DECENTRALIZED DECISION MAKING
IN A HYPOTHESIS TESTING ENVIRONMENT

by

Jason D. Papastavrou

B.S., Mathematics, Massachusetts Institute of Technology
(1984)

M.S., Electrical Engineering, Massachusetts Institute of Technology
(1986)

Submitted to the Department of
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 1990

© Jason D. Papastavrou, 1990

The author hereby grants to MIT permission to reproduce and to
distribute copies of this thesis document in whole or in part.

Signature of Author: ..

Department of Electrical Engineering and Computer Science
May 21, 1990

Certified by: ..

Michael Athans
Thesis Supervisor

Accepted by: .................................................................

Arthur C. Smith, Chairman
Committee on Graduate Students
DECENTRALIZED DECISION MAKING
IN A HYPOTHESIS TESTING ENVIRONMENT

by
Jason D. Papastavrou

Submitted to the Department of Electrical Engineering and Computer Science on May 21, 1990 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The distributed detection problem is considered, in which a set of decision makers (DMs) receive observations of the environment and transmit finite-valued messages to other DMs according to prespecified communication protocols. All DMs make decisions, in order to maximize a measure of organizational performance. Given the DMs and the communication resources, the problem is to find an architecture for the organization which remains optimal for a variety of operating conditions (if it exists). It is shown that even for very small organizations this problem is quite complex, because the optimal architecture depends on variables external to the team like the prior probabilities of the hypotheses and the misclassification costs.

The tandem team which consists of $N$ DMs and performs binary hypothesis testing using binary messages is considered. Necessary and sufficient conditions are derived for the probability of error to go asymptotically to zero. The result is generalized for multiple hypotheses and multiple messages. Two easily implementable suboptimal decision schemes are also considered. The trade off between the complexity of the decision rules and their performance is examined, and numerical results are presented.

We analyze the characteristics and the properties of the individual decision maker, and we demonstrate that they cannot be directly generalized for the team, unless assumptions on the team coordination and/or training are imposed.

Since the "quest for optimality" in the problems in this framework is associated with great computational and inherent complexity, we suggest that simple suboptimal solutions, which take into consideration the particular characteristics of the problem be employed. To demonstrate this approach, the issues involved with the reduction of an $M$-ary hypothesis testing problem into a sequence of simpler hypothesis testing subproblems are addressed.

Thesis Supervisor: Dr. Michael Athans
Title: Professor of Systems Science and Engineering
ACKNOWLEDGEMENTS

I am greatly indebted to my thesis supervisor Prof. Michael Athans whose influence and philosophy have shaped much more than this document. I was converted to engineering, was shown the proper way to do research and, to use one of his favorite expressions, was taught how to distinguish the forest from the trees. I have been extremely privileged to have associated with him not only as a student, but also as a person for the past six and a half years.

I am grateful to my thesis reader Prof. John Tsitsiklis, who has been a constant source of support, motivation and research ideas; without his contributions this thesis would be very different and it would have taken much longer to complete.

I would like to thank my other thesis reader Prof. Pierre Humblet for his helpful comments, as well as Prof. Amedeo Odoni for his valuable advice and help through my years in MIT. I would also like to thank my fellow Athens College alumni Dr. Alex Levis and Dr. John Pararas who contributed in making my life a lot easier with discussions and assistance over a wide range of issues. Many thanks for the good times to the members of the Greek soccer team (I will always treasure the trophy from our 1984-85 championship) and to my colleagues in the office, and in particular to Dr. Petros Kapasouris and Prof. Tony Rodriguez. I also appreciate the help and cheerful attitude of Fifa Monserrate.

Last, but definitely not least, I am grateful to Prof. Elias Gyftopoulos for his concern, guidance and advice. If I can put into use all the things that I learned from him, I will not only end up a better scientist, but a much better person as well.

Finally, I would like to thank my parents for their continuous encouragement and strong support through my endless academic years.

This research was conducted at the M.I.T. Laboratory for Information and Decision Systems and was supported by the National Science Foundation under grant NSF/IRI-
8902755 (under a subcontract from the University of Connecticut) and by the Office of Naval Research under the grant ONR/N00014-84-K-0519.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>I</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>II</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>IX</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>XIII</td>
</tr>
</tbody>
</table>

## CHAPTER 1. INTRODUCTION AND OVERVIEW

1.1. Problem Definition ................................................. 1
1.2. Literature Review .................................................. 5
1.3 Contributions of Thesis ............................................. 7
1.4 Outline of Thesis ................................................... 10

## CHAPTER 2. AN ENGINEERING APPROACH TO MODELING A CLASS OF DECISION MAKING PROBLEMS

2.1. Introduction .......................................................... 13
2.2. Modeling Decision Making .......................................... 13
2.3. The Search for the Illusive Expert ................................ 17
2.4. The Decentralized Hypothesis Testing Framework ................. 25
2.5. Summary ............................................................... 31

## CHAPTER 3. THE DISTRIBUTED HYPOTHESIS TESTING MODEL

3.1. Introduction .......................................................... 32
3.2. The General Framework .............................................. 32
    3.2.1. Hypotheses and Prior Probabilities .......................... 32
3.2.2. The Decision Makers .............................................. 33
3.2.3. The Fusion Center .................................................. 33
3.2.4. The Communication Messages ................................. 34
3.2.5. The Costs Structure .............................................. 34
3.2.6. The Normalized Probability of Error ......................... 35
3.2.7. Notation .............................................................. 36
3.3. The ROC Curve ......................................................... 36
3.4. Some Special ROC Curves .......................................... 42
  3.4.1. Discrete Distributions ........................................ 42
  3.4.2. Exponential Distributions .................................... 44
  3.4.3. Gaussian Distributions with Equal Variances .............. 45
  3.4.4. Conic ROC Curves .............................................. 45
3.5. The Conditional Independence Assumption .................... 47
3.6. Summary ............................................................... 47

CHAPTER 4. SMALL TEAM PROBLEMS

4.1. Introduction ........................................................... 48
4.2. Tandem versus Parallel Architecture ............................. 50
4.3. Two DMs in Tandem .................................................. 53
  4.3.1. Problem Formulation and Solution .......................... 53
  4.3.2. Architecture Comparisons .................................... 56
  4.3.3. The Analysis ..................................................... 59
  4.3.4. An Example ...................................................... 65
  4.3.5. The Number of Messages ...................................... 69
  4.3.6. Performance Bounds ........................................... 72
  4.3.7. Special Probability Distributions ............................ 75
CHAPTER 5. LARGE TEAM PROBLEMS

5.1. Introduction ............................................................................. 112

5.2. The Optimal Infinite Tandem Team ............................................. 115
   5.2.1. Problem Formulation .......................................................... 115
   5.2.2. A Special Example ............................................................ 117
   5.2.3. The Main Result ............................................................... 125
   5.2.4. The Necessary and Sufficient Conditions ......................... 130
   5.2.5. Generalizations ............................................................... 131

5.3. The "Selfish" Suboptimal Tandem Team .................................... 132
   5.3.1. Problem Formulation ........................................................ 132
   5.3.2. The Main Result ............................................................... 134
   5.3.3. Generalizations ............................................................... 146

5.4. The "Identical" Suboptimal Decision Scheme ............................. 147
   5.4.1. Problem Formulation ........................................................ 147
CHAPTER 6. THE TEAM ROC CURVE AND RANDOMIZED DECISION RULES

6.1. Introduction ................................................................. 165
6.2. Randomized Decision Rules: An Example ................................. 167
   6.2.1. The Example .................................................. 167
   6.2.2. Deterministic Decision Rules ................................ 169
   6.2.3. Independent Joint Randomization .............................. 170
   6.2.4. Dependent Joint Randomization .............................. 175
6.3. Smooth ROC Curves ...................................................... 176
   6.3.1. Strictly Concave Primary DM and Piecewise Linear Consulting DM ... 177
   6.3.2. Piecewise Linear Primary DM and Strictly Concave Consulting DM ... 180
   6.3.3. Both DMs Strictly Concave .................................. 182
6.4. Discussion ................................................................... 186
   6.4.1. Summary .......................................................... 186
   6.4.2. The Neyman-Pearson Formulation .............................. 189
   6.4.3. The Significance of the Non-Concave Team ROC Curve .......... 190

CHAPTER 7. A MULTIPLE HYPOTHESIS TESTING PARADIGM

7.1. Introduction .................................................................. 192
7.2. Problem Definition and Solution ..................................... 193
7.3. The Suboptimal Decision Scheme ......................................................... 196
  7.3.1. The Normalized Prior Probabilities ................................................. 196
  7.3.2. The Binary Decision Tree .............................................................. 199
  7.3.3. The Suboptimal Algorithm ............................................................. 200
  7.3.4. Discussion ....................................................................................... 202
  7.3.5. Selecting the Primary DM ................................................................. 203
  7.3.6. Constructing the Binary Decision Tree ............................................ 204
  7.3.7. Assigning the Consulting DMs ......................................................... 208
  7.3.8. Improving on the Suboptimal Decision Scheme............................... 211

7.4. A Multiattribute Hypothesis Testing Problem ...................................... 213

7.5 Summary ............................................................................................... 215

CHAPTER 8. CONCLUDING REMARKS

8.1. Summary and Conclusions .................................................................... 217

8.2. Directions for Future Research and Concluding Remarks ....................... 221

REFERENCES ............................................................................................ 225
## LIST OF FIGURES

### CHAPTER 1. INTRODUCTION AND OVERVIEW

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Three Decision Makers <em>in Tandem</em></td>
<td>4</td>
</tr>
<tr>
<td>1.2</td>
<td>The <em>Two Consultant</em> Architecture</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Adding a New Decision Maker</td>
<td>4</td>
</tr>
</tbody>
</table>

