score* [124]. An evidence strength value of one means that the subject has assigned the correct outcome full decision confidence. An evidence strength value of zero means that the subject has assigned the correct outcome outcome no decision confidence. In this thesis, we consider the sample average of evidence strengths, $\overline{\xi}_i$ as a measure of the i^{th} subject's decision and confidence assessment performance. Subjects with higher average evidence strengths will tend to give better confidence assessments than those with lower average evidence strengths.

When the output fusion center is a set of beliefs aggregated in the form of a subjective probability assignment, a performance metric similar to average evidence strength can be used. Recall from Chapter 2 however, that the fusion operators from Dempster-Shafer (DS) theory in general produce a set of belief/plausibility ranges that can be used to describe a *class of subjective probability assignments* supported by the fused evidence. We quantify fusion operator performance as the nearness of its combined subjective probability assignments to one which assigns the correct alternative $u^* \in \mathcal{A}$ full belief (*i.e.*, a subjective probability of one).

The result is a class of evidence strengths defined by the intervals $[\xi_{\text{Bel}}, \xi_{\text{Pl}}]$, where

$$\xi_{\text{Bel}} = \xi(u^*, \text{Bel}(u^*)|u^*), \tag{4.7}$$

and

$$\xi_{\text{Pl}} = \xi(u^*, \text{Pl}(u^*)|u^*). \tag{4.8}$$

The functions Bel$(\cdot)$ and Pl$(\cdot)$ are evaluated based on an aggregated belief mass assignment (BMA) after fusion.The lower envelope ξ_{Bel} can be thought of as a measure of the *accuracy* of the combination operator, and the size of the interval $(\xi_{\text{Pl}} - \xi_{\text{Bel}})$ can be thought of as the *precision* of the combination operator. Accurate belief combination operators will tend assign subjective probabilities close to one to the correct outcome, resulting in values of ξ_{Bel} close to one. Precise belief combination operators will tend to produce more specific evidence, resulting in values of $(\xi_{\text{Pl}} - \xi_{\text{Bel}})$ being close to zero. Since $\text{Bel}(u^*) \leq \text{Pl}(u^*) \leq 1$ [43], it follows that $\xi_{\text{Bel}} \leq \xi_{\text{Pl}} \leq 1$. Hence systems with high accuracy

(*i.e.*, ξ_{Bel} close to one) will also be very precise (*i.e.*, $(\xi_{\text{Pl}} - \xi_{\text{Bel}})$ close to zero). In the Bayesian case, $(\xi_{\text{Pl}} - \xi_{\text{Bel}}) = 0$ since $\text{Bel}(u^*) = \text{Pl}(u^*) = \mathcal{P}(u^*)$ [43].

4.2 Discounted Decisions and Confidences

There are many ways to take into account the performance of the subject when performing fusion. For example, if decision error rates (*i.e.*, false alarm and missed detection rates) can be estimated, then the Chair and Varshney Fusion rule can be applied[1] using their decisions. When decisions and confidences are used, a *discounting operation* can be used to take the average source evidence strength into account. Recall from Chapter 3 that Ω describes the set of alternatives for a possible phenomenon. For BMAs, the discounting operation used here is well known [59] and can be represented mathematically for any $X \subseteq \Omega$ as

$$m_i(X;\alpha_i) = \begin{cases} \alpha_i m_i(X) & X \neq \Omega \\ \alpha_i m_i(X) + (1-\alpha_i) & X = \Omega \end{cases} \tag{4.9}$$

where $m_i(\cdot)$ and α_i are the BMA and the discount rate associated with the i^{th} subject. For subjective probability assignments, this operation can be similarly written [123, 125] as

$$\mathcal{P}_i(\omega;\alpha_i) = \alpha_i \mathcal{P}_i(\omega) + |\Omega|^{-1}(1-\alpha_i). \tag{4.10}$$

Here, these discounting operations are used to scale the evidence present in the belief mass or subjective probability assignment according to a subject's discount rate α_i. When $\alpha_i = 1$, the BMA (or subjective probability assignment) is unmodified. When $\alpha = 0$, the BMA (or subjective probability) is transformed into a vacuous belief mass assignment (or a subjective probability assignment with equiprobable alternatives).

4.3 Fusion Study 1: Different Fusion Operator Input Considerations

A subset of this study's results were previously published in [123].

[1] Optimal decision performance however may not necessarily be guaranteed as the conditional probabilities of the sources may not be stationary or independent conditioned on the hypotheses.

4.3.1 Motivation

As mentioned in Chapter 2, Bayesian epistemology and DS theory are popular tools for combining beliefs (*i.e.*, decisions and decision confidence assessments). To use the combination operators from either area, subject decisions u_i and confidence assessments p_i must be translated into BMAs (or in the case of Bayes' Rule, subjective probability functions). In this study, we seek to determine how such operator input constructions affect the performance of the fusion operator when working with human opinions. For Bayes' rule of probability combination, we consider a subjective probability assignment that assigns probability s_i towards the subject's choice u_i and divides the remainder equally amongst the remaining alternatives (for the general case of M alternatives over the set Ω)[2]. That is,

$$\mathcal{P}_i(\omega) = \begin{cases} s_i & \omega = u_i \\ \frac{1}{M-1}(1-s_i) & \text{otherwise} \end{cases}. \tag{4.11}$$

To implement the operators from Dempster-Shafer theory, a BMA must be constructed. Here we make use of the simple support function, defined similarly as

$$m_i(X) = \begin{cases} s_i & X = u_i \\ 1-s_i & X = \Omega \\ 0 & \text{otherwise} \end{cases}. \tag{4.12}$$

In this study, we seek to investigate the performance of a fusion operator in terms of the value s_i and the use of discounting (*i.e.*, Equations 4.9 and 4.10).

4.3.2 Experimental Setup

Decisions and confidence assessments were simulated from six subjects under the line length discrimination task and city population size discrimination tasks using the 2DSD models given in [44]. For the line length discrimination task, subjects were simulated as comparing a pair of 32.00 and

[2]For the binary case, $\Omega = \mathcal{A}$ and $M = 2$. We define these equations for M alternatives for use later in Chapters 5 and 6.

32.27 millimeter long lines. For the city population size discrimination task, subjects were simulated as comparing U.S. city pairs which exhibited a population rank difference between 10 and 18 (*e.g.*, Houston, TX versus Baltimore, MD). For both decision tasks, subjects were simulated to provide decisions and decision confidence assessments when focusing on either accurate or fast responses. The following four BMAs (or subjective probability) construction cases were examined.

1. **Decisions Only:** Each subject's source strength s_i was estimated beforehand by estimating the subject's proportion of correct decisions over 2,000 simulations. The subject's simulated confidence value p_i was not used.

2. **Confidences Only:** Each subject's source strength s_i was taken as the simulated confidence value p_i.

3. **Detection Rate Discounting:** Each subject's source strength s_i was taken as the simulated confidence value p_i. The resulting simple support function was then discounted using expression 4.9 (or in the Bayesian case, expression 4.10). The discount rates for each subject α_i was taken as their proportion of correct decisions over 2,000 simulations.

4. **Evidence Strength Discounting:** Each subject's source strength s_i was taken as the simulated confidence value p_i. The resulting simple support function was then discounted using expression 4.9 (or in the Bayesian case, expression 4.10). The discount rate for each subject α_i was taken as the subject's average evidence strength $\overline{\xi}_i$, estimated beforehand over 2,000 simulations.

Simulated decisions and confidence assessments were used to generate subjective probability assignments and simple support functions for each subject according to the four fusion operator input cases mentioned above. The following five fusion operators were implemented: Bayes' rule of probability combination [97], Dempsters Rule of Combination (DRC) [43], Yagers Rule [94], Proportional Conflict Redistribution Rule #5 (PCR5) [97], and the Consensus Operator [29]. The combination order of each source was fixed across all fusion rules, such that the first source of [44] was combined with the second, and then the third, and so on. A total of 2,000 simulation trials were conducted for

each case, and the average evidence strengths associated with each fusion operator were determined.

4.3.3 Results

Figures 4.1 through 4.4 each show the average evidence strength values for each of the five fusion operators when using each of the four BMA (or subjective probability) construction cases discussed above. Figures 4.1 and 4.2 each show the performance of the fusion operators on the line length discrimination task when subjects were simulated to provide accurate and fast responses respectively. Figures 4.3 and 4.4 show the performance of the fusion operators on the city population size discrimination task, also when subjects were simulated to provide accurate and fast responses respectively. In each of these four figures, horizontal lines are included to represent the best and worst average evidence strength values amongst the subjects. The height of each colored bar represents the minimum average evidence strength produced after fusion (*i.e.*, $\overline{\xi}_{\mathrm{Bel}}$). The height of the colored and clear bars represents the maximum average evidence strength after fusion (*i.e.*, $\overline{\xi}_{\mathrm{Pl}}$). As discussed earlier, higher colored bars represent higher accuracy fusion rules, whereas smaller clear bars represent more precise fusion rules.

Based on the parameters of the 2DSD models used in this study, it is possible that subjects can provide conflicting decisions and decision confidence assessments (*i.e.*, assess full belief in conflicting alternatives). Therefore, it is not surprising that Bayes' rule of probability combination and DRC could not be used in the "Confidences Only" BMA formulation case (see [59] for a more detailed discussion). It is also not surprising that many fusion operators exhibit inferior performance in the "Confidences Only" case, since it is the only one in which the statistical performance of the subjects are not taken into account. Placing an increased focus on fast responses did not change the overall performance of the fusion operators significantly. Additional decision time in these cases most likely does not make a considerable difference in subject decision and confidence assessment quality because of the difficulty level of the tasks. In terms of the fusion operators, Yager's rule was found to exhibit the worst performance across all simulation cases, followed very closely by the consensus operator. In many cases, the minimum average evidence strength of Yager's rule was lower than the average evidence strength of the worst performing source. Excluding Bayes' rule of probability

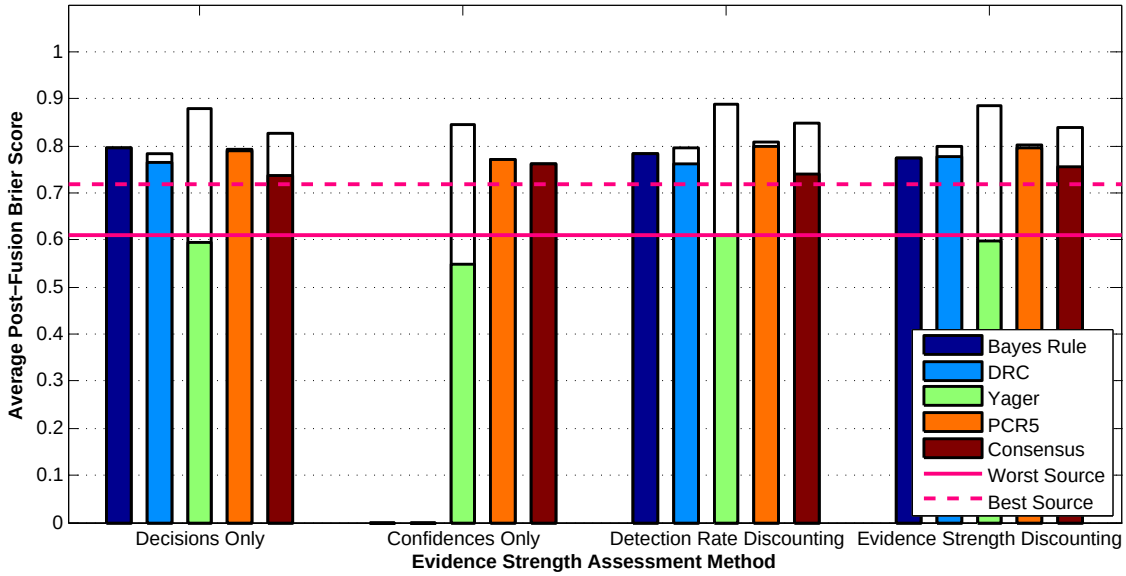


Figure 4.1: Average evidence strength of fusion operators on the line length discrimination task of [44], showcasing four BMA (or subjective probability) construction cases. Subjects were simulated as comparing a 32.00 millimeter line with a 32.27 millimeter line while focusing on providing accurate responses. Higher colored bars indicate better accuracy performance. Smaller clear bars indicate better precision performance.

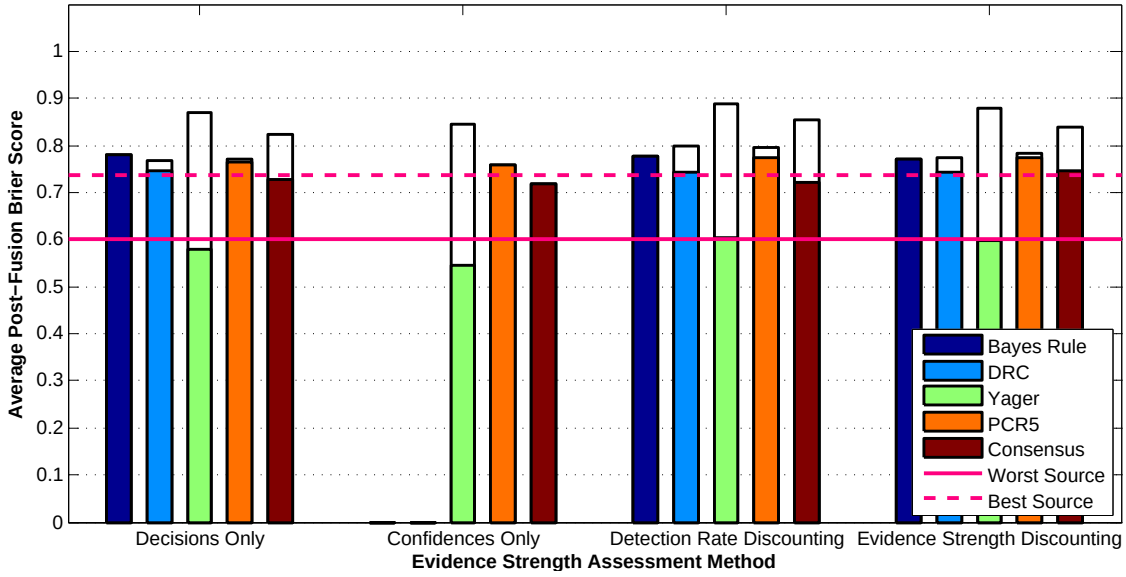


Figure 4.2: Average evidence strength of fusion operators on the line length discrimination task of [44], showcasing four BMA (or subjective probability) construction cases. Subjects were simulated as comparing a 32.00 millimeter line with a 32.27 millimeter line while focusing on providing fast responses. Higher colored bars indicate better accuracy performance. Smaller clear bars indicate better precision performance.

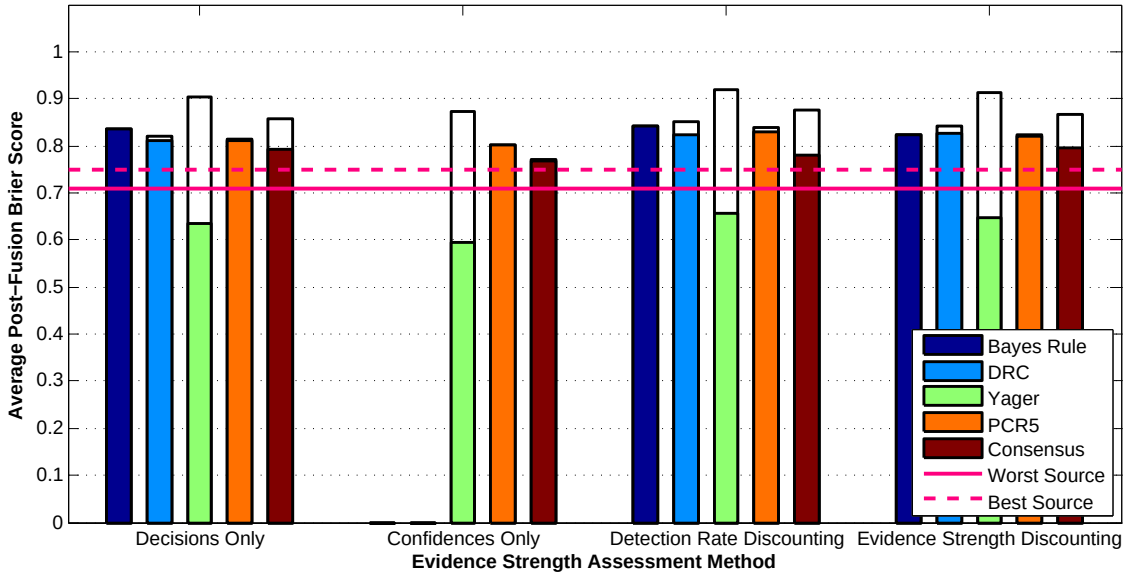


Figure 4.3: Average evidence strength of fusion operators on the city population size discrimination task of [44], showcasing four BMA (or subjective probability) construction cases. Subjects were simulated as comparing U.S. city pairs having a population rank difference between 10 and 18 while focusing on providing accurate responses. Higher colored bars indicate better accuracy performance. Smaller clear bars indicate better precision performance.

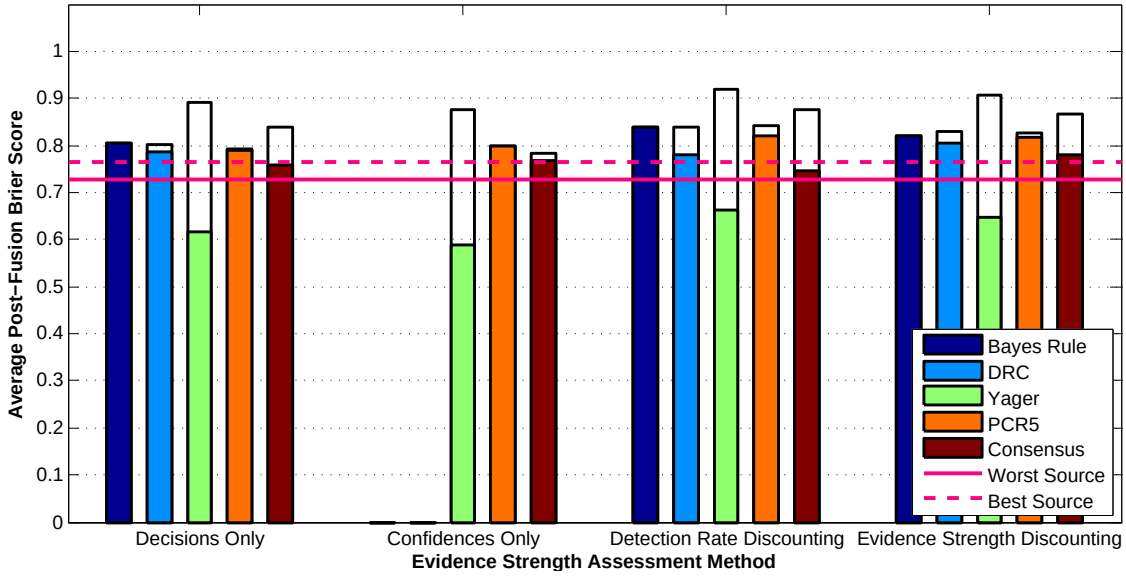


Figure 4.4: Average evidence strength of fusion operators on the city population size discrimination task of [44], showcasing four BMA (or subjective probability) construction cases. Subjects were simulated as comparing U.S. city pairs having a population rank difference between 10 and 18 while focusing on providing fast responses. Higher colored bars indicate better accuracy performance. Smaller clear bars indicate better precision performance.

combination, the PCR5 fusion operator was found to exhibit the best precision performance across all simulation cases.

4.4 Fusion Study 2: Inclusion of Superior/Inferior Sources

A subset of this study's results were previously published in [123].

4.4.1 Motivation

In the previous study, the average evidence strength values associated with the fusion operators were relatively close. This could possibly be attributed to the number of subjects combined by each of the fusion operators. In the next study, we further investigate fusion operator performance using the six 2DSD models of [44] as the number of subject responses included in the fusion operator increases. This is important, as monotonically increasing fusion performance with respect to the number of responses in the combination is a very desirable property [126]. More specifically, we investigate fusion performance with an increasing number of responses from the best and worst performing subjects (*i.e.*, subjects with the highest and lower average evidence strength values). The performance of a fusion operator in these two cases showcase how quickly a fusion operator can incorporate "good" information and also how robust a fusion operator can be towards "bad" information.

4.4.2 Experimental Setup

An experimental setup similar to Study 1 (Section 4.3.2) was used here to simulate fusion performance on the line length and city population size discrimination tasks of [44]. Simulated decisions and confidence assessments were used to generate subjective probability assignments and simple support functions for each of the six 2DSD subject models as given in [44]. Two of the four subjective probability and BMA construction cases described in Section 4.3.2 were investigated: the Confidences Only case and the Evidence Strength Discounting case. In addition to responses from the six 2DSD subjects, an increasing number of responses were simulated from the best and worst performing sources. The same five fusion operators from Study 1 were implemented, namely Bayes' rule of probability combination [97], Dempsters Rule of Combination (DRC) [43], Yagers Rule [94],

Proportional Conflict Redistribution Rule #5 (PCR5) [97], and the Consensus Operator [29]. The combination order was again fixed across all fusion rules (see [123] for a detailed explanation). For the line length discrimination task, subjects were again simulated as comparing a pair of 32.00 and 32.27 millimeter long lines. For the city population size discrimination task, subjects were again simulated as comparing U.S. city pairs which exhibited a population rank difference between 10 and 18 (*e.g.*, Houston, TX versus Baltimore, MD). For both decision tasks, subjects were again simulated to provide decisions and decision confidence assessments when focusing on either accurate or fast responses. A total of 2,000 simulation trials were conducted for each case.

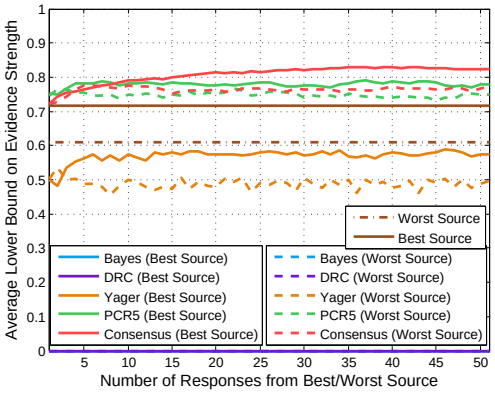
4.4.3 Results

Figures 4.5 through 4.12 each show the the accuracy and precision performance for each of the five fusion operators versus the number of responses included from either the best or worst performing source in final fused result. Recall that the accuracy performance of the fusion operator is given as $\overline{\xi}_{\text{Bel}}$ (higher is better). The precision performance of the fusion operator is given as $(\overline{\xi}_{\text{Pl}} - \overline{\xi}_{\text{Bel}})$. For the subplots which show accuracy performance, the average evidence strengths of the best and worst performing sources are shown for comparison. Figures 4.5 and 4.6 show the accuracy and precision performance of each fusion operator on the line length discrimination task with subjects focused on accurate responses and fusion operator inputs constructed using the "Confidences Only" and "Evidence Strength Discounting" cases from Study 1 respectively. Figures 4.7 and 4.8 show fusion operator performance on the line length discrimination task for the "Confidences Only" and "Evidence Strength Discounting" cases, but with subjects focused on fast responses. Figures 4.9 and 4.10 show fusion operator performance on the city population size discrimination task with subjects focused on accurate responses and fusion operator inputs constructed using the "Confidences Only" and "Evidence Strength Discounting" cases. Finally, Figures 4.11 and 4.12 show fusion operator performance on the city population size discrimination task for the "Confidences Only" and "Evidence Strength Discounting" cases, but with subjects focused on fast responses.

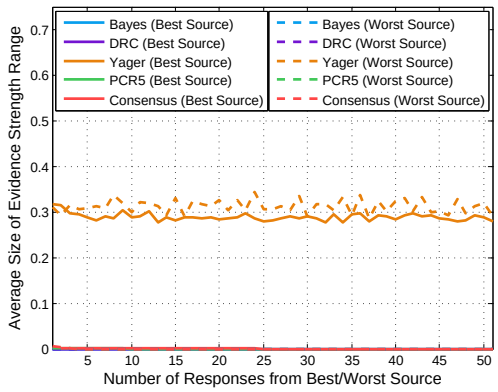
As in Study 1 (Section 4.3.3), performance relating the Bayes' rule of probability combination and DRC are not pictured for the "Confidences Only" case, since neither of these fusion operators

can be used when the BMAs and subjective probabilities being combined can be contradictory. Also in the "Confidences Only" case, the precision performance of every fusion operator excluding Yager's rule appears to approach the best possible precision performance very quickly (*i.e.*, after including approximately five sources in the final fused result). This trend in precision performance is observed for both the line length and city population size discrimination tasks, regardless of a fast or accurate response focus. The accuracy performance of the operators in the "Confidences Only" case was observed to change considerably when including an increasing number of responses from either the best or worst source in the final fused result (*e.g.*, Figures 4.5a and 4.11a). This trend was observed across all pairs of decision tasks and response foci, excluding the city population size discrimination task with accurate responses. It seems logical that the "Confidences Only" case would have instances in which the accuracy performance of the final fused result would be greatly affected by the performance of the sources, since no a priori knowledge of average source performance is considered, nor are there any heuristics employed to identify inferior sources. Finally, the accuracy performance of PCR5 and Yager's rule in the "Confidences Only" case stop increasing after about five or ten sources are included in the final fused result. The accuracy performance of the consensus operator appears to be increasing when an increasing number of the best sources are included in the combination. When an increasing number of the worst sources are included, the accuracy performance of the consensus operator stops increasing.

In the "Evidence Strength Discounting" case, the accuracy and precision performance of all five fusion rules generally increases over the "Confidences Only" case. Almost all of the fusion operators eventually approach the best possible precision performance. These two observations occurred for all possible decision tasks and response foci pairings. For the line length discrimination task, including an increasing number of the best or worst performing subject exhibits minimal differences in accuracy performance, regardless of the response focus. A similar trend was observed for the city population size discrimination task when subjects were simulated to provide fast responses. Excluding Yager's rule, each of the fusion operators exhibited higher average evidence strengths than that of the best source included in the final fused result. The accuracy performance of PCR5 and the consensus

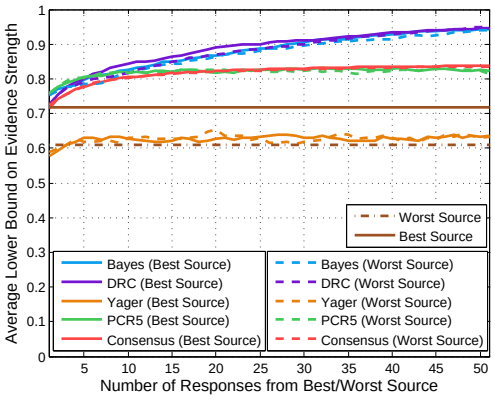


(a) Accuracy performance

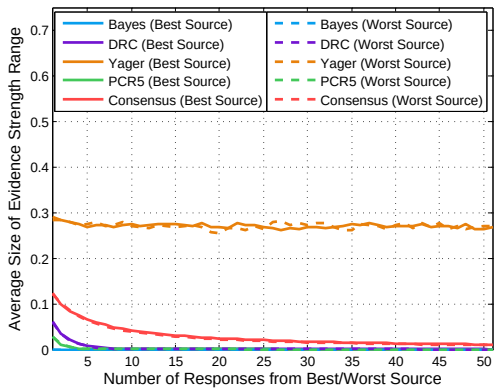


(b) Precision performance

Figure 4.5: Fusion performance on the line length discrimination task (accuracy focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Confidences Only" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.

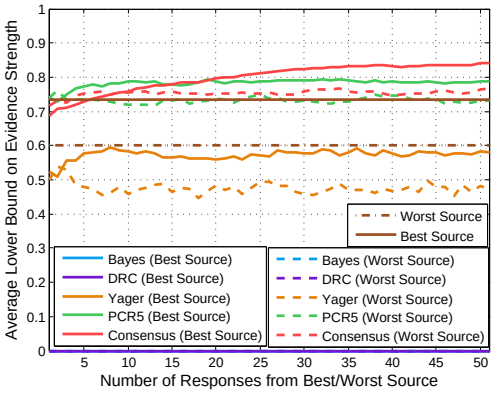


(a) Accuracy performance

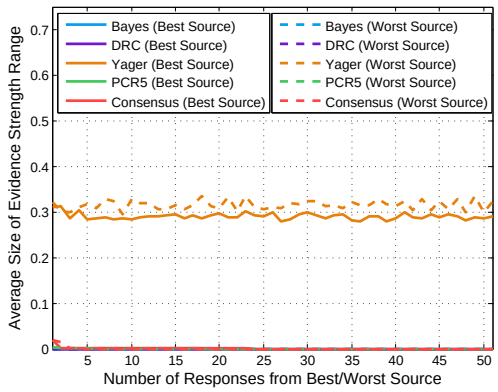


(b) Precision performance

Figure 4.6: Fusion performance on the line length discrimination task (accuracy focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Evidence Strength Discounting" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.

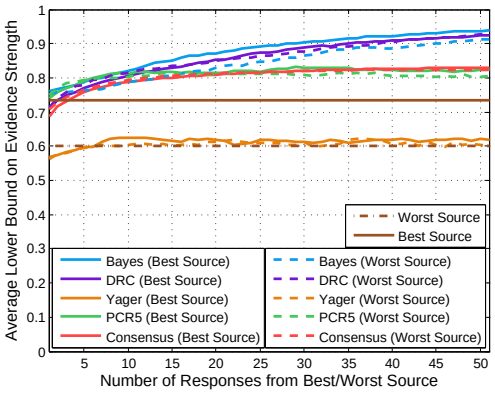


(a) Accuracy performance

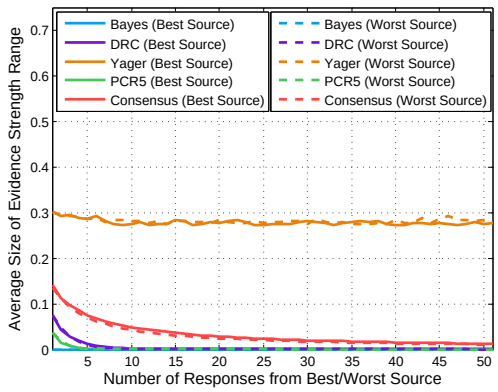


(b) Precision performance

Figure 4.7: Fusion performance on the line length discrimination task (speed focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Confidences Only" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.

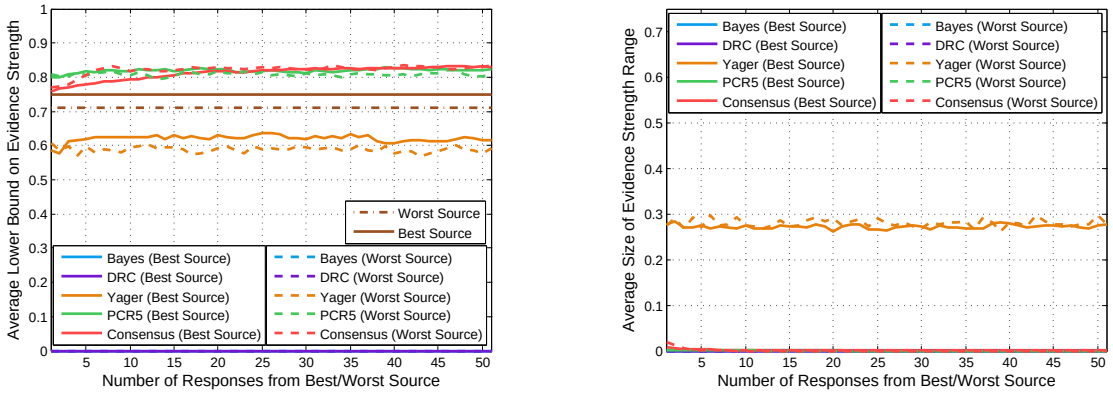


(a) Accuracy performance



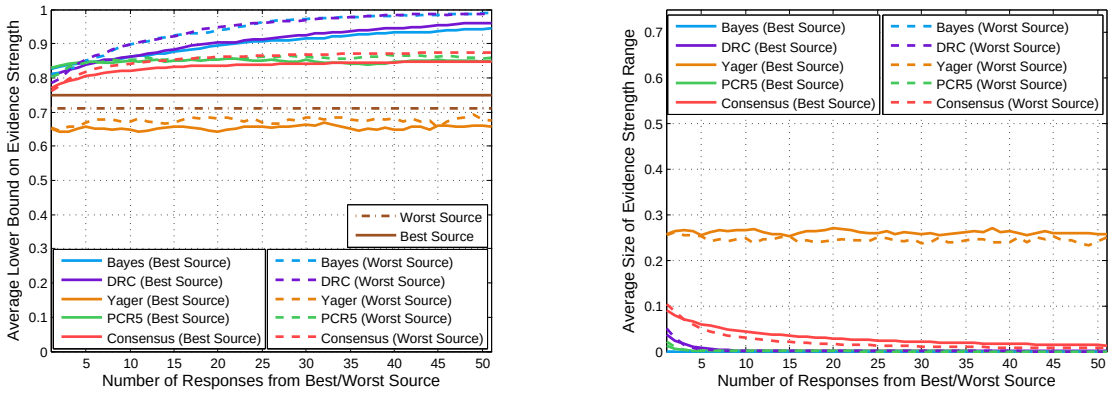
(b) Precision performance

Figure 4.8: Fusion performance on the line length discrimination task (speed focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Evidence Strength Discounting" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.