### CHAPTER 3. THE DISTRIBUTED DETECTION MODEL

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Optimal Solution to the Detection <em>Problem 3.1</em></td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>Ranking DMs Unequivocally (DM A better than DM B)</td>
<td>41</td>
</tr>
<tr>
<td>3.3</td>
<td>Ranking DMs: Better Depends on Threshold $\eta$</td>
<td>41</td>
</tr>
<tr>
<td>3.4(a)</td>
<td>ROC Curve for Discrete Distributions <em>without</em> Randomization</td>
<td>43</td>
</tr>
<tr>
<td>3.5(b)</td>
<td>ROC Curve for Discrete Distributions <em>with</em> Randomization</td>
<td>43</td>
</tr>
</tbody>
</table>

### CHAPTER 4. SMALL TEAM PROBLEMS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Two DMs <em>in Tandem</em></td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>Two DMs <em>in Parallel</em>.</td>
<td>51</td>
</tr>
<tr>
<td>4.3(a)</td>
<td>Operating Points of Typical <em>Primary DM</em></td>
<td>55</td>
</tr>
<tr>
<td>4.3(b)</td>
<td>Operating Point of Typical <em>Consulting DM</em></td>
<td>55</td>
</tr>
<tr>
<td>4.4</td>
<td>Different Configurations for the Two DM Tandem Team</td>
<td>57</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary of the Analysis</td>
<td>58</td>
</tr>
<tr>
<td>4.6</td>
<td>The Six Cases for the Consulting DM</td>
<td>58</td>
</tr>
<tr>
<td>4.7</td>
<td>The 'Restricted' DMs of the Analysis</td>
<td>60</td>
</tr>
<tr>
<td>4.8</td>
<td>The DMs for <em>Example 4.1</em></td>
<td>66</td>
</tr>
</tbody>
</table>
Figure 4.9. Probability Distributions for the DMs of Figure 4.8 ........................................ 66
Figure 4.10(a). Team ROC Curves for Both Configurations and Binary Messages .... 67
Figure 4.10(b). Close Up of Team ROC Curves of Figure 4.11 ................................. 67
Figure 4.11(a). Team ROC Curves for Both Configurations and Ternary Messages... 70
Figure 4.11(b). Close Up of Team ROC Curves of Figure 4.11(a) .............................. 70
Figure 4.12. The DMs of Example 4.2 ............................................................................. 73
Figure 4.13. Probability Distributions for the DMs of Figure 4.12 ............................ 73
Figure 4.14. The First Geometric Approach ................................................................. 79
Figure 4.15. The Second Geometric Approach ............................................................. 80
Figure 4.16(a). Identical DMs with Non-Identical Decision Rules ......................... 89
Figure 4.16(b). The 'Worse' Identical DMs with Non-Identical Decision Rules .... 89
Figure 4.17. Probability Distributions that Maximize the Absolute Bound .......... 93
Figure 4.18. Probability Distributions that Maximize the Relative Bound ............. 95
Figure 4.19. The Two Consultant or V-Architecture ................................................. 97
Figure 4.20. The Three DM Tandem Architecture ..................................................... 97
Figure 4.21. The Three DM Parallel Architecture ..................................................... 97
Figure 4.22. The Three DM Asymmetrical Architecture ........................................... 98
Figure 4.23. ROC Curve of DM with Binary Observations ..................................... 100
Figure 4.24. Probability Distributions for DM of Figure 4.23 ............................... 100
Figure 4.25. Team ROC Curves for DMs of Figure 4.23 ........................................... 101
Figure 4.26. DM who Does Not Result in a Dominant Architecture .................... 103
Figure 4.27. Probability Distributions for DM of Figure 4.26 ............................... 103
Figure 4.28(a). Team ROC Curves for DMs of Figure 4.26 ..................................... 104
Figure 4.28(b). Close Up of Team ROC Curves of Figure 4.28(a) ............................. 104
Figure 4.29. ROC Curve for DM Resulting in Dominant Tandem Architecture .... 107
Figure 4.30. Probability Distribution for DM of Figure 4.29 ..................................... 107
Figure 4.31. Team ROC Curves for DMs of Figure 4.29 ........................................... 108
CHAPTER 5. LARGE TEAM PROBLEMS

Figure 5.1. The Team Consisting of $N$ DMs in Tandem ......................... 116
Figure 5.2. A Special ROC Curve ........................................... 118
Figure 5.3. ROC Curve with Infinite Initial Slope ($m_0 = \infty$) ............ 127
Figure 5.4. ROC Curve with Infinite Initial Slope and Non-Zero Final Slope .... 135
Figure 5.5. ROC Curve with Initial Slope: $m_0 = \infty$ and Final Slope: $m_1 = 0$ .... 137
Figure 5.6. ROC Curves of the Worse DM of Step 3 and of the Original DM .... 139
Figure 5.7. ROC Curve of the Worse DM when Eq.(5.33) is Not True .......... 140
Figure 5.8. The Identical Optimal Decision Scheme .......................... 148
Figure 5.9. Ergodic Markov Chains for Identical Decision Scheme .......... 152
Figure 5.10. $P^e(10)$: Error Probability for a Ten DM Team ($N = 10$) vs. $\eta$ .... 157
Figure 5.11. ROC Curves for the Three Teams of $N = 2$ DMs .................. 157
Figure 5.12. ROC Curves for the Three Teams of $N = 6$ DMs .................. 158
Figure 5.13. ROC Curves for the Three Teams of $N = 10$ DMs ................. 158
Figure 5.14. Performance Comparison of Centralized and Optimal Team ....... 159
Figure 5.15. Performance Comparison of Optimal and Suboptimal Team ...... 160
Figure 5.16. Optimal Team ROC Curves for $N = 1, 2, 6, \text{ and } 10$ DMs .... 160
Figure 5.17. $P^e(N)$: Probability of Error for $N$ DM Team ($\eta = 3$) ......... 162
Figure 5.18. Performance Comparison of Centralized and Optimal Team ....... 162
Figure 5.19. Performance Comparison for Optimal and Suboptimal Team ...... 163

CHAPTER 6. THE TEAM ROC CURVE AND
AND RANDOMIZED DECISION RULES

Figure 6.1. ROC Curves for the DMs of the Example in 6.2.1 ............ 168
Figure 6.2. Probability Distributions for the DMs of Figure 6.1 ................. 168
Figure 6.3(a). Team ROC Curve with Independent and Dependent Randomization .... 172
CHAPTER 7. A MULTIPLE HYPOTHESIS TESTING PARADIGM
LIST OF TABLES

CHAPTER 4. SMALL TEAM PROBLEMS

Table 4.1. The Team ROC Curves of Figure 4.10 ............................................. 68
Table 4.2. Configuration Comparison for Binary Messages ................................. 68
Table 4.3. The Team ROC Curves of Figure 4.11 ............................................. 71
Table 4.4. Configuration Comparison for Ternary Messages ............................... 71
Table 4.5. Configuration Comparison for Example 2 ............................................ 74
Table 4.6. Probability Distributions for 'Worse' DMs of Figure 4.16(b) ............... 89
Table 4.7. Probability of Error of Team of Identical DMs of Figure 4.17 ............... 93
Table 4.8. Probability of Error of Team of Identical DMs of Figure 4.18 ............... 96
Table 4.9. The Team ROC Curves of Figure 4.25 ............................................. 101
Table 4.10. The Team ROC Curves of Figure 4.28 .......................................... 105
Table 4.11. Architecture Comparisons; Optimal Architecture Depends on η ............ 106
Table 4.12. The Team ROC Curves of Figure 4.31 ............................................. 108

CHAPTER 5. LARGE TEAM PROBLEMS

Table 5.1. Probability Distributions for the DM of Figure 5.2 ............................... 118
Table 5.2. Probability Distributions for the DM of Figure 5.4 ............................... 136
Table 5.3. Probability Distributions for the DM of Figure 5.6 ............................... 139
Table 5.4. Probability Distributions for the DM of Figure 5.7 ............................... 140
CHAPTER 6. THE TEAM ROC CURVE AND RANDOMIZED DECISION RULES

Table 6.1. Team ROC Curve for DMs of Figure 6.1 ........................................... 169
Table 6.2. The Non-Concave Portions of the ROC Curve ............................... 174

CHAPTER 7. A MULTIPLE HYPOTHESIS TESTING PARADIGM

Table 7.1. Distance Matrix for Example 6.1 .................................................... 207
CHAPTER 1

Introduction and Overview

1.1. PROBLEM DEFINITION

The main difference between the decision making process of an organization and that of an individual is that organizational decision making occurs in a decentralized or distributed manner. As organizations grow, it becomes very quickly apparent that the decision making process cannot be handled in a centralized way. It is not necessary to reach the level of large organizations, for which decentralization is dictated by geographical constraints, in order to find organizations which operate in a decentralized framework.

There exist decision makers, or even groups of decision makers, within the organization which do not possess all the relevant information and/or ability required to make a good decision. Moreover, constraints on the communication capacity may exist which do not allow for centralized decision making. But, even if we accept that modern technological advances provide organizations with essentially infinite communication capacity (afforded by fiber optic LANs) and infinite data storing capacity, a single decision maker may not be able to process, filter and digest all the available information fast enough in order to come to a decision.

For example, the CEO of McDonald's corporation is not particularly interested to know how many pickles were consumed, or french fries and burgers were sold the night before in the outlet at Central Square in Cambridge, Massachusetts. In fact, the CEO may not be interested even in the total annual sales of the outlet, or even in the total annual sales of all the outlets in Massachusetts. In order to design the company's marketing strategy, the CEO may be more interested in the total sales in New England. On the other hand,
someone needs to know how many pickles were consumed in the Central Square outlet, in order to make sure that no shortage occurs. The decentralized process has to be designed so that the outlet manager knows when new supplies need to ordered, and so that the CEO's desk is not filled with thousands receipts from the company outlets across the world.

Furthermore, human decision makers tend to prefer sharing the burden and responsibility associated with important decisions; factors like risk aversion and regret can severely hinder the centralized decision making process. Finally, the decentralized decision making process is more robust because it is not dependent on the performance and the communication channels of a (single) centralized decision making entity.

The decentralization of the decision making process is therefore necessary. Each decision maker in the organization will be faced with his own personal problem, which the decision maker will have to tackle based on the partial information available to him. Thus, the complex problem of the organization will be decomposed into several smaller coupled and possibly overlapping subproblems.

The objective function of each decision maker is characterized by his own abilities, beliefs, goals and loyalty to the organization; this implies that factors other than the optimization of the performance of the organization will influence the partial decisions made. The decentralized decision making process has to address the issues of the superior decomposition of the organization into smaller decision making entities, in a way that the optimization of each subproblem also optimizes the total goal of the organization as a whole. *The objective of this thesis is to study the effects of the decentralization on the decision making process, to understand its properties, and to obtain, in a systematic way, design guidelines for decentralized organizations.*

---

1 Throughout this thesis, the pronouns 'he', 'his' and 'him' will refer to the term 'decision maker' which is employed in an androgynous form encompassing both sexes.
In order to achieve our goals, several simplifying assumptions were made. First, there exists a clear cut and well defined team\(^2\) objective, which is the minimization of a given cost function. The decision makers are assumed to be \textit{perfect team players}; they cooperate with each other and their only objective is the minimization of the same team cost\(^3\). Moreover, all the probabilistic measures of the model are known a priori by all the decision makers. Each decision maker is allowed a very limited number of alternative actions, for the combinatorial complexity of the problems to be kept under control. Our objective is to analyze decentralized organizations in order to determine the advantages and the disadvantages that they exhibit relative to centralized organizations. In fact, a convenient and desirable property of the problems in our framework is that their centralized counterparts are trivial, so that all the difficulties encountered arise because of the decentralization.

Despite the above simplifying assumptions, the framework of this research is extremely rich and provides for a huge number of interesting examples which will improve our understanding of decentralized decision making and of decision making in general. The architectures of small teams are analyzed in order to study the factors which influence the performance of these architectures. For example, consider two teams which consist of three decision makers, the \textit{tandem} (serial) team (Figure 1.1) and the \textit{two consultant} team (Figure 1.2). We would like to determine which architecture achieves superior performance. Does the superior architecture depend on the quality of the decision makers which constitute the team? Does it depend on the communication protocols of the decision makers? Does it depend on the prior uncertainties and on the team objective? Moreover, given the decision makers and the team architecture, we would like to determine the optimal way of placing

\(^2\) Since in this research all decision makers are assumed to be \textit{perfect team players}, the terms 'organization' and 'team' (of decision makers) can and will be used alternatively.
\(^3\) Thus, game theoretic problems involving different cost functions are not to be addressed in this thesis.
Figure 1.1. Three Decision Makers in Tandem

Figure 1.2. The Two Consultant Architecture

Figure 1.3. Adding a New Decision Maker

the decision makers within the architecture; that is, we would like to determine the optimal configuration for the team.

We also examine the issues involved with the modification of the architecture of a given team. Suppose that a new decision maker is to be added to an existing team in order to improve its performance (Figure 1.3). We would like to study the effects of the addition of the new decision maker. If there exists a choice for the new decision maker, who should be selected? In which position should the new decision maker be assigned? How should
the existing members of the team modify their behavior to incorporate for the new addition? And, how are these modifications affected when the team becomes very large (infinite)?

Throughout this research it will be evident that even small decentralized problems (for which the combinatorial complexity is not a factor) exhibit significant complexity as compared to their centralized counterparts, and sometimes give rise to somewhat counterintuitive results. In order to overcome these difficulties, intuitive and easily implementable (suboptimal) decision rules are presented; bounds are obtained for them and the trade off between complexity and performance is established. To demonstrate this trade off, an example will be presented in which a complex multiple hypothesis testing problem is broken down to a series of simple binary hypothesis testing problems.

In summary, our research goal is to study decentralized decision making and to derive building blocks and design guidelines for an organization, so that its performance be optimized.

1.2. LITERATURE REVIEW

The detection problem was the initial motivating example, and the Bayesian formulation was first considered in [TS81], where the optimality of constant thresholds strategies was established; this was generalized and formalized in [T89a]. It was shown that, contrary to intuition, the optimal decision rules of two identical decision makers in a symmetric parallel team should not in general be identical or even symmetric. Several generalizations of the basic detection model have appeared. In [E82] and [ET82] different architectures of small teams were presented and the optimal decision rules for the decision makers were derived. Also two conjectures about the optimal team architecture were formulated; the first stated that in the serial (tandem) two decision maker architecture, the better decision maker should make the final team decision, and the second stated that the optimal architecture for a team consisting of three decision makers is the two consultant
architecture (Figure 1.3). In [S86] the problem in [TS81] is generalized for any number of
decision makers and for any number of hypotheses. In [CV86] and in [CV88] the same
problem is considered, but a fusion center responsible for the team decision is introduced.
In [HV88] the Neyman-Pearson version of the same problem was presented and in [TP89]
the fusion center was replaced by a decision maker.

The parallel architecture with identical sensors has been analyzed in [R87], [RN87],
[HV88] and in [T88], where asymptotic results were established as the number of the team
members grows to infinity. The Neyman-Pearson formulation of similar problems is
considered in [S86a], [HV86], [R87], [TV87] and [VT88] where different team
architectures are compared; different team architectures, for the Bayesian case, are also
compared in [E82], [RN87a] (numerically) and [PA88] (analytically), where the first
conjecture of [E82] was disproved. In [PA86] the communications between the decision
makers are constrained by a communication cost. In [PT88] a particular class of decision
makers is studied and numerical results were obtained. In [P89] the effect of increased
communication between two decision makers was examined.

Asymptotic results for the parallel team and binary hypothesis testing appeared in
[T86] and were generalized for multiple hypothesis testing in [T88]. Results in the context
of automata and learning with finite memory obtained in [C69], [HC70] and [K75] can be
interpreted to yield asymptotic results for the tandem team. These findings, but for much
smaller teams, are substantiated by extensive numerical results in [R87].

Decentralized sequential decision problems in which the sensors have a choice between
stopping to make the final decision, and receiving more information appear in [TH87],
[T89b], [P86] and [HR89]; the results of [HR89] are not correct since simple examples
contradicting them exist in [T86].

The computational complexity of decentralized problems is examined in [T84], [TA85]
and [PT86]; NP-completeness is established for several classes of such problems.
Furthermore, considerable research effort has been devoted into trying to combine the
normative decision models with descriptive decision models developed by psychologists into more realistic and accurate models of human decision making (for example, [GS66], [SF74], [P85] and [K89]).

Finally, we would like to point out that an excellent and thorough overview of the field was presented in [T89].

1.3. CONTRIBUTIONS OF THESIS

This thesis has two major goals. The first goal is to advance the state of the art of decentralized decision making in a hypothesis testing environment. For this, we tried to address different questions than the ones in the literature, where usually the main objective has been to derive the optimal decision rules for different architectures, under different conditions. Throughout this research, we do not give any emphasis to obtaining the optimal decision rules because we assume that they are known (and indeed they are). We try to address the problems from a higher level, the team designer's point of view; that is, given the optimal decision rules, we try to determine the optimal team architecture, its properties and the factors which influence it.

The second goal is to establish the trade off between the team performance and the complexity of the decision rules. The optimal decision rules are hard to compute; therefore, we suggest some simple suboptimal decision schemes and try to bound the deterioration of the team performance. This is accordance with some of the most important existing theories on human decision making which indicate that human decision makers prefer to make a decision that is good enough, instead of spending much effort in trying to make the optimal decision (for example, the theories of satisfiability and of elimination by aspects).

The long range objective of decision engineering is to improve decision making by accurately modeling it. This research should be a part of a normative theory of decision making which has to be developed eventually. This theory should serve a dual objective; it
should provide decision makers with a tool to evaluate, and hence improve, the quality of their decisions and, it should provide behavioral scientists with a series of hypotheses, which will be investigated with experiments or in practice, and will result in more accurate and realistic models of decision making. In this context, we can state the following contributions of our research:

Chapter 2.

We present an overview of the theories of decision making, in order to discuss the position of the field of decision engineering in the general framework of modeling human decision making.

Chapter 3.

The distributed hypothesis testing model is defined.

Chapter 4.

1. We try to determine whether an optimal team architecture exists. We show that the optimal architecture for teams consisting of three or more decision makers depends on the particular decision makers involved, on the configuration of the decision makers in a given architecture, on the communication that the decision makers are allowed to exchange and on variables which are external to the team, like the prior probabilities of the hypotheses and the detection costs. Therefore, there are no generalizations that can be made about the optimal team architecture.

2. The two conjectures of [E82], which had been often supported with numerical results (for example, in [R87]), were disproved. The first claimed that if a tandem team consists of two decision makers, one better than the other, then the final team decision should be made by the better decision maker. The second conjecture claimed that the optimal architecture for any team which consists of three decision makers is the two consultant architecture (Figure 1.2). We obtain simple examples which disprove the conjectures by developing a
procedure to analyze these types of problems. Thus, it should be clear that *counterintuitive* results exist in this framework.