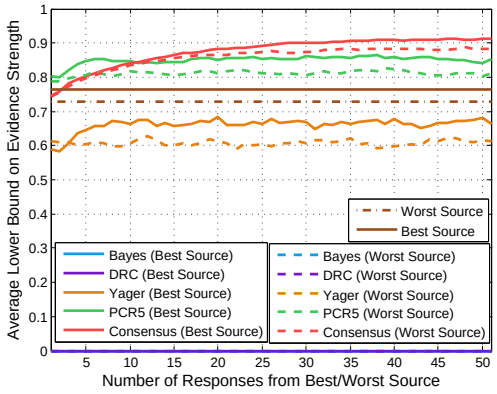
(a) Accuracy performance **(b)** Precision performance

Figure 4.9: Fusion performance on the city population size discrimination task (accuracy focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Confidences Only" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.

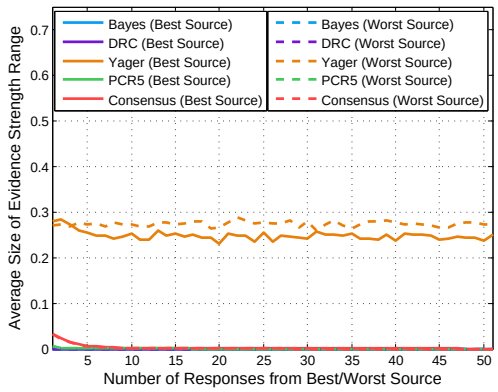


(a) Accuracy performance **(b)** Precision performance

Figure 4.10: Fusion performance on the city population size discrimination task (accuracy focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Evidence Strength Discounting" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.

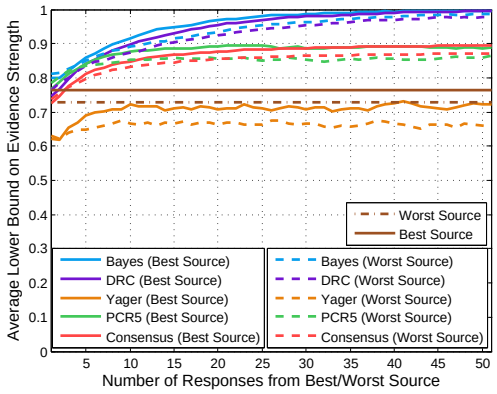


(a) Accuracy performance

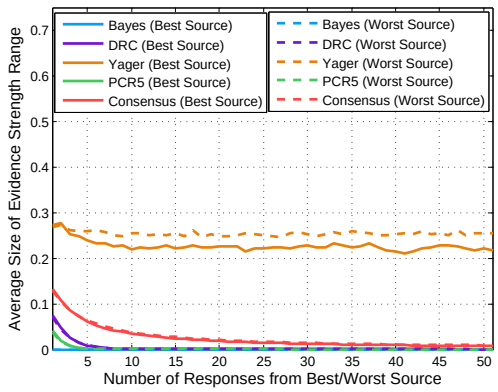


(b) Precision performance

Figure 4.11: Fusion performance on the city population size discrimination task (speed focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Confidences Only" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.



(a) Accuracy performance



(b) Precision performance

Figure 4.12: Fusion performance on the city population size discrimination task (speed focus) of [44] when including an increasing number of responses from better or worse performing sources in the combination. Subject BMAs (or subjective probability assignments) for each fusion operator formed using simulated subject decision and confidence values according to the "Evidence Strength Discounting" case from Study 1. (a) Accuracy performance (*i.e.*, minimum average evidence strength) versus number of best/worst source responses. Higher is better. (b) Precision performance (*i.e.*, evidence strength interval size). Lower is better.

operator were observed to stop increasing after approximately ten to fifteen sources were included in the final fused result. On the other hand, Bayes' rule of probability combination and DRC were observed to exhibit similar, monotonically increasing accuracy performance until eventually they both approached the best possible accuracy performance.

Similar to Study 1 (Section 4.3.3), Yager's rule exhibited the worst accuracy and precision performance across all cases. In fact, in almost all of the cases examined the accuracy performance of Yager's rule was strictly lower than the average evidence strength of the worst performing source in the final fused result. This seems logical for the same reasons discussed earlier in Study 1. Yager's rule resolves conflict amongst the sources conservatively by regarding it as total uncertainty towards the alternatives. Although this method of conflict management may seem logical, many times conflict can be relegated to shortcomings in the omniscience[3] of the sources being fused.

4.5 Fusion Study 3: Subjective Confidence and Reliability

A subset of this study's results were previously published in [127].

4.5.1 Motivation

As discussed in the problem overview of Chapter 2, the output of a fusion operator can be an aggregated set of beliefs towards each of the alternatives on some phenomenon, or a global decision that takes into account all of the data from each source. In Studies 1 and 2, we considered the former by assessing the performance of a set of fusion operators without an accompanying decision rule. In this study, we examined the performance of a set of fusion operators while implementing a set of decision rules (*i.e.*, threshold tests) in order to produce global decisions from each fusion operator. This study compared four types of fusion operators that can produce global decisions: (1) operators that use human-subject decisions (*e.g.*, the k-out-of-N majority rule); (2) operators that use subject decisions and their error rates (*e.g.*, the Chair and Varshney fusion rule); (3) operators that use subject decisions and confidence assessments (*e.g.*, Yager's rule and PCR5); and (4) operators that use subject decisions, confidence assessments, and their average evidence strength (*e.g.*, Dempster's rule of combination and Bayes' rule of probability combination). For

[3] See [59] for a more detailed discussion of such qualities.

these fusion operator types, 2DSD human response models were used to determine the performance gains possible by incorporating decision confidences and the accuracy of decision confidences into the fusion operator. Although it seems intuitive that including subjective confidences and/or reliabilities should produce a performance benefit, we seek to emphasize the importance of quantifying this benefit in order to keep it in context with the implementation complexities associated with each fusion operator.

4.5.2 Experimental Setup

The 2DSD human response models from [44] were simulated on the line length and city size discrimination tasks. Subjects were simulated to focus on producing accurate decisions and confidence assessments. The decision performance of six fusion operators were simulated: (1) the k-out-of-N majority rule [54], (2) the Chair and Varshney fusion rule [54], (3) Yager's rule [94], (4) PCR5 [97], (5) Bayes' rule of probability combination [97], and (6) DRC [43]. The k-out-of-N majority rule and the Chair and Varshney fusion rule were implemented according to Definitions 2.4 and 2.5 (Chapter 2). Yager's rule and PCR5 were implemented according to the "Confidences only" fusion operator input construction case that was discussed in Section 4.3.2. Bayes' rule of probability combination and DRC were implemented according to the "Evidence Strength Discounting" fusion operator input construction case that was also discussed in Section 4.3.2. The decision rules and corresponding false alarm and correct detection rates for the k-out-of-N majority rule and the Chair and Varshney fusion rule are described in Section 2.3. For the remaining fusion operators, the pignistic probability transform, $\mathrm{BetP}(\cdot)$, defined in Equation (2.34) was implemented to produce subjective probability assignments. Using the resulting subjective probability assignments, a global decision u_0 was generated for each fusion rule such that

$$u_0 = \begin{cases} +1 & \mathrm{BetP}(H_1) \geq \lambda \\ -1 & \text{otherwise} \end{cases} \tag{4.13}$$

where $\lambda \in [0, 1]$ is a decision threshold that controls the balance between errors due to false alarms and missed detections. For these fusion rules, the corresponding false alarm and correct detection rate associated with u_0 are given as

$$p_{F_0} = P(u_0 = +1|H_0) = P(\text{BetP}(H_1) \geq \lambda|H_0), \tag{4.14}$$

$$p_{D_0} = P(u_0 = +1|H_1) = P(\text{BetP}(H_1) \geq \lambda|H_1) \tag{4.15}$$

Notice here that we maintain the format for the global decision u_0 as set forth for the k-out-of-N majority rule and the Chair and Varshney fusion rule. When $u_0 = +1$, the global decision is H_1. When $u_0 = -1$, the global decision is H_0.

Subject decisions and confidence assessments were simulated over three difficulty levels (*i.e.*, stimuli) for each task. For the line length discrimination task, the stimuli were a pair of 32.00 and 32.27 millimeter long lines, a pair of 32.00 and 32.59 millimeter long lines, and a pair of 32.00 and 33.23 millimeter long lines. For the city population size discrimination task, the stimuli were pairs of cities that differed in population rank by 1 and up to 9 (e.g., New York, N.Y. versus Los Angeles, C.A.), by 10 and up to 18 (e.g., Houston, T.X. versus Baltimore, M.D.), and by 19 and up to 29 (e.g., Detroit M.I. versus Cleveland, O.H.). For each simulation trial, a set of 24 simulated human responses were generated by simulating 4 decision and confidence assessment pairs from each of the six subjects according to the 2DSD parameter sets in [44, Tables 3 and 6]. Each fusion operator was implemented by randomly selecting one response at a time from the human response set. Once all of the responses were fused, the decision rule of each operator was implemented for a given threshold value. In total, 10,000 trials were conducted for the second stimuli (located on the right) being correct (*i.e.*, H_1) and 10,000 first stimuli (located on the left) being correct (*i.e.*, H_0). For each fusion operator, the corresponding thresholds were varied for the trials in which H_1 was true and H_0 was true, to generate sets of false alarm and detection rates. These were used to generate empirical receiver operating characteristic (ROC) curves for each fusion operator, and the corresponding normalized areas under the curves (AUCs) [128]. Higher AUC values represent fusion

operators which can discriminate better between alternatives (*i.e.*, producing higher detection rates for a given false alarm rate).

4.5.3 Results

Figure 4.13 shows the normalized AUCs for each fusion operator versus the number of responses present in the combination for the line length discrimination task of [44]. Figure 4.14 also shows the normalized AUCs for each fusion operator versus the number of responses present in the combination, but for the city population size discrimination task of [44]. The subplots of Figure 4.13 and Figure 4.14 show the normalized AUC values observed for each of the three task difficulty levels simulated. Error bars show the estimation error at the 95% confidence level. For a small number of responses (*e.g.*, two), fusion operators which incorporate confidences yielded higher normalized AUCs (*i.e.*, Bayes' rule, DRC, Yager's rule, and PCR5). As the number of responses in combination is increased (*e.g.*, greater than 5), the fusion operators which incorporate decisions, confidence assessments, and reliabilities (*i.e.*, Bayes' rule and DRC) yielded the highest normalized AUCs, followed very closely by the Chair and Varshney fusion rule and then by the k-out-of-N majority rule for both tasks and all difficulty levels. Similar discriminability performance between the k-out-of-N majority rule and the Chair and Varshney fusion rule are due to the sources exhibiting similar detection and false alarm rates. Also as the number of responses in combination increases, the fusion rules which incorporate confidences and decisions alone (*i.e.*, Yager's rule and PCR5) produce the lowest AUC values. Furthermore, these normalized AUC values were observed to stop increasing, whereas the AUC values for the remaining fusion operators continue to increase. These results support the notion that as the number of sources increases, the importance of decision self-assessment diminishes. In fact, the inclusion of the self-assessment seems to have caused "confusion" and degradation in the cases where the subjects' ability to give good confidence assessments was not taken into account. This confusion may occur because the subjects in the line length and city size discrimination task are not proficient in providing accurate decision confidence values.

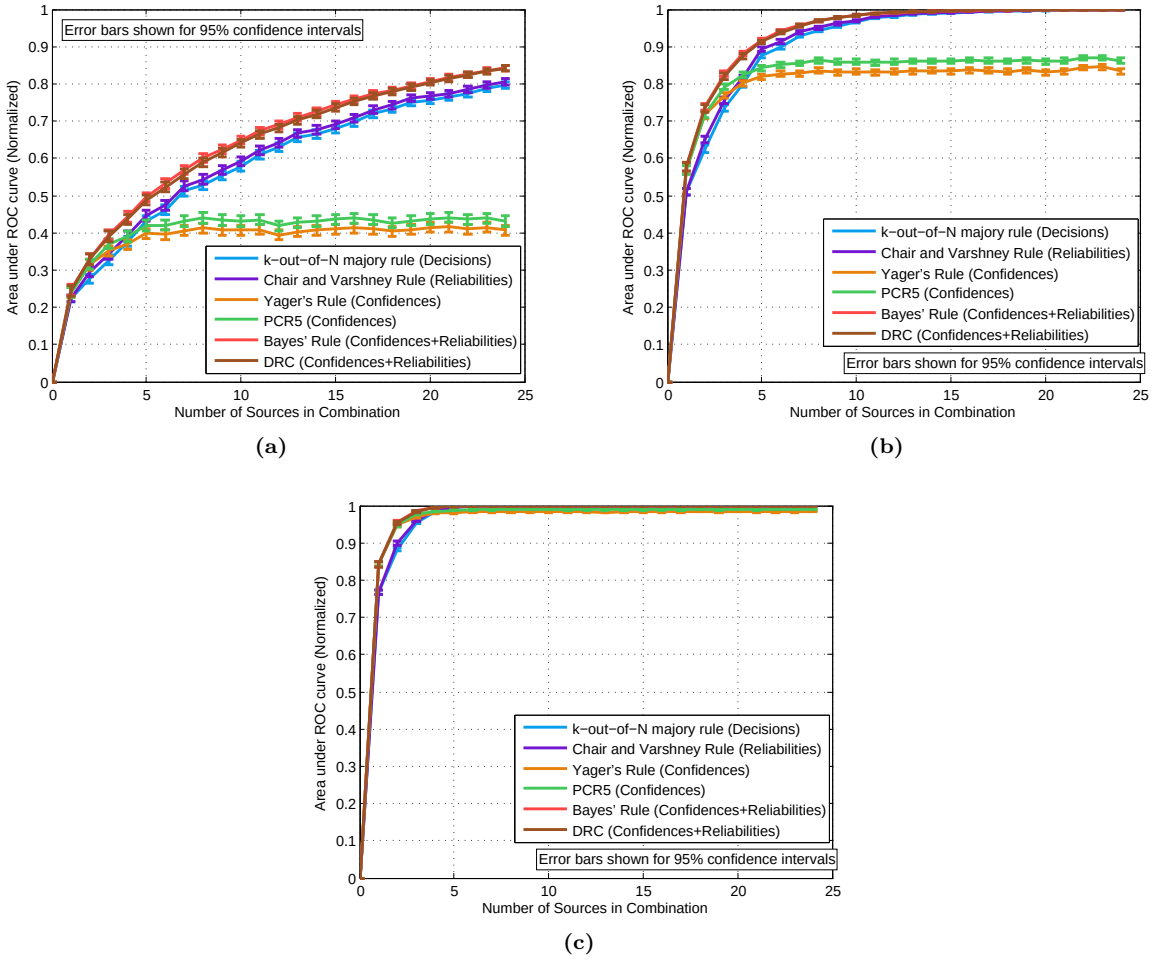


Figure 4.13: Normalized area under the ROC curve versus the number of sources included in combination for the line length discrimination task of [44]. Error bars show estimation error at the 95% confidence level. (a) Comparing 32.00 versus 32.27 millimeter long lines (b) Comparing 32.00 versus 32.59 millimeter long lines. (c) Comparing 32.00 versus 33.23 millimeter long lines.

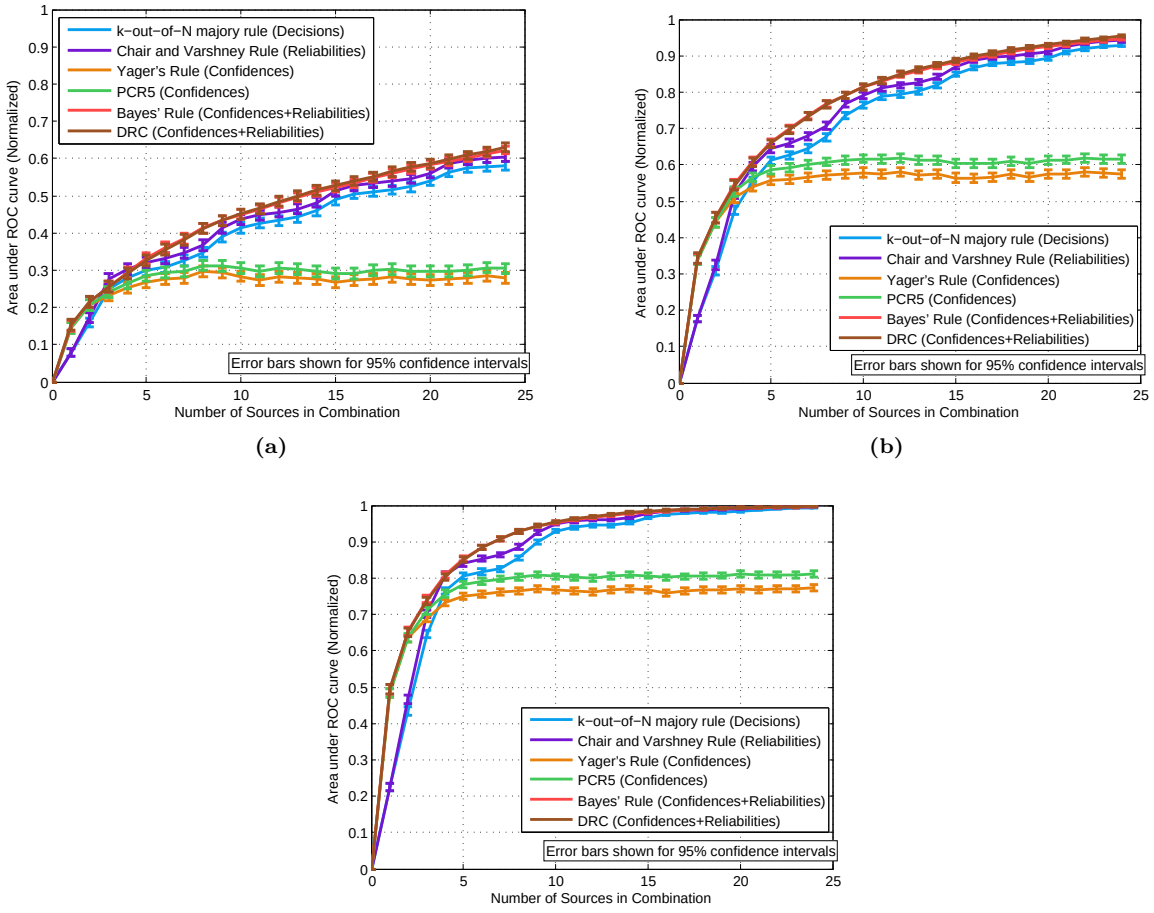


Figure 4.14: Normalized area under the ROC curve versus the number of sources included in combination for the city population size discrimination task of [44]. Error bars show estimation error at the 95% confidence level. (a) Comparing cities differing in population rank by 1 and up to 9 (e.g., New York, N.Y. versus Los Angeles, C.A.) (b) Comparing cities differing in population rank by 10 and up to 18 (e.g., Houston, T.X. versus Baltimore, M.D.). (c) Comparing cities differing in population rank by 19 and up to 29 (e.g., Detroit M.I. versus Cleveland, O.H.).

4.6 Fusion Study 4: Probability Transformation Performance

A subset of this study's results were previously published in [129].

4.6.1 Motivation

As mentioned in Chapter 2, the Dempster-Shafer (DS) theory is a popular tool in the information fusion community [19,21,27,130–132]. As opposed to the use of subjective probabilities (*i.e.*, Bayesian epistemology [42]), a Dempster-Shafer approach employs a normalized measure of evidence (*i.e.*, a BMA) on a powerset of alternatives [43]. To facilitate decision making, *probability transformations* are used to generate a subjective probability supported by a given belief mass assignment. There exists several Dempster-Shafer theory based fusion rules [59], as well as a large number of probability transformations (*e.g.*, [104, 105, 133]). These transforms are usually evaluated through the use of hypothetical examples and by measuring the amount of entropy present in the resulting transformed probabilities for a given set of evidence. Such information has been characterized using the probability information content (PIC) mertic [105,134], which is defined for a subjective probability assignment $\mathcal{P}(\cdot)$ generated by a probability transformation as

$$PIC_{\mathcal{P}(\omega)} = 1 + \frac{1}{\log M} \sum_{\omega \in \Omega} \mathcal{P}(\omega) \log \mathcal{P}(\omega) \tag{4.16}$$

where $M = |\Omega|$. A low PIC value represents a subjective probability assignment where the alternatives are close to being equiprobable (*i.e.*, a difficult decision). A high PIC value represents a subjective probability assignment where one of the alternatives is close to having probability one (*i.e.*, an easier decision). There have been studies however that suggest that the PIC metric alone may not be sufficient in evaluating the performance of a probability transformation (*e.g.*, [135]). This study makes use of the 2DSD human response models from [44] to investigate these claims in terms of the effect that the choice of probability transform has on the statistical performance of the fusion operator (*i.e.*, decision error rates and AUC values).

4.6.2 Experimental Setup

Similar to the setup of Study 3, a pool of 24 human responses were generated for the line length and city population discrimination tasks of [44], with subjects focused on providing accurate responses. For the line length discrimination task, we let H_1 denote the hypothesis that the second line presented to the subject is longer than the first, and let H_0 denote the hypothesis that the first line presented to the subject is longer than the second. For the city population size discrimination task, we let H_1 denote the hypothesis that the second city presented to the subject has a higher population than the first and let H_0 denote the hypothesis that the first city presented has a higher population than the second. For each task, a total of 10,000 trials were conducted for H_1 being true and 10,000 trials for H_0 being true. The subject responses were combined two at a time using Yager's rule [94]. Since Yager's rule is not associative [59], the combination order was randomized by choosing from the subject response pool with equal probability. The final combined BMAs were then each transformed into a subjective probability assignments $\mathcal{P}(\cdot)$ by using one of six probability transformations, namely $BetP$ (Equation (2.34)), $PrPl$ (Equation (2.35)), $PrNPl$ (Equation (2.36)), $PrHyb$ (Equation (2.37)), and $DSmP$ (Equation (2.39)). A decision rule similar to Equation (4.13) was used, that is

$$u_0 = \begin{cases} +1 & \mathcal{P}(H_1) \geq \lambda \\ \\ -1 & \text{otherwise} \end{cases}, \tag{4.17}$$

where the threshold $\lambda \in [0, 1]$ is chosen to obtain a desired global decision false alarm and correct detection rate. Again, when $u_0 = +1$, the global decision is H_1, and when $u_0 = -1$, the global decision is H_0. The threshold test of Equation (4.17) was used to estimate false alarm rates and detection rates for varying threshold values λ using the 10,000 simulated responses with H_1 true and the 10,000 responses with H_0 true after applying Yager's rule and each of the five probability transforms. The resulting sets of false alarm and detection rates were then used to determine empirical ROC curves, and the resulting normalized AUC values for each probability transform.

Table 4.1: Average PIC for each probability transformation (H_1 true, line length discrimination task of [44]).

Line Length Difference	$BetP$	$PrPl$	$PrNPl$	$PrHyb$	$DSmP$
0.27 mm	0.56	0.63	0.51	0.68	0.83
0.59 mm	0.67	0.72	0.64	0.76	0.88
1.23 mm	0.85	0.87	0.84	0.89	0.95
1.87 mm	0.93	0.94	0.92	0.95	0.98

Table 4.2: Average PIC for each probability transformation (H_1 true, city population size discrimination task of [44]).

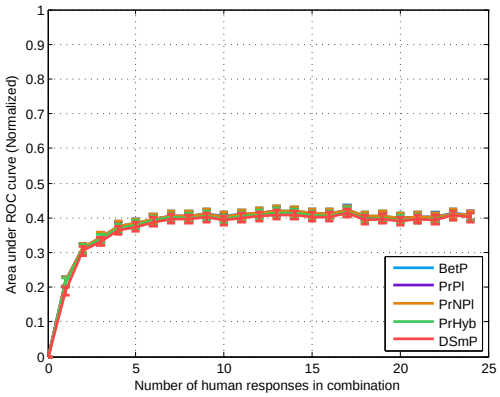
City Rank Difference	$BetP$	$PrPl$	$PrNPl$	$PrHyb$	$DSmP$
1 - 9	0.49	0.57	0.43	0.64	0.80
10 - 18	0.53	0.61	0.48	0.67	0.82
19 - 29	0.60	0.67	0.56	0.73	0.85
30 - 43	0.68	0.73	0.64	0.78	0.89

4.6.3 Results

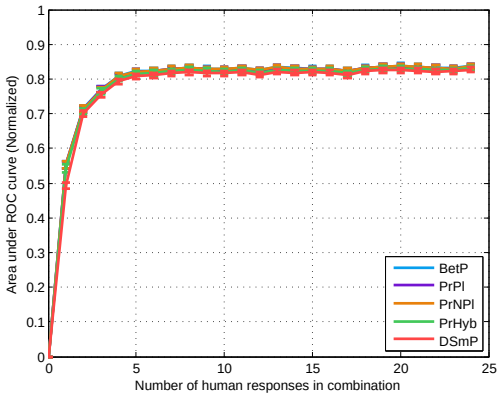
Tables 4.1 and 4.2 show the average PIC values after applying each of the five probability transformations to the combined BMAs resulting from the line length discrimination task and the city size discrimination task. Higher PIC values are usually considered better [105, 134, 136]. The results in Tables 4.1 and 4.2 show increasing PIC values as task difficulty decreases, which seems reasonable. For each difficulty level, the PIC values follow the same trend with $PrNPl$ having the lowest PIC and $DSmP$ having the highest PIC. These trends supports the notion that DSmP produces subjective probabilities which are the most committed towards one of the alternatives (*i.e.*, having the lowest entropy) [105].

Figures 4.15 and 4.17 show the normalized AUCs versus the number of human responses present in combination for all five probability transforms on the line length discrimination task and the city size discrimination task. Each subplot of Figures 4.15 and 4.17 shows the observed normalized AUC values for the four task difficulty levels simulated here. As expected, normalized AUC values increase as the task becomes easier. For any given difficulty level however, all five probability transformations exhibited statistically insignificant differences between normalized AUC values. This indicates that the overall discriminating performance of each probability transform is actually same.

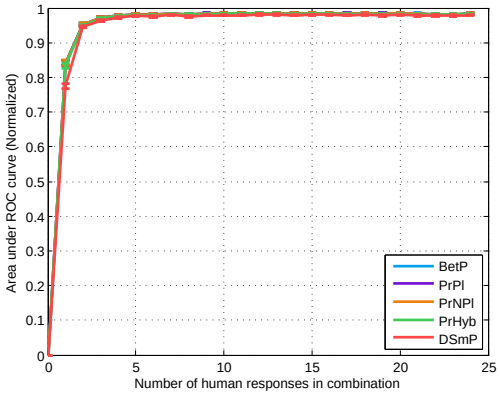
Figures 4.16 and 4.18 show the ROC curves after combination for all five probability transforms



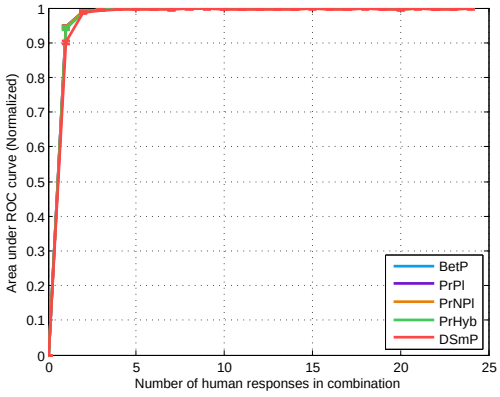
(a) Line length difference 0.27 mm



(b) Line length difference 0.59 mm

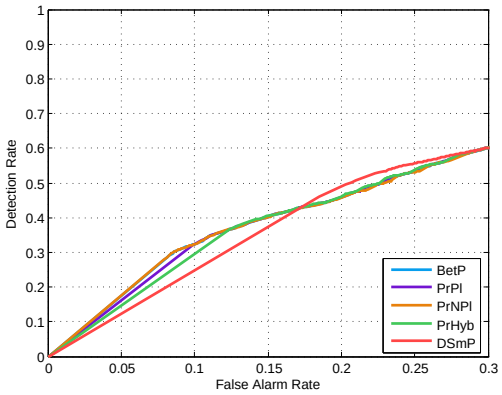


(c) Line length difference 1.23 mm

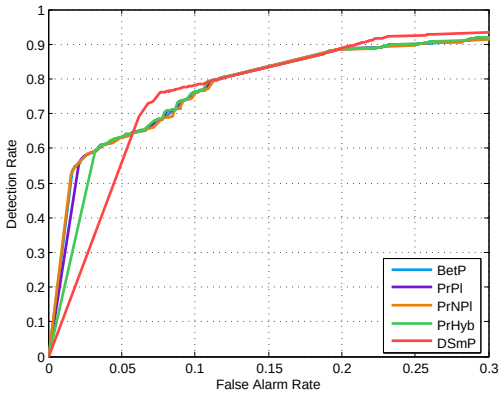


(d) Line length difference 1.87 mm

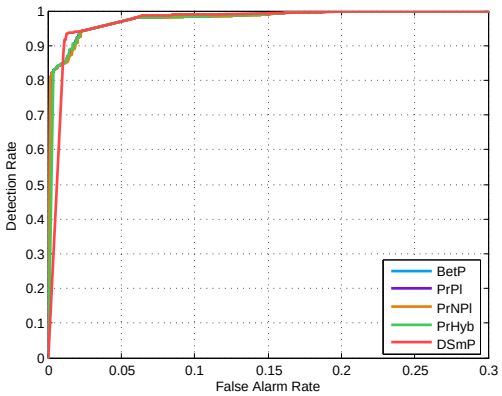
Figure 4.15: Normalized area under the ROC curve (AUC) versus the number of sources present in combination, for each difficulty level of the line length discrimination task of [44]. Different lines represent the five different probability transforms investigated by this work. Error bars shown for the 95% confidence intervals. In each of the four difficulty levels, all five probability transforms are nearly overlapping.



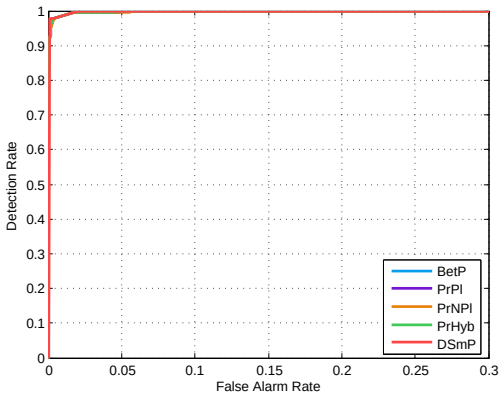
(a) Line length difference 0.27 mm



(b) Line length difference 0.59 mm

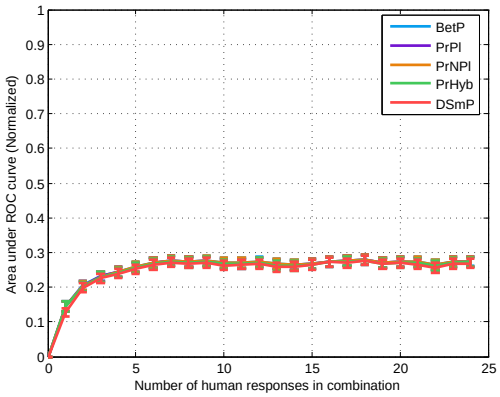


(c) Line length difference 1.23 mm

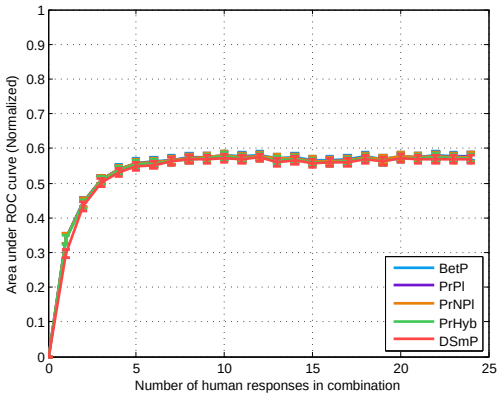


(d) Line length difference 1.87 mm

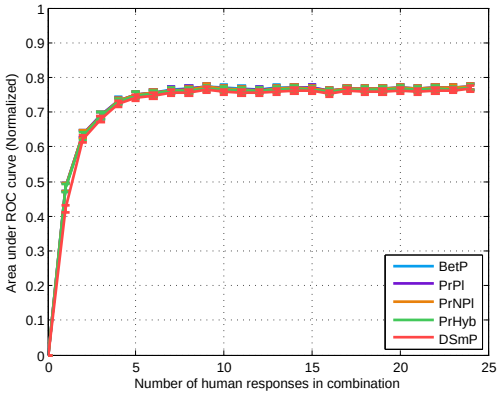
Figure 4.16: ROC curves for each difficulty level of the line length discrimination task of [44], showing false alarm rates up less than 0.30. Different lines represent the five different probability transforms investigated.