3. If the better decision maker is restricted to be the primary decision maker in the two decision maker tandem team, the percentage deterioration of the team performance is unbounded.

4. We analyze the team which consists of two decision makers in parallel and obtain the second order optimality conditions. We pay particular attention to the case of identical decision makers and obtain tight bounds on the deterioration of the team performance if both decision makers are restricted to employing identical decision rules.

5. The main contribution of this chapter is the conclusion that, in general, the optimal architectures of teams which consist of two or more decision makers depend not only on the characteristics of the team itself, but also on the environment in which the team operates.

**Chapter 5.**

1. The team which consists of an infinite number of identical decision makers in tandem is analyzed. Conditions on the individual decision maker for the team probability of error to asymptotically go to zero are obtained.

2. A simple 'selfish' suboptimal decision scheme is presented and conditions on the individual decision maker are obtained for the team probability of error to asymptotically go to zero. These conditions suggest that the suboptimal decision scheme be not always asymptotically optimal.

3. A second suboptimal decision scheme under which all decision makers (except the first one) are restricted to employing identical decision rules is presented and again conditions on the individual decision maker are obtained for the team probability of error to asymptotically go to zero. This decision scheme is shown to be asymptotically optimal.
4. Numerical results which compare the decision schemes are obtained. The performances of the schemes suggest that it could be worthwhile to sacrifice optimality for the benefit of the reduced computational complexity.

5. In spite of the mathematical proofs, the main contribution of this chapter is that it exposes the inefficiencies of the tandem architecture and that it demonstrates that the suboptimal decision scheme, as compared to the optimal one, performs very well.

Chapter 6.

We demonstrate that the Receiver Operating Characteristic curve of any team does not have to be concave, unless additional conditions on the coordination between decision makers are imposed. We show that, while this result does not affect the Bayesian formulation of the distributed hypothesis testing problems, it is important in the Neyman-Pearson formulation of the same problems.

Chapter 7.

In order to further demonstrate the trade off between team performance and computational complexity, we suggest a suboptimal satisficing 'intuitive' scheme. This takes advantage of the particular structure of the problem in order to break down a complex multiple hypothesis testing problem into a series of simpler problems.

1.4. OUTLINE OF THESIS

In this section we present an overview of the thesis and try to establish the unity and continuity that exists between the various subsequent chapters.

In chapter 2, we present some thoughts on modeling decision making, and in particular human decision making, from an engineering point of view. We discuss the position of decision engineering relative to the traditional theories of decision making. We propose the development of a normative decision theory, which will be the first step to the
development of more realistic and accurate normative/prescriptive models. We present the reasons which make essential the analysis and understanding of decentralized decision making. This chapter provides the motivation for the rest of the research; we also discuss some motivating examples for the particular problems which will be analyzed in the sequel.

In chapter 3, we introduce the main distributed hypothesis testing model. For the case of binary hypothesis testing, we extensively discuss the Receiver Operating Characteristic (ROC) curve which provides a complete description of the decision makers. We present and analyze the properties of the ROC curve which are going to be explored in the subsequent proofs.

In chapter 4, we begin to analyze the simplest possible teams; we discuss small team problems, that is problems of teams which consist of two or three decision makers. We compare the two possible architectures for the teams which consist of two decision makers, the tandem architecture and the parallel architecture, and show that the tandem is the superior (dominant) architecture. We disprove the conjecture claiming that the better decision maker should make the team decision in the tandem architecture. We further analyze the parallel architecture and obtain the second order conditions for optimality; we pay particular attention to the case where both decision makers are identical and obtain tight bounds on the deterioration of the team performance if both decision makers are restricted to employing identical decision rules. Finally, we compare the architectures of the teams which consist of three decision makers and demonstrate that the optimal architecture depends not only on the particular characteristics of the team, but also on the environment in which the team operates. Therefore, we disprove the conjecture claiming that the two consultant architecture is the superior architecture.

In chapter 5, we obtain asymptotic results for the tandem team. We obtain conditions on the individual decision maker for the team probability of error to asymptotically go to zero as the number of decision makers in the team grows to infinity. We introduce two easily implementable suboptimal decision schemes and again obtain conditions for the team
probability of error to go to zero. We also perform numerical studies to compare the performance of the decision schemes.

In chapter 6, we discuss the effects of deterministic and randomized decision rules on the performance of the team. We show that, unlike the Receiver Operating Characteristic curve of the individual decision maker, the team Receiver Operating Characteristic curve does not have to be concave; it can be made concave if decision rules with dependent randomization can be employed by the decision makers. We analyze the consequences of these result on the Bayesian and the Neyman-Pearson formulation of the problems.

Throughout this research, it becomes clear that the 'search' for optimality requires great analytical and computational power. The inherent complexity of these problems is such that even small instances of them are very hard to solve optimally. Moreover, no generalizations can be made, since the optimal architecture of organizations with three or more decision makers depends also on the environment. Thus, it is worthwhile to design computationally simple suboptimal decision schemes which take into consideration the particular characteristics of the environment in which the team operates. In order to demonstrate this approach, in chapter 7, we present a complex multiple hypothesis testing problem and discuss the issues involved with the generation of a good and simple suboptimal solution for it.

Finally, in chapter 8, we summarize the results and conclusions of this thesis. We demonstrate how these results can be incorporated in the general field of decision engineering, and suggest some directions for future research.
CHAPTER 2

An Engineering Approach to Modeling
A Class of Decision Making Problems

2.1. INTRODUCTION

In this chapter we present a quick and by no means complete overview of the theories of decision making. We emphasize the parts of the theories which are related to this thesis. We try to show where and how the relatively new field of decision engineering fits within the general framework of decision theories and we propose the guidelines for a normative decision theory which when developed will provide a powerful tool for improving the decision making process. We also discuss the reasons for selecting the decentralized hypothesis testing framework for this research and present the motivation for our specific problems. For a more complete overview of the decision making theories the reader is referred to a paper by Einhorn and Hogarth [EH81] and to two books, one by Kahneman, Slovic and Tversky [KS82], and one by Raiffa [R68].

2.2. MODELING DECISION MAKING

Decision making research has been the meeting point of psychologists, philosophers, sociologists, economists, organizational theorists, statisticians and, more recently, of operational researchers and engineers. They all work together trying to develop a theory which will improve the quality of decision making both of individuals and of organizations. Gaining a better understanding of how good decisions are made is an essential prerequisite to the goal of increased productivity and improved allocation of resources.
There are two distinct points of view in modeling decision making, which are reflected in the two schools of thought which have been developed: the organizational decision theory and the behavioral decision theory. In spite of our familiarity with developments in both theories, we do not claim to be experts in either field; thus we will use the words of Ungson and Braunstein [UB82, p. 89] to compare the two:

"Behavioral decision theory is essentially cognitive and generally uses experimental methods; organizational decision theory is primarily theoretical and naturally oriented, examining rather conspicuous individual and social phenomena.... The decision maker, as seen by some behavioral theorists, is portrayed as constantly deviating from normative models provided by statistical decision theory. In contrast, organizational decision theory pictures the decision maker purposefully participating in a messy world of bounded rationality, intentional nondecisions, convenient inconsistencies, and legitimated biases."

Behavioral decision theorists gather large amounts of data, analyze them and construct a model which fits the data. The models usually are not very sophisticated from a mathematical perspective. There are many difficulties with and shortcomings to this approach. Collecting data on real life decisions through questionnaires or interviews often leads to biased results. Moreover, designing laboratory experiments with humans does not yield reliable results as people tend to behave differently in the lab. Connolly [UB82, p. 47] states that real humans in their real ecologies tend to act more and err less than do as laboratory subjects. Humans tend to overestimate the probabilities of unlikely events and to underestimate the probabilities of very likely events. They tend to ignore their a priori information, to ignore sample sizes, to simplify their choices and to bias their decision by their familiarity with the questions being asked. These biases are also encountered by the 'bayesian estimate' models (for example, [P85]). These show that the updates in beliefs performed by human decision makers are monotonically related to the updates predicted by Bayes's rule, but they tend to be more heavily influenced by their prior information.
Tversky and Kahneman [TK80] state that humans tend to overestimate the ability of decision models to predict behavior; when faced with biases they prefer to search for plausible explanations of their behavior, rather than revise their models.

Even careful screening of the data may lead scientists to inaccurate models because of what engineers call 'unmodeled dynamics'; that is, overlooking reasons which make humans behave according to a particular pattern and attributing the behavior to the wrong stimulus. Also there exist certain parameters which can not be quantified and thus are lost in the analysis.

Tversky [T72] formulated the elimination by aspects theory. This states that humans make decisions by playing complicated versions of the twenty questions game. They introduce one desired criterion at the time and they eliminate all the options which do not meet the criterion. This procedure greatly simplifies the decision process, but unfortunately its outcome can be considerably far from the optimal. Tversky and Kahneman [TK86] also showed that decision making depends on the way on which the alternatives are presented to the decision makers. In fact, Kahneman's results state that decision makers do not always behave rationally, because their utility and objective functions are not always well behaved.