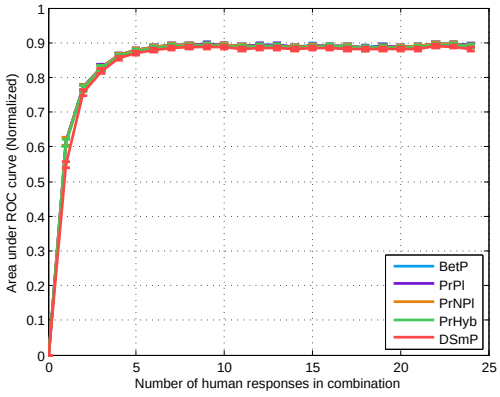
(a) City rank difference 1 - 9



(b) City rank difference 10 - 18

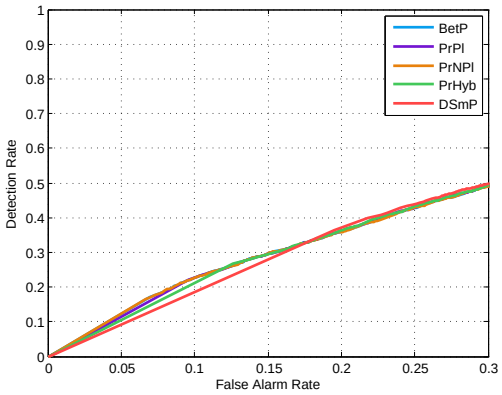


(c) City rank difference 19 - 29

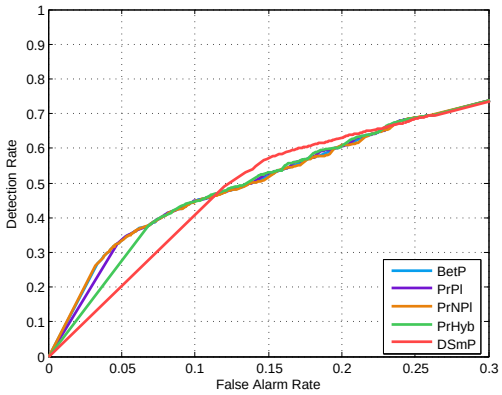


(d) City rank difference 30 - 43

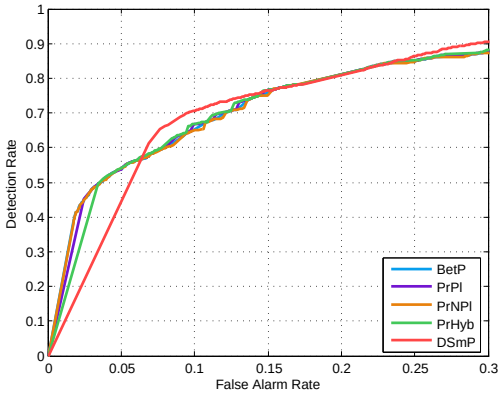
Figure 4.17: Normalized area under the ROC curve (AUC) versus the number of sources present in combination, for each difficulty level of the city population size discrimination task of [44]. Different lines represent the five different probability transforms investigated by this work. Error bars shown for the 95% confidence intervals. In each of the four difficulty levels, all five probability transforms are nearly overlapping.



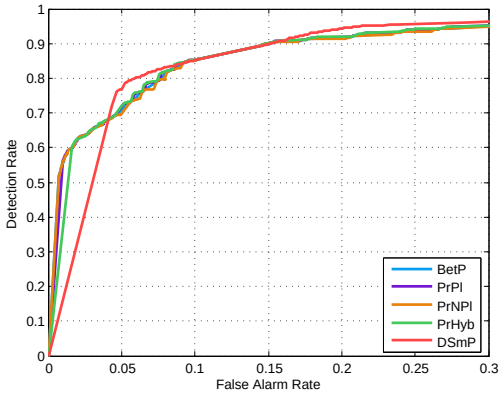
(a) City rank difference 1 - 9



(b) City rank difference 10 - 18



(c) City rank difference 19 - 29



(d) City rank difference 30 - 43

Figure 4.18: ROC curves for each difficulty level of the city population size discrimination task of [44], showing false alarm rates up less than 0.30. Different lines represent the five different probability transforms investigated.

on the line length discrimination task and the city size population task. Each subplot of Figures 4.16 and 4.18 shows ROC curves for the four task difficulty levels simulated here. The ranges of the graph axes correspond to "reasonable" false alarm rates (*i.e.*, up to 0.30)). Again, the overall shape of the ROC for all probability transforms improves as the tasks become easier (supporting the results shown in Figure 4.17). For lower false alarm rates (*e.g.*, less than 0.07), $BetP$, $PrPl$, $PrNPl$, and $PrHyb$ produce similar detection rates, which are all higher than those produced by $DSmP$. For higher false alarm rates (*e.g.*, greater than 0.10), $DSmP$ produces higher detection rates over the remaining four probability transforms. As false alarm rates become even higher (*e.g.*, greater than 0.25), similar performance is observed for all probability transforms. For the hardest and easiest variations of both tasks, these performance gains become less apparent. These observations support the conclusions reached by [135]; depending on the acceptable error rate for a specific task, a higher PIC value may not necessarily indicate better decision performance by itself. According to the average PIC values, the easiest decisions can be made when using $DSmP$, but at a cost of slightly reduced detection performance at small fixed false alarm rates.

4.7 Chapter Summary

In this chapter, we have shown how the performance of fused decisions and/or beliefs produced by soft fusion operators on binary decision tasks can be assessed using two-stage dynamic signal detection (2DSD). Specifically, we have investigated four case studies to assess the performance of fusion operators from Bayesian epistemology and Dempster-Shafer (DS) theory, in addition to assessing the performance of various probability transformations. In the first study, several BMA (or subjective probability) construction cases were investigated along with a selection of fusion operators. It was observed that fusion operators which incorporate some sort of statistical knowledge regarding the reliability of the subjects perform better than those which do not. These results were further examined in the second study, in which an increasing number of responses from the best or worst performing source were included in the fusion step. Bayes' rule of probability combination and Dempster's rule of combinatioin (DRC) were both observed to exhibit monotonically increasing performance with respect to the number of responses in combination when some sort of source dis-

counting was performed. In most cases, source discounting was observed to make the performance operators less sensitive to variability in the reliability of each of the subjects. In the third study, the effects of confidences and reliability were assessed on a selection of fusion operators that incorporated decision rules. When the reliability of confidences was not considered in the fusion operator, increasing the number of responses in the fusion operator resulted in inferior performance. When discounting was performed using subject average evidence strengths, the incorporation of decision confidences slightly improved fused decision performance over the k-out-of-N majority rule and Chair and Varshney fusion rule was observed. Finally, in the fourth study we evaluated the performance of a variety of probability transformations from DS theory alongside a probability transformation performance metric known as the probability information content (PIC). It was determined that a higher PIC value may not necessarily indicate better decision performance by itself. For our specific results, highest average PIC values were obtained when using the $DSmP$ probability transform, but at a cost of slightly reduced detection performance at small fixed false alarm rates.

Chapter 5: Multihypothesis Fusion Operator Simulation Methods

As mentioned in Chapter 3, the majority of human response models developed by cognitive psychologists so far are focused on two-alternative forced choice (TAFC) tasks. Theoretical and experimental investigation into multihypothesis human decision-making models has only recently become an active topic of research [137]. There has been very little investigation of the development of multihypothesis human decision-making models that also account for the generation of decision confidence assessments. Furthermore, vague[1] human decision making models have seen little investigation. In this chapter, we present a method for generating multihypothesis human decisions and decision confidence assessments. This method aggregates simulated pairwise-successive comparisons using binary human decision making models (*i.e.*, two-stage dynamic signal detection (2DSD)). We then discuss how this pairwise-successive comparison method can be used to generate vague decisions and confidence assessments. The dynamics of the aggregation method is then assessed using sample simulations on multihypothesis extensions of the line length and city population size discrimination tasks. Finally, we use the proposed pairwise-successive comparison method to simulate the performance of several fusion operators defined in Chapter 2.

5.1 Pairwise-Successive Comparisons

A more specific version of the pairwise successive comparison aggregation method was discussed in [125] for precise responses on the line length discrimination task modeled by the authors of [44]. We present the general form of the aggregation method and detail its relationship to a given TAFC human decision-making model (*i.e.*, 2DSD).

5.1.1 Aggregation of TAFC Responses

Consider some phenomenon in which an individual is presented with a set of M stimuli, each having some commonly quantifiable feature. We label a given instance of these M features as

[1] We define vague human decision-making in Section 2.1.

$\{z_1, \ldots, z_M\} = \mathcal{Z} \subset \mathbb{R}^M$, and further assume that each one of these features has a distinct value (*i.e.*, $z_i = z_j$ if and only if $i = j$ for any $i, j \in \{1, \ldots, M\}$). Define a decision task consisting of M alternatives $\{\omega_1, \ldots, \omega_M\} = \Omega$ such that ω_i corresponds to the decision that feature $z_i \in \mathcal{Z}$ is the most apparent. For example, in an M-ary line length discrimination task z_i denotes the length of the i^{th} line among M enumerated lines, and hence ω_i corresponds to decision that z_i is the longest length line among the line lengths given in $\mathcal{Z}$. For an M-ary city population size discrimination task z_i denotes the population of the i^{th} city among M enumerated cities, and hence ω_i corresponds to the decision that z_i is the most populated city among the cities given in $\mathcal{Z}$.

Regardless of the decision task, we denote the actual most apparent feature as $z_{i^*} \in \mathcal{Z}$, and hence ω_i corresponds to the choice that $z_{i^*} = z_i$. The goal of the pairwise successive comparison aggregation method is to construct a subjective probability assignment on Ω by use of simulated pairwise comparisons of the elements $z_i \in \mathcal{Z}$. Given that each feature $z_i \in \mathcal{Z}$ has a distinct value, we can write the subjective probability of ω_i as

$$\mathcal{P}(\omega_i) = \mathcal{P}(z_{i^*} = z_i | z_{i^*} \in \mathcal{Z}) = \frac{\mathcal{P}(z_{i^*} = z_i)}{\sum_{\hat{i}=1}^{M} \mathcal{P}(z_{i^*} = z_{\hat{i}})}. \tag{5.1}$$

Let $z_i \succ z_j$ be a binary comparison denoting that feature z_i is more apparent than z_j for any $i, j \in \{1, \ldots, M\}$. Thus, expression 5.1 can be rewritten as

$$\mathcal{P}(\omega_i) = \frac{\mathcal{P}(\bigcap_{\substack{j \in \{1,\ldots,M\} \\ j \neq i}} z_i \succ z_j)}{\sum_{\hat{i}=1}^{M} \mathcal{P}(\bigcap_{\substack{j \in \{1,\ldots,M\} \\ j \prec i}} z_{\hat{i}} \succ z_j)}. \tag{5.2}$$

Assuming that beliefs are assessed independently between each binary comparison, $z_i \succ z_j$, the pairwise successive comparison aggregation method can be finally written as

$$\mathcal{P}(\omega_i) = \frac{\prod_{\substack{j \in \{1,\ldots,M\} \\ j \neq i}} \mathcal{P}(z_i \succ z_j)}{\sum_{\hat{i}=1}^{M} \prod_{\substack{j \in \{1,\ldots,M\} \\ j \neq \hat{i}}} \mathcal{P}(z_{\hat{i}} \succ z_j)}, \tag{5.3}$$

where the subjective probability functions $\mathcal{P}(z_i \succ z_j)$ can be obtained by simulating decision and decision confidence assessments using a TAFC human response model for each possible pairing of

$i, j \in \{1, \ldots, M\}$. Equivalently, every value of $\mathcal{P}(\omega_i)$ can be determined by first simulating TAFC human response models to compute the numerators of Equation 5.3, and then normalizing them such that they sum to one.

5.1.2 Decision-Making and Confidence Assessment

After aggregating the subjective probabilities $\mathcal{P}(\omega_i)$ for each $\omega_i \in \Omega$, a vague decision $A \subset \Omega$ can be generated by choosing the one which maximizes the value of $\mathcal{P}(A)$ for a given level of vagueness, or *imprecision level*, $l = |A|$. That is,

$$A = \arg\max_{\substack{A \subset \Omega \\ |A| = l}} \left(\sum_{\omega_i \in A} P(\omega_i) \right), \tag{5.4}$$

with a corresponding decision confidence value of $p_A = \sum_{\omega_i \in A} P(\omega_i)$. When $l = 1$, this rule reduces to the one proposed in [125] for the line length discrimination task. Since 2DSD models choose from a finite set of confidence values [44], the following three cases can occur: A is unique, A is not unique, or A does not exist because the denominator of Equation (5.3) is zero. In the second case, a single vague decision A can be made by choosing one of the maximizing sets of Equation (5.4) at random, assuming each is equally likely. In the third case, a decision cannot be reached and a "no decision" state is returned.

5.1.3 Imprecise Evidence Strength

To assess the quality of the information in a subject's vague decision, the evidence strength metric ξ as defined in Equation (4.6) needs to be altered. We define a more generalized metric known as *imprecise evidence strength* ξ' such that

$$\xi'(A, p_A | \omega^*) = \begin{cases} 1 - (1 - p_A)^2 & \omega^* \in A \\ 1 - p_A^2 & \omega^* \notin A \end{cases}, \tag{5.5}$$

where ω^* denotes the correct alternative. Notice that if a decision is precise (*i.e.*, $|A| = 1$), then $\xi'(\cdot) = \xi(\cdot)$. As opposed to average evidence strength, the average imprecise evidence strength $\overline{\xi'}$

does not fully characterize the decision-making performance of the subject. For example, subjects may exhibit high values of $\overline{\xi'}$ but yield very vague decisions. Therefore, $\overline{\xi'}$ only measures the net amount of decision confidence provided by a subject which does not rule out the correct alternative.

5.1.4 Assumptions and Limitations

The pairwise successive comparison method for simulating M-ary human responses assumes that subjects consider every possible pair stimuli. In diffusion models of human decision making, the difficulty of various stimuli are captured through different drift rates, δ [44,45]. The beliefs produced by the proposed M-ary extension methodology are therefore dependent on the various drift rates used in generating each of the binary responses of the pairwise successive comparisons. Any pairwise successive comparison involving two stimuli which are clearly distinct will correspond to large drift rate values and consequently lead to exponentially fast response times [44]. In reality however, human subjects may not even make use of a pairwise comparison technique for alternatives which are clearly different. Furthermore, this method assumes that drift rates are known (or can be reliably estimated) for every possible pair of stimuli.

5.2 Simulation of Subject Performance

Before continuing, we present two sample subject simulations in order to emphasize the properties of the proposed successive pairwise comparison technique. The results of the first simulation were published earlier in [125].

5.2.1 Decision-Making on Varying Numbers of Alternatives

In this first set of subject simulations, we simulated the performance of M-ary decisions using the proposed successive pairwise comparison technique and the 2DSD parameter tuples for subjects #1 through #6 on the line length discrimination task modeled in [44]. Similar to the simulation presented in Chapter 3, a 2DSD optional stopping model with trial variability on the drift rate, δ, and initial condition, L_0, was used. All six subject tuples were simulated under the accuracy focus. The line lengths presented to the simulated subjects were $\mathcal{Z} = \{32, 32+d, 32+2d, \ldots, 32+(M-1)d\}$ where M was the number of lines being compared and d was the incremental length difference

between lines, in millimeters. Subject decisions and confidence assessments were generated using the successive pairwise comparison aggregation method of Equation (5.3). The evidence strengths of each subject were determined and averaged over 10,000 trials from $d = 0.01$ to $d = 1.0$ in increments of 0.01 and for $M = 2$, 4, 6, and 8. For each subject, trials which produced the "no decision" state were repeated until a decision and confidence value were reached.

For the line length discrimination task performed in [44], only five stimuli were presented to the subjects: a 32.00 millimeter long line paired with either a 32.27, 32.59, 33.23, 33.87, or a 34.51. To simulate responses on additional line length pairs, we performed linear interpolation to estimate the corresponding mean drift rates ν. Let Δl represent the length difference between each line, such that

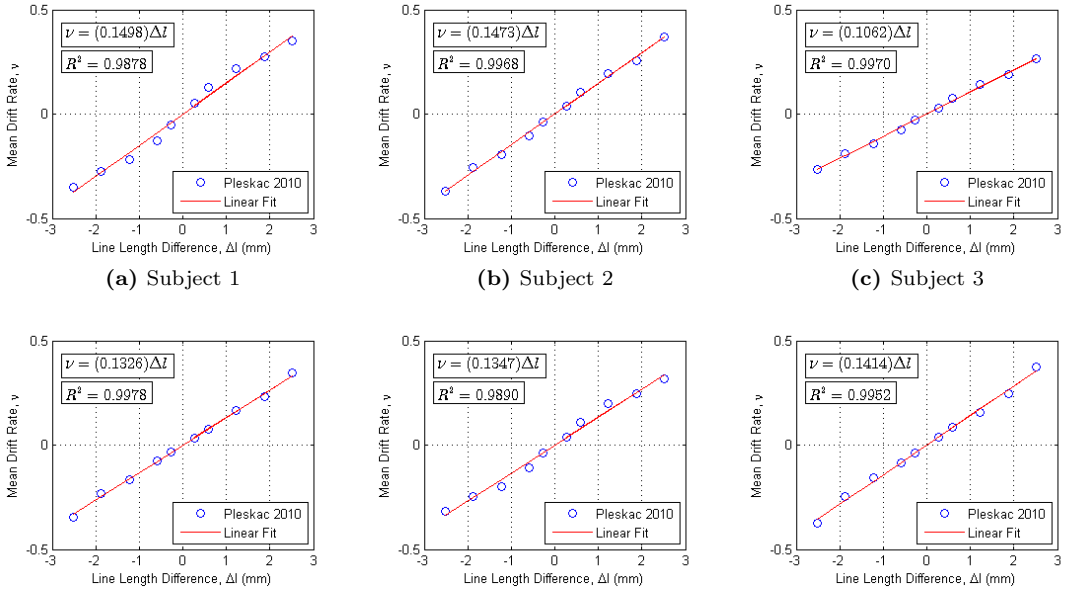
$$\Delta l = l_R - l_L, \tag{5.6}$$

where l_R and l_L represent the lengths of the right and left lines as presented to the subjects. Linear regression was applied to the coordinate pairs $(\Delta l, \nu)$ for each subject of the line length discrimination task in [44] (Figure 5.1). All subject drifts rates appear to follow a linear relationship. With this relationship in mind, the value of ν used to simulate a comparison of lines having a length difference of Δl is determined such that

$$\nu = \nu_m \Delta l. \tag{5.7}$$

Here ν_m is the slope of the linear fit as given for each subject in Figure 5.1.

Figure 5.2 shows the average evidence strength, ξ, of each subject versus the incremental line length difference, d. Evidence strengths for different numbers of alternatives M for each subject are also shown. As expected, increasing the perceptual difficulty of the task (*i.e.*, decreasing d) decreased subject performance. For large enough d (*e.g.*, $d > 0.60$), increasing the number of alternatives was found to have little effect on subject performance. For smaller d (*e.g.*, $d < 0.40$), increasing the number of alternatives caused the largest decrease in performance when going from $M = 2$ to $M = 4$ alternatives. For $M > 4$ however, the subject performance was similar to the $M = 4$ case. This outcome seems logical, as increasing the number of alternatives without changing



(a) Subject 1 **(b)** Subject 2 **(c)** Subject 3

(d) Subject 4 **(e)** Subject 5 **(f)** Subject 6

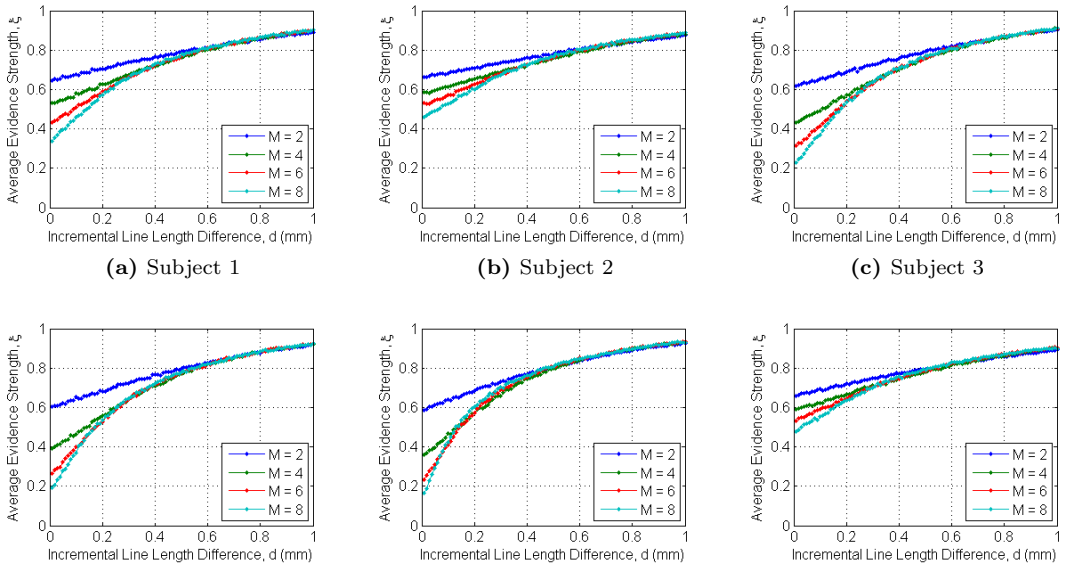
Figure 5.1: Linear fits of subject mean drift rates versus line length differences for the line length discrimination task as presented in [44]. Equations and R^2 values shown for each subject.

the task difficulty will result in some alternatives being easier to rule out than others.

Figure 5.3 shows the chance of a no decision case occurring for each simulated subject, estimated over 10,000 trials. For the subjects that exhibited large evidence strengths across all values of M and d, the chances of a no decision case occurring were found to be low (e.g., Subjects 1, 2, and 6). For the subjects that exhibited small evidence strengths at small values of d, the chances of a no decision case were found to be high (e.g., Subjects 3, 4, and 5).

5.2.2 Vague Decision-Making

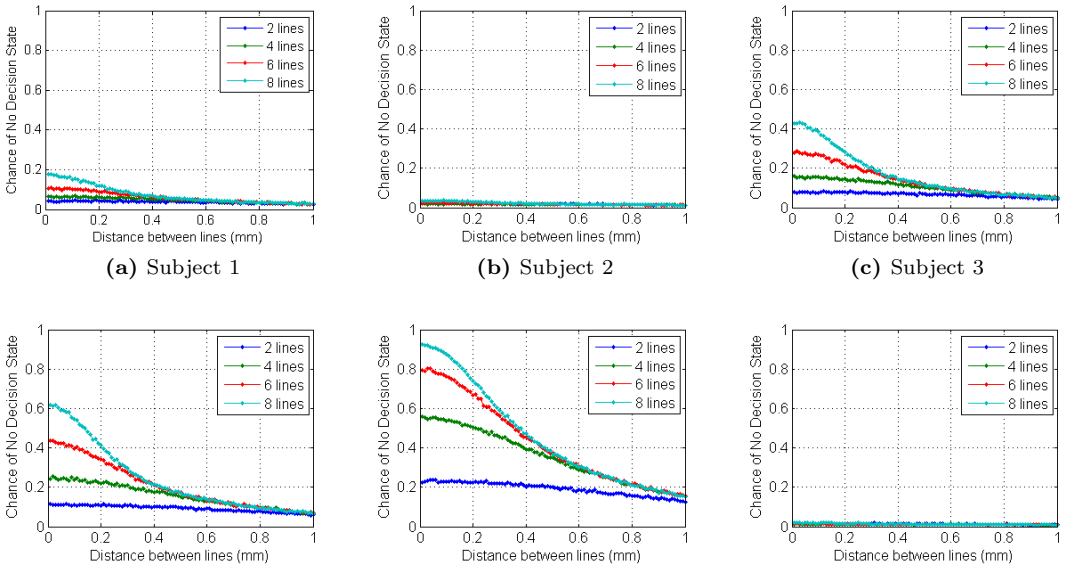
In this second set of subject simulations. we simulated the performance of vague human decision-makers using subjects from the line length and city population size discrimination tasks of [45]. For the line length discrimination task, the authors of [45] modeled 40 human decision makers comparing line lengths in a manner similar to [44]. For the city population size discrimination task, the authors of [45] modeled 91 human decision makers comparing city populations sizes ranked according to the 2010 United States Census Bureau. The decision, confidence assessment, and response time statistics from all subjects for both tasks were used by the authors of [45] to estimate parameters



(a) Subject 1 **(b)** Subject 2 **(c)** Subject 3

(d) Subject 4 **(e)** Subject 5 **(f)** Subject 6

Figure 5.2: Simulated averages of evidence strengths, ξ, for all six 2DSD subject models from [44, Tables 3 and 6] under the 2DSD M-ary human response simulator for the line length discrimination task versus the incremental line length difference, d. Average evidence strengths shown for $M = 2, 4, 6, 8$ alternatives. Averages obtained over 10,000 trials of the M-ary extension algorithm for each subject.



(a) Subject 1 **(b)** Subject 2 **(c)** Subject 3

(d) Subject 4 **(e)** Subject 5 **(f)** Subject 6

Figure 5.3: Estimated chance of no decision case for all six 2DSD subject models from [44, Tables 3 and 6] under the 2DSD M-ary extension algorithm. Averages obtained over 10,000 trials for each subject.

for a 2DSD interrogation model using state-dependent decay and drift rate attenuation. For the line length discrimination task, subject decisions were cued after 250 milliseconds, followed by an interjudgment period of 50 milliseconds before a confidence assessment was cued. For the city population size discrimination task, subject decisions were cued after 1600 milliseconds, followed by an interjudgment period of 500 milliseconds before a confidence assessment was cued. We used the corresponding parameter sets of [45] to simulate M-ary decisions and confidence assessments using Equation (5.3) and Equation (5.4). For the line length discrimination task, decisions were simulated for choosing the $l = 1, 2, 3$ longest lines amongst a set of 9.60, 9.65, 9.72, and 9.73 millimeter long lines. When simulating line length comparisons, pairs of lines which did not have a corresponding drift rate given by [45] were estimated using a similar technique described in Section 5.2.1. For the city population size discrimination task, decisions were simulated for choosing the $l = 1, 2, 3$ most populated city amongst the following four cities: Houston (TX), Philadelphia (PA), Las Vegas (NV), and Aurora (CO). These four cities were ranked as being the 4^{th}, 5^{th}, 31^{st}, and 56^{th} most populated United States cities according to the 2010 United States Census estimate. A total of 10,000 decisions and decision confidence assessments were simulated for each subject and each task, repeating instances in which the "no decision" state was returned.

Figure 5.4 and Figure 5.5 show grouped bar graphs for the average imprecise evidence strength for varying imprecision levels l of the simulated decision A for the line length and city population size discrimination tasks respectively. For a given imprecision level l, the spectrum of colored bars represents the different subjects. For both tasks, the evidence strength of the subjects appears to increase as the subjects are allowed to make less specific decisions (*i.e.*, higher imprecision levels). These trends seem logical, as requiring less specific decisions from subjects should in general lead to higher amounts of subjective confidence being assessed to a correct (*i.e.*, non-contradictory) decision.

Figures 5.6, 5.8, and 5.10 show the normalized histograms for the observed subject decisions on the line length discrimination task while instructing subjects to choose the longest, two longest, and three longest lines respectively. Figures 5.7, 5.9, and 5.11 also show similar normalized histograms,

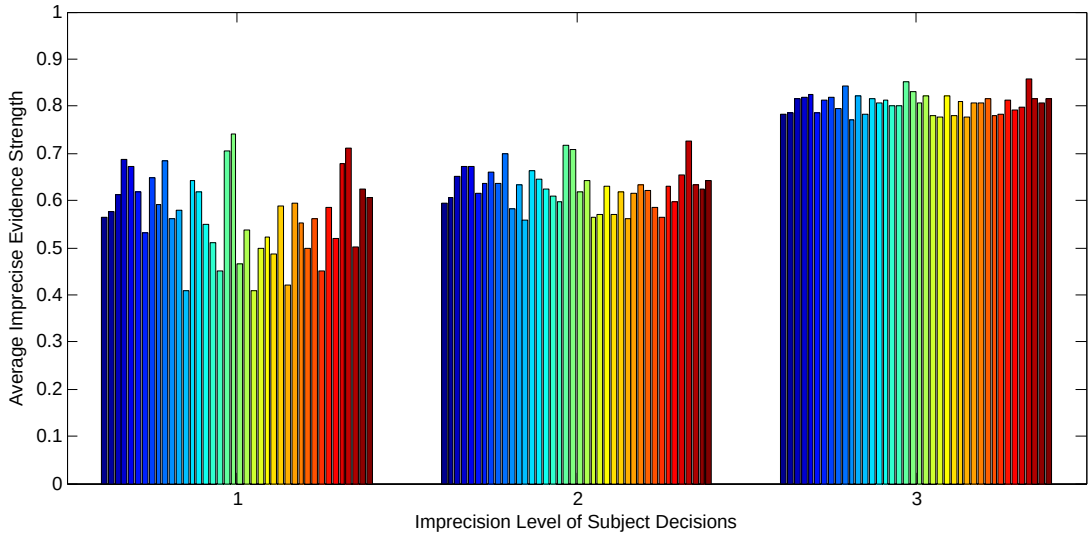


Figure 5.4: Average imprecise evidence strength for each subject on the line length discrimination task of [45]. The spectrum of colored bars represents the average imprecise evidence strength of the 40 subjects when simulated to select the l longest lines amongst a set of 9.60, 9.65, 9.72, and 9.73 millimeter long lines, where $l = 1, 2, 3$. Subject decisions and confidence assessments were simulated over 10,000 trials.

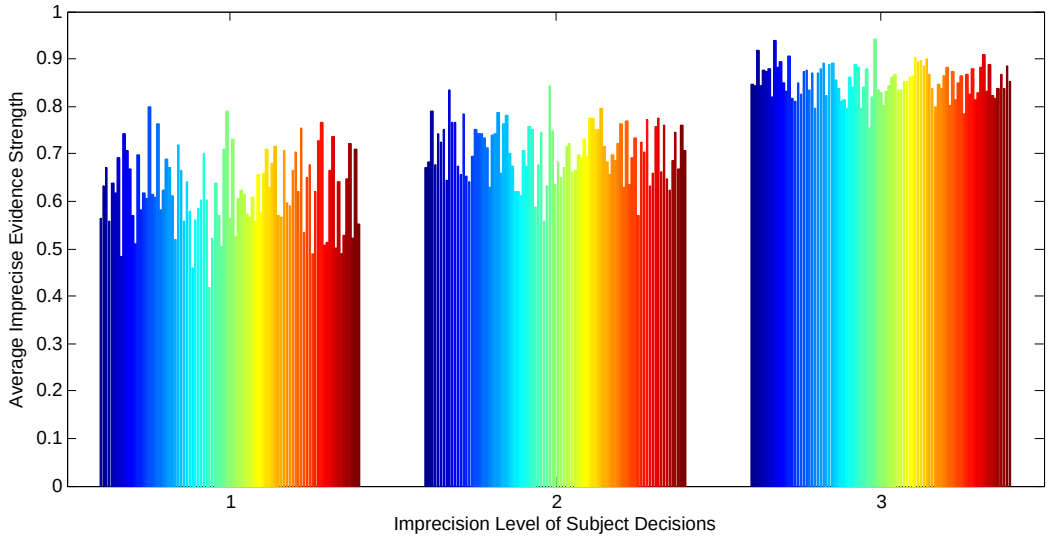


Figure 5.5: Average imprecise evidence strength for each subject on the city population size discrimination task of [45]. The spectrum of colored bars represents the average imprecise evidence strength of the 91 subjects when simulated to select the l most populated cities among Houston (TX), Philadelphia (PA), Las Vegas (NV), and Aurora (CO), where $l = 1, 2, 3$. Subject decisions and confidence assessments were simulated over 10,000 trials.

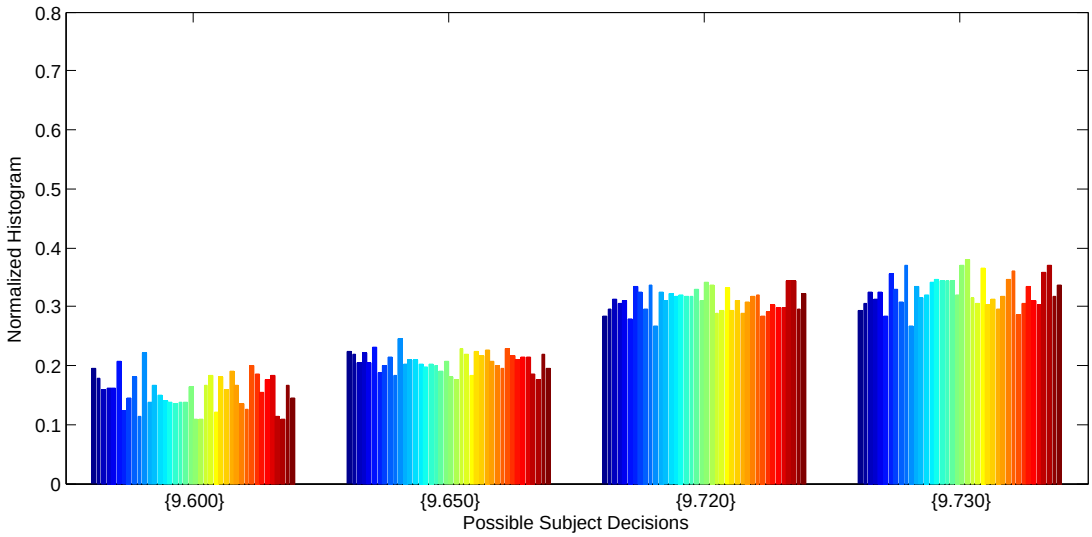


Figure 5.6: Normalized histogram of subject decisions on the line length discrimination task of [45]. A total of 40 subjects were simulated to select the longest line amongst a set of 9.60, 9.65, 9.72, and 9.73 millimeter long lines. Subject decisions and confidence assessments were simulated over 10,000 trials.

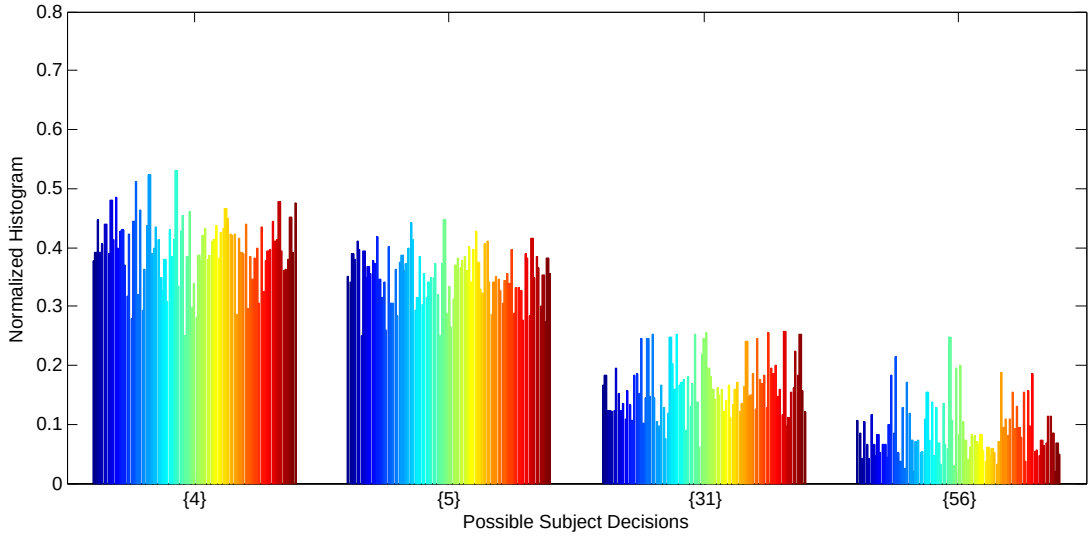


Figure 5.7: Normalized histogram of subject decisions on the city population size discrimination task of [45]. A total of 91 subjects were simulated to select the most populated city among Houston (TX), Philadelphia (PA), Las Vegas (NV), and Aurora (CO). These four cities were ranked the 4^{th}, 5^{th}, 31^{st}, and 56^{th} most populated United States cities respectively, according to the 2010 US Census estimate. Subject decisions and confidence assessments were simulated over 10,000 trials.

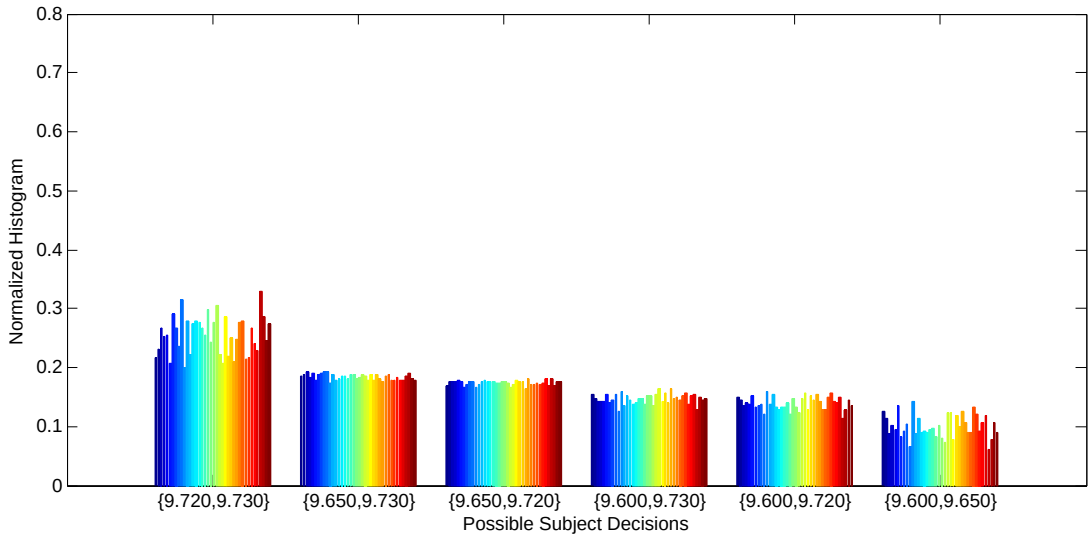


Figure 5.8: Normalized histogram of subject decisions on the line length discrimination task of [45]. A total of 40 subjects were simulated to select the two longest lines amongst a set of 9.60, 9.65, 9.72, and 9.73 millimeter long lines. Subject decisions and confidence assessments were simulated over 10,000 trials.

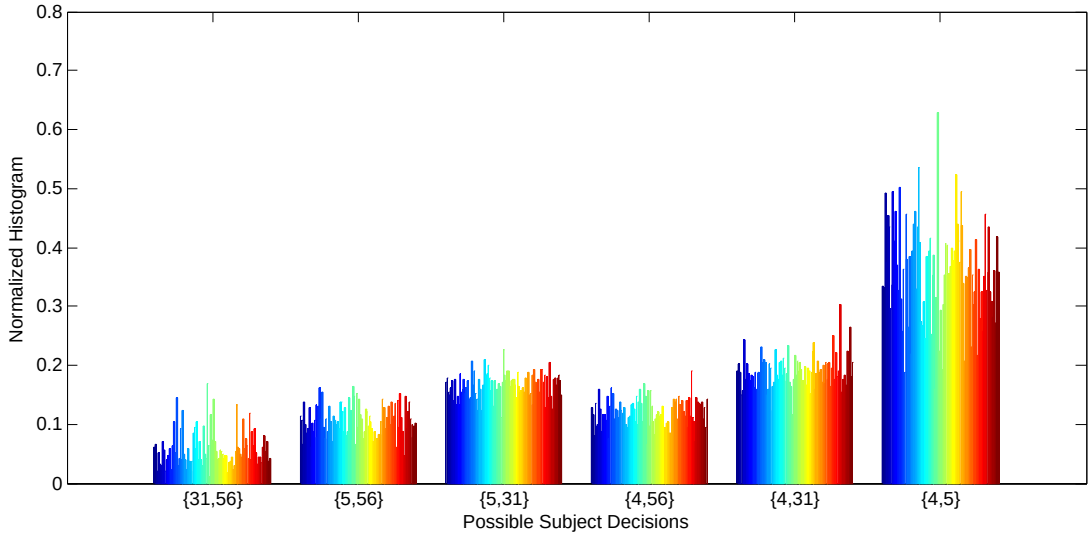


Figure 5.9: Normalized histogram of subject decisions on the city population size discrimination task of [45]. A total of 91 subjects were simulated to select the two most populated cities among Houston (TX), Philadelphia (PA), Las Vegas (NV), and Aurora (CO). These four cities were ranked the 4^{th}, 5^{th}, 31^{st}, and 56^{th} most populated United States cities respectively, according to the 2010 US Census estimate. Subject decisions and confidence assessments were simulated over 10,000 trials.

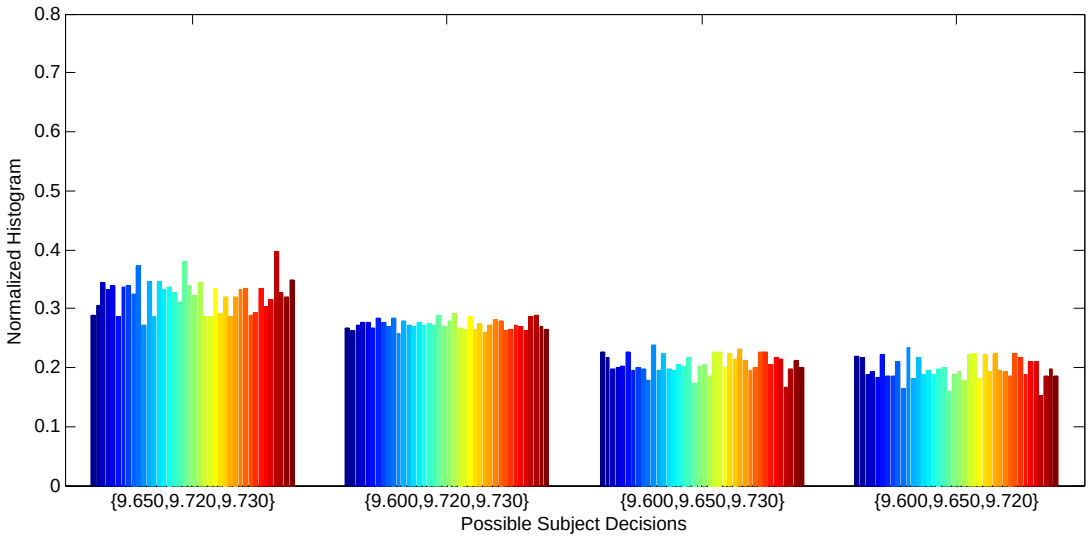


Figure 5.10: Normalized histogram of subject decisions on the line length discrimination task of [45]. A total of 40 subjects were simulated to select the three longest lines amongst a set of 9.60, 9.65, 9.72, and 9.73 millimeter long lines. Subject decisions and confidence assessments were simulated over 10,000 trials.

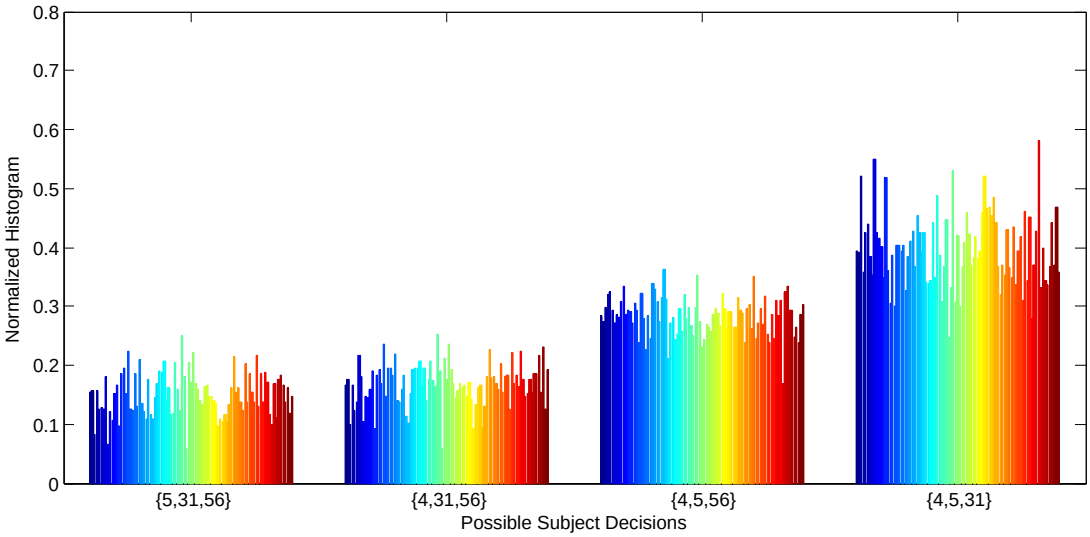


Figure 5.11: Normalized histogram of subject decisions on the city population size discrimination task of [45]. A total of 91 subjects were simulated to select the three most populated cities among Houston (TX), Philadelphia (PA), Las Vegas (NV), and Aurora (CO). These four cities were ranked the 4^{th}, 5^{th}, 31^{st}, and 56^{th} most populated United States cities respectively, according to the 2010 US Census estimate. Subject decisions and confidence assessments were simulated over 10,000 trials.

but for the city population size discrimination task while instructing subjects to chose the largest, two largest, and three largest cities respectively. As expected, the subject decisions which occur most frequently are those which support the correct stimuli. This trend is observed for both decision tasks, and for each of the simulated imprecision levels. Furthermore, for the $l = 2$ imprecision level (*i.e.*, choosing the two longest lines or most populated cities), we note that the most frequently occurring decisions have markedly higher normalized frequencies of occurrence. This trend is again logical, since in both simulated decision tasks, the two longest lines or most populated cities are very similar while also markedly different from the remaining stimuli.

5.3 Fusion Study 1: Fusion with Varying Numbers of Alternatives

A subset of this study's results were previously published in [125].

5.3.1 Motivation

As discussed in the beginning of this chapter, many important soft and hard/soft fusion tasks consist of subjects and sensors discriminating between multiple alternatives. Hence, in this first study we seek to observe the statistical trends in a selection of information fusion operators that have been proposed for use in soft and hard/soft data fusion applications. More specifically, we focus on a selection of fusion operators in which the determination of statistical performance is very difficult in closed form (*e.g.*, Dempster-Shafer theory operators).

5.3.2 Experimental Setup

The proposed successive pairwise comparison method of Equation (5.3) and Equation (5.4) was used to generate M-ary decisions and confidence assessments for the line length discrimination task of [44] where a 2DSD optional stopping model using trial variability on the drift rate, δ, and initial condition, L_0, was used. A total of 10,000 simulation trials were conducted. For each of these trials, a pool of thirty-six human decisions and confidence assessments were generated by simulating six responses from each of the six subject parameters estimated in [44] associated with the line length discrimination task. Furthermore, subjects were simulated to focus on providing accurate decisions and decision confidence assessments. The line lengths presented to the simulated subjects were

$\mathcal{Z} = \{32, 32+d, 32+2d, \ldots, 32+(M-1)d\}$ where M was the number of lines being compared and d was the incremental length difference between lines, in millimeters. For this study, we set $d = 0.20$ millimeters and let $M = 2, 4, 8$.

Each of the decision and confidence assessment pairs were used to construct subjective probability assignments and belief mass assignments (BMA) according to the "Confidences Only" and "Evidence Strength Discounting" construction cases described in Section 4.3.2. When a subject returned the "no decision state", an equiprobable subject probability assignment or a vacuous BMA was generated. The resulting subjective probability assignments and BMAs were used to implement a set of five fusion operators. They are Bayes' rule of probability combination [55], Dempster's Rule of Combination (DRC) [43], Yager's Rule [94], Dubois and Prade's rule (DPR) [138], and the Proportional Conflict Redistribution Rule #5 [97]. These fusion operators were used to combine subjective probabilities or BMAs two at a time, and hence the combination order used was randomized for each simulation trial. The performance of these operators were assessed according to the metrics discussed in Section 4.1.2. Namely, accuracy performance was taken as the average value of ξ_{Bel} (Equation 4.7) and precision performance was taken as the average value of $1 - (\xi_{\text{Pl}} - \xi_{\text{Bel}})$ (Equations 4.7 and (Equation 4.8). As a reminder, accurate belief combination operators will tend to assign probability one to the correct outcome, resulting in values of ξ_{Bel} close to one. Precise belief combination operators will tend to produce more specific evidence, resulting in values of $(\xi_{\text{Pl}} - \xi_{\text{Bel}})$ being close to zero, and hence values of $1 - (\xi_{\text{Pl}} - \xi_{\text{Bel}})$ being close to one.

5.3.3 Results

Figure 5.12 shows the accuracy performance for each of the five fusion operators simulated here (*i.e.*, ξ_{Bel}) versus the number of sources in the combination. The precision performance for each of the five fusion operators simulated here (*i.e.*, $1 - (\xi_{\text{Pl}} - \xi_{\text{Bel}})$) is shown in Figure 5.13. The six subfigures associated with each of these two figures represent different combinations of the number of alternatives ($M = 2, 4, 8$) and the two fusion operator input construction cases, "Confidences Only" and "Evidence Strength Discounting."

When source discounting is not performed, Bayes' rule of probability combination and DRC could

Table 5.1: Summary of M-ary soft fusion performance results for the experiment setup defined in Section 5.3.2. Results shown in terms of the average post-fusion accuracy, ξ_{Bel}, toward the correct alternative and for each BMA/subjective probability construction case. These trends were observed in the two, four, and eight alternative decision tasks simulated.

Performance Order	Confidences Only	Evidence Strength Discounting
1 (Best)	PCR5	Bayes/DRC
2	Yager/DPR	PCR5
3	—	Yager/DPR
4	—	—
5 (Worst)	—	—

Table 5.2: Summary of M-ary soft fusion performance results for the experiment setup defined in Section 5.3.2. Results shown in terms of the average post-fusion precision, $1-(\xi_{\text{Pl}}-\xi_{\text{Bel}})$, toward the correct alternative and for each BMA/subjective probability construction case. These trends were observed in the two, four, and eight alternative decision tasks simulated.

Performance Order	Confidences Only	Evidence Strength Discounting
1 (Best)	PCR5	Bayes
2	Yager/DPR	DRC/PCR5
3	—	Yager/DPR
4	—	—
5 (Worst)	—	—

not be used, for similar reasons discussed in Section 4.3.3; the chances of any two simulated subjects presenting totally conflicting evidence was non-negligible, resulting in the possibility of a division by zero in the equations for Bayes' rule of probability combination and for DRC. A summary of the fusion results are shown in Table 5.1 for the average accuracy performance and in Table 5.2 for the average precision performance of each operator. We also make note of the following observations.

- When source discounting was performed using average source evidence strength, Bayes' rule of probability combination and DRC exhibited similar accuracy performance. The number of alternatives was also observed to have a stronger impact on the accuracy performance when source discounting was not performed (Figure 5.12).

- Similar to the subject performance results (Figure 5.2), the largest decrease in accuracy performance occurred when going from 2 to 4 alternatives. The decrease was smaller when going from 4 to 8 alternatives (Figure 5.12).

- When source discounting was performed, similar performance was observed by PCR5, Bayes'

rule of probability combination and DRC, as long as there were twelve or less human responses in the combination. When more than twelve human responses were included in the combination, Bayes' rule of probability combination and DRC exhibited higher accuracy performance than PCR5 (Figures 5.12b, 5.12d, and 5.12f).

- When source discounting was performed, PCR5 and DRC precision increased as the number of sources present in the combination increased. Eventually the precision converged to one for both PCR5 and DRC. This convergence was observed to occur more quickly with PCR5 than with DRC. Additionally, it was observed that increasing the number of alternatives decreased the rate of convergence for both PCR5 and DRC (Figures 5.13b, 5.13d, and 5.13f).

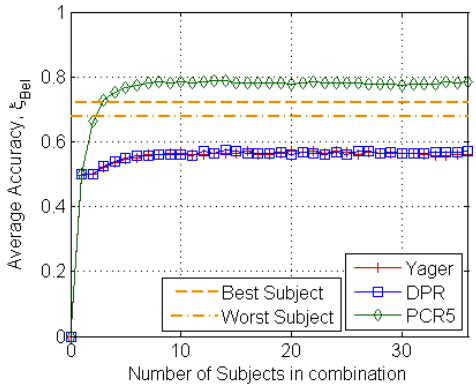
- When source discounting was not performed, the precision performance of PCR5 was found to be the same regardless of the number of line length task alternatives (Figures 5.13a, 5.13c, and 5.13e).

- Yager's rule and Dubois and Prade's rule exhibited inferior accuracy performance in all cases. Both rules exhibited accuracy performance that was worse than the worst single source present in the combination. Furthermore, Yager's rule and Dubois and Prade's rule also exhibited the lowest precision performance.

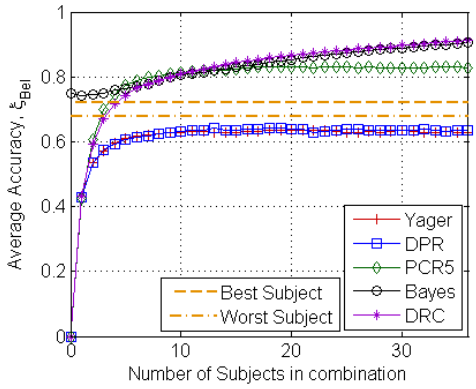
5.4 Fusion Study 2: Fusion of Vague Decisions and Confidences

5.4.1 Motivation

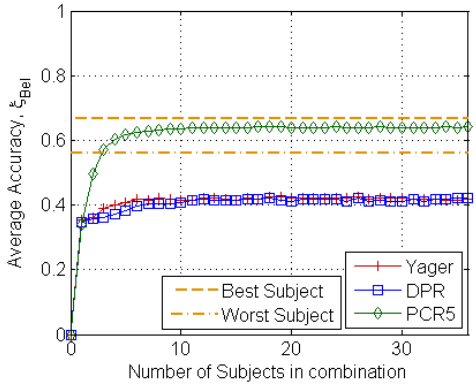
In all of the studies thus far, the performance of Bayes' rule and Dempster's rule of combination (DRC) has been similar. Several studies have investigated the proper usage of each of these combination operators (*e.g.*, [61–64]). The conclusions were not consistent owing to the use of different examples/counterexamples and Monte Carlo simulations based on different assumptions. For example, the results of [63] suggest similar accuracy performance between Bayes' rule and DRC, and further that Bayes' rule exhibits faster convergence time in a target tracking application when compared to DRC. The work in [64] however suggests that as the imprecision in the combined evidence



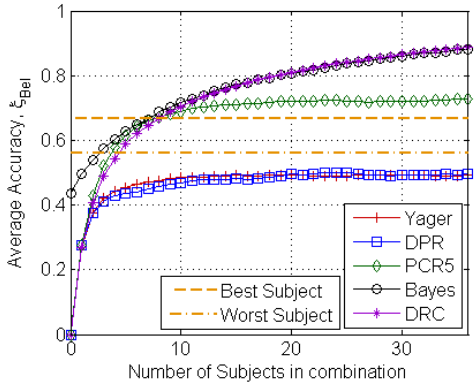
(a) "Confidences Only", $M = 2$



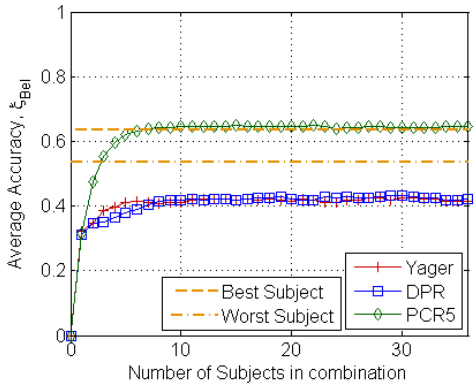
(b) "Evidence Strength Discounting", $M = 2$



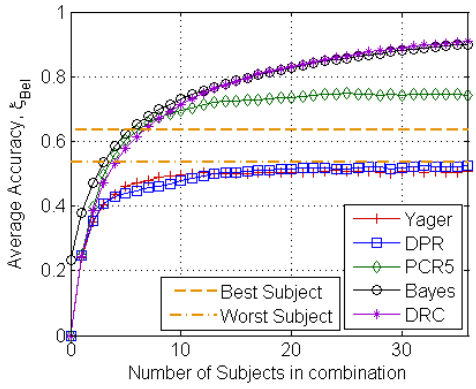
(c) "Confidences Only", $M = 4$



(d) "Evidence Strength Discounting", $M = 4$

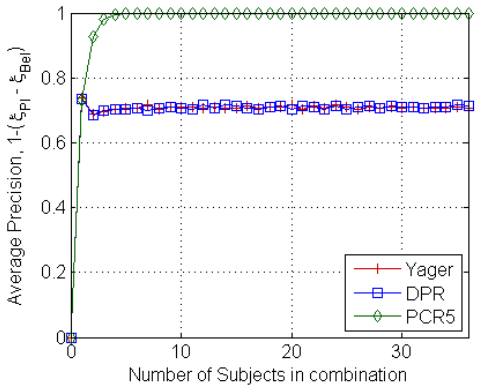


(e) "Confidences Only", $M = 8$

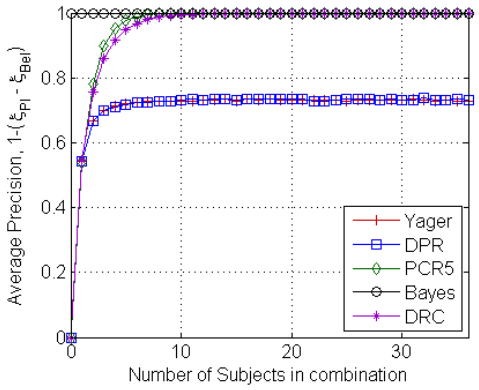


(f) "Evidence Strength Discounting", $M = 8$

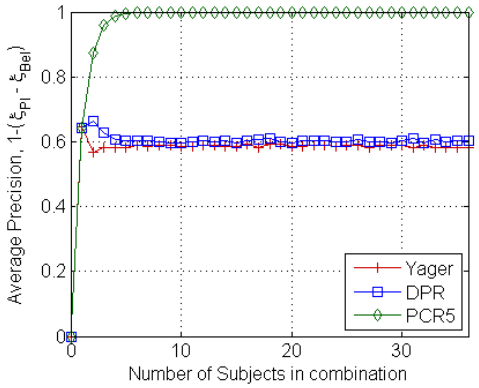
Figure 5.12: Average accuracy performance (*i.e.*, ξ_{Bel}) for each of the five fusion methods versus the number of sources present in combination (higher is better). Results simulated over 10,000 trials for the M-ary line length discrimination task with $M = 2, 4, 8$. The evidence strengths for the best and worst subjects in the combination are also shown for comparison.



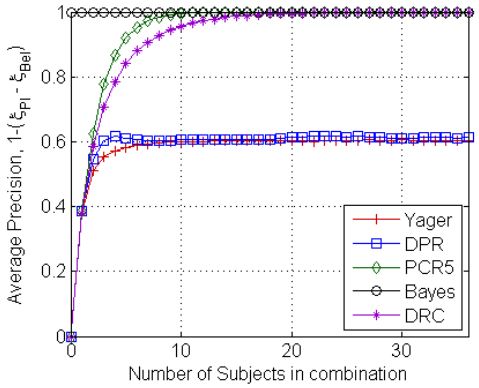
(a) "Confidences Only", $M = 2$



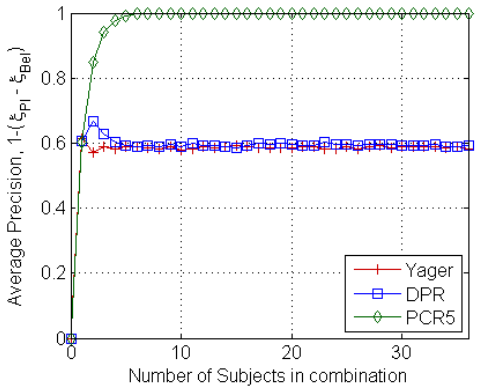
(b) "Evidence Strength Discounting", $M = 2$



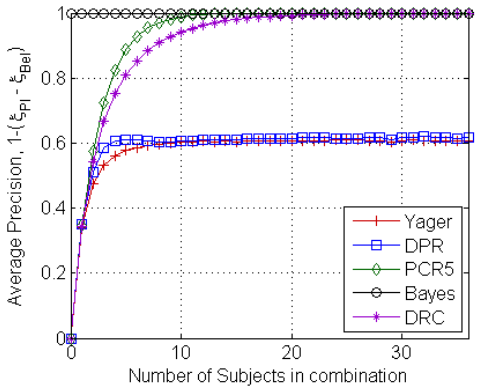
(c) "Confidences Only", $M = 4$



(d) "Evidence Strength Discounting", $M = 4$



(e) "Confidences Only", $M = 8$



(f) "Evidence Strength Discounting", $M = 8$

Figure 5.13: Average precision performance (*i.e.*, $1 - (\xi_{\text{Pl}} - \xi_{\text{Bel}})$ for each of the five fusion methods versus the number of sources present in combination (higher is better). Results simulated over 10,000 trials for the M-ary line length discrimination task with $M = 2, 4, 8$.

increases, the Dempster-Shafer theory rules tend to perform better than rules from Bayesian epistemology. In this section, we seek to investigate these claims further for soft fusion operators that use vague decisions. This is an important simulation case, as one of the main motivations of using a Dempser-Shafer (DS) theoretic approach over a Bayesian approach is in the purported ability of fusion operators from DS theory to handle vague information better [43, 59].

5.4.2 Experimental Setup

The proposed successive pairwise comparison method of Equation (5.3) and Equation (5.4) was used to generate vague decisions and confidence assessments in a manner similar to that used in Section 5.2.2 using the line length and city population size discrimination tasks of [45]. Specifically, all 40 human decision makers on the line length discrimination task and the first 40 of the 91 human decision makers on the city population size discrimination task were simulated using the 2DSD parameters estimated by the authors of [45]. A total of 10,000 simulation trials were conducted. For each trial, decisions and confidence assessments were generated using Equation (5.3) and Equation (5.4). For the line length discrimination task, decisions were simulated for choosing the $l = 1, 2, 3$ longest lines amongst a set of 9.60, 9.65, 9.72, and 9.73 millimeter long lines. When simulating line length comparisons, pairs of lines which did not have a corresponding drift rate given by [45] were estimated using linear interpolation of the drift rates estimated in [45] (Section 5.2.1). For each of the pairwise successive comparisons, the 2DSD parameter sets that were used cued simulated decisions approximately 250 milliseconds after initial presentation of the stimuli, and decision confidence assessments approximately 50 milliseconds after declaration of a decision. For the city population size discrimination task, decisions were simulated for choosing the $m = 1, 2, 3$ most populated city amongst the following four cities: Houston (TX), Philadelphia (PA), Las Vegas (NV), and Aurora (CO). These four cities were ranked as being the 4^{th}, 5^{th}, 31^{st}, and 56^{th} most populated United States cities respectively. For each of the pairwise comparison simulations, the 2DSD parameter sets that were used cued simulated decisions approximately 1600 milliseconds after initial presentation of the stimuli, and decision confidence assessments approximately 500 milliseconds after declaration of a decision.

Each of the subject decision and confidence assessment pairs were used to construct subjective probability assignments and belief mass assignments (BMAs) similar to the "Confidences Only" and "Evidence Strength Discounting" construction cases described in Section 4.3.2. In the "Confidences Only" case, subjective probability assignments and BMAs were constructed for the i^{th} simulated human decision-maker such that the belief attributed to his or her decision, $A_i \subseteq \Omega$, is equal to his or her decision confidence, $p_{A_i} \in [0, 1]$. Specifically, the subjective probabilities for each subject, $\mathcal{P}_i(\cdot)$, were constructed such that

$$\mathcal{P}_i(\omega) = \begin{cases} \frac{1}{|A_i|} p_{A_i} & \omega \in A_i \\ \\ \frac{1}{|\Omega| - |A_i|}(1 - p_{A_i}) & \text{otherwise} \end{cases} . \tag{5.8}$$

Similarly, simply supported BMAs for each subject, $m_i(\cdot)$, were constructed such that

$$m_i(X) = \begin{cases} p_{A_i} & X = A_i \\ 1 - p_{A_i} & X = \Omega \\ 0 & \text{otherwise} \end{cases} . \tag{5.9}$$

When a subject returned the "no decision state", an equiprobable subjective probability assignment or a vacuous BMA was generated. For the "Evidence Strength Discounting" case, the average imprecise evidence strength for each subject was used as their corresponding discount rates in order to implement the source discounting methods described by Equation 4.10 for subjective probability assignments and Equation 4.9 for BMAs. We also considered a third subjective probability assignment and BMA construction case known as "Intersubject Conflict Discounting." Average intersubject conflict discounting is presented in [139], and again implements the discounting operation of Equation (4.10) and Equation (4.9). For a given simulation trial, the discount rate used for each subject is defined as the average distance of the current subject's BMA (or subjective probability assignment) from those produced by the other subjects. Formally, the intersubject conflict discount

rate α_i is defined for the i^{th} subject amongst a pool of N subjects as

$$\alpha_i = 1 - \frac{1}{N-1} \sum_{\substack{j \in \{1,\ldots,N\} \\ j \neq i}} \sqrt{\frac{1}{2}(\mathbf{m}_i - \mathbf{m}_j)^T \mathbb{D}(\mathbf{m}_i - \mathbf{m}_j)}. \tag{5.10}$$

Here $\mathbf{m}_i$ and $\mathbf{m}_j$ are BMAs written in vector form, and $\mathbb{D}$ is a $2^M \times 2^M$ matrix whose elements are constructed such that

$$\mathbb{D}_{A,B} = \begin{cases} 1 & A = B = \emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \text{otherwise} \end{cases}. \tag{5.11}$$

In the above formulation, the sets A and B are elements of the power set corresponding to the element order used in construction of the BMA vectors $\mathbf{m}_i$ and $\mathbf{m}_j$.

When a subject returned the "no decision state," an equiprobable subjective probability assignment or a vacuous BMA was generated. The resulting subjective probability assignments and BMAs were used to implement a set of six fusion operators. They are Bayes' rule of probability combination [55]; Dempster's Rule of Combination (DRC) [43]; Yager's Rule [94]; Dubois and Prade's rule (DPR) [138]; the Proportional Conflict Redistribution Rule #5 [97]; and Murphy's combination rule [99, 100]. These fusion operators were used to combine subjective probabilities or BMAs two at a time. The combination order used was randomized for each simulation trial. For the Dempster-Shafer theory operators, combined BMAs were used to generate subjective probabilities using the pignistic probability transform defined in Equation (2.34). The pignistic probabilities generated from each fusion operator towards the correct alternative were averaged over all of the simulation trials. Additionally, the uncertainty towards the correct alternative of Equation (2.16) was calculated from each combined BMA and averaged over all of the simulation trials.

5.4.3 Results

Figures 5.14 through 5.16 show the performance of the six (6) fusion operators investigated in this study versus the number of subject responses included in the final fused result for the M-ary line length discrimination task ($M = 4$). Specifically, Figure 5.14 shows the performance of the six (6) fusion operators when subjects were simulated to choose a single most correct alternative ($l = 1$),

Table 5.3: Best performing fusion operator associated with the experiment setup defined in Section 5.4.2. Results shown in terms of the average post-fusion pignistic probability towards the correct alternative and for each BMA/subjective probability construction case and imprecision level. The observed trends were the same for the line length discrimination and city population tasks.

Imprecision Level ($l = \|A\|$)	Confidences Only	Evidence Strength Discounting	Intersubject Conflict Discounting
$l = 1$	PCR5/Yager/DPR	Bayes	Bayes
$l = 2$	PCR5/Yager/DPR	Bayes/DRC	Bayes
$l = 3$	PCR5/Yager/DPR	Bayes/DRC	Bayes/DRC

Figure 5.15 shows the performance of the six (6) fusion operators when subjects were simulated to chose the $l = 2$ most correct alternatives, and Figure 5.16 shows the performance of the six (6) fusion operators when subjects were simulated to chose the $l = 3$ most correct alternatives. Similar plots are shown for the M-ary city population size discrimination task ($M = 4$) in Figures 5.17 through 5.19. Each of these figures consist of 6 subplots. In subplots (a), (c), and (e) of each figure, the average post-fusion pignistic probability generated for true alternative is shown versus the number of subject responses included in the final fused result. Subplots (b), (d), and (f) of each figure show the average post-fusion uncertainty for the true alternative versus the number of subject responses included in the final fused result. For both sets of subplots, the performance of each fusion operator is shown for the "Confidences Only," "Evidence Strength Discounting," and "Intersubject Conflict Discounting" fusion operator input construction cases (Section 5.4.2). For this study, we consider fusion operators that give higher pignistic probabilities towards the correct outcome to be more accurate, and fusion operators that give lower average uncertainty towards the correct outcome to be more precise.