Organizational decision theorists tend to start with general models and then either move towards simplicity, in which case much of the phenomenon of interest is lost, or move towards accuracy, in which case the models become too complicated to follow and appreciate. They tend to view decision makers as agents with rationality, that is as having nice utility functions. This means that decision makers are well behaved and try to optimize their decisions. Edwards [UB82, p. 320] stated that the "dirty word rationality" implies orderliness and good sense. But Herbert Simon [S76] has stated that human decision makers rarely behave in this way because people do not have the "wits to maximize." In fact several studies exist in the literature which demonstrate that human decision makers do not always behave rationally (in spite of their good intentions) and that they have strange utility functions (for example, Ellsberg's paradox [E61]). Humans can be faced with the
exact same question, under the same circumstances and still make different decisions; an example of this is an intelligent rational human playing chess against a computer and employing different opening moves in every game. This brings us to a very interesting and widely accepted theory of Simon; the theory of *satisficing* [S76].

This theory claims that human decision makers do not search for the optimal decision, but rather for a decision which is 'good enough.' Simon sees humans as agents of *bounded or limited rationality*. Connolly [UB82] states that the bounds of rationality depend on several factors like the expertise of the people involved, the importance of the intellectual task and the availability of relevant tools. People will be satisfied with a decision which yields even a small improvement compared to the present and will try to avoid uncertainty. They will usually settle for a smaller expected pay-off if this implies a smaller variance. People are happy to look at a drastically simplified model of the world and to examine just two alternatives, if these meet some minimal requirements: the same course and one new course. In fact, the philosophy of this theory is quite similar to the philosophy of the *elimination by aspects* theory discussed above.

In this overview we tried to demonstrate that the normative tools of utility theory and probability theory, although extremely useful, cannot alone describe a theory of decision making. Despite Tversky and Kahneman's [TK86] warning that "the dream for constructing a theory that is acceptable both descriptively and normatively appears unrealizable", the field of decision engineering according to March and Shapira [UB82, p. 113] is trying "to bring together how people act and how we believe they ought to act." This should be the area of interaction between behavioral decision theory and organizational decision theory. For behavioral decision theorists decision engineering is a tool which assists decision makers in avoiding errors and consequently in modifying their behavior; for organizational decision theorists it represents the use of artificial intelligence.

Decision engineering should not try to replace knowledge and experience; rather, it should try to complement them by analyzing both errors and correct decisions. Simon
[S81] states that "... accumulation of experience may allow people to behave in ways that are very nearly optimal in situations to which their experience is pertinent, but will be of little help when genuinely novel situations are presented." In our opinion, decision engineering should have two objectives: first, to assist decision makers by presenting to them, if not the optimum decision, at least an intelligent decision and second, to indicate where decision makers repeatedly make errors, especially if their decisions are believed to be 'common sense.' In order to achieve these goals a systematic methodology has to be developed. According to Athans [A87]:

"... technology approaches all of the above problems in a completely intuitive and qualitative way. As a consequence, there does not exist even a methodology that can be used to understand in a precise manner the complex cause-and-effect relationships inherent in a (decision) process and to describe them using a minimal set of primitives, measures of performance, and measures of effectiveness."

Developing such a methodology should be a priority of decision engineering. Levis and his students (for example, [BL82], [LB83], [L88] and [L88a]) have been working towards building a general framework in which the realization of this goal may be possible.

2.3. THE SEARCH FOR THE ILLUSIVE EXPERT

In the previous section, we loosely described decision engineering, looked at its place among decision theories and presented one of its most immediate goals. In this section, we try to complete the picture by presenting the need for the development of new normative decision theories, discuss the key concept of the expert decision maker and propose guidelines to a normative decision theory which could fill the existing gap that exists in the field of decision engineering. Finally, we discuss the relationship of this research with the proposed framework.
The answers to some very important questions which organizations constantly face should be the focus of research. What human and machine limitations cause bad decisions and how can they be remedied? How can we ensure that good and relevant information is appropriately incorporated on the bearing of decisions? How can irrelevant information be detected and filtered out? How should communication flow within an organization and how should it be presented and stored in order to maximize its effectiveness? How do inconsistencies between decision makers influence individual and team decisions? How are the decisions of individuals affected when they become members of a team? How are their aspirations, beliefs and skills affected? How can a decision maker be improved in order to perform his assigned task more successfully? How should an organization grow and how do vertical and parallel architectures compare? How should an organization deal with the increased workload of one of its members? How should an organization be designed in order to be resilient in case of failure of a decision maker or of a communication link? How should an organization be designed to effectively face a competitor or an enemy?

In order to answer these and other related questions we have to develop new theories. Theories are developed painstakingly slowly, but then again Rome was not built in one day. We are only in the beginning, but immediate attention should be given to this end. As Einhorn and Hogarth [EH86, p. 25] stated:

"We believe it is time to move beyond the tidy experiments and axiomatizations built upon the explicit lottery. The real world involves ambiguous probabilities and utilities, context and framing effects, 'illusions of control', and superstitions. Given the richness of the phenomena before us, our biggest risk would be to ignore them."

In the same time, we need to develop ways to extrapolate from the results of our models to real world situations. Since there exist great varieties of individuals and organizations, the theory should not depend on specific forms of the decision making agents; rather it should provide ways to aggregate the results. As an individual becomes a
member of a team he adapts and changes his personal objectives in order to adjust to the requirements and goals of the organization. The team members' decisions are also influenced by the information and communication network within the team. Decision makers can and should behave very differently as team members than as individuals. The theory should employ the human intelligence and ingenuity, along with the computing power, to incorporate all of the above aspects of decision making.

We need to build models which are simple, but allow the study of particular aspects of decision making and can be augmented to study the interrelation of more than one of these aspects. It is important that these models separate issues of optimal decision making from issues of biases and beliefs. Simon [S57, p. 198] claims that decision theorists should "... construct a simplified model of the real situation in order to deal with it. (The decision maker) behaves rationally with respect to this model, and such behavior is not even approximately optimal with respect to the real world. To predict his behavior we must understand the way in which this simplified model is constructed." These normative models are very useful for providing specific decision making paradigms. Then, behavioral decision theorists can design experiments based on the paradigms in order to determine exactly how the decisions of intelligent human decision makers differ from the decision of the normative models, and in order to subsequently construct descriptive models. The next step will be to combine the normative and descriptive models into normative/descriptive models which will make use of the findings of empirical research in order to construct better and more robust mathematical models of decision making.

From our everyday life experience, it is very well known that when people or organizations are faced with important decisions or when they try to analyze effects of decisions that were made, they resort to experts. It does not matter if it is the President of the United States who turns to a State Department thinktank to assist him in the design of the foreign policy for the Middle East, or if it is NASA that turns to a committee of scientists to determine the reasons for the Challenger disaster, or a couple who turns to a
counsellor in order to find a solution to its marital problems; they all try to gain a deeper understanding of the causes and the effects of their decisions, in order to improve the quality of their decision making process. Therefore, it would only seem natural that any decision theory, which tries to improve the quality of decision making, deals extensively with the concept of the expert, in particular, the theory has to quantify the properties that characterize the expert, to determine the process for the development of a decision maker into an expert, and to address the issues involved with the interaction among experts.

We are now ready to present a characterization of the normative decision theory for teams which needs to be developed. We believe that it adequately summarizes our discussion and motivates the research which has to be done for the above questions to be answered. Paraphrasing Athans [A81]:

The normative team decision theory is a theory should stipulate the following two processes:

1. The development of a decision maker into an expert decision maker and,
2. The development of a team of expert decision makers into an expert team of expert decision makers.

As was stated above, this characterization immediately raises a very important point. Terms like expert decision makers, expert teams and expert systems are frequently encountered in the literature. But what do we really mean by expert? In the following discussion we present our meaning of the expert decision maker; the notion can be generalized to include other disciplines like teams or systems.

We believe that for the purpose of modeling decision making the word expert has two disjoint, but complementary meanings associated with it. The first meaning is a relative meaning which is very close to the everyday use of the word. When there are several decision makers to perform a certain task, the decision maker(s) who can best perform the task is (are) considered to be expert(s) for that particular task. The concept of the expert is
obviously relative to the field of decision makers. For example, the players in the MIT Varsity Football team are experts in playing football when viewed as members of the MIT student community, but these same players are definitely not experts in playing football when viewed as members of the NCAA football players.

The second meaning that we associate with the word expert is absolute. We consider a decision maker who has to perform a certain task to be an expert in the absolute sense if he performs the task to the best of his ability; that is, we consider a decision maker to be an expert in this sense if he performs at his potential. Every organization has many decision makers who are inherently different and who cannot perform the same tasks with the same degree of success. Simon [S81, p. 36] asserts that "what a person cannot do he will not do, no matter how hard he wants to do it." Thus, an organization can not and should not expect a decision maker to perform over and above his limited abilities; rather, it must be totally satisfied by having him perform the task at his level best. Then, a decision maker is considered to be an expert among a given group of decision makers in performing a given task, if he is considered to be an expert in both the relative and the absolute sense.

The objective of the organizations in this theory should be the minimization of errors and not the elimination of errors. Eliminating errors could backfire because only by making errors does one learn in real life. According to Einhorn [E85]: "The acceptance of error to make less error is likely to be a safer and more accurate strategy over a wide range of practical situations." The acceptance of error as a means of learning from it raises issues of training for the decision makers. An organization can improve the performance of its members through training, but even then their limited abilities (i.e., bounded rationality) constrain them to a 'saturation' point beyond which no improvement can be achieved. Training increases the potential of decision makers by developing their familiarity with particular situations. It can increase the ability of an individual to perform a task and it can increase the ability of an individual to communicate and cooperate with others. Hence, it can increase the level of expertise of a decision maker in both the relative and the absolute
sense. Therefore, an important part of the theory, which is not addressed in this thesis, has to be devoted to the modeling and the analysis of the training process.