Similar to the results of Section 5.3.3, Bayes' rule of probability combination and DRC could not be used when source discounting was not performed (*i.e.*, all of the "Confidences Only" fusion operator input construction cases). A summary of the best performing fusion operators in each case is shown in Table 5.3 in terms of average post-fusion pignistic probabilities towards the correct alternative and in Table 5.4 in terms of the average post-fusion uncertainty measure towards the correct alternative. We make the following additional observations.

Table 5.4: Best performing fusion operator associated with the experimental setup defined in Section 5.4.2. Results shown in terms of the average post-fusion uncertainty measure towards the correct alternative and for each BMA/subjective probability construction case and imprecision level. The observed trends were the same for the line length discrimination and city population tasks.

Imprecision Level ($l = \|A\|$)	Confidences Only	Evidence Strength Discounting	Intersubject Conflict Discounting
$l = 1$	PCR5	Bayes	Bayes
$l = 2$	PCR5	Bayes	Bayes
$l = 3$	PCR5	Bayes	Bayes

- Increasing the vagueness of the decision task generally leads to a decrease in the average pignistic probability and an increase in the average uncertainty measure towards the correct alternative. This trend was observed for both the M-ary line length and city population size discrimination tasks.

- Out of the three (3) fusion operator input construction cases, the "Evidence Strength Discounting" case produced the highest average pignistic probabilities and lowest average uncertainty measures towards the correct alternative. For the M-ary line length discrimination task, the average pignistic probabilities when going from the "Evidence Strength Discounting" case to the "Intersubject Conflict Discounting" case was minimal. Also for the M-ary line length discrimination task, it took the fusion operators an additional five (5) subjects in the "Intersubject Conflict Discounting" to reach the average uncertainty towards the correct alternative achieved by the "Evidence Strength Discounting" case. For the M-ary city population size discrimination task, the decreases in performance were much more significant when decision vagueness was increased.

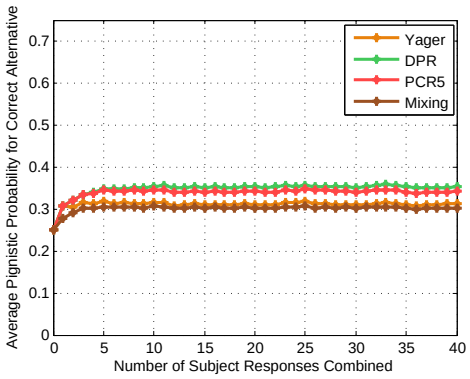
- In six (6) out of the twelve (12) cases where Bayes' rule of probability combination and DRC could be used (*i.e.*, the "Evidence Strength Discounting" and "Intersubject Conflict Discounting" cases), Bayes' rule produced higher pignistic probabilities towards the correct alternative than DRC. In the remaining six (6) cases, the average pignistic probabilities towards the correct alternative for Bayes' rule and DRC were indistinguishable. The next highest average pignistic probabilities were then produced by PCR5, followed by DPR, Yager's rule, and finally

Murphy's combination rule. This performance ordering of the fusion operators was similarly observed for the average uncertainty towards the correct outcome (*i.e.*, Bayes' rule and DRC yielded the lowest uncertainties, followed by PCR5, and so on).

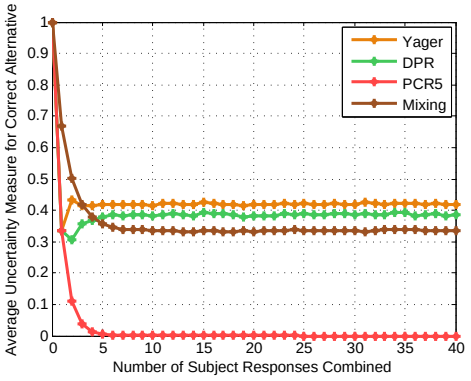
- Similar to the results of Section 5.3.3, the use of source discounting with Bayes' rule of probability combination and DRC were observed to produce monotonically increasing belief in the true outcome with respect to the number of sources included in fusion. The belief towards the true outcome for all of the remaining fusion operators was observed to stop increasing after a certain number of sources were included in fusion.

5.5 Chapter Summary

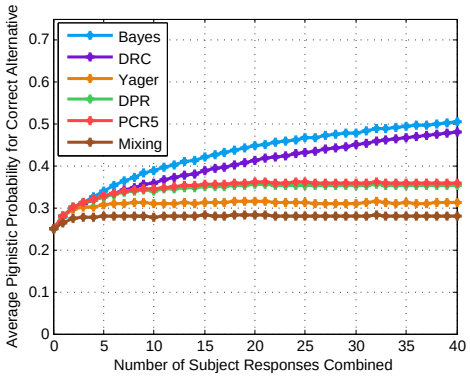
In this chapter, a method for simulating precise and vague human decisions and decision confidence assessments using pairwise successive comparisons was presented. After discussing the general definition of the pairwise successive comparison technique, two-stage dynamic signal detection (2DSD) parameters from [44] and [45] were used to simulate the performance of a selection of fusion operators on a M-ary line length discrimination task and a M-ary city population size discrimination task. In the first study, we observed the performance of a set of fusion operators when increasing the number of decision alternatives. In the second study, we observed the performance of a set of fusion operators when increasing the level of vagueness present in subject decisions while holding the number of alternatives constant. In both studies, Bayes' rule of probability combination [55] exhibited similar performance to DRC [43]. In fact, we observed that Bayes' rule produced slightly higher post fusion subjective probabilities towards the correct alternative in half of the vague fusion simulation cases investigated where Bayes' rule and DRC could be used. The worst fusion performance was given by Yager's rule [94], Dubois and Prade's rule [138], and Murphy's combination rule [100].



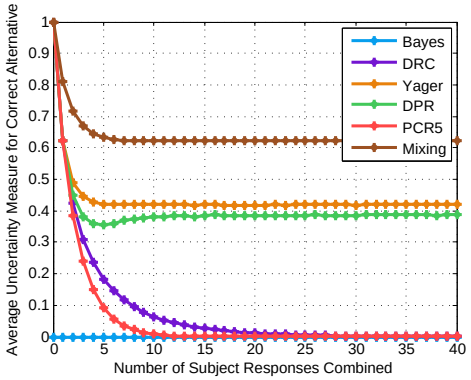
(a) Average pignistic probability for true alternative after fusion, "Confidences Only" operator input construction case



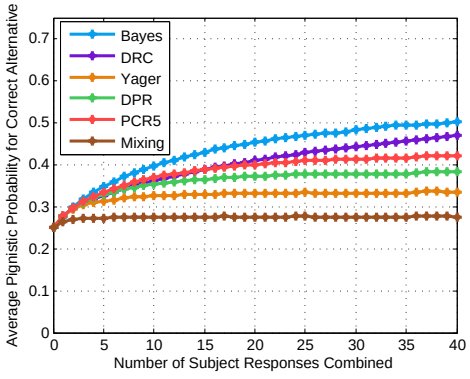
(b) Average uncertainty for true alternative after fusion, "Confidences Only" operator input construction case



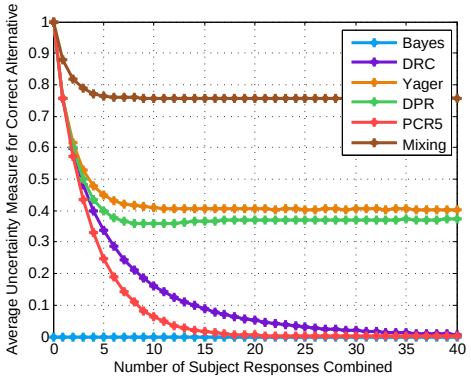
(c) Average pignistic probability for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(d) Average uncertainty for true alternative after fusion, "Evidence Strength Discounting" operator input construction case

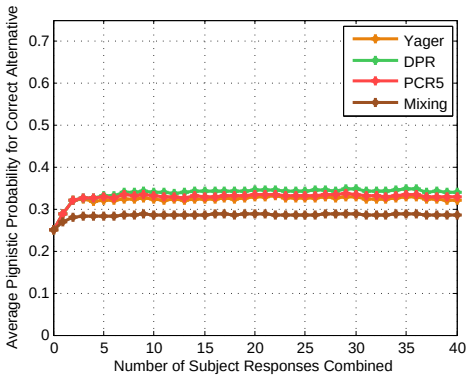


(e) Average pignistic probability for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

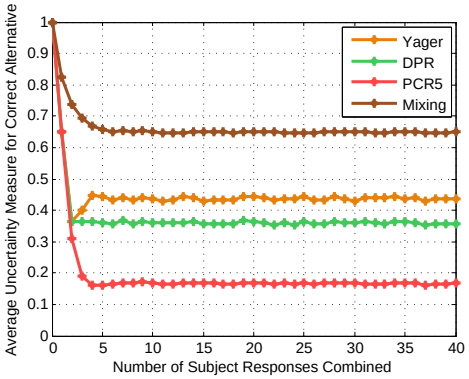


(f) Average uncertainty for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

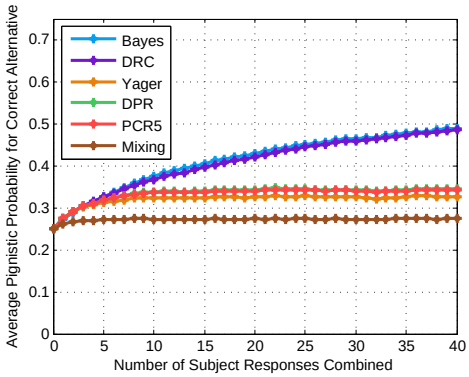
Figure 5.14: Performance of each of the six fusion operators versus the number of subjects included in the combination for the M-ary line length discrimination task ($M = 4$). Subjects were simulated to chose the l most correct alternatives, where $l = 1$. Subplots show average pignistic probability and average uncertainty for the true alternative after performing fusion, and in terms of the fusion operator input construction case used.



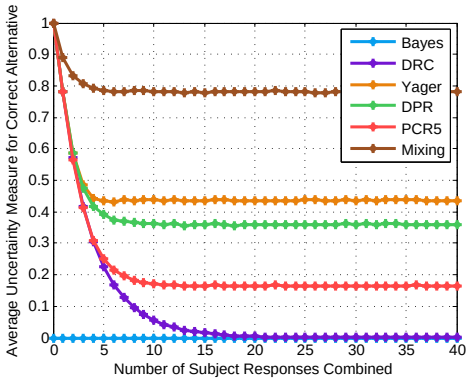
(a) Average pignistic probability for true alternative after fusion, "Confidences Only" operator input construction case



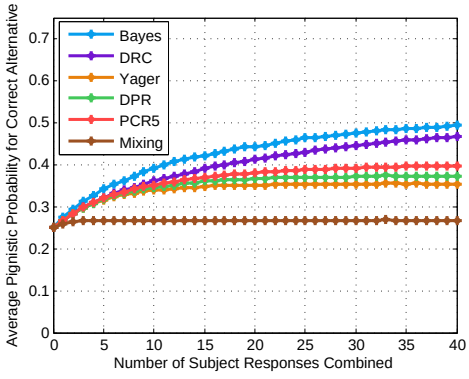
(b) Average uncertainty for true alternative after fusion, "Confidences Only" operator input construction case



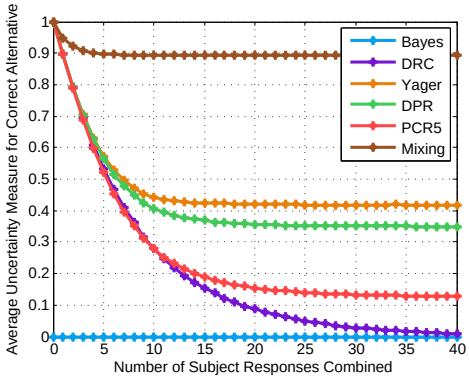
(c) Average pignistic probability for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(d) Average uncertainty for true alternative after fusion, "Evidence Strength Discounting" operator input construction case

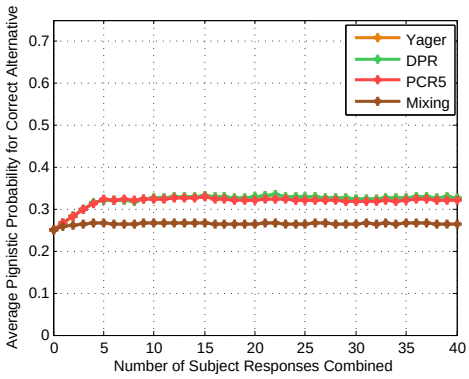


(e) Average pignistic probability for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

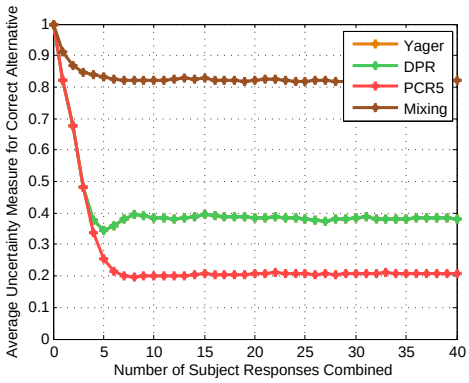


(f) Average uncertainty for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

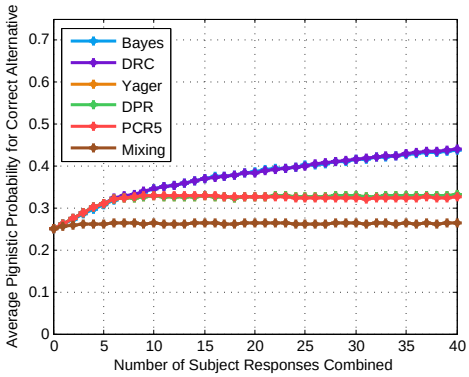
Figure 5.15: Performance of each of the six fusion operators versus the number of subjects included in the combination for the M-ary line length discrimination task ($M = 4$). Subjects were simulated to chose the l most correct alternatives, where $l = 2$. Subplots show average pignistic probability and average uncertainty for the true alternative after performing fusion, and in terms of the fusion operator input construction case used.



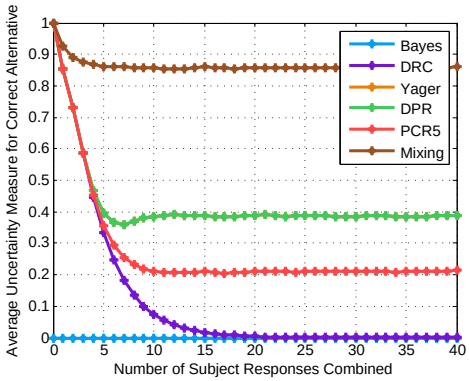
(a) Average pignistic probability for true alternative after fusion, "Confidences Only" operator input construction case



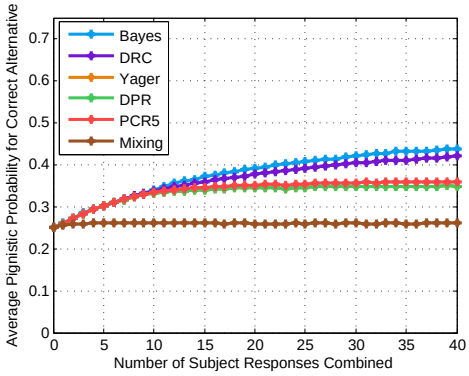
(b) Average uncertainty for true alternative after fusion, "Confidences Only" operator input construction case



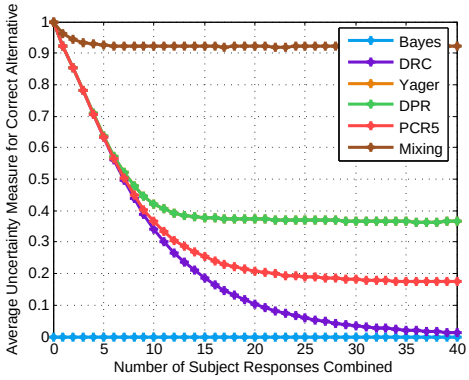
(c) Average pignistic probability for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(d) Average uncertainty for true alternative after fusion, "Evidence Strength Discounting" operator input construction case

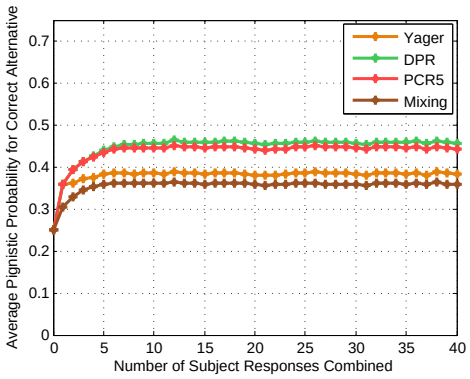


(e) Average pignistic probability for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

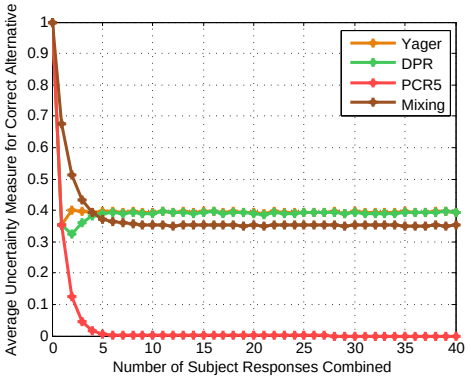


(f) Average uncertainty for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

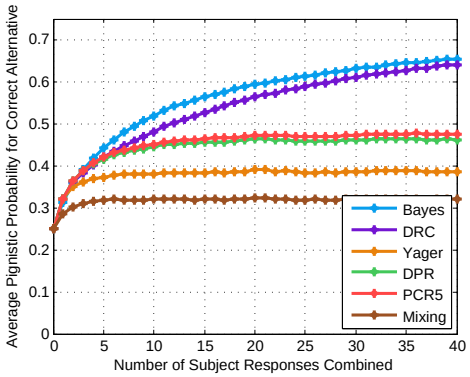
Figure 5.16: Performance of each of the six fusion operators versus the number of subjects included in the combination for the M-ary line length discrimination task ($M = 4$). Subjects were simulated to chose the l most correct alternatives, where $l = 3$. Subplots show average pignistic probability and average uncertainty for the true alternative after performing fusion, and in terms of the fusion operator input construction case used.



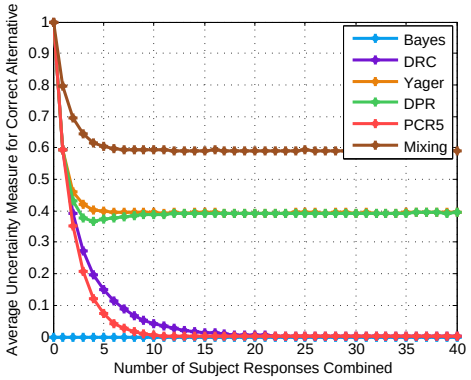
(a) Average pignistic probability for true alternative after fusion, "Confidences Only" operator input construction case



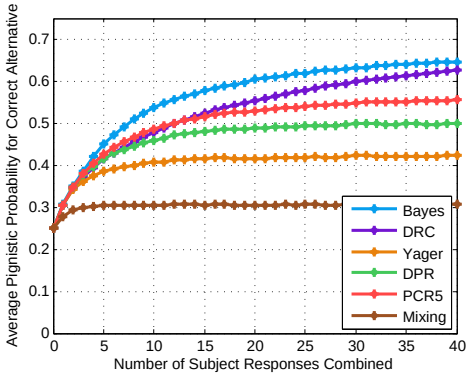
(b) Average uncertainty for true alternative after fusion, "Confidences Only" operator input construction case



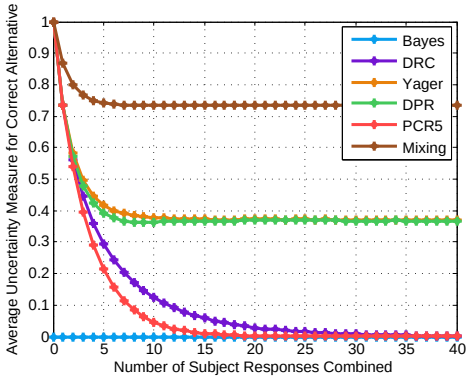
(c) Average pignistic probability for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(d) Average uncertainty for true alternative after fusion, "Evidence Strength Discounting" operator input construction case

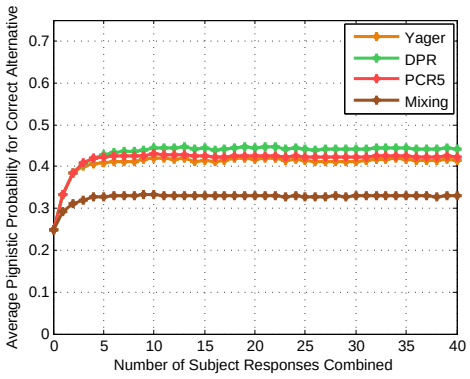


(e) Average pignistic probability for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

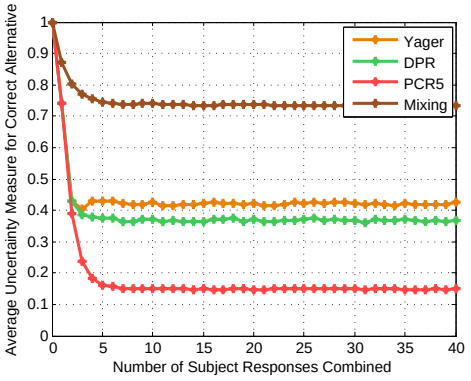


(f) Average uncertainty for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

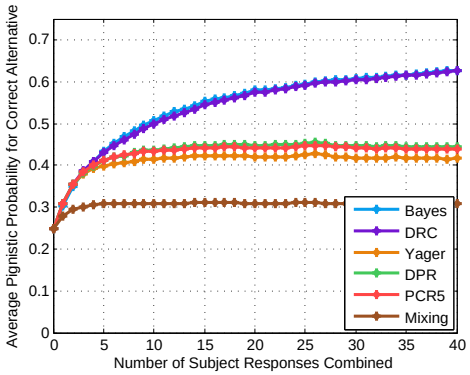
Figure 5.17: Performance of each of the six fusion operators versus the number of subjects included in the combination for the M-ary city population size discrimination task ($M = 4$). Subjects were simulated to chose the l most correct alternatives, where $l = 1$. Subplots show average pignistic probability and average uncertainty for the true alternative after performing fusion, and in terms of the fusion operator input construction case used.



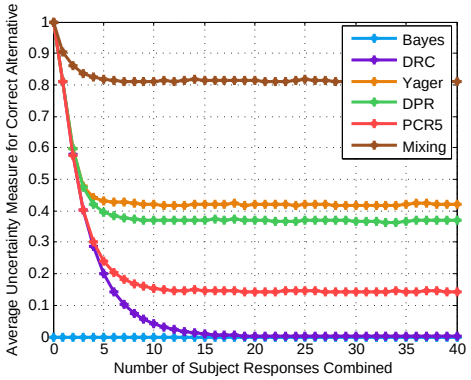
(a) Average pignistic probability for true alternative after fusion, "Confidences Only" operator input construction case



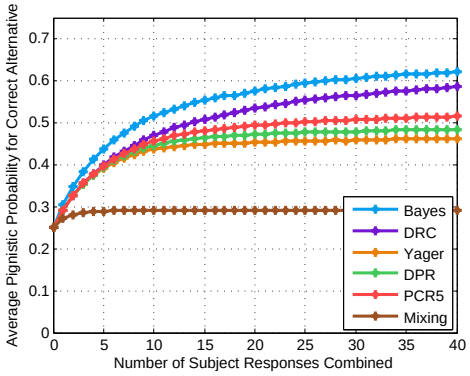
(b) Average uncertainty for true alternative after fusion, "Confidences Only" operator input construction case



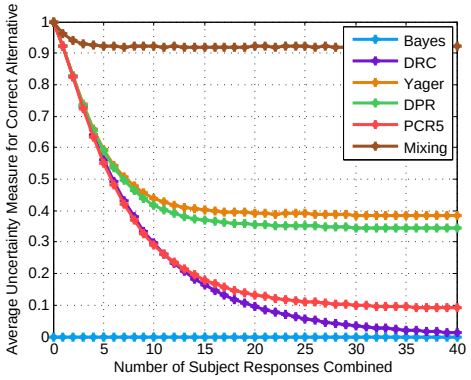
(c) Average pignistic probability for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(d) Average uncertainty for true alternative after fusion, "Evidence Strength Discounting" operator input construction case

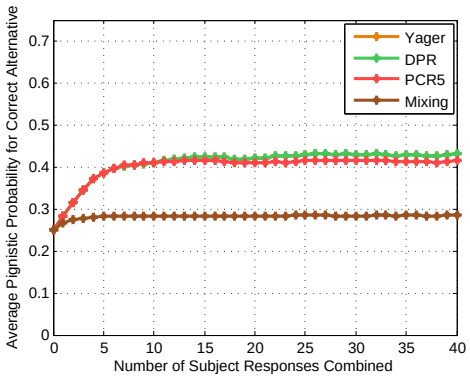


(e) Average pignistic probability for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

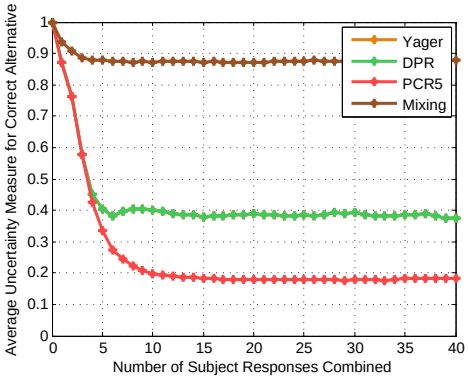


(f) Average uncertainty for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

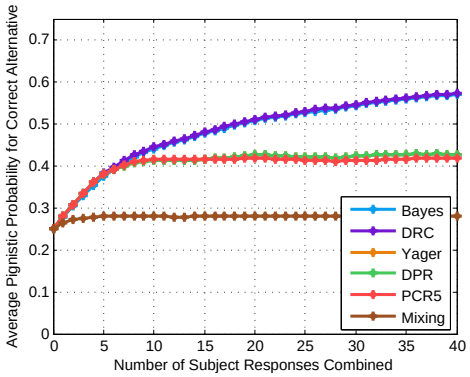
Figure 5.18: Performance of each of the six fusion operators versus the number of subjects included in the combination for the M-ary city population size discrimination task ($M = 4$). Subjects were simulated to chose the l most correct alternatives, where $l = 2$. Subplots show average pignistic probability and average uncertainty for the true alternative after performing fusion, and in terms of the fusion operator input construction case used.



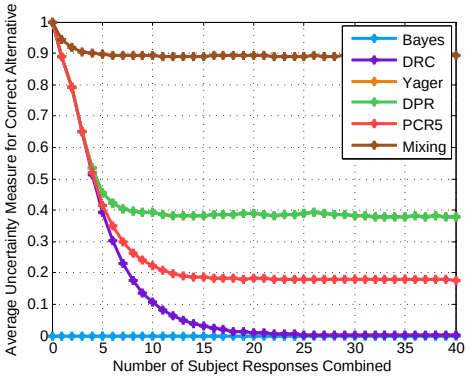
(a) Average pignistic probability for true alternative after fusion, "Confidences Only" operator input construction case



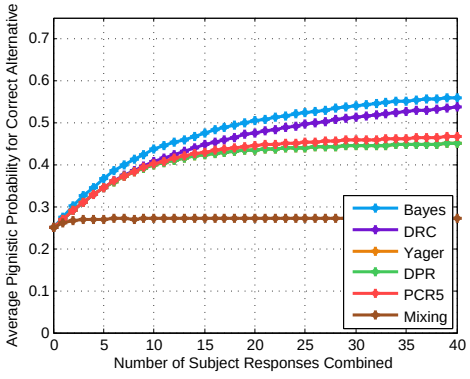
(b) Average uncertainty for true alternative after fusion, "Confidences Only" operator input construction case



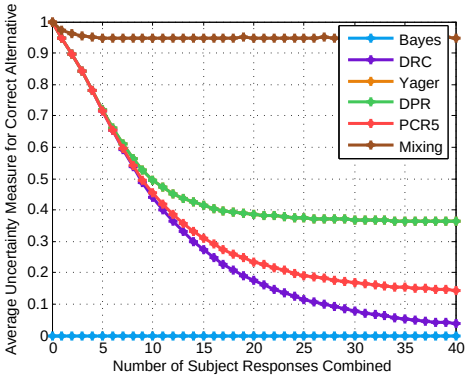
(c) Average pignistic probability for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(d) Average uncertainty for true alternative after fusion, "Evidence Strength Discounting" operator input construction case



(e) Average pignistic probability for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case



(f) Average uncertainty for true alternative after fusion, "Intersubject Conflict Discounting" operator input construction case

Figure 5.19: Performance of each of the six fusion operators versus the number of subjects included in the combination for the M-ary city population size discrimination task ($M = 4$). Subjects were simulated to chose the l most correct alternatives, where $l = 3$. Subplots show average pignistic probability and average uncertainty for the true alternative after performing fusion, and in terms of the fusion operator input construction case used.

Chapter 6: Simulation of Hard and Soft Fusion Operators

In the previous chapters, we considered probabilistic models of human decision making when evaluating statistical performance trends of "soft" fusion operators. In this chapter, two-stage dynamic signal detection (2DSD) models are used to simulate fusion operator performance when combining "hard" and "soft" decisions and decision confidence assessments on the random dot movement (RDM) decision task investigated by the authors of [45]. We begin this chapter by summarizing the experimental setup of the RDM task used in [45]. Then, we describe a Naive Bayes pattern recognition algorithm [128] for the RDM decision task. This algorithm is used in an experimental setup similar to the one used by the authors of [45] in order to generate a database of "hard" data on the RDM decision task. The database is then used to simulate "soft" data using 2DSD human decision making models. Finally, we use the resulting "hard" and "soft" data sets to consider two fusion cases. In the first case, the "hard" sensor is trained before performing fusion of "hard" and "soft" data. In the second case, the "hard" sensor is trained online by feeding back the global decision after performing "hard" and "soft" fusion. In both cases, we implement a set of fusion operators from Chapter 2 and compare their performances.

6.1 RDM Task Hard Sensor Construction

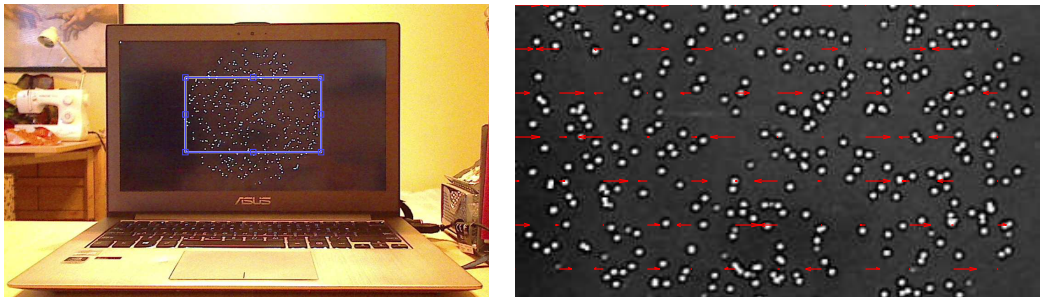
As discussed in Chapter 3 (Section 3.3.3), the random dot movement (RDM) decision task investigated by the authors of [45] consisted of human subjects positioned in front of a computer screen and presented with a circular field of moving dots. A subset of the dots were moving in a random direction, while the remaining dots were moving either to the right or to the left of the field. Subjects were instructed to choose which of the two directions they believed the dots were moving in and assess their decision confidence on a subjective probability scale. The chances of a dot moving in the correct direction, termed in [45] as *motion coherence*, quantifies the complexity of a given RDM trial. Higher values of motion coherence indicate that there are less dots moving randomly, making

it easier to determine the direction that the dot field is moving in. The authors of [45] chose motion coherences for each trial of the RDM experiment from from the values 0.02, 0.04, 0.08, 0.16, and 0.32. The field movement direction was also chosen randomly during each RDM trial.

Simulated RDM animations were coded by the authors of [45] using MATLAB and Psychtoolbox [140]. We used this same MATLAB code to produce RDM task stimuli, and to simulate the fusion of hard/soft data. RDM fields were displayed on the screen of an Asus UX32V laptop at a resolution of 1920x1080 pixels. The hard sensor was a Naive Bayes classifier [128] that used an estimate of the time/space horizontal movement of the RDM field to output a subjective probability assignment on which direction the RDM field was moving in. Then, the same RDM fields were used to simulate human decision and confidence assessments using the 2DSD models. These models were simulated using the model parameters estimated for each subject in [45] in order to produce decision and decision confidence assessments towards which direction the RDM field was moving in. In the following subsections, we detail the feature extraction and classifier algorithms used to construct the RDM hard sensor.