The normative decision theory should first address the issues of developing decision makers into experts in the absolute sense (i.e., performing at their potential). Then it should address the issues of organizing, given certain exogenous constraints, the decision makers as the best possible (expert) team; thus developing the decision makers to experts in the relative sense with respect to the tasks which need to be performed within the team. The exogenous constraints could and should not only be quantitative constraints (i.e., constraints on the number and the quality of the decision makers, on the available communication links and technology), but also constraints which take the form of issues of resiliency in cases of failures of decision makers and in cases of 'enemy' (adverse) interference.

Both issues to be addressed by the normative decision theory are extremely difficult and complex, but it seems to us that the second (i.e., team organization) is even more difficult than the first because of the large number of different ways in which an organization can be set up and hence of the explosive combinatorial complexity that these problems exhibit. Each domain of the normative decision theory can be researched independently, but only development in both will bring together a complete theory which will in turn result in an optimal organization, that is a true expert team of experts.

This thesis is a modest attempt in the second goal of the proposed theory: the optimal design of an organization. We assume that the decision makers are experts in the absolute sense, that they are perfect team players and try to develop them into an expert team. The main characteristic that distinguishes, and in fact makes more difficult, the decision making process of an organization from that of an individual is decentralization. We study elementary problems of decentralization which need to be analyzed, solved and understood if a normative theory is ever to be developed. We try to construct organizational 'building blocks' to be used in the design of larger organizations. We also present these simple
problems to clearly demonstrate that, despite the restrictive and ideal assumptions of the decision makers being perfect team players and performing at their potential, there exist several counterintuitive results and many difficulties in the development of the theory. Some of the problems examined are directly related to 'real life' situations; for this, we propose the following two examples:

Example 2.1. Consider a nuclear powerplant and suppose that there exist two identical devices for detecting a possible radioactive leak, which have been installed in identical places in the powerplant. As soon as one of the devices detects something, an alarm sounds and the nuclear reactors are shut off immediately as a precaution. The plant is required to operate at a probability of detection of a radioactive leak of at least $1 - \alpha$ (for some small $\alpha > 0$). The decision rules of the two devices should be designed so that the detection requirement is met, but with the smallest possible probability of false alarm because shutting off the nuclear reactors is very costly.

Because of the perfect symmetry of the problem, it seems intuitive that both devices should employ identical optimal decision rules and hence operate at the same level of probability of detection, namely $1 - (\alpha)^{0.5}$. But, it has been shown [T88] that this is not always the case (and in section 4.4 conditions for this will be derived), which implies that in general it is optimal for the detection 'team' to have the devices employ distinct and asymmetric decision rules.

Suppose that these detection devices operate optimally and that they are examined by decision scientists. When the scientists realize that the devices employ different decision rules, they may conclude that either the devices are not employing the optimal decision rules, or that the devices are indeed employing the optimal decision rules, but are not identical because of small differences in their materials or construction. Thus, the scientists may attribute the devices' nonidentical operation to the wrong reasons and they might not
consider the strongly counterintuitive possibility that the devices are not only identical, but are also employing different, yet optimal, decision rules.

The normative decision theory will assist decision scientists by indicating that such counterintuitive results may occur; thus it should change their perceptions towards 'intuitively optimal' behavior and consequently influence their decision theories.

Example 2.2. Consider Boston's Logan International Airport. Logan airport has two air-traffic corridors for take offs and for landings; depending on the weather conditions, one of the two corridors is selected. As soon as the planes land, they need to be taxied to the gate. Similarly, they need to be taxied from the gate to the take off corridor. In this example, we deal with the airport ground control. Logan airport has five different terminals and there are many roads for the planes to be taxied to and from. There exist several crossroads (danger points) which need to be checked, and almost always there exists too much traffic for a single controller to alone be able to perform the job safely and efficiently. Therefore, more than one ground controllers are on service at the same time; in fact, there are usually three or four.

It should be clear that the team of ground controllers has a common objective, the safe and efficient taxiing of the incoming and outgoing planes, and that their jobs are highly dependent. Suppose that the shift consists of three different controllers. It is very interesting to find the best possible 'team architecture' for it. Should all three controllers work in parallel? Should two of them work in parallel and one act as the supervisor? Or should there be two supervisors who assist the other controller when he can not handle the job alone? We should note here that authorities of Logan airport indeed had such a case study performed a few years ago.

And even when the optimal architecture for the team of the ground controllers has been determined, other interesting questions arise. How should two controllers working in parallel share the job: should each one be responsible for specific roads (zone) or should he
be responsible for specific planes (man-to-plane)? What should each controller communicate to his supervisor and vice versa? Is the optimal architecture dependent on the number of active airplanes (work load)? Is the optimal architecture dependent on the particular controllers involved? Clearly, the optimal team architecture needs to be robust with respect to the controllers in service, because it does not seem feasible and intelligent to modify it every time the controllers' shift changes.

All these questions will be addressed in section 4.5 for teams consisting of up to three decision makers and it will be shown that the truly optimal architecture depends not only on the particular characteristics of the team, but also on the environment in which the team operates. This suggests that we may often have to abolish the search of optimality, in favor of some simple but efficient suboptimal solution; a conclusion which is in accordance with the theory of satisficing.

In this thesis we employ a decentralized hypothesis testing framework to study examples like the ones described in order to move towards the primary goal, the optimal normative design of an organization.

2.4. THE DECENTRALIZED HYPOTHESIS TESTING FRAMEWORK

There are two major reasons for selecting the decentralized hypothesis testing framework for this research. First, problems in this framework are very simple to describe, so that decision scientists without great mathematical sophistication can understand them and incorporate the conclusions into their own research. In fact, we should emphasize that problems in this framework look deceptively simple. We began our research by trying to formally prove the most basic and 'obvious' results which existed in the literature as conjectures and we were surprised by the difficulties which we encountered, their inherent complexity and the counterexamples which we derived. Moreover, since the centralized
counterparts of the problems in this framework can be trivially solved, it is evident that all the difficulties arise because of the decentralization. Thus as intended, our research will focus on the effects of the decentralization on the performance of an organization.

The second major reason for employing the decentralized hypothesis testing framework for the development of a normative decision theory is that decentralized hypothesis testing is in itself a very interesting subject with applications. As Lt. General James A. Abrahamson stated [N86] some of SDI's immediate goals are command, control and communications concepts on architecture and design of the National Test Bed, target discrimination, space surveillance and high endoatmospheric detection. Although the complete normative decision theory will have to be developed for the first of these goals to be realized, a decentralized hypothesis testing theory will be sufficient for the realization of the others. As another C$^3$ expert wrote [A87]: "We need novel theoretical and algorithmic advances in the distributed versions of multiple hypothesis testing …"

Before beginning the discussion of our model we would like to point out that problems of this type have been shown to be NP-complete [TA85], [PT86]. These results identify the combinatorial difficulties that exist. But, like in the case of the Travelling Salesman Problem (TSP), these problems have so many important practical applications so that merely identifying the difficulties associated with them is not enough. We have to develop new mathematical techniques to solve them or to at least obtain satisfactory heuristics. As Simon suggested [S79]: "… because of the complexity of the environment, one has but two alternatives: to build optimal models by making simplifying assumptions or to build heuristic models that maintain greater environmental realism."

The purpose of this thesis is not to demonstrate the difficulties which arise because of the combinatorial explosion; the NP-completeness is a testament to these. In most well known NP-complete problems the difficulties arise only because of the combinatorial explosion. For example, it is trivial to optimally solve the TSP for up to four or five nodes.
But, our problems are different because difficulties arise even in the simplest versions. Hence except in chapter 7, in order to keep the combinatorial explosion under control we will test only a small number of hypotheses (two or three), employ a small number of decision makers (up to three) and use limited communications, so that we concentrate on the difficulties which arise because of the inherent complexity of optimizing the decision rules. Only by understanding these difficulties and overcoming them, we will be able to make educated generalizations in order to build good heuristics and eventually achieve a truly optimal solution.

Many problems have been studied in a framework similar to the one we employ (for example, [TS81], [E82], [T88]) and many results have been obtained and conjectures (generalizations) drawn. We consider problems in which the decision makers are given and the optimum architecture for the organization satisfying certain requirements is requested. The environment consists of discrete hypotheses which occur with prior probabilities known to all the decision makers. The decision makers receive uncertain observations of the environment. Then preliminary decisions are made and communications take place until a final team decision is reached. The team incurs a cost which depends on the final team decision and on the true hypothesis. The costs are known a priori by all the team members; thus, our problems are problems of perfect rational team training.

As already stated, we do not deal with problems of developing decision makers into experts in the absolute sense. We assume throughout that the decision makers perform at their potential which is fixed and cannot be improved. Following Simon's advice we make great simplifications. In this context the decision makers can be seen as 'perfect decision robots.' That is, they do not experience conflicts, defensive avoidance, regrets, hopes, aspirations, emotions, coercion, confrontations or any of the conditions which make modeling human decision making so difficult. The decision makers behave like computer processors who follow a code and strive to make decision to minimize the expected team cost. The negative of this expected team cost can be seen as the global team utility function.
This cost/utility function describes completely the whole persona of the decision makers since this is the only factor that influences their decision.

Since we have to start building from the ground up, we study problems for teams consisting of two decision makers, the simplest possible teams. The first such problem is to compare the parallel and the tandem architectures for the two decision maker team. We will formally show that, as expected, the tandem architecture is superior in this case as it achieves greater centralization. This is the first attempt to deal with one of the most important problems in organizational design: parallel (horizontal) versus tandem (vertical) diversification. It is known from experience that for larger organizations a combination of the two is optimal. A strictly tandem architecture will result in decisions made down the command chain to hardly influence the final team decision and thus the potential of decision makers not to be fully utilized; moreover, it is not a fault-tolerant architecture because a failure in one of the links will result in the loss of the potential of all the decision makers below it. On the other hand, in a strictly parallel architecture team members will not be able to assist other team members and no degree of centralization will be achieved; moreover this organization will not be able to carry out one of its main goals which is supervision and control, and many conflicts and indecisions will arise because no decision maker will be able to exercise decisiveness and authority (according to a Greek proverb: "In places where many roosters crow, the dawn comes late", meaning that when too many people have opinions the job never gets done).