6.1.1 Feature Extraction

A Logitech Webcam Pro 9000 was positioned approximately 60 centimeters from the screen of the laptop displaying RDM fields. For a given RDM field, 5 video frames were acquired at a speed of 10 frames per second and a resolution of 1280x1024 pixels (I420). For the first video acquisition, a cropping window was manually specified to eliminate background interference (Figure 6.1a). Each frame was then cropped and converted to gray scale. A block matching function from the MATLAB computer vision toolbox [141] was used to determine the horizontal components of movement between adjacent video frames for 25x25 pixel subsections. The specific parameters used for the block matcher consisted of a maximum block displacement of 14 pixels (horizontal or vertical) and a three-step search method. An example of the estimated horizontal components of movement between two adjacent frames is shown in Figure 6.1b for each block of 25x25 pixels. The horizontal components of movement for each block were averaged for each of the adjacent frames in the acquired video, resulting in an average horizontal component of movement between frames. Finally, these inter-frame



(a) Sample RDM image with crop window **(b)** Results of block matching

Figure 6.1: Sample images acquired by the RDM decision task hard sensor.

horizontal components of movement were averaged, resulting in a time/space average of horizontal movement.

6.1.2 Classifier Construction

Let $x \in \mathbb{R}$ denote the time/space average of horizontal movement acquired for a given video, and let H_0 and H_1 represent the hypotheses that the RDM field is moving to the left or right respectively. For a single feature, the Naive Bayes pattern recognition algorithm [128] returns the a posteriori probabilities $P(H_i|x)$ for $i = 1, 2$ using Bayes rule. That is,

$$P(H_i|x) = \frac{P(x|H_i)P(H_i)}{\sum_{j=1}^{2} P(x|H_j)P(H_j)} \tag{6.1}$$

where the model parameters, $P(x|H_i)$ and $P(H_i)$, are denoted the likelihood and a priori class probabilities respectively. A decision can be generated by selecting the hypothesis which yields the maximum a posteriori probability (MAP). For the RDM task simulated here, the a priori class probabilities $P(H_i)$ are known by construction of the experiment. For simplicity, we assume that the likelihoods $P(x|H_i)$ are normally distributed with differing means and variances. That is, $P(x|H_0) \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and $P(x|H_1) \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Although a non-normal distribution may prove to be a better fit of the likelihoods, the implementation of the fusion operators will not change and hence their performance trends should be similar.

6.2 Fusion Study 1: Fusion with Trained Hard Sensors

6.2.1 Motivation

In chapters 4 and 5, we observed that the use of discounting techniques improved the performance of many fusion operators. In these cases, we also observed that Bayes' rule of probability combination [55] and Dempster's rule of combination (DRC) [43] yielded superior fusion performance, and further that the performance of these operators was monotonic with respect to the number of responses included in fusion. In this study, we consider a similar approach but while also fusing the a posteriori probabilities of the hard sensor described in Section 6.1. The hard sensor is trained offline using a known set of training examples. The results of this study should give insight as to how the quality of the final fused result increases when incorporating both hard and soft data over hard data alone.

6.2.2 Experimental Setup

A database of 15,000 RDM task stimuli was generated using the Psychtoolbox code of [45] by randomly choosing a motion coherence from the set 0.02, 0.04, 0.08, 0.16, and 0.32, and randomly choosing a direction of movement (*i.e.*, left or right). The likelihood functions of the RDM hard sensor, $P(x|H_0)$ and $P(x|H_1)$ were estimated using 5,000 of the 15,000 stimuli in the generated database. Then, the remaining 10,000 stimuli were used to simulate a posteriori probabilities from the hard sensor, and decisions and confidence assessments from 14 2DSD human decision making models. The 2DSD human decision making models were those estimated by the authors of [45]. We simulated 2DSD human decisions having a decision response time of 800 milliseconds and an interjudgment time of 500 milliseconds.

The hard sensor a posteriori probabilities were combined with the decisions and confidence assessments of the human decision makers using Bayes' rule of probability combination [55] (Equation 2.13), Dempster's rule of combination (DRC) [43] (Equation 2.25), Yager's rule [94] (Equation 2.27), the Proportional Conflict Redistribution Rule #5 (PCR5) [97] (Equation 2.30), and Murphy's combination rule [99, 100] (Equation 2.33). When needed, hard sensor belief mass assignments (BMAs) were constructed as simple support functions (Equation 2.20) focused on the hypothesis with the higher a posteriori probability. BMAs and subjective probabilities for the soft

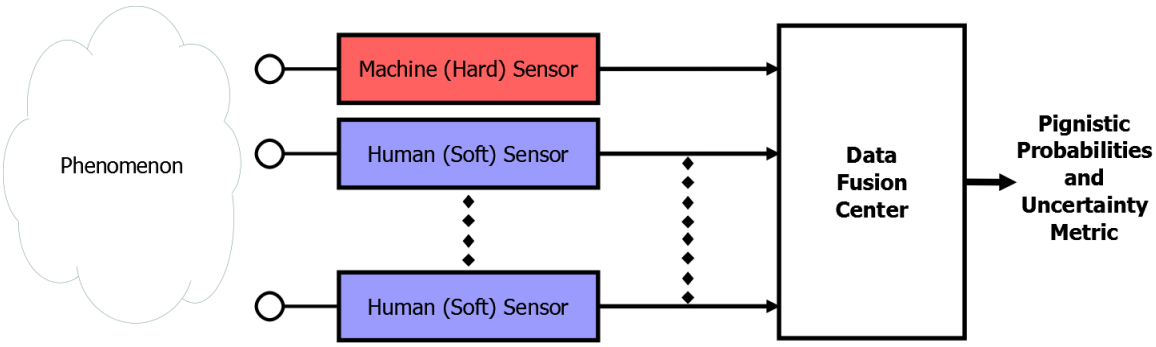


Figure 6.2: Diagram of a hard/soft fusion system where the hard sensor is trained offline.

sensors were constructed in four different manners. The first two cases were the the "Confidences Only" and "Evidence Strength Discounting" construction cases described in Section 4.3.2. The third case was the "Intersubject Conflict discounting" construction case described in Section 5.4.2. The fourth case investigated by this study was known as the "Hard Sensor Discounting" construction case, which involved choosing a discount rate α_i' for subjects $i = 1, 2, \ldots, 14$ such that

$$\alpha_i = 1 - \sqrt{\frac{1}{2}(\mathbf{m}_i - \mathbf{m}_H)^T \mathbb{D}(\mathbf{m}_i - \mathbf{m}_H)}, \tag{6.2}$$

where $\mathbf{m}_i$ and $\mathbf{m}_H$ are BMAs written in vector form with $\mathbf{m}_H$ being the BMA or subjective probability of the hard sensor. The matrix $\mathbb{D}$ is of size $2^M \times 2^M$ whose elements are given in Equation 5.11.

All BMAs and subjective probabilities were combined in pairs of two, with the hard sensor BMA or subjective probability being first in the combination (Figure 6.2). The soft sources were then included in the combination in a random order. For the Dempster-Shafer theory operators, combined BMAs were used to generate subjective probabilities using the pignistic probability transform defined in Equation (2.34). The pignistic probabilities associated with the correct alternatives generated from each fusion operator were averaged over all of the simulation trials. The uncertainty metric associated with the correct alternatives of Equation (2.16) was also calculated from each fusion operator and averaged over all of the simulation trials. Finally, the correct classification rate for each fusion operator was determined by applying a MAP decision rule to the pignistic probabilities produced by each fusion operator.

6.2.3 Results

Figures 6.3 through 6.5 show the performance of the five fusion operators investigated in this study. Figure 6.3 shows the average pignistic probability associated with to the correct alternative versus the number of sources included in the combination. Figure 6.4 shows the average uncertainty measure associated with the correct alternative versus the number of sources included in the combination. Figure 6.5 shows the correct classification rate when applying a MAP decision rule to the pignistic probabilities produced after fusion versus the number of sources included in the combination. Also shown in Figure 6.5 is the MAP correct classification rate associated with the best and worst soft sources, and the trained hard source. In each of these three figures, the four subplots correspond to the four BMA and subjective probability cases investigated, namely "Confidences Only" (Section 4.3.2), "Evidence Strength Discounting" (Section 4.3.2), "Intersubject Conflict Discounting" (Section 5.4.2), and "Hard sensor discounting" (Section 6.2.2). Better performing fusion operators will have higher pignistic probabilities associated towards the correct alternative, lower uncertainty, and higher correct classification rates.

Similar to the results of Section 5.3.3, Bayes' rule of probability combination and DRC could not be used when source discounting was not performed (*i.e.*, all of the "Confidences Only" fusion operator input construction cases). A summary of the fusion results are shown in Table 6.1 for the average post-fusion pignistic probability toward the correct alternative and Table 6.2 for the average post-fusion uncertainty metric toward the correct alternative. In terms of MAP correct classification rates, the performance of Bayes' rule, DRC, Yager's rule, and PCR5 were the same in three out of the four BMA construction cases. In the hard sensor discounting case, the performance order for MAP correct classification rates was Bayes' rule, the mixing rule, Yager's rule, and DRC/PCR5. We make note of the following additional observations.

- For the "Evidence Strength Discounting" and "Intersubject Conflict Discounting" cases, Bayes' rule of probability combination and DRC exhibit the highest post-fusion average subjective probabilities toward the correct alternative. The subjective probabilities associated with Bayes' rule of probability combination were observed to be be monotonically increasing with respect

Table 6.1: Summary of hard/soft fusion performance results for the experiment setup defined in Section 6.2.2. Results shown in terms of the average post-fusion pignistic probability toward the correct alternative and for each BMA/subjective probability construction case.

Performance Order	Confidences Only	Evidence Strength Discounting	Intersubject Conflict Discounting	Hard Sensor Discounting
1 (Best)	Yager/PCR5	Bayes/DRC	Bayes/DRC	Bayes
2	Mixing	Yager/PCR5	Yager/PCR5	DRC/Yager/PCR5
3	—	Mixing	Mixing	Mixing
4	—	—	—	—
5 (Worst)	—	—	—-	—

Table 6.2: Summary of hard/soft fusion performance results for the experiment setup defined in Section 6.2.2. Results shown in terms of the average post-fusion uncertainty metric towards the correct alternative and for each BMA/subjective probability construction case.

Performance Order	Confidences Only	Evidence Strength Discounting	Intersubject Conflict Discounting	Hard Sensor Discounting
1 (Best)	PCR5	Bayes	Bayes	Bayes
2	Yager	DRC/PCR5	DRC/PCR5	DRC/PCR5
3	Mixing	Yager	Yager	Yager
4	—	Mixing	Mixing	Mixing
5 (Worst)	—	—	—-	—

to the number of subjects included in the combination. The subjective probabilities associated with DRC were observed to be monotonically increasing in the "Evidence Strength Discounting" and "Intersubject Conflict Discounting" cases.

- Bayes' rule of probability combination exhbited the highest post fusion pignistic probabilities towards the correct alternative in the"Hard Sensor Discounting" case. The performance of Bayes' rule of probability combination was very similar to its performance in the "Intersubject Conflict Discounting" case.

- Across all fusion operator input construction cases, the average post-fusion uncertainty measure of the correct alternative associated with DRC and and PCR5 converges to zero, indicating that their post fusion BMAs eventually converge to subjective probability assignments.

- The Mixing combination rule was observed to produce the lowest post-fusion average subjective probabilities across all BMA construction cases.

- The average post-fusion uncertainty measures associated with Yager's rule and the Mixing

combination rule were observed to approach steady-state values. The steady-state values associated with Yager's rule were observed to be the highest in the "Confidences Only" and "Evidence Strength Discounting" cases. The steady-state values associated with the Mixing combination rule were observed to be the highest in the "Intersubject Conflict Discounting" and "Hard sensor discounting" cases. These results are logical as Yager's rule will tend to increase the imprecision present in the post-fusion result according to the conflict amongst the sources. The mixing combination rule does not incorporate a conjunction operation, and therefore it has no method of reducing the imprecision in its post-fusion BMAs.

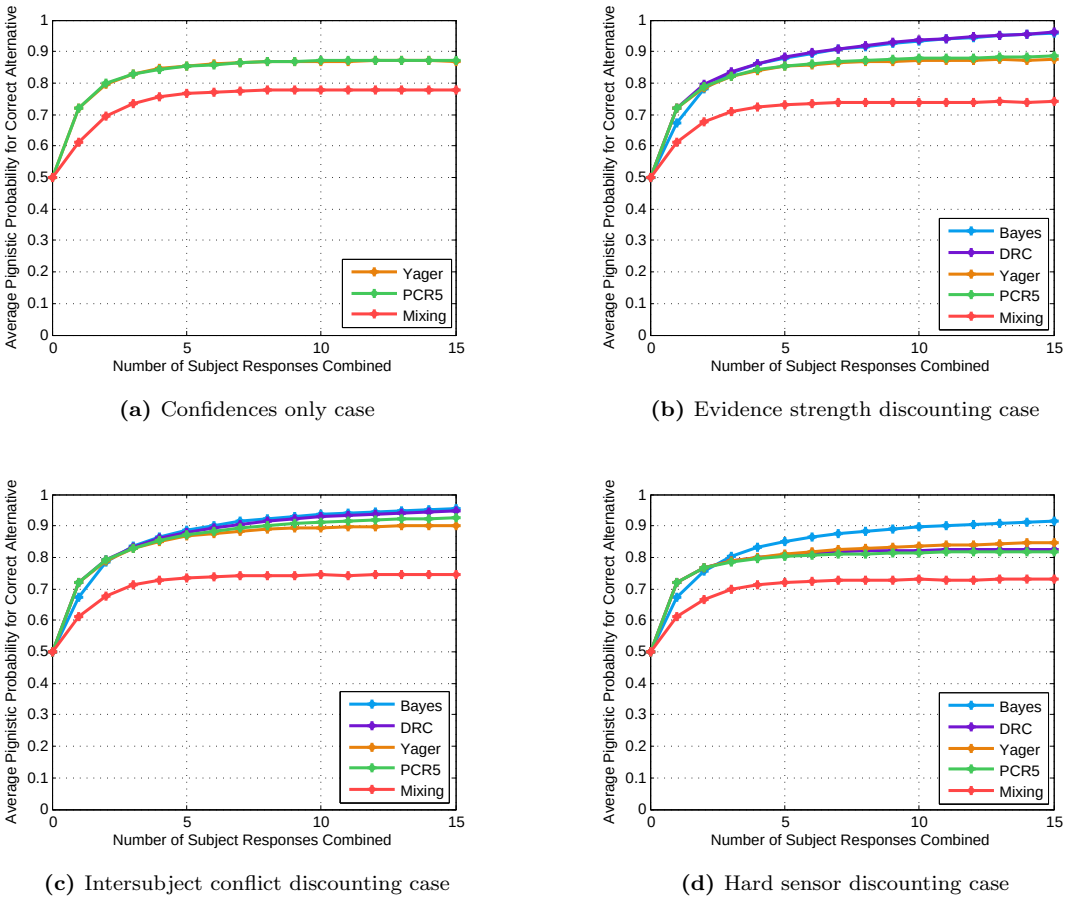
- In general, all five of the fusion operators investigated produced MAP correct classification rates between 85 and 95%. The highest MAP correct classification rates are equally achieved by Bayes rule of probability combination, DRC, PCR5, and Yager's rule in the "Evidence Strength Discounting" and "Intersubject Conflict Discounting" cases.

- The Mixing combination rule exhibited the lowest MAP correct classification rates of the five fusion operators in the "Confidences Only," "Evidence Strength Discounting," and "Intersubject Conflict Discounting" cases. However, the Mixing combination rule exhibited the second highest MAP correct decision rates in the "Hard Sensor Discounting" case.

- The best performing fusion operators exhibited 3-4% higher MAP correct classification rates than the best performing source for the "Evidence Strength Discounting" and "Intersubject Conflict Discounting" cases. In the "Confidences Only" and "Hard Sensor Discounting" cases, the best performing fusion operator exhibited MAP correct classification rates approximately equal to the best performing source in the combination.

- All fusion operators in the "Confidences Only" and "Hard Sensor Discounting" cases were not observed to exhibit MAP correct classification rates higher than the best performing source.

Based on the above observations, Bayes' rule of probability combination would be the best overall fusion operator for this specific RDM decision task when applying evidence strength discounting or intersubject conflict discounting. Bayes' rule of probability combination was among the fusion



(a) Confidences only case

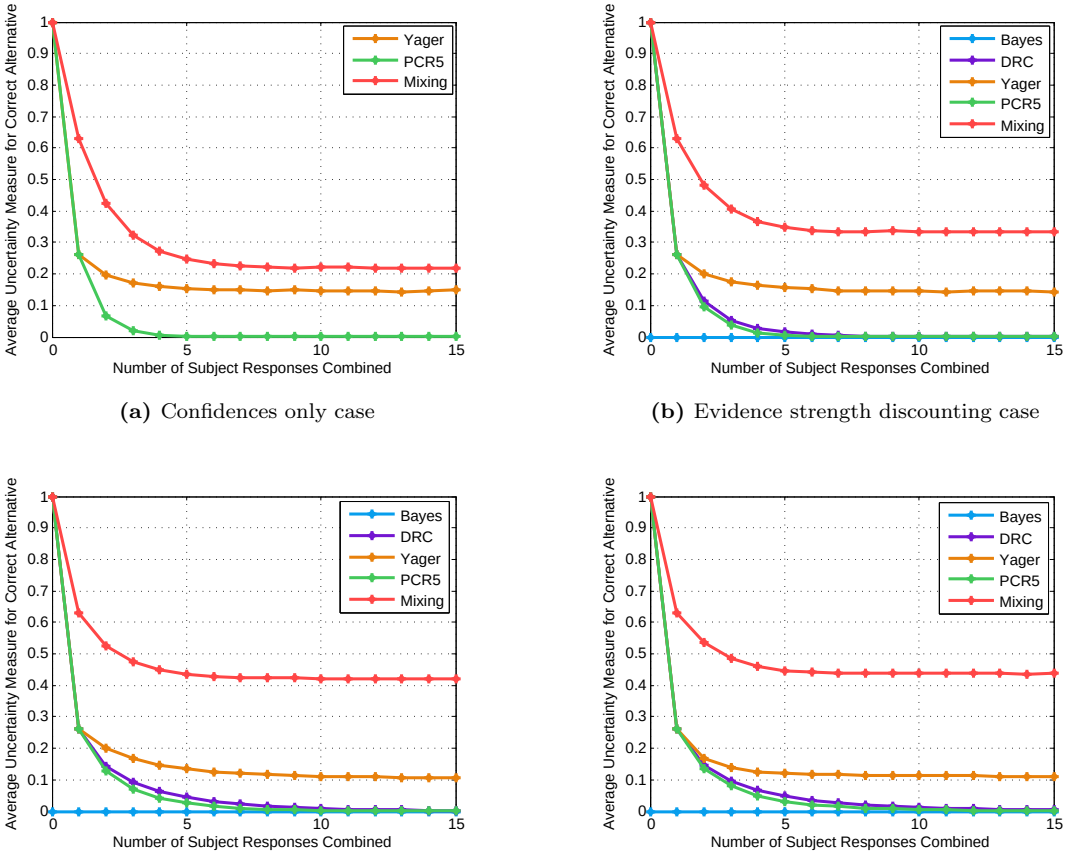
(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.3: Average post combination pignistic probability towards correct alternative versus the number of responses included in the fusion operator. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of the soft sources. Results averaged over 10,000 simulation trials.

operators that were observed to produce the highest post-fusion average subjective probabilities towards the correct alternatives and the highest MAP correct classification rates. Although similar performance was acheived by DRC and PCR5 in some cases, they exhibit higher computational complexity since the number of required operations to use them in fusion are on the scale of $2^{|\Omega|}$, whereas the computations required for Bayes' rule are on the order of $|\Omega|$.



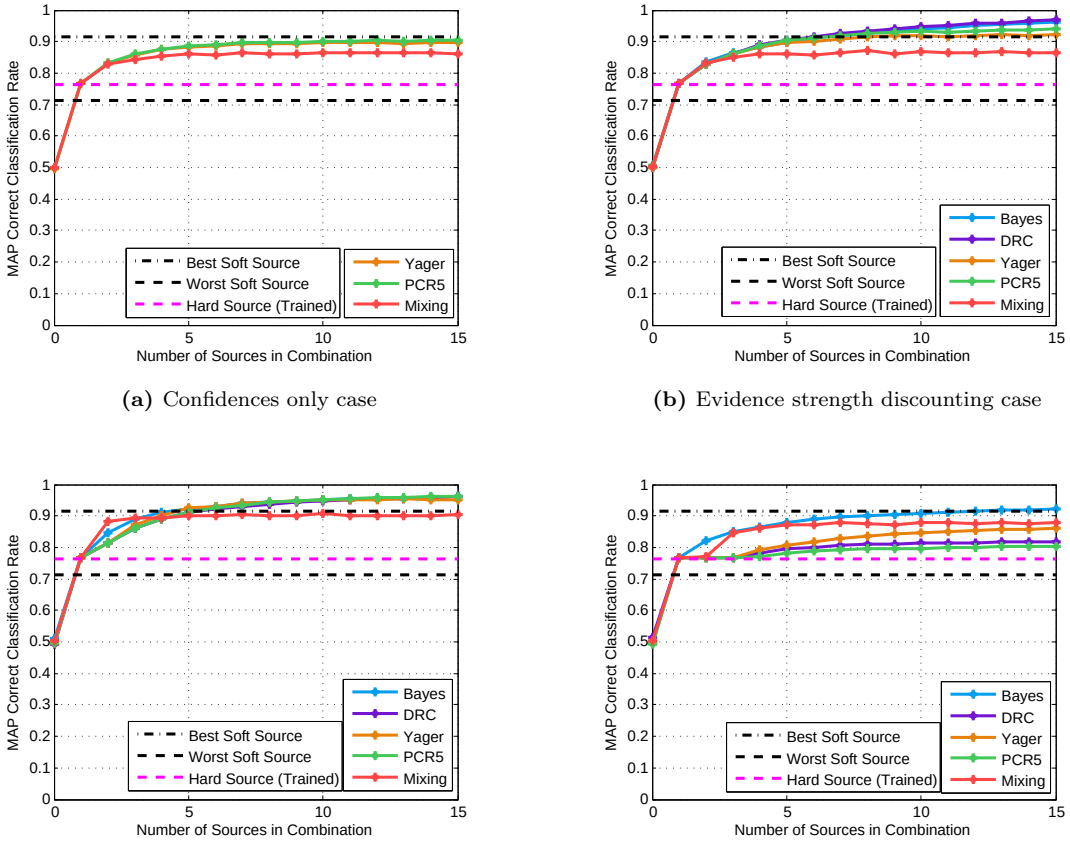
(a) Confidences only case

(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.4: Average post combination uncertainty towards correct alternative versus the number of responses included in the fusion operator. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of the soft sources. Results averaged over 10,000 simulation trials.



(a) Confidences only case

(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.5: Estimated correct classification rate using a maximum a posteriori (MAP) decision rule versus the number of responses included in the fusion operator. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of the soft sources. Results averaged over 10,000 simulation trials. MAP correct classification rates of the best and worst soft sources and the trained hard source are shown for comparison.

6.3 Fusion Study 2: Online Training of Hard Sensors

6.3.1 Motivation

In this study, we investigate the performance of a hard and soft fusion operator when its output is used to perform online training of the hard sensors included the fusion. This is an important case to evaluate, as a training period for the hard sensors may not always be feasible. A diagram of the fusion-enabled online training system investigated in this study is shown in Figure 6.6. Similar to Section 6.2, a hard sensor is designed using a Naive Bayes pattern recognition system where the likelihoods are estimated as normal random variables. For a given observation, the a posteriori probabilities produced by the hard sensor are combined at a data fusion center with the BMAs or subjective probabilities of the simulated human decision-makers (depending on the fusion operator employed at the fusion center). For the Dempster-Shafer theoretic fusion operators, a probability transform is used on the final fused result to produce subjective probabilities towards each of the possible alternatives. Then, a MAP decision rule is applied to generate a global decision. The global decision produced by the data fusion center is fedback to the hard sensor. The observed feature set of the hard sensor and the fed back decision constitute a single training example that is used to update the likelihood function parameter estimates (*i.e.*, mean and variance) of the hard sensor online using Welford's algorithm [142, 143]. To apply Welford's algorithm, at least 3 training examples are needed for each likelihood function, $P(x|H_i)$. If there were ever less than 3 available training examples associated with either alternative, H_0 or H_1, the hard sensor was constructed to output equiprobable a posteriori probabilities.

6.3.2 Experimental Setup

A total of 20,000 RDM task stimuli were generated in the same manner as Section 6.2.2. The 20,000 stimuli were separated into 1,000 Monte Carlo simulation trials in which 20 consecutive RDM stimuli were used to train the hard sensor and perform fusion. In order to implement the online training architecture of Figure 6.6, each RDM stimulus was used to generate decisions and confidence assessments from the first 5 of the 14 2DSD human decision making models estimated in [45]. We again simulated 2DSD human decisions having a decision response time of 800 milliseconds and

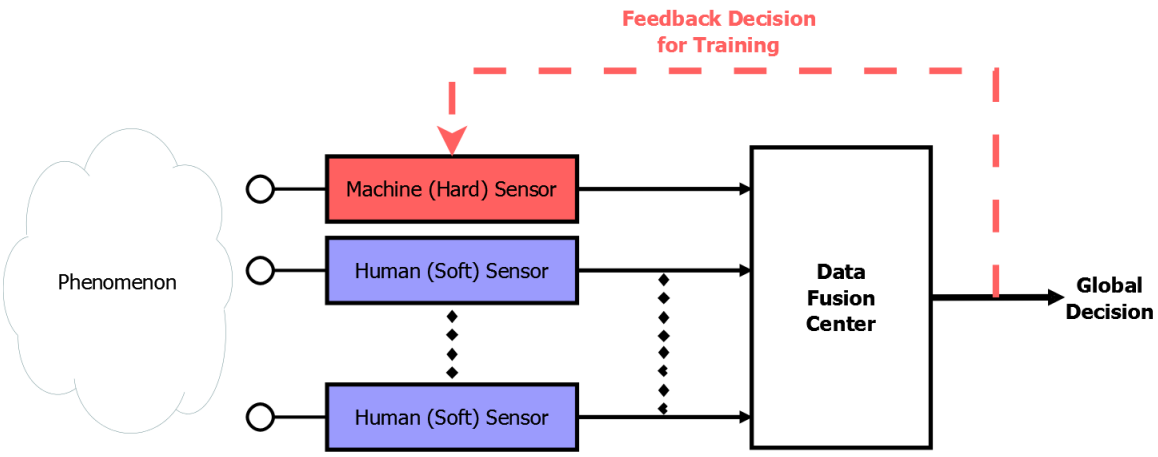


Figure 6.6: Diagram of a fusion system that applies online training of a single hard sensor using feedback of the global decision.

an interjudgment time of 500 milliseconds. The same five fusion operators from the first hard and soft fusion study (Section 6.2.2) were used here, namely Bayes' rule of probability combination [55], Dempster's rule of combination (DRC) [43], Yager's rule [94], the Proportional Conflict Redistribution Rule #5 (PCR5) [97], and Murphy's combination rule [99, 100]. This study evaluated the "Confidences Only" (Section 4.3.2), "Evidence Strength Discounting" (Section 4.3.2), "Intersubject Conflict Discounting" (Section 5.4.2), and "Hard sensor discounting" (Section 6.2.2) cases.

All BMAs and subjective probabilities were combined in pairs of two, with the hard sensor BMA or subjective probability being first in the combination. The soft sources were then included in the combination in a random order. For the Dempster-Shafer theory operators, combined BMAs were used to generate subjective probabilities using the pignistic probability transform defined in Equation (2.34). These post fusion subjective probabilities were used to generate the same three performance metrics discussed alongside the first hard and soft fusion study (Section 6.2.2), namely the average pignistic probability towards the correct outcome, the average uncertainty metric towards the correct outcome, and the estimated MAP decision rule correct classification rate.

6.3.3 Results

Figures 6.7 through 6.10 show the performance of the five fusion operators and four fusion operator input construction cases investigated in this study. Figure 6.7 shows the estimated MAP correct

classification rate of the online-trained hard sensor versus the sequential number of stimuli presented to the hard and soft sources. For comparison, the MAP correct classification rate for a hard sensor trained offline is also shown in Figure 6.7. Similar to Section 6.2.3, the average pignistic probability associated with correct alternative, the average uncertainty metric associated with the correct alternative, and the estimated post-fusion MAP correct classification rate are shown in Figures 6.8 through 6.5 versus the sequential number of stimuli presented to the hard and soft sources. Also shown in Figure 6.10 is the MAP correct classification rate associated with the best and worst soft sources, and the offline trained hard source. In each of these four figures, the four subplots correspond to the four BMA and subjective probability cases investigated, namely "Confidences Only" (Section 4.3.2), "Evidence Strength Discounting" (Section 4.3.2), "Intersubject Conflict Discounting" (Section 5.4.2), and "Hard sensor discounting" (Section 6.2.2).

In summary, Bayes' rule of probability combination, DRC, Yager's rule, and PCR5 performed equally as well in terms of post-fusion average pignistic probability towards the correct alternative and MAP correct classification rate. In terms of the average post-fusion uncertainty metric towards the correct alternative, Bayes' rule of probability combination, DRC, and PCR5 exhibited similar performance, followed by Yager's rule and then the mixing combination rule. The following paragraphs discuss these performance trends in more detail.

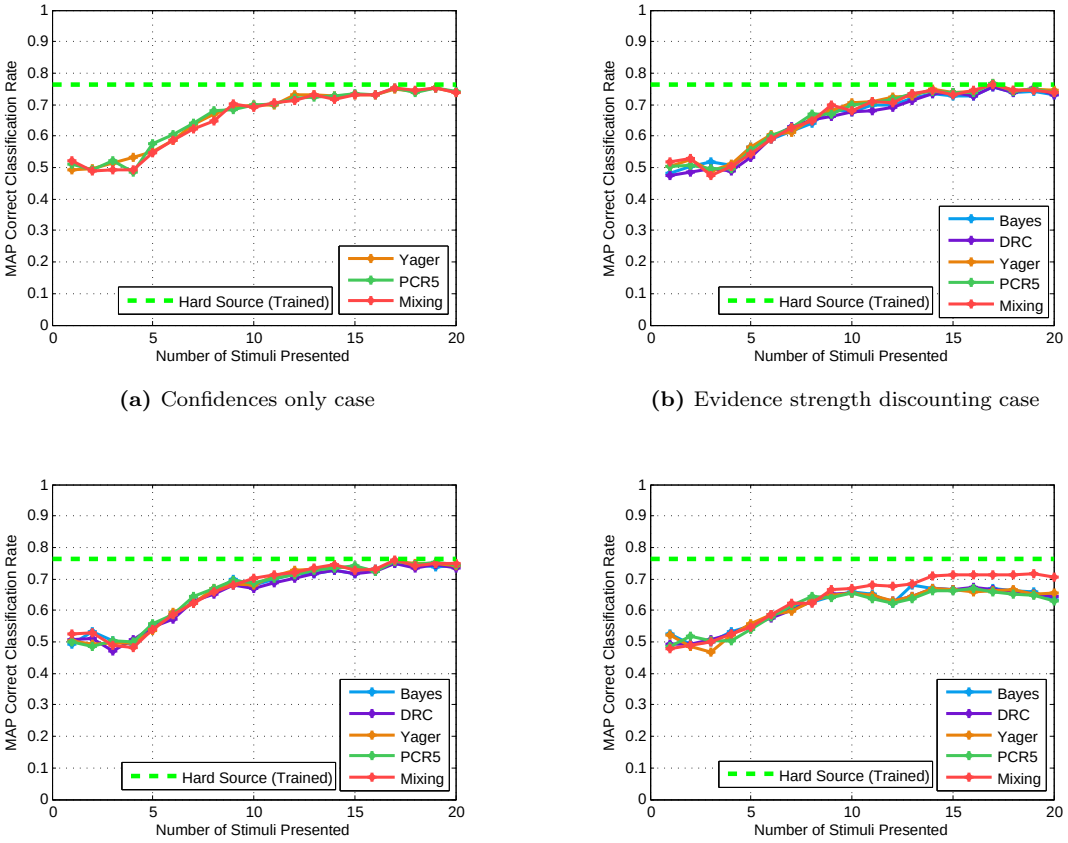
Excluding the "Hard Sensor Discounting" case, the five fusion operators and remaining three fusion operator input construction cases were observed to exhibit similar trends in performance. In these cases, after approximately 15 iterations the MAP correct classification rate of the online-trained hard sensor approached the MAP correct classification rate obtained by the offline trained hard sensor (Figure 6.7). The subjective probabilities towards the correct alternative, uncertainty metric, and MAP correct classification rate associated with the output of the data fusion operators also exhibited fairly consistent performance, even at the earlier stages of training (*e.g.*, when less than five stimuli presented to the hard sensor).

In the "Hard Sensor Discounting" case, the five fusion operators and the online-trained hard sensor produced markedly lower performance. After presenting 20 stimuli to the hard and soft

sensors, the MAP correct classification rate of the online-trained hard sensor was found to be lower than that of the offline trained hard sensor (Figure 6.7d). Furthermore, the subjective probabilities towards the correct alternative, uncertainty metric, and MAP correct classification rate associated with the output of the data fusion operators were observed to decrease as the number of stimuli presented to the hard and soft sensors increased. The mixing combination rule was observed to produce MAP correct classification rates close to the best soft source present in the combination (Figure 6.10d). The remaining four fusion operators were observed to eventually converge to a MAP correct classification rate similar to the worst soft source present in the combination.