We then consider more closely the two decision maker tandem team. We assume that communications between the two decision makers are constrained; otherwise we would have a known centralized problem. One decision maker makes a temporary decision and transmits it to the other who then makes the final team decision. It is conjectured ([E82] and [R87]) in the literature that if one decision maker is better than (i.e., an expert relative to) the other, then the better decision maker should make the final team decision. If this were to be true it would indicate that whenever a decision maker has to consult an expert on a case,
his optimal action would be to present to the expert his own assessment of the situation and let the expert make the decision. Unfortunately, we found some special cases in which this result does not hold true, but we still try to analyze it as it should prove useful to building heuristics.

The next logical problem to examine was discussed in Example 2.1 above. The team consists of two identical decision makers in parallel, a perfectly symmetric problem and we try to derive conditions for the optimal decision rules of both decision makers to be the same or in other words for the risk to be split evenly between the two. In our effort to construct heuristics we study the effects on the team cost of constraining the decision makers to employ identical decision rules. A tolerable increase in cost would greatly simplify things because it would imply that we can constrain similar decision makers on the same organizational level to employ the same decision rules without severe degradation in team performance; the obvious trade off between complexity and performance and an application of the satisficing theory. We should note here that it has been shown in [T88] that as the number of the identical decision makers grows to infinity, it is optimal for all the decision makers to employ the same decision rule. Therefore the heuristic decision rule just proposed besides being simple and intuitive, it has the extra advantage of being asymptotically optimal.

The following important question arises: when should a new decision maker be added to an organization in parallel and when in tandem? To study this we add a third decision maker to the team and study the different architectures. It was conjectured in the literature that the best architecture for three decision makers was the 'A' shaped architecture in which a particular decision maker is designated to make the final team decision after he has been consulted by the other two. We show that this is not always correct as in some cases the tandem architecture can be optimal. We try to analyze the situations in order to 'update' the conjecture so that it can be used in the development of heuristics.
In this spirit, we also examine the infinite tandem team and obtain both the optimal and suboptimal, but asymptotically optimal decision rules. We want to study the effects of the addition of a new decision maker to the existing team; in particular, we want to examine how the existing decision makers of the team modify their behavior to account for the addition of the new decision maker. The simple structure of the tandem team makes this possible. Moreover, by comparing the performance of the optimal and of the suboptimal decision rules, we try to establish the trade-off between the complexity of the decision rules and the performance of the team.

We then examine the effects of employing randomized decision rules as opposed to deterministic decision rules. We also analyze the properties that characterize individual decision makers and try to establish whether they also characterize teams of decision makers; our objective is to obtain convenient descriptions for these teams in order to generalize our results for larger organizations.

Our next goal is to demonstrate that the computational complexity of these problems considerably increases when we move from binary to multiple hypothesis testing problems. For this, we employ a team which consists of a single supervisor and a number of consultants. Not only does the number of decision rules grow, but also they become increasingly more difficult to compute. This suggests that the way to proceed is to brake the complex hypothesis testing problem down to smaller (possible binary) hypothesis testing problems. But decision theorists like March and Shapiro [UB82, p. 109] claim that this is indeed how humans make decisions: "Organizational theory has emphasized a small number of theoretically crucial simplifications. These include a simplification of preferences to two-value functions ..." In fact expert detection systems, not to mention computers, operate by making binary decisions. They first need to decide whether a target is present or not; then, in case a target has been detected, they need to decide whether it is large or small, quick or slow in order to determine the target's type. Thus, research should focus on how
to effectively break down complex multiple hypothesis problems into a series of simpler binary problems.

These are only a few examples which attempt to demonstrate the type of questions that need to be asked, the problems which need to be formulated and the analysis which needs to be performed in order to be able to draw conclusions and make useful generalizations. There still remains a vast number of issues which have to be addressed and we believe that the proposed framework is an adequate testbed. Only by proceeding step by step in this context, a complete normative decision theory will be built and the dreams of an expert surveillance system and of optimal decision making will eventually be realized.

2.5. SUMMARY

We presented a brief overview of the existing theories of decision making to demonstrate that the need for a normative theory in the field of decision engineering emanates from them. We also presented guidelines for this theory, discussed the concept of the expert decision maker and explained the reasons for selecting the distributed hypothesis testing model which we employ in this research, together with the motivation for the specific problems we analyzed.
CHAPTER 3

The Distributed Hypothesis Testing Model

3.1. INTRODUCTION

In this chapter we define the model that is going to be employed and present several facts which will facilitate our subsequent analysis. A team consists of decision makers (DMs) who receive noisy observations of the environment, make decisions and communicate messages according to specified protocols. The objective of the team is to perform hypothesis testing and minimize a given Bayesian cost function which takes into account different costs for hypothesis misclassification.

In the following section, we present the main variables of our model and introduce the notation. In section 3.3, we define the Receiver Operating Characteristic (ROC) curve and examine its properties. In section 3.4, we present some special cases of ROC curves. Finally in section 3.5, we analyze the very important conditional independence assumption.

3.2. THE GENERAL FRAMEWORK

3.2.1. Hypotheses and Prior Probabilities

We assume that there are $M$ distinct hypotheses $H_0, ..., H_{M-1}$. Each hypothesis $H_m$ occurs with a positive prior probability $p_m = P(H_m), m = 0, ..., M-1$; thus, the true state of the environment can be viewed as a discrete random variable $H$ which is realized by $H_m$ with probability $p_m$. 
3.2.2. The Decision Makers

Each team consists of $N$ DMs. Each DM $n$ receives an observation $y_n$ which is a random variable in a set $Y_n$. The distribution of $(y_1, ..., y_N)$ conditioned on $H_m$ is known by all DMs or by the system designer, for all $m$.

DMs communicate between them according to communication protocols which have been prespecified by the designer of the team. We call all the DMs which communicate a message to DM $n$ the immediate predecessors of DM $n$, and, the DMs to which DM $n$ communicates his message (if any exist) the immediate successors of DM $n$. Each DM $n$, upon receiving his own measurement $y_n$ and the messages from his immediate predecessors, makes his decision and communicates a message to his immediate successors if any exist; otherwise he declares the hypothesis he considers to be true. The function which maps the observation of DM $n$ and the messages received by DM $n$ to a prespecified number of bit decision is called the decision rule of DM $n$; this function can be either deterministic or randomized. Formally, let $u_n$ denote the decision of DM $n$ which takes values in a set \{1, 2, ..., $D$\} and let $PD(n)$ denote the number of the immediate predecessors of DM $n$, $n = 1, ..., N$. Then, the decision rule $u_n$ of DM $n$ is a function $\gamma_n : Y_n \times \{1, 2, ..., D\}^{PD(n)} \rightarrow \{1, 2, ..., D\}$. Let $\gamma = (\gamma_1, ..., \gamma_N)$. Moreover, let $\Gamma_i$ be the set of all possible decision rules of DM $n$ and $\Gamma = \Gamma_1 \times ... \times \Gamma_N$.

3.2.3. The Fusion Center

A fusion center is a DM who receives messages from his immediate predecessors, but no observation of his own, and then has to make a decision based upon them. Thus, the fusion center can be viewed as a DM with an extremely unreliable observation; therefore, it does not have to be treated in a different context.
3.2.4. The Communication Messages

As was mentioned above, every DM communicates a message which consists of a given number of bits. In certain problems a particular interpretation of the messages can exist; for example, in the case of $M$-ary hypothesis testing and $M$-ary messages each message $m$ could indicate that the corresponding hypothesis $H_m$ is true. But, we do not a priori assign to them any interpretations; we provide each DM with his specified number of communication bits and then allow the DMs to select the optimal decision rules and hence use the messages as efficiently and effectively as possible.

REMARK. In a more general formulation, we could allow each $u_n$ to take values in a set $\{1, 2, ..., D_n\}$, where the $D_n$s are different positive integers.

3.2.5. The Cost Structure

We assume throughout that the objective of the team is the minimization of the expectation of a cost function. We are given a cost function $J: D^N \times \{H_0, ..., H_{M-1}\} \rightarrow R$, with $J(u_1, ..., u_N, H_m)$ being the cost of deciding $u_1, ..., u_N$ when $H_m$ is the true hypothesis. The problem is thus to find the optimal strategy $\gamma \in \Gamma$ which minimizes $J(\gamma) = E[J(u_1, ..., u_N, H)]$ where $E(\cdot)$ denotes the expectation over all the relevant random variables. Therefore, the cost function is a Bayesian cost function which assigns different costs for hypothesis misclassifications.

In some problem formulations, DMs have to incur a communication cost in order to receive messages from other DMs. This cost gets included in the minimization by proper augmentation of the cost function.