6.4 Chapter Summary

In this chapter, we have used the 2DSD model parameters from [45] on the RDM decision task to simulate the performance of a variety of hard and soft fusion schemes. After describing a feature extraction and pattern recognition algorithm for a RDM hard sensor, we investigated two hard and soft fusion studies. In the study, we considered a hard sensor trained offline and observed how the post combination average subjective probability towards the correct alternative, average uncertainty towards the correct alternative, and the MAP correct classification rate changed with respect to the number of soft sources included in the combination. It was determined that Bayes' rule of probability combination was the best overall fusion operator for this specific task when applying evidence strength discounting or intersubject conflict discounting. In the second fusion study, we considered a hard sensor trained online using global decisions produced by a variety of fusion operators and fusion operator input considerations. The "Hard Sensor Discounting"' was observed as the worst performing fusion operator input construction case, as it lead to MAP correct detection rates significantly lower than the best source in the combination. This is not surprising, as this input construction case will falsely discount the information from the soft sources based on incomplete training information at the hard source. The remaining three fusion operator input construction cases exhibited similar trends in performance for all five the fusion operators evaluated. Because of the similar levels of performance, we consider the best performing fusion rule in these cases as Bayes' rule of probability combination when applying either "Evidence Strength Discounting" or



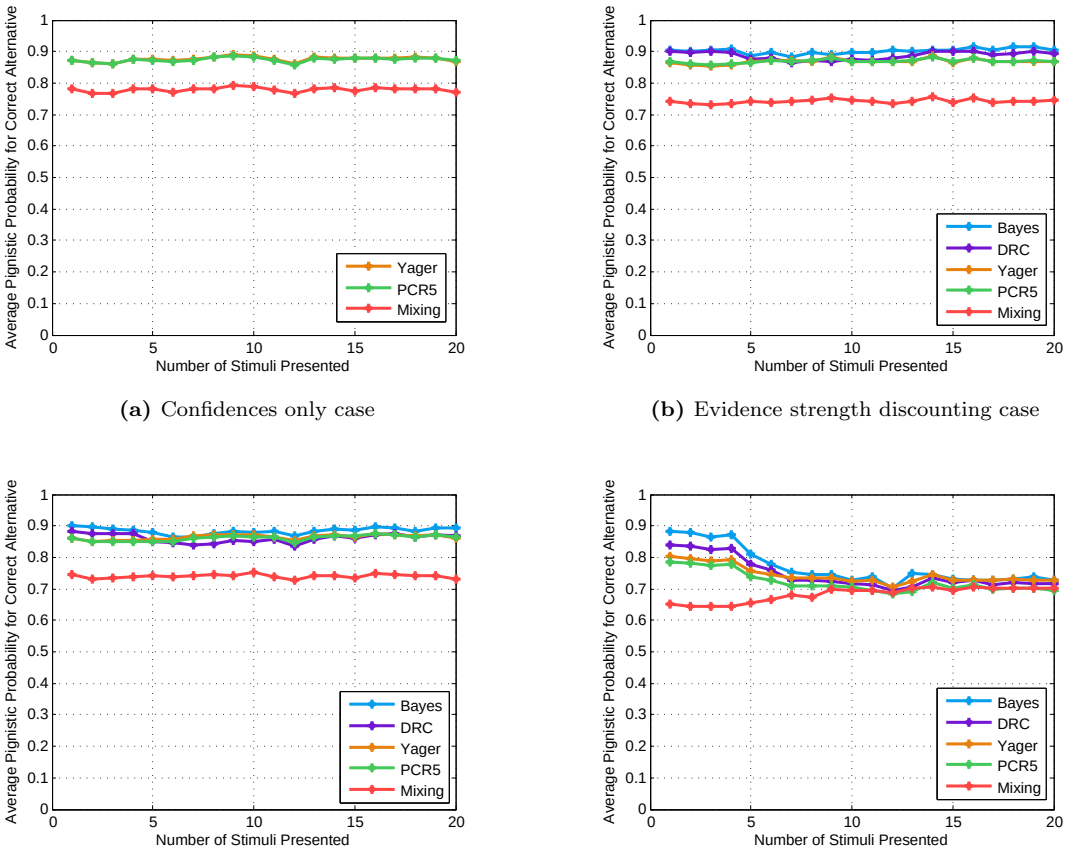
(a) Confidences only case

(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.7: Estimated MAP correct classification rate of the RDM task hard sensor trained online (Figure 6.6) versus the number of stimuli presented. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of five soft sources. Results averaged over 10,000 simulation trials. MAP correct classification rates of the offline-trained hard source is shown for comparison.



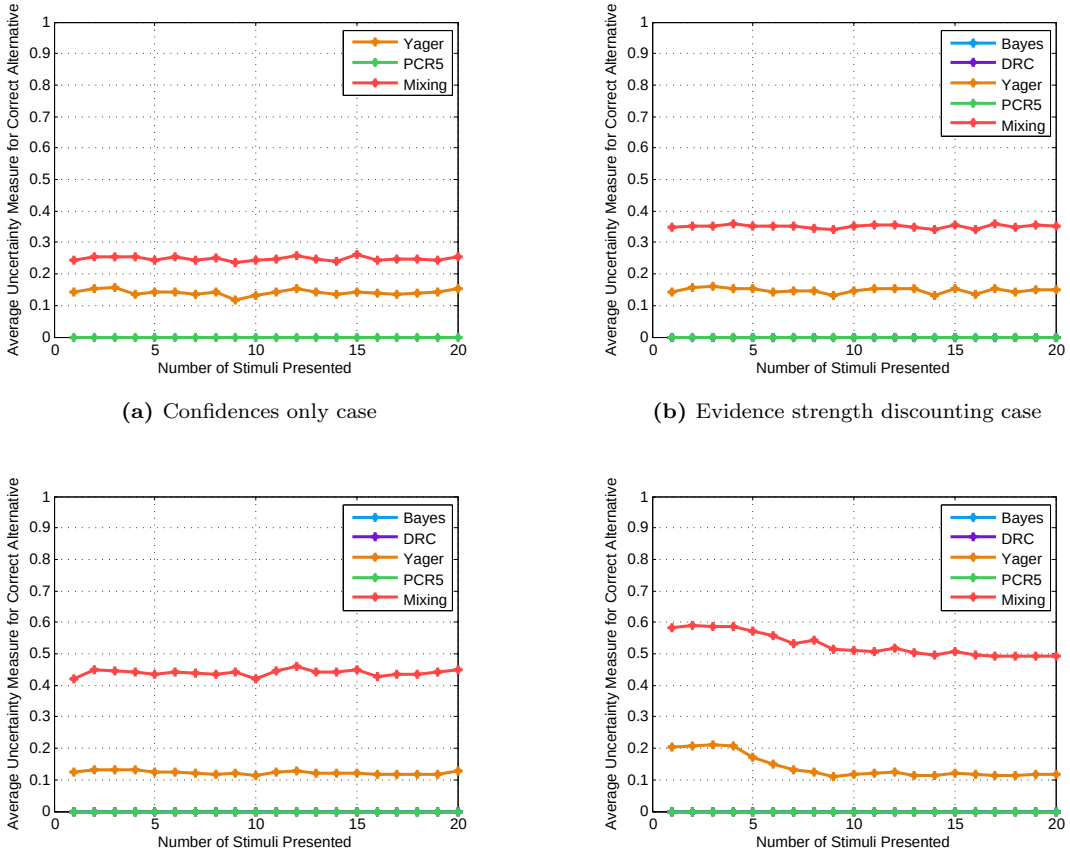
(a) Confidences only case

(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.8: Average post combination pignistic probability towards correct alternative versus the number of stimuli presented to the hard and soft sources when training the RDM hard sensor online. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of five soft sources. Results averaged over 10,000 simulation trials.



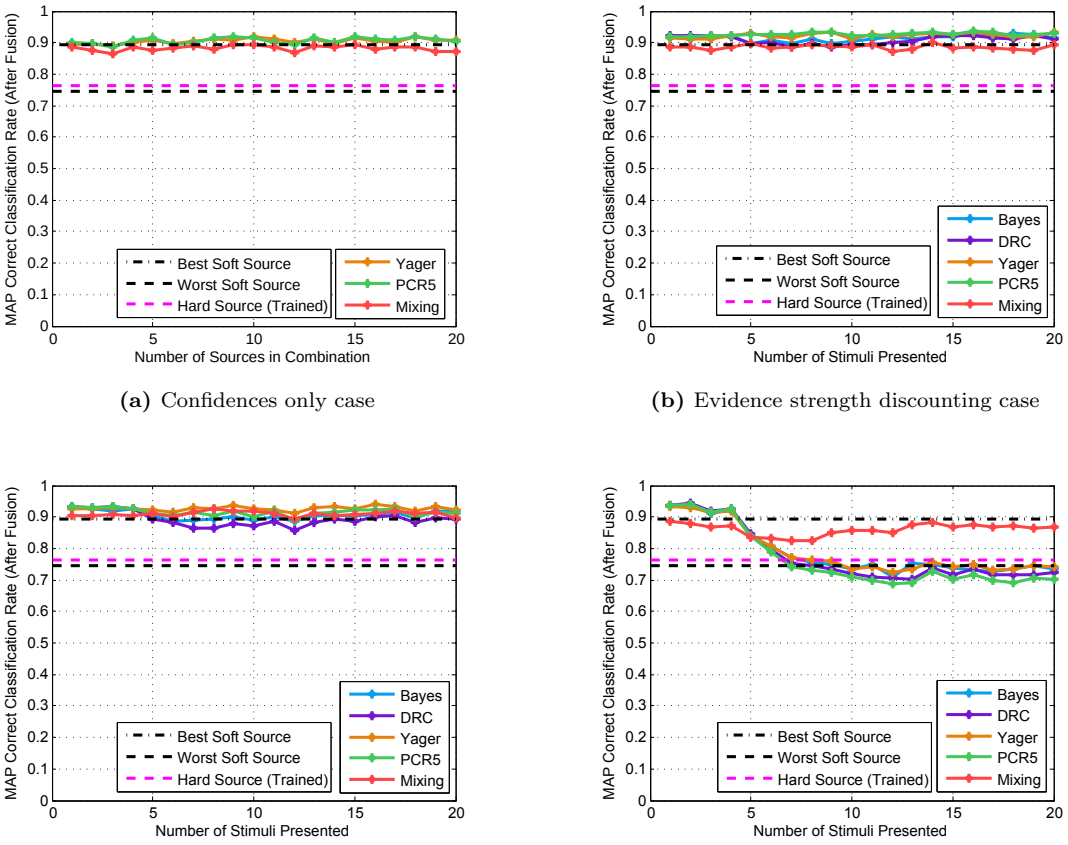
(a) Confidences only case

(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.9: Average post combination uncertainty towards correct alternative versus the number of stimuli presented to the hard and soft sources when training the RDM hard sensor online. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of of five soft sources. Results averaged over 10,000 simulation trials.



(a) Confidences only case

(b) Evidence strength discounting case

(c) Intersubject conflict discounting case

(d) Hard sensor discounting case

Figure 6.10: Estimated post combination correct classification rate using a maximum a posteriori (MAP) decision rule versus the number of stimuli presented to the hard and soft sources when training the RDM hard sensor online. The different lines represent the fusion operators being investigated. In all simulation cases, the hard sensor was the first source in the combination, followed by a random permutation of five soft sources. Results averaged over 10,000 simulation trials. MAP correct classification rates of the best and worst soft sources and the trained hard source are shown for comparison.

"Intersubject Conflict Discounting" (depending on the availability of a priori information). Even though a selection of the Dempster-Shafer theory fusion operators exhibit similar performance in some cases, Bayes' rule of probability combination exhbits lower implementation complexity as it is calculated over the set of alternatives, as opposed to the powerset of alternatives.

Part III

Epilogue

Chapter 7: Areas of Future Work

The work presented in this thesis suggests future studies for evaluating the statistical the performance of soft and hard/soft data fusion operators. Specifically, we have shown through a number of examples that human decision-making models from cognitive psychology can be used to estimate the statistical performance of a soft or hard/soft fusion operator. In this chapter, we discuss the natural progression of this work towards new research opportunities for cognitive psychology and data fusion researchers. The ares of future work are separated into two categories: (1) research opportunities in cognitive psychology, motivated by the needs of soft and hard/soft fusion researchers and (2) research opportunities in soft and hard/soft fusion motivated by the human response probability models developed by cognitive psychologists.

7.1 Cognitive Psychology Research Areas

7.1.1 Cognitive Models of Multihypothesis and Vague Decision-Making

As discussed in Chapter 3, the majority of research performed by cognitive psychologists in modeling human decision-making dynamics has been on two-alternative, forced choice (TAFC) tasks [44]. Research on modeling multihypothesis human decision-making is still relatively new, and studies combining experimental results and theoretical analyses have only begun surfacing over the past few years [137]. There has been little effort towards the development of multihypothesis human response models that account for the generation of decision confidence assessments and also the generation of vague decisions. As described in Chapters 1 and 2, these types of decision tasks are of importance to researchers in soft and hard/soft data fusion. In this thesis (Chapter 5), we have proposed methods for simulating vague human decisions using available human decision-making models on binary decision tasks. There is however a need for developing a stochastic model of vague human decision-making and confidence assessment on multihypothesis decision tasks. Such a model should be able to capture the statistical connections between time pressure, the number of decision

alternatives, and task difficulty to the observed decision, confidence assessment, and response time statistics associated with a human decision-maker.

We briefly outline how a formal cognitive model of precise/imprecise human decision-making and confidence assessment on M-ary decision tasks could be constructed using leaky competing accumulators. The *leaky competing accumulator* (LCA) model of human decision-making was initially proposed by Usher and McClelland in [144] and has been discussed as a natural candidate for extending the drift diffusion model of human decision-making to M-ary decision tasks [137]. Consider a multihypothesis decision task defined by the alternatives $\omega \in \Omega$, where $|\Omega| = M$. A LCA model consists of M time-dependent accumulators, denoted $x_\omega(t)$, that update in a manner similar to a drift diffusion process (Figure 7.1). Specifically, each accumulator increments in finite time steps of Δt such that $x_\omega(t + \Delta t) = \max\left(x_\omega(t) + \Delta x_\omega(t + \Delta t), 0\right)$. The incremental evidence $\Delta x_\omega(t + \Delta t)$ for each alternative $\omega \in \Omega$ is defined as

$$\Delta x_\omega(t + \Delta t) = \left(\delta_\omega - \kappa_\omega x_\omega(t) - \beta_\omega \sum_{\substack{\hat{\omega} \in \Omega \\ \hat{\omega} \neq \omega}} x_{\hat{\omega}}(t) \right) \Delta t + \sqrt{\Delta t} \epsilon_\omega(t + \Delta t) \tag{7.1}$$

where δ_ω is the evidence accumulation rate, κ_ω is the evidence decay rate, β_ω is the lateral inhibition rate due to evidence accumulation for the other alternatives, and $\epsilon_\omega(\cdot)$ is a white noise process with fixed variance. Furthermore, the values of δ_ω are usually constrained such that $\sum_{\omega \in \Omega} \delta_\omega = 1$. The decision rule for the LCA model is to declare ω when $x_\omega(t) > \theta_\omega$, where θ_ω is a response threshold towards declaring ω. There is no known closed form solution (*i.e.*, distribution) that summarizes the LCA model [137], and hence parameter estimation must be performed using Monte Carlo methods [137].

Using the LCA model as a starting point, there are a few ways to take into account the generation of subjective confidences on imprecise decisions. One method would be to apply the same logic of two-stage dynamic signal detection (2DSD) to the LCA model. Specifically, a set of M accumulators would be simulated as described above until at least one of them crosses the response criterion threshold θ_ω. Then, each accumulator would continue for a specified interjudgment period, after

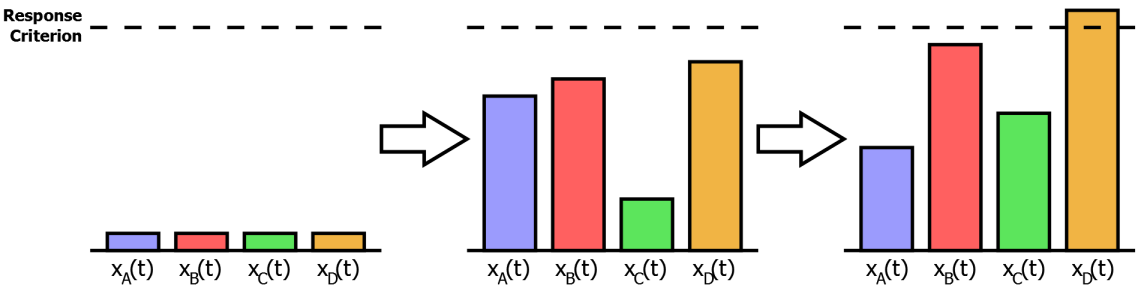
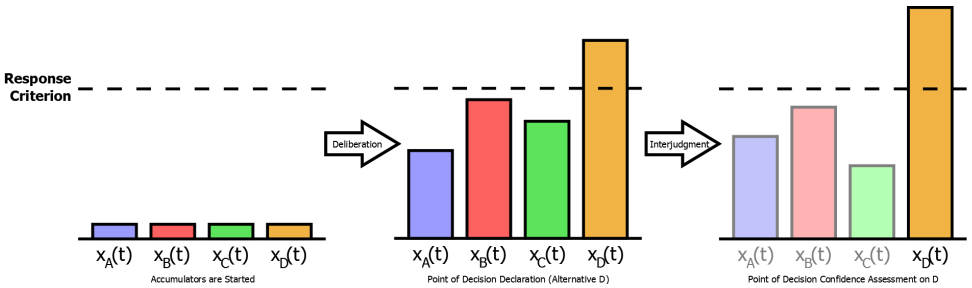


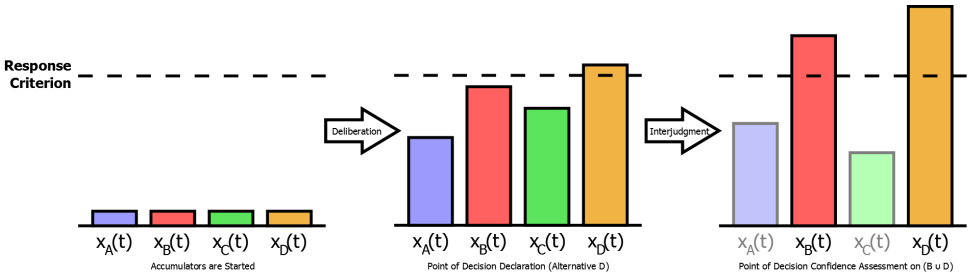
Figure 7.1: Conceptual diagram of the LCA model proposed in [144] for modeling human decision-making dynamics on multihypothesis tasks. Evidence is accumulated over time towards a set of alternatives, denoted here as $\Omega = \{A, B, C, D\}$. Evidence towards each alternative can also decrease by a self-decaying process (*i.e.*, leakage), or by evidence accumulated towards the other alternatives (*i.e.*, lateral inhibition). A decision is made when enough evidence has been accumulated towards one of the alternatives (*i.e.*, it passes a threshold depicted here as the response criterion).

which the final values of all accumulators above the response criterion threshold would be used in a binning operation to produce a decision confidence assignment. If more than one accumulator is above the response threshold at the end of the interjudgment period, then the final decision of the model would be the disjunction of the corresponding alternatives. A graphical example of such a 2DSD/LCA hybrid model is shown in Figure 7.2 for a decision task with four alternatives, hypothetically labeled as A, B, C, and D. The accumulators are simulated for a deliberation period until at least one crosses the response criterion threshold (in this case, alternative "D"). The accumulators are continued for a fixed interjudgment period, after which their final values are used to assess a decision and decision confidence assessment. In Figure 7.2a, only one accumulator is above the response criterion at the end of the interjudgment period, and hence the 2DSD/LCA hybrid model would produce a single decision (namely, alternative "D") with some confidence rating. In Figure 7.2b, two accumulators are above the response criteria at the end of the interjudgment period, and hence the model would produce a vague decision (namely, the disjunction $B \cup D$) with some confidence rating.

Although this extension of the LCA model may seem straightforward, the generalization of the 2DSD model of human decision making to multihypothesis tasks is non-trivial. In the most general case, the number of parameters required in a 2DSD/LCA hybrid would be an increasing



(a) 2DSD/LCA hybrid example showing a decision of alternative "D" declared.



(b) 2DSD/LCA hybrid example showing a vague decision of the disjunction "$B \cup D$" declared.

Figure 7.2: Conceptual diagram of a possible 2DSD/LCA hybrid model of vague human responses on a M-ary decision task. Evidence is accumulated over time towards a set of alternatives, denoted here as $\Omega = \{A, B, C, D\}$, until at least one crosses some threshold (*i.e.*, the response criterion). Then, the accumulators are run for a specified interjudgment period. Any accumulators which are above the threshold are included as the simulated vague decision, and the final values are used in a binning operation to produce a decision confidence assessment.

function of not only the number of possible confidence values per alternative, but also for the number of alternatives themselves. Furthermore, the absence of closed form expressions for the LCA model makes parameter estimation per subject computationally complex as all of the probability distributions must be estimated via Monte Carlo methods. Much research (*i.e.*, theoretical and experimental) needs to be conducted in order to understand what parameter estimation constraints for such a 2DSD/LCA hybrid model are required to produce the best fitting model while at the same time minimizing dimensionality and computational complexity. With this in mind, it would also be interesting to analyze experimentally the model accuracy of the vague decision-making and confidence assessment extension methods presented in Chapter 5 with respect to a 2DSD/LCA hybrid model. Although simulating multihypothesis responses using pairwise successive comparisons of TAFC subtasks may not prove as accurate as a direct model of multihypothesis decision-making, it may prove to be a preferred method of response simulation because of the reduced complexity in

parameter estimation.

7.1.2 Fatigue, Stress, and Anxiety Models

As discussed in Chapter 3, The 2DSD human response models used in this thesis were constructed over highly controlled TAFC tasks (*i.e.*, the line length, city population size, and random dot movement discrimination tasks). Although these specific decision tasks have proven useful in determining the cognitive underpinnings of decision and confidence assessment generation under time constraints, they are limited in determining the effects of environmental constraints that exist in practical fusion scenarios such as stress, fatigue, and anxiety. There have been some investigations geared towards determining what cognitive mechanisms are affected by fatigue and anxiety. The study in [145] investigated conditions for which an anxiety-related threat bias will and will not manifest using drift diffusion models on TAFC tasks. In [146], drift diffusion models on TAFC tasks are again used to model the effects of fatigue on the human decision-making process in the case of sleep deprivation. Both of these studies however, do not address how anxiety and/or fatigue affect the generation of subjective confidences. Furthermore, experimental and theoretical analyses considering anxiety and fatigue have not yet been conducted on multihypothesis (or vague) decision tasks. It would be beneficial to develop a series of models (either 2DSD or 2DSD/LCA hybrid) to better understand how stress, fatigue, and anxiety affect the performance of human decision makers, and ultimately the performance of the data fusion center. Such research would be used to inform soft and hard/soft fusion researchers in the design of fusion systems which are most appropriate for the tasks at hand.

7.1.3 Response Models on Practical Applications

Although the decision tasks currently used by cognitive psychologists researchers in decision field theory have been proven useful in assessing the underpinnings of many cognitive processes, they are detached from applications which have direct relevance to the data fusion research community. In this thesis we have discussed the decision tasks given in [44, 45], namely the line length, city population size, and random dot movement discrimination tasks. Indirectly, the line length discrimination task can be thought of as representing perceptual decision-making problems in which a subject's decision

involves differentiating between two (or more) visual stimuli. The random dot movement task also can be thought of as representing a perceptual decision making task, but differs from the line length task in that it involves detecting the presence of a stimuli amongst noise. The city population size discrimination task can be thought of as representing intellectual decision-making problems, in which a subject must draw upon prior knowledge and experience to chose the most correct stimuli. Investigations on more applicable decision tasks exist, and involve topics as textual processing [145] and purchasing decisions [147].

There are several relevant soft and hard/soft fusion applications that can be readily modeled using stochastic models of human decision-making. We briefly outline two of these tasks while describing their real-world significance, application to fusion, and general methods in which human response models can be developed.

Handwriting Forgery Detection

Authenticating signatures is a common task in the field of forensics document analysis [148], in which a trained forensic document examiner (FDE) uses known true samples of a subject's handwriting to determine if a questioned signature or document is authentic. FDEs usually undergo years of training in order to become proficient in extracting and matching trends in handwriting [149], and as a result their services are usually expensive and in high demand. For these reasons, forgery detection has been heavily investigated over the past decade as a machine learning problem (*e.g.*, [3,149,150]). There has been a substantial amount of work on applying data fusion techniques to combine multiple classifiers (e.g., [3, 151–153]), however there has been little work concerning the fusion of document examiner opinions (*i.e.*, soft fusion), and similarly, the fusion of document examiner opinions with opinions generated via a forgery detection classifier (*i.e.*, hard/soft fusion). In fact, a recent state of the art review noted that handwriting examination continues to be predominantly focused on using expert opinions [154]. The review goes on to state that this trend is not surprising, as the role of expert opinions in "high stakes" systems usually considers automation as only a part of the overall decision-making process.

The forgery detection problem not only presents itself as an ideal candidate for soft and hard/soft

fusion, but it can also be readily modeled using human decision making models from cognitive psychology that operate on TAFC tasks. In a hypothetical experimental setup, subjects would be instructed to observe a set of authentic and questioned handwriting samples, declare that the questioned samples are "Authentic" or "Not Authentic", and assess their confidence on a subjective probability scale. The difficulty of the forgery detection task scales with the level of similarity between the questioned and authentic handwriting samples (*i.e.*, higher signature similarity yields higher task difficulty). Many metrics have been proposed for assessing image similarity (*e.g.*, the study in [154]), implying that stimulus difficulty could be quantified in a straightforward manner. The same logic could probably also be applied in modeling subjects responding in writer identification tasks, in which an unknown document must be matched against one of many sample of handwriting from possibly different authors.

Facial Recognition in Unconstrained Crowds

Many threat response scenarios, such as the recent 2013 Boston Marathon bombing [155], have highlighted the need for automated systems that can use surveillance footage to detect persons-of-interest among large, unconstrained crowds. As opposed to a lineup of suspects, unconstrained crowds typically involve a very large number of faces that may have non-ideal poses, thus making it difficult to apply standard facial recognition algorithms. Multiple images of large crowds prove very difficult for law enforcement officials and analysts to parse, motivating the need for algorithms which can accurately isolate and detect the faces of persons-of-interest. Unfortunately, facial recognition in such crowds is challenging for existing facial detection methods and databases [156]. As an example, the study in [157] evaluated the performance of existing facial detection algorithms on being able to detect the Boston Marathon bombings perpetrators. Their results showed that the facial recognition algorithms that were investigated tended to rank images of the perpetrators with lower scores than images of many other people, sometimes on the order of hundreds or even thousands. Alongside the development of improved facial detection/recognition systems [156], the problem of facial recognition in unconstrained crowds may benefit from the use of hard/soft fusion operators to combine the outputs of facial recognition software suites in addition to the opinions of bystanders

and law enforcement officials.

There are a few different task structures in which we can envision casting the facial recognition in unconstrained crowds problem. In a TAFC task construction, subjects would be shown images of a person (or persons) of interest and asked to answer "Suspects Present" or "Suspects Not Present" when presented with multiple images of unconstrained crowds. In a multihypothesis construction, subjects could be asked to answer as to which (if any) of the persons of interest are present in a given image. As opposed to the other tasks discussed in the thesis, task difficulty for a subject making decisions on unconstrained crowds might be difficult to assess directly. There are many factors that may make it more difficult for a subject to render a decision, such as the number of people present in the image, the number of faces which are partially or totally obstructed, or the overall quality of the image itself. Thus for this task, it would be very important to address the impact of each possible source of task difficulty to determine its effects both on the subjects and also on the performance of the fusion operators being investigated.

7.2 Data Fusion Research Areas

7.2.1 Performance Bounds for Soft and Hard/Soft Fusion Operators

Recall from Chapter 2 that we assumed the human sources in a soft or hard/soft fusion system provided responses over a set of alternatives in the form of decisions and a probabilistic decision confidence assessments. We noted that this assumption was reasonable, since other information extraction or data association algorithms could be used on data from a human to arrive at a decision and decision confidence assessment (*e.g.*, natural language processing algorithms). When viewed in this manner, the soft and hard/soft data fusion problem can be considered similar to a distributed detection (*i.e.*, decision fusion) problem. Many well-known decision fusion operators exist, and satisfy some sort of optimality criterion (*e.g.*, minimizing a risk function or maximizing a correct detection rate for some fixed error rate constraint). A few of these decision fusion operators have been discussed in this thesis (*e.g.*, the Chair and Varshney Fusion rule [54]). The challenge in using many of these optimal decision fusion operators however, lies in the fact that they require knowledge (or partial knowledge) of the probability models governing the information sources in a

fusion system. It would seem natural to attempt to use the human response models developed by cognitive psychologists. However, such an approach may prove to be impractical due to the complicated nature of the human response models from decision field theory. Furthermore, implementing the optimal fusion rule becomes increasingly difficult as the number of alternatives increase because of the increased complexity for both the human response models and the optimal fusion rule.

Nevertheless, if the optimal decision fusion rule exists and is applicable to a given soft or hard/soft fusion scenario then it is possible to implement it in very specific and highly controlled situations (*e.g.*, simulation). In such situations, the optimal fusion rule gives an "upper bound" on the performance possible over any decision fusion rules that can be used. Other practical and easily implementable fusion operators could be tested in order to see how close to the optimal fusion rule they perform. It may then be possible to chose a fusion operator that yields statistical properties and performance which are closest to the optimal rule. For this approach to be feasible, a larger number of cognitive psychology models and decision tasks must be evaluated to create a diverse set of case studies. Additionally, methods and computational tools for systematically implementing such human response models on new fusion rules in a straight forward manner would need to be developed and proposed to the data fusion research community.

7.2.2 Adaptive Approaches to Optimal Soft and Hard/Soft Fusion

As mentioned in the previous section, the optimal fusion rule for a given soft or hard/soft fusion scenario may not be known or be practical to implement in real-world scenarios. Alternatively, it may be possible to develop heuristic-based soft and hard/soft fusion operators that are asymptotically optimal (*i.e.*, that approximate the optimal rule in the long run). Several decision fusion operators that make use of adaptive learning algorithms exist for traditional binary decision fusion scenarios (*e.g.*, see [158, Chapter 4] for a full summary). There have also been studies that address the multihypothesis adaptive decision fusion problem [159] and studies that attempt to develop adaptive methods for vague decisions using Dempster-Shafer (DS) theory [160]. It would be interesting to first evaluate a selection of these adaptive fusion operators on soft and hard/soft fusion scenarios, using a selection of human response-models from cognitive psychology. Then, assuming a general form of

human response probability models from decision field theory, one could investigate the possibility of developing adaptive fusion rules that learn (or indirectly learn) the relevant parameters of the human response models in order to approximate an optimal fusion rule.

7.2.3 Additional Simulation Areas

There are a number of additional simulations and case studies that could also be performed using the content of this thesis as a starting point. We have focused on parallel fusion schemes that perform "one-shot" data combinations. Alternatively, it would be very interesting to investigate the performance of "sequential" data fusion operators. This is an important area to investigate, especially since "response time" is an important element of the human response models from cognitive psychology. We have also assumed in this thesis that the subjects being combined were all honest in their responses. There has been recent interest in analyzing the performance of data fusion operators in light of malicious (or faulty) sources of information; such information sources have been referred to as *Byzantines* [161]. When choosing amongst alternatives, for example, a Byzantine information source would attempt to determine the most correct alternative and then select an incorrect alternative in an attempt to lower the overall performance of the fusion center. It is possible to perform set of fusion studies similar to this thesis while incorporating a few example Byzantine strategies from the literature into the human response models.

Chapter 8: Concluding Remarks

In this thesis, we have discussed how probabilistic models of human decision-making can be used in the evaluation of various soft and hard/soft fusion operators. Specifically, we have used a model of human decision-making, confidence assessment, and response time known as two-stage dynamic signal detection (2DSD) to simulate the statistical performance of fusion operators from detection and estimation theory, Bayesian epistemology, and Dempster-Shafer theory when used in a variety of soft and hard/soft fusion cases. Unlike the use of examples, counterexamples, thought experiments, or direct human testing, we have demonstrated how models of human decision-making and confidence assessment can be used to flexibly evaluate the performance of fusion operators in a statistically meaningful manner..

In addition to fusion operator performance, 2DSD parameter sets from [44] and [45] were used to evaluate the performance of different fusion operator input considerations, the value of including human subjective confidences over decisions alone, and the performance of various probability transformation operators from Dempster-Shafer theory. We presented four soft fusion studies in Chapter 4 and 2 hard/soft fusion studies in Chapter 6. The results in Chapter 4 show improvements in fusion performance when discounting accordingly to subject decisions and confidence assessment reliability. These results also showed that the importance of decision confidence self-assessment decreased as the number of human responses included in the fusion combination increased. Similar results were observed for the cases of hard/soft fusion evaluated in Chapter 6.

In Chapter 5, we considered the problem of soft fusion again, but focusing on multihypothesis (M-ary) decision tasks. Since multihyptohesis decision-making and confidence assessment modeling has not yet seen significant attention by cognitive psychologists, we proposed a pairwise successive comparison technique for generating M-ary human decision making from binary models of human decision making. This method was also used to simulate imprecise human decisions and decision confidence assessments. We then presented two fusion studies in Chapter 5. In the first study, we

investigated the performance of a few fusion operators as the number of alternatives was increased. We observed that the performance losses were more negligible as the number of sources were increased. In the second study, we evaluated the performance of the same set of fusion operators while considering vague human decisions and confidence assessments. It was observed that the overall performance of the fusion operators decreased as the human decisions became more vague.

Some of the performance trends exhibited by the fusion operators discussed in this dissertation were very similar across all of the studies in Chapters 4 through 6. For example, fusion operators which regard source conflict as uncertainty tended to produce inferior fusion performance (*e.g.*, Yager's rule and Dubois and Prade's rule). Bayes's rule of probability combination and Dempster's rule of combination (DRC) were observed to exhibit monotonically increasing belief towards the true outcome with respect to the number of human decision and confidence assessment pairs present in the combination. Although their usage requires that the sources cannot assign probability zero or one to an alternative, this was shown to be easily circumvented by some of the methods investigated in Chapters 4 through 6. Across every fusion simulation case investigated, it was found that Bayes's rule of probability combination yielded fusion performance that was either as good as or better than DRC. These results were especially apparent in Chapter 5 when simulating the fusion of vague human decisions and confidence assessments, an area of soft and hard/soft fusion that Dempster-Shafer theory operators have been heavily proposed for use in. Combined with its reduced computational complexity, Bayes' rule of probability combination may prove to be a more appropriate operator for soft and hard/soft fusion scenarios, however further investigation is needed.

Finally, we have presented a few interdisciplinary areas for future work in Chapter 7. We have discussed how the logic of 2DSD could be used in extending existing models M-ary human decision-making, a selection of fusion-relevant cognitive mechanisms that warrant investigation, and a set of fusion-relevant decision tasks which could be straightforwardly modeled by cognitive psychologists. We have also discussed how knowledge of the forms of the actual human decision-making and confidence assessment probabilistic models could be used in the design of an optimal soft or hard/soft fusion operator, in addition to near-optimal and adaptive variants.

Soft and had/soft data fusion is an area of research that could provide many potential benefits to a variety of societally relevant applications (*e.g.*, defense, forensics, medicine, and disaster response). By definition, this problem is inherently interdisciplinary and hence addressing it will depend heavily on the cooperation of data fusion and cognitive psychology researchers. We believe that the studies provided by this dissertation motivate the need for further investigation in appropriate methods for designing and evaluating the performance of data fusion systems which incorporate human opinions.

Bibliography

[1] D. L. Hall, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, pp. 6–23, 1997.

[2] H. Wang, T. Kirubarajan, and Y. Bar-Shalom, "Precision large scale air traffic surveillance using imm/assignment estimators," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 255–266, Jan 1999.

[3] M. Fuentes, S. Garcia-Salicetti, and B. Dorizzi, "On line signature verification: Fusion of a hidden markov model and a neural network via a support vector machine," in *8th International Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 253–258.

[4] V. Calhoun and T. Adali, "Feature-based fusion of medical imaging data," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 711–720, Sept 2009.

[5] M. Kam, X. Zhu, and P. Kalata, "Sensor fusion for mobile robot navigation," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 108–119, 1997.

[6] D. Obradovic, H. Lenz, and M. Schupfner, "Fusion of sensor data in siemens car navigation system," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 43–50, Jan 2007.

[7] D. L. Hall, M. McNeese, J. Llinas, and T. Mullen, "A framework for dynamic hard/soft fusion," in *Information Fusion*, 2008, pp. 1–8.

[8] J. Avouc, D. Huscher, D. Furst, C. Opitz, O. Distler, Y. Allanore, K. Ahmadi-Simab, C. Albera, F. Behrens, M. olster *et al.*, "Expert consensus for performing right heart catheterisation for sususpect pulmonary arterial hypertension in sysystem sclerosis: a delphi consensus study with cluster analysis," *Annals of the rheumatic diseases*, 2013, sourced data.

[9] J. K. Hammitt and Y. Zhang, "Combining experts' judgments: Comparison of algorithmic methods using synthetic data," *Risk Analysis*, vol. 33, pp. 109–120, 2013, simple statistical model.

[10] E. Merkle and M. Steyvers, "A psychological model for aggregating judgments of magnitude," in *Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 2011, pp. 236–243.

[11] P. A. Morris, "Combining expert judgments: A Bayesian approach," *Management Science*, vol. 23, no. 7, pp. pp. 679–693, 1977. [Online]. Available: http://www.jstor.org/stable/2630848

[12] K. Sambhoos, J. Llinas, and E. Little, "Graphical methods for real-time fusion and estimation with soft message data," in *Information Fusion, 2008 11th International Conference on*, 2008.

[13] R. T. Clemen and R. L. Winkler, "Aggregation of expert probability judgments," in *Advances in Decision Analysis: From Foundations to Applications*, W. Edwards, R. F. M. Jr, and D. von Winterfeldt, Eds. Cambridge University Press, 2007, vol. 7, ch. 9, pp. 154–176. [Online]. Available: http://www.usc.edu/dept/create/assets/001/50849.pdf

[14] C. W. Granger and R. Ramanathan, "Improved methods of combining forecasts." *Journal of Forecasting*, vol. 3, no. 2, pp. 197 – 204, 1984. [Online]. Available: http://www.library.drexel.edu/cgi-bin/r.cgi?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=6143042&site=ehost-live

[15] S. Kaplan, "'Expert information' versus 'expert opinions'. Another approach to the problem of eliciting/ combining/using expert knowledge in {PRA}," *Reliability Engineering & System Safety*, vol. 35, no. 1, pp. 61 – 72, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/095183209290023E

[16] I. Albert, S. Donnet, S. Guihenneuc-Jouyaux, Chantal an Low-Choy, K. Mengersen, and J. Rousseau, "Combining expert opinions in prior elicitation," *Bayesian Analysi*, vol. 7, pp. 503–532, 2012.

[17] N. B. Abdallah, N. Mouhous-Voyneau, and T. Denaeux, "Combining statistical and expert evidence using belief functions: Application to centennial sea level estimation taking into account climate change," *International Journal of Approximate Reasoning*, 2013, uses historical/preexisting data.

[18] L. Scholten, A. Scheidegger, P. Reichert, and M. Maurer, "Combining expert knowledge and local data for improved service life model of water supply networks," *Environmental Modeling and Software*, vol. 42, pp. 1–16, 2013, uses preexisting data sets.

[19] T. Wickramarathne, K. Premaratne, M. Murthi, M. Scheutz, S. Kubler, and M. Pravia, "Belief theoretic methods for soft and hard data fusion," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2388–2391.

[20] M. Pravia, O. Babko-Malaya, M. Schneider, J. White, C.-Y. Chong, and A. Willsky, "Lessons learned in the creation of a data set for hard/soft information fusion," in *Proceedings of the 12th International Conference on Information Fusion*, July 2009, pp. 2114–2121.

[21] S. Acharya and M. Kam, "Evidence combination for hard and soft sensor data fusion," in *Proceedings of the 14th International Conference on Information Fusion*, July 2011, pp. 1–8.

[22] R. N. nez, T. L. Wickramarathne, K. Premaratne, M. N. Murthi, S. Kübler, M. Scheutz, and M. A. Pravia, "Credibility assessment and inference for fusion of hard and soft information," in *Advances in Design for Cross-Cultural Activities Part I*, D. M. Nicholson, Ed. CRC Press, 2012, pp. 96–105.

[23] S. Merten, "Employing data fusion in cultural analysis and counterinsurgency in tribal social systems," *Strategic Insights*, vol. 8, no. 3, August 2009. [Online]. Available: http://edocs.nps.edu/npspubs/institutional/newsletters/strategic%20insight/2009/mertenAug09.pdf

[24] Hard/soft data fusion research has broad potential. Penn State College of Information Sciences and Technology. [Online]. Available: http://ist.psu.edu/news/hard-soft-data-fusion-research-has-broad-potential

[25] B. Khaleghi, A. Khamis, F. Karray, and S. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, January 2013.

[26] M. Pravia, R. K. Prasanth, P. Arambel, C. Sidner, and C.-Y. Chong, "Generation of a fundamental data set for hard/soft information fusion," in *Proceedings of the 11th International Conference on Information Fusion*, June 2008, pp. 1–8.

[27] K. Premaratne, M. Murthi, J. Zhang, M. Scheutz, and P. Bauer, "A dempster-shafer theoretic conditional approach to evidence updating for fusion of hard and soft data," in *Proceedings of the 12th International Conference on Information Fusion*, July 2009, pp. 2122–2129.

[28] J. Dezert, P. Wang, A. Tchamova *et al.*, "On the validity of dempster-shafer theory," in *Proceedings of the 15th International Conference on Information Fusion*, 2012.

[29] A. Josang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, pp. 279–311, 2001.

[30] L. A. Zadeh, "A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 7, no. 2, pp. 85–90, 1986.

[31] J. Graham, D. Hall, and J. Rimland, "A coin-inspired synthetic dataset for qualitative evaluation of hard and soft fusion systems," in *Proceedings of the 14th International Conference on Information Fusion*, July 2011, pp. 1–8.

[32] J. Graham, W. R. Grace, and A. K. Sridhara, "Beyond SYNCOIN: Informational requirements for the three-block war," in *17th International Conference on Information Fusion*, July 2014, pp. 1–9.

[33] K. Date, G. Gross, S. Khopkar, R. Nagi, and K. Sambhoos, "Data association and graph analytical processing of hard and soft intelligence data," in *17th International Conference on Information Fusion*, July 2013, pp. 404–411.

[34] G. A. Gross, K. Date, D. R. Schlegel, J. J. Corso, J. Llinas, R. Nagi, and S. C. Shapiro, "Systemic test and evaluation of a hard+soft information fusion framework: Challenges and current approaches," in *17th International Conference on Information Fusion*, July 2014, pp. 1–8.

[35] K. Date, G. A. Gross, and R. Nagi, "Test and evaluation of data association algorithms in hard+soft data fusion," in *17th International Conference on Information Fusion*, July 2014, pp. 1–8.

[36] B. Park, A. Johannson, and D. Nicholson, "Crowdsourcing soft data for improved urban situation assessment," in *Proceedings of the 16th International Conference on Information Fusion*, 2013.

[37] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz, "Enabling effective human-robot interaction using perspective-taking in robots," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 35, no. 4, pp. 460–470, 2005.

[38] M. Cao, A. Stewart, and N. Leonard, "Convergence in human decision-making dynamics," *Systems & Control Letters*, vol. 59, no. 2, pp. 87–97, 2010.

[39] A. Stewart, M. Cao, and N. E. Leonard, "Steady-state distributions for human decisions in two-alternative choice tasks," in *American Control Conference (ACC), 2010.* IEEE, 2010, pp. 2378–2383.

[40] A. Stewart, M. Cao, A. Nedic, D. Tomlin, and N. E. Leonard, "Towards human–robot teams: Model-based analysis of human decision making in two-alternative choice tasks with social feedback," *Proceedings of the IEEE*, vol. 100, no. 3, pp. 751–775, 2012.

[41] H. Van-Trees, *Detection, estimation, and modulation theory.* John Wiley & Sons, Inc, 1968.

[42] W. Talbott, "Bayesian epistemology," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed. Stanford University, 2011. [Online]. Available: http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/

[43] G. Shafer, *A Mathematical Theory of Evidence.* Princeton University Press: Princeton, NJ, 1976.

[44] T. Pleskac and J. Busemeyer, "Two-stage dynamic signal detection: A theory of choice, decision time, and confidence," *Psychological review*, vol. 117, no. 3, pp. 864–901, July 2010.

[45] Y. Shuli, T. J. Pleskac, and M. D. Zeigenfuse, "Dynamics of post-decisional processing of confidence," *Journal of Experimental Psychology: General*, In Press.

[46] Z. Wang and G. J. Klir, *Generalized Measure Theory.* Springer Science+Business Media, LLC, 2009.

[47] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, Inc., 1965.

[48] P. K. Varshney, *Distributed Detection and Data Fusion.* Springer-Verlag, New York, Inc., 1997.

[49] F. Ramsey, *The Foundations of Mathematics and other Logical Essays.* London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company, 1926, ch. VII: Truth and Probability, pp. 156–158, 1999 electronic edition.

[50] B. de Finetti, "Foresight: Its logical laws, its subjective sources," *Annales de l'Institut Henri Poincar*, vol. 7, pp. 1–68, 1937, translated by H. E. Kyburg, Jr. and H. E. Smokler in Studies in Subjective Probability, Robert E. Krieger Publishing Company, 1980.

[51] T. Bayes, "An essay towards solving a problem in the doctrine of chances," *Philosophical Transactions*, vol. 53, pp. 370–418, 1763.

[52] S. M. Stigler, *The History of Statistics: The measurement of uncertainty before 1900.* Harvard University Press, 1986.

[53] I. Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and statistical inference*, 2nd ed. Cambridge University Press: Cambridge, UK, 1975.

[54] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, pp. 98–101, 1986.

[55] M. Daniel, "On probabilistic transformations of belief functions," Institute of Computer Science: Academy of Sciences of the Czech Republic, Tech. Rep. 934, Sep 2005.

[56] G. Shafer, "Perspectives on the theory and practice of belief functions," *International Journal of Approximate Reasoning*, vol. 4, no. 5-6, pp. 323–363, 1990.

[57] L. A. Zadeh, "On the validity of dempster's rule of combination of evidence," University of California, Berkely, Tech. Rep. 79/24, 1979.

[58] R. Haenni, "Shedding new light on zadeh's criticism of dempster's rule," in *7th International Conference on Information Fusion (FUSION)*, 2005, pp. 879–884.

[59] P. Smets, "Analyzing the combination of conflicting belief functions," *Information Fusion*, vol. 8, no. 4, pp. 387–412, 2007.

[60] J. Pearl, "Reasoning with belief functions: An analysis of compatibility," *International Journal of Approximate Reasoning*, vol. 4, no. 5-6, pp. 363–389, 1990.

[61] G. Shafer, "Rejoinder to comments on "perspectives on the theory and practice of belief functions"," *International Journal of Approximate Reasoning*, vol. 3, pp. 445–480, 1992.

[62] J. Pearl, "Rejoinder to comments on "reasoning with belief functions: An analysis of compatibility"," *International Journal of Approximate Reasoning*, vol. 6, no. 3, pp. 425–443, 1992.

[63] D. Buede and P. Girardi, "A target identification comparison of bayesian and dempster-shafer multisensor fusion," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, no. 5, pp. 569–577, September 1997.

[64] J. Braun, "Dempster-shafer theory and bayesian reasoning in multisensor data fusion," *Proceedings of the SPIE*, vol. 4051, pp. 255–266, 2000.

[65] S. Ben Chaabane, F. Fnaiech, M. Sayadi, and E. Brassart, "Relevance of the dempster-shafer evidence theory for image segmentation," in *Signals, Circuits and Systems (SCS), 2009 3rd International Conference on*, nov. 2009, pp. 1 –4.

[66] Z. Chunjiang and D. Yong, "Color image edge detection using dempster-shafer theory," in *Artificial Intelligence and Computational Intelligence, 2009. AICI '09. International Conference on*, vol. 3, nov. 2009, pp. 476 –479.

[67] ——, "A modified sobel edge detection using dempster-shafer theory," in *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, oct. 2009, pp. 1 –4.

[68] A. Bouakache, R. Khedam, N. Abbas, Y. Ait Abdesselam, and A. Belhadj-Aissa, "Multi-scale satellite images fusion using dempster shafer theory," in *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, april 2008, pp. 1 –6.

[69] H. Ip and J. Ng, "Human face recognition using dempster-shafer theory," in *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 2, nov 1994, pp. 292 –295 vol.2.

[70] Y. Sugie and T. Kobayashi, "Media-integrated biometric person recognition based on the dempster-shafer theory," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 4, 2002, pp. 381 – 384 vol.4.

[71] D. Kisku, J. Sing, and P. Gupta, "Multibiometrics belief fusion," in *Machine Vision, 2009. ICMV '09. Second International Conference on*, dec. 2009, pp. 37 –40.

[72] A. Yazdani, T. Ebrahimi, and U. Hoffmann, "Classification of eeg signals using dempster shafer theory and a k-nearest neighbor classifier," in *Neural Engineering, 2009. NER '09. 4th International IEEE/EMBS Conference on*, 29 2009-may 2 2009, pp. 327 –330.

[73] R. Conte, M. Longo, S. Marano, V. Matta, and F. Velardi, "Fusing evidences from intracranial pressure data using dempster-shafer theory," in *Digital Signal Processing, 2007 15th International Conference on*, july 2007, pp. 159 –162.

[74] R. Medina, M. Garreau, C. Navarro, J. Coatrieux, and D. Jugo, "Reconstruction of three-dimensional cardiac shapes in biplane angiography: a fuzzy and evolutionary approach," in *Computers in Cardiology 1999*, 1999, pp. 663 –666.

[75] E. Straszecka and J. Straszecka, "Medical diagnosis support using uncertainty and imprecision measures," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 6, oct. 2004, pp. 5835 – 5840 vol.6.

[76] P. Vannoorenberghe, O. Colot, and D. De Brucq, "Dempster-shafer's theory as an aid to color information processing. application to melanoma detection in dermatology," in *Image Analysis and Processing, 1999. Proceedings. International Conference on*, 1999, pp. 774 –779.

[77] Z. Heng-Hao, M. Xiu-yun, and L. Zao-zhen, "Improved algorithm based on dempster-shafer theory: Using in the integrated navigation," in *Information Science and Engineering (ICISE), 2009 1st International Conference on*, dec. 2009, pp. 3808 –3811.

[78] F. Golshani, E. Cortes-Rello, and S. Ahluwalia, "Ruta-100: a dynamic route planner," in *Tools with Artificial Intelligence, 1992. TAI '92, Proceedings., Fourth International Conference on*, nov 1992, pp. 418 –423.

[79] S. Izri and E. Brassart, "Uncertainties quantification criteria for multi-sensors fusion: Application to vehicles localisation," in *Control and Automation, 2008 16th Mediterranean Conference on*, june 2008, pp. 457 –462.

[80] S. Pashazadeh and M. Sharifi, "Reliability assessment under uncertainty using dempster-shafer and vague set theories," in *Computational Intelligence for Measurement Systems and Applications, 2008. CIMSA 2008. 2008 IEEE International Conference on*, july 2008, pp. 131 –136.

[81] W. Miao and Y. Liu, "Information system security risk assessment based on grey relational analysis and dempster-shafer theory," in *Mechatronic Science, Electric Engineering and Computer (MEC), 2011 International Conference on*, aug. 2011, pp. 853 –856.

[82] F. Dong, S. Shatz, and H. Xu, "Inference of online auction shills using dempster-shafer theory," in *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, april 2009, pp. 908 –914.

[83] H. Fang, Z. Wang, X. Gu, and K. Chakrabarty, "Ranking of suspect faulty blocks using dataflow analysis and dempster-shafer theory for the diagnosis of board-level functional failures," in *European Test Symposium (ETS), 2011 16th IEEE*, may 2011, pp. 195 –200.

[84] R. Srivastava and T. Mock, "Evidential reasoning for webtrust assurance services," in *System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on*, vol. Track5, 1999, p. 10 pp.

[85] N. Pal and S. Ghosh, "Some classification algorithms integrating dempster-shafer theory of evidence with the rank nearest neighbor rules," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 31, no. 1, pp. 59 –66, jan 2001.

[86] J. Boston, "A signal detection system based on dempster-shafer theory and comparison to fuzzy detection," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 30, no. 1, pp. 45 –51, feb 2000.

[87] W. Zhao, T. Fang, and Y. Jiang, "Data fusion using improved dempster-shafer evidence theory for vehicle detection," in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, vol. 1, aug. 2007, pp. 487 –491.

[88] T. Chen and V. Venkataramanan, "Dempster-shafer theory for intrusion detection in ad hoc networks," *Internet Computing, IEEE*, vol. 9, no. 6, pp. 35 – 41, nov.-dec. 2005.

[89] Z. Yun, Y. Xuelian, C. Minglei, and W. Xuegang, "Radar target recognition based on multiple features fusion with dempster-shafer theory," in *Electronic Measurement Instruments (ICEMI), 2011 10th International Conference on*, vol. 1, aug. 2011, pp. 243 –247.

[90] P. Smets, "Belief functions: The disjunctive rule of combination and the generalized bayesian theorem," *International Journal of Approximate Reasoning*, vol. 9, pp. 1–35, 1993.

[91] J. Y. Jaffray, "Linear utility theory for belief functions," *Operations Research Letters*, vol. 8, no. 2, pp. 107–112, 1989.

[92] M. Liggins, D. Hall, and J. Llinas, *Handbook of multisensor data fusion: theory and practice*. CRC, 2001, vol. 22.

[93] P. Smets, "Decision making in the tbm: The necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, pp. 133–147, 2005.

[94] R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, pp. 93–138, 1987.

[95] P. Smets, "Decision making in a context where uncertainty is represented by belief functions," in *Belief Functions in Business Decisions*, R. P. Srivastava and T. J. Mock, Eds. Physica-Verlag, 2002, pp. 17–61.

[96] D. Dubois and H. Prade, "On several representations of an uncertain body of evidence," in *Fuzzy Information and Decision Processes*, E. S. M. Gupta, Ed. Elsevier Science Ltd, 1982, pp. 167–181.

[97] F. Smarandache and J. Dezert, *Advances and Applications of DSmT for Information Fusion.* American Research Press, 2009, vol. III.

[98] A. Josang, "The consensus operator for combining beliefs," *Artificial Intelligence*, vol. 142, pp. 157–170, 2002. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.1508

[99] K. Sentz and S. Ferson, "Combination of evidence in dempster-shafer theory," SANDIA Tech Report, Tech. Rep. SAND2002-0835, 2002.

[100] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision Support Systems*, vol. 29, pp. 1–9, 2000.

[101] F. Dietrich and C. List, "Probabilistic opinion pooling," Paris School of Economics, Tech. Rep., March 2014. [Online]. Available: http://personal.lse.ac.uk/list/PDF-files/OpinionPoolingReview.pdf

[102] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," *Uncertainty in AI*, vol. 5, pp. 29–39, 1990.

[103] P. Smets and R. Kennes, "The transferrable belief model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[104] J. Sudano, "Yet another paradigm illustrating evidence fusion," in *Proceedings of the 9th International Conference on Information Fusion*, 2006.

[105] J. Dezert and F. Smarandache, "A new probabilistic transformation of belief mass assignment," in *Proceedings of the 11th International Conference on Information Fusion*, 2008, pp. 1–8.

[106] M. L. Krieg, "A tutorial on Bayesian belief networks," Australian Department of Defense, Electronics and Surveillance Research Laboratory, Tech. Rep. DSTO-TN-0403, 2001. [Online]. Available: http://www.dsto.defence.gov.au/publications/2424/DSTO-TN-0403.pdf

[107] K. Wu, W. Tang, K. Z. Mao, G. Ng, and L. O. Mak, "Semantic-level fusion of heterogenous sensor network and other sources based on Bayesian network," in *Proceedings of the 17th International Conference on Information Fusion*, July 2014.

[108] R. Fagin and J. Y. Halpern, "A new approach to updating beliefs," in *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, 1990.

[109] R. P. S. Mahler, "Random sets: Unification and computation for information fusion – a retrospective assessment," in *Proceedings of the 7th International Conference on Information Fusion*, 2004.

[110] R. Ratcliff, "Diffusion and random walk models," in *International encyclopedia of the social and behavioral sciences.* Oxford, England: Elsevier, 2001, vol. 6, pp. 3668–3673.

[111] M. Stone, "Models for choice reaction time." *Psychometrika*, vol. 25, pp. 251–260, 1960.

[112] D. R. J. Laming, *Information Theory of Choice Reaction Time.* Wiley, New York, 1968.

[113] S. W. Link and R. A. Heath, "A sequential theory of psychological discrimination," *Psychometrika*, vol. 40, pp. 77–105, 1975.

[114] R. Ratcliff, "A theory of memory retrieval," *Psychological review*, vol. 83, pp. 59–108, 1978.

[115] R. A. Heath, "Random-walk and accumulator models of psychophysical discrimination–a critical evaluation," *Perception*, vol. 13, pp. 57–65, 1984.

[116] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. Cohen, "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks." *Psychological review*, vol. 113, no. 4, pp. 700–765, 2006.

[117] P. L. Smith, "A note on the distribution of response time for a random walk with gaussian increments," *Journal of Mathematical Psychology*, vol. 34, pp. 445–459, 1990.

[118] T. Van Zandt, "Roc curves and confidence judgments in recognition memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 26, no. 3, p. 582, 2000.

[119] D. Vickers, *Decision Processes in Visual Perception*. New York, NY: Academic Press, 1979.

[120] A. Heathcote, S. Brown, and D. Mewhort, "Quantile maximum likelihood estimation of response time distributions," *Psychonomic Bulletin & Review*, vol. 9, no. 2, pp. 394–401, 2002.

[121] R. Ratcliff and J. N. Rouder, "Modeling response times for two-choice decisions," *Psychological Science*, vol. 9, pp. 347–356, 1998.

[122] R. Ratcliff and P. L. Smith, "A comparison of sequential sampling models for two-choice reaction time," *Psychological review*, vol. 111, no. 2, pp. 333–367, April 2004.

[123] D. J. Bucci, S. Acharya, and M. Kam, "Simulating human decision making for testing soft and hard/soft fusion algorithms," in *Proceedings of the 47th Annual Conference on Information Sciences and Systems (CISS)*, 2013.

[124] G. Brier, "Verification of forecasts expressed in terms of probability," *Monthly weather review*, vol. 78, pp. 1–3, 1950.

[125] D. J. Bucci, S. Acharya, and M. Kam, "Performance simulation of M-ary soft fusion systems using binary models of human responses," in *17th International Conference on Information Fusion (FUSION)*, 2014, pp. 1–8.

[126] J. N. Tsistsiklis, "Decentralized detection," in *Advances in Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI, 1993, vol. 2, pp. 297–344.

[127] D. J. Bucci, S. Acharya, T. J. Pleskac, and M. Kam, "Subjective confidence and source reliability in soft data fusion," in *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS)*, March 2014.

[128] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, Inc, 2001.

[129] D. J. Bucci, S. Acharya, T. J. Pleskac, and M. Kam, "Performance of probability transformations using simulated human opinions," in *17th International Conference on Information Fusion (FUSION)*, 2014, pp. 1–8.

[130] A. Moro, E. Mumolo, M. Nolich, K. Terabayashi, and K. Umeda, "Improved foreground-background segmentation using dempster-shafer fusion," in *8th International Symposium on Image and Signal Processing and Analysis*, 2013, pp. 72–77.

[131] J. Li, Y. Wang, and Z. J. Mao, "A new multisensor information fusion model using dempster-shafer theory," *Applied Mechanics and Materials*, vol. 475, pp. 415–418, 2014.

[132] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using dempster–shafer evidence theory," *Information Fusion*, vol. 8, no. 4, pp. 379–386, 2007.

[133] F. Cuzzolin, "A geometric approach to the theory of evidence," *IEEE Transactions on Systems Man and Cybernetics, Part C*, vol. 38, p. 2008, 2008.

[134] J. Sudano, "Pignistic probability transforms for mixes of low-and high-probability events," in *4th International Conference on Information Fusion*, 2001.

[135] D. Han, J. Dezert, C. Han, and Y. Yang, "Is entropy enough to evaluate the probability transformation approach of belief function?" in *Proceedings of the 13th Conference on Information Fusion*, 2010, pp. 1–7.

[136] J. J. Sudano, "The system probability information content (pic) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities," in *Proceedings of the 5th International Conference on Information Fusion.* IEEE, 2002, pp. 1277–1283.

[137] F. P. Leite and R. Ratcliff, "Modeling reaction time and accuracy of multiple-alternative decisions," *Attention, Perception, and Psychophysics*, vol. 72, no. 1, pp. 246–273, 2010.

[138] D. Dubois and H. Prade, "A set-theoretic view of belief functions: Logical operators and approximations by fuzzy sets," in *Classical Works of the Dempster-Shafer Theory of Belief Functions*, ser. Studies in Fuzziness and Soft Computing, L. L. Ronald Yager, Ed. Springer-Verlag, Berlin, 2008, vol. 219, pp. 375–410.

[139] A. Martin, A. L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Proceedings of the 11th International Conference on Information Fusion*, 2008, pp. 1–8.

[140] D. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, pp. 433–436, 1997.

[141] (2014) Block mathcing: Estimate motion between images or video frames. MathWorks. [Online]. Available: http://www.mathworks.com/help/vision/ref/blockmatching.html

[142] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, 3rd ed. Boston: Addison-Wesley, 1998, vol. 2.

[143] B. P. Welford, "Note on a method for calculating corrected sums of squares and products," *Techonometrics*, vol. 4, no. 3, pp. 419–420, 1962.

[144] M. Usher and J. L. McClelland, "On the time course of perceptual choice: The leaky competing accumulator model," *Psychological Review*, vol. 108, pp. 550–592, 2001.

[145] C. N. White, R. Ratcliff, M. W. Vasey, and G. McKoon, "Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis," *Emotion*, vol. 10, no. 5, pp. 662–677, 2010.

[146] R. Ratcliff and H. P. A. Van Dongen, "Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 27, pp. 11 285–11 290, July 2011.

[147] I. Krajbich, D. Lu, C. Camerer, and R. A., "The attentional drift-diffusion model extends to simple purchasing decisions," *Frontiers in Psychology*, vol. 3, pp. 1–18, June 2012.

[148] R. Huber and A. Headrick, *Handwriting Identification: Facts and Fundamentals.* CRC Press, 1999.

[149] H. Srinivasan, S. N. Srihari, and M. J. Beal, "Machine learning for signature verification," in *Proceedings of the International Conference on Vision, Graphics, and Image Processing.* Springer LNCS 4338, December 2006, pp. 761–775.

[150] M. Puri, S. N. Srihari, and Y. Tang, "Bayesian network structure learning and inference methods for handwriting," in *12th International Conference onD ocument Analysis and Recognition (ICDAR).* IEEE, 2013, pp. 1320–1324.

[151] Y. S. Huang and C. Y. Suen, "A method of combining multiple experts for the recognition of unconstrained handwritten numerals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 90–94, 1995.

[152] V. Mottl, M. Lange, V. Sulimova, and A. Yermakov, "Signature verification based on fusion of on-line and off-line kernels," in *International Conference on Pattern Recognition*, 2008, pp. 1–4.

[153] A. F. R. Rahman and M. C. Fairhurst, "A novel pair-wise recognition scheme for handwritten characters in the framework of a multi-expert configuration," in *Image Analysis and Processing*, 1997, pp. 624–631.

[154] S. N. Srihari and K. Singer, "Role of automation in the examination of handwritten items," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2012, pp. 619–624.

[155] F. Konkel. (2013, April) Boston probe's big data use hints at the future. FCW: The Business of Federal Technology. [Online]. Available: http://fcw.com/articles/2013/04/26/big-data-boston-bomb-probe.aspx

[156] J. Barr, K. W. Bowyer, and P. J. Flynn, "The effectiveness of face detection algorithms in unconstrained crowd scenes," in *IEEE Winter Conference on Applications of Computer Vision*, 2014, pp. 1020–1027.

[157] J. C. Klontz and A. K. Jain, "A case study on unconstrained facial recognition using the boston marathon bombins suspects," Michigan State University, Tech. Rep. MSU-CSE-13-4, May 2013.

[158] X. Zhu, "Decision-level data fusion," Ph.D. dissertation, Drexel University, September 1997.

[159] B. Liu, A. Jeremic, and K. M. Wong, "Blind adaptive algorithm for M-ary distributed detection," in *International Conference on Acoustics, Speech, and Signal Processing*, 2007, pp. 1025–1028.

[160] E. Côme, L. Oukhellou, T. Denoeux, and P. Aknin, "Learning from partially supervised data using mixture models and belief functions," *Pattern recognition*, vol. 42, no. 3, pp. 334–348, 2009.

[161] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 65–75, September 2013.

Vita

Name: Donald J. Bucci

Degrees: M.S. in Electrical Engineering, Drexel University, Philadelphia, PA

B.S. in Engineering Science, The College of New Jersey, Ewing, NJ

Publications: Lester, C. S., D. J. Bucci, R. Measel, et al., "Performance of Reconfigurable Antennas in a Below-Decks Environment", *Antennas and Wireless Propagation Letters*, In Press.

Measel, R., C. S. Lester, D. J. Bucci, et al., "Reconfigurable Antennas in Highly Multipath Environments", *IEEE International Symposium on Antennas and Propagation*, July, 2014.

Bucci, D. J., S. Acharya, and M. Kam, "Performance of M-ary Soft Fusion Systems Using Simulated Human Responses", *17th International Conference on Information Fusion*, July, 2014.

Bucci, D. J., S. Acharya, T. J. Pleskac, and M. Kam, "Performance of Probability Transformations Using Simulated Human Opinions", *17th International Conference on Information Fusion*, July, 2014.

Measel, R., D. J. Bucci, C. S. Lester, et al., "A MATLAB Platform for Characterizing MIMO-OFDM Communications with Software-Defined Radios", *International Communications Quality and Reliability Workshop*, May, 2014.

Lester, C. S., R. Measel, D. J. Bucci, et al., "Effects of Reconfigurable Antennas On Wireless Network Performance Within A Ticonderoga-Class Engine Room", *American Society of Naval Engineers Electric Machines Technology Symposium*, May, 2014.

Wanuga, K., R. Measel, C. S. Lester, D. J. Bucci, et al., "Performance Evaluation of MIMO OFDM Systems in On-Ship Below-Deck Environments", *Antennas and Wireless Propagation Letters*, vol. 13, pp. 173-176, 2014.

Bucci, D. J., S. Acharya, and M. Kam, "Simulating Human Decision Making for Testing Soft and Hard/Soft Fusion Algorithms", *47th Annual Conference on Information Sciences and Systems*, March, 2013.

Acharya, S., D. J. Bucci, and M. Kam, "Representation and Fusion of Conditionally Refined Opinions using Evidence Trees", *15th International Conference on Information Fusion*, July, 2012.

Bucci, D. J., M. Rotella, et al., "A Pressure Controlled, Hand-Assistive Exoskeleton for Actuated Pinch and Grasp", *Trenton Computer Festival, IEEE GOLD Region 9 Paper Competition (Winner)*, April, 2009.

Bucci, D. J., S. Fathima, and B. F. BuSha, "A two-bit binary control system for an orthotic, hand-assistive exoskeleton", *IEEE 35th Annual Northeast Bioengineering Conference*, April, 2009.