We make special notes for the cases where the Neyman-Pearson formulation is also considered; see chapter 6.
3.2.6. The Normalized Probability of Error

In most cases in our research the cost function depends explicitly on a single team decision, namely \(u_t\), which implies that the whole team decision process is summarized into \(u_t\). A particular example of this is the minimum probability of error cost function, which assigns a cost of one to an erroneous team decision and no cost (i.e., cost of zero) to a correct team decision; it is defined as:

\[
P(e) = \sum_{j=0}^{M-1} \sum_{j \neq k}^{M-1} P(u_t = k \mid H_j) = 1 - \sum_{j=0}^{M-1} P(u_t = j \mid H_j) \tag{3.1}
\]

Another example of such a cost function which we often employ is the normalized probability of error cost function. It is a variant of the probability of error, which takes into account the prior probabilities and different costs for hypothesis misclassification; it is defined as follows:

\[
P^e = \sum_{k=0}^{M-1} \sum_{j=0}^{M-1} c_{kj} P(u_t = k \mid H_j) \tag{3.2}
\]

where for every \(k, j = 0, ..., M-1\):

\[
c_{kj} = \frac{p_j [J(u_t = k, H_j) - J(u_t = j, H_j)]}{\sum_{r=0}^{M-1} \sum_{\nu=0}^{M-1} [J(u_t = r, H_\nu) - J(u_t = \nu, H_\nu)]} \tag{3.3}
\]

We will show that this cost structure takes a special form for the case of binary \((M=2)\) hypothesis testing. Moreover, it should be clear that we consider the Bayesian formulation of the problems.
3.2.7. Notation

We denote by \( P(A) \) the probability of an event \( A \). If \( u \) is a random variable and \( A \) is an event, the conditional probability of \( A \) given \( y \) is denoted by \( P(A \mid y) \); this is a random variable because it is a function of the random variable \( y \). We denote by \( P(A \mid y = y_0) \) the value of the random variable \( P(A \mid y) \) when the realization of \( y \) is \( y_0 \).

Let \( \lfloor x \rfloor \) denote the integer part of \( x \) and \( \lceil x \rceil \) denote the smallest integer greater than or equal to \( x \).

3.3. THE ROC CURVE

Consider the following trivial detection problem:

**Problem 3.1.** Given the conditional distributions \( P(y \mid H) \) for \( H = H_0, H_1 \), the prior probabilities \( p_0 \) and \( p_1 \) (\( p_0 + p_1 = 1 \)) and costs \( J(u, H) \), for \( u = 0, 1 \), and \( H = H_0, H_1 \), find the optimal decision strategy for a single DM who receives the observation \( y \) to minimize his expected cost.

Define the likelihood ratio as:

\[
\Lambda(y) = \frac{P(y \mid H_1)}{P(y \mid H_0)} \tag{3.4}
\]

and the threshold of the test:

\[
\eta = \frac{p_0 [J(1, H_0) - J(0, H_0)]}{p_1 [J(0, H_1) - J(1, H_1)]} \tag{3.5}
\]

The solution of the above problem is given by the following likelihood ratio test with constant threshold (Figure 3.1):

\[
\Lambda(y) \begin{cases} 
\geq \eta & \text{if } u = 1 \\
< \eta & \text{if } u = 0
\end{cases} \tag{3.6}
\]
Figure 3.1. Optimal Solution to the Detection Problem 3.1.

$P_F$: Probability of False Alarm; $P_D$: Probability of Detection

The likelihood ratio is a random variable; if we denote by $P(A \mid H)$ the probability density function of $A$ when $H$ is true, if it exists, we can define the probability of false alarm as:

$$P_F(\eta) = \int_{\eta}^{-} P(A \mid H_0) \, dA$$

(3.7a)

and the probability of detection as:

$$P_D(\eta) = \int_{\eta}^{-} P(A \mid H_1) \, dA$$

(3.8a)

Equivalently, if we denote by $P(A \mid H)$ the probability mass function of $A$ when $H$ is true, if it exists, we can define the probability of false alarm as:

$$P_F(\eta) = \sum_{A \geq \eta} P(A \mid H_0)$$

(3.7b)

and the probability of detection as:
\[ P_d(\eta) = \sum_{\Lambda \geq \eta} P(\Lambda | H_1) \] (3.8b)

The ROC curve is the plot of \( P_D \) versus \( P_F \), with \( \eta \) as the varying parameter. It is usually defined in an open-form, by the two parametric equations as the following set:

\[
\text{ROC} \equiv \{(P_F, P_D) : P_F = P_F(\eta), P_D = P_D(\eta), 0 \leq \eta < \infty\}
\]

If the ROC curve of an individual DM is continuous, then it is concave and goes through (0,0) and (1,1). If the ROC curve of an individual DM is discontinuous, then the upper bound of its convex hull\(^1\) is concave and goes through (0,0) and (1,1) (as shown in section 3.4.1, the discontinuous ROC curve of an individual DM becomes both continuous and concave when randomized decision rules are allowed). Under some additional assumptions, this property will be generalized for team ROC curves in chapter 6. Also, two ROC curves may intersect at other points besides (0,0) and (1,1).

**REMARK.** Randomized decision rules are the procedure in which two or more likelihood ratio tests are mixed in some probabilistic manner. In the team level, there exist two types of randomized decision rules: decision rules with **dependent randomization** and decision rules with **independent randomization**. These decision rules and their implications on the team performance are discussed extensively in chapter 6.

One very important property of continuous ROC curves, which we are going to explore in our subsequent analysis, is that the slope of the tangent (if it exists) to a ROC curve at a particular point is equal to the value of the threshold \( \eta \) required to achieve the \( P_F \) and \( P_D \) of that point (Figure 3.1):

\[
\frac{dP_D}{dP_F} = \frac{dP_D/d\eta}{dP_F/d\eta} = \eta
\] (3.9)

\(^1\) For a complete definition of the convex hull see [BS79].
At the points where an ROC curve is not differentiable (i.e., the slope of the tangent to the curve does not exist), the value of \( \eta \) required to achieve the \( P_F \) and \( P_D \) of that point is a subgradient\(^2\) of the ROC curve at \((P_F, P_D)\).

The ROC curve is a convenient tool for describing DMs and it is also a convenient tool for ranking DMs (for binary hypothesis testing problems). The higher the ROC curve the better the DM since a higher probability of detection corresponds to the same level of probability of false alarm. Thus, we can give the following definitions:

**DEFINITION 3.1a.** Consider two DMs \( a \) and \( b \) with associated upper bounds of the convex hulls of their ROC curves \( f_a: [0, 1] \to [0, 1] \) and \( f_b: [0, 1] \to [0, 1] \), respectively. In the *Bayesian* framework, we say that DM \( a \) is better than DM \( b \) if:

\[
f_a(P_F) \geq f_b(P_F) \quad \text{for every } P_F \in [0, 1]
\]

and:

\[
f_a(P_F) > f_b(P_F) \quad \text{for some } P_F \in [0, 1].
\]

**REMARK.** Formalizing the above definition makes it sound more complex than it really is. If both ROC curves to be compared are concave, then the "upper bounds of the convex hulls of their ROC curves" may be replaced by a simple "ROC curves"; this is the case throughout this thesis, excluding chapter 6. We will return to this point in chapter 6, where the above Definition 3.1a will be explained and justified.

**DEFINITION 3.1b.** Consider two DMs \( a \) and \( b \) with associated ROC curves \( f_a: F_a \to D_a \) and \( f_b: F_b \to D_b \) respectively \( (F_i, D_i \subseteq [0, 1], \text{ for } i = a, b) \). In the *Neyman-Pearson* framework, we say that DM \( a \) is better than DM \( b \) if:

\[
F_b \subseteq F_a
\]

and:

\[
f_a(P_F) \geq f_b(P_F) \quad \text{for every } P_F \in F_b
\]

and:

\[
f_a(P_F) > f_b(P_F) \quad \text{for some } P_F \in F_b.
\]

\(^2\) For a complete definition of the subgradient see [BS79].
**REMARK 1.** In other words, for DM $a$ to be better than DM $b$, the ROC curve of DM $a$ has to be defined everywhere in the region where the ROC curve of the DM $b$ is defined, and to have probability of detection at least as high as that of DM $b$.

**REMARK 2.** Definitions 3.1a and 3.1b are also valid when the ROC curves to be compared are ROC curves of teams of DMs, as opposed to ROC curves of individual DMs.

The DMs in Figure 3.2 can be unequivocally ranked; DM $A$ is better than DM $B$. But, this is not true when the ROC curves of two DMs intersect. In the Bayesian formulation, which DM is better depends on factors which are external to the team, namely the threshold $\eta$ (i.e., prior probabilities and costs). In Figure 3.3, DM $A$ is better for thresholds larger than the slope of the common tangent to the two ROC curves (not shown in Figure 3.3), DM $B$ is better for thresholds smaller than the slope of the common tangent and both DMs achieve the same performance for thresholds equal to the slope of the common tangent. In the Neyman-Pearson formulation, which DM is better depends on the level of the probability of false alarm at which the DMs operate. In Figure 3.3, DM $A$ is better for probabilities of false alarm smaller than $P_F(\eta^*)$, DM $B$ is better for probabilities of false alarm greater than $P_F(\eta^*)$ and both DMs achieve the same performance for probability of false alarm equal to $P_F(\eta^*)$.

**REMARK 3.** We can now rewrite the normalized probability of error for the binary hypothesis testing case, using the notation just introduced:

$$P_e = c P_F + (1 - c) (1 - P_D)$$  \hspace{1cm} (3.10)

with:

$$c = \frac{\eta}{1 + \eta}$$  \hspace{1cm} (3.11)
Figure 3.2. Ranking DMs Unequivocally (DM A better than DM B)

Figure 3.3. Ranking DMs: Better DM depends on Threshold $\eta$
**REMARK.** The **threshold** of the test \( \eta \) is a threshold on the likelihood ratio axis. This defines in turn, through Eqs. (3.7) and (3.8), an **operating point** \((P_F, P_D)\) on the ROC curve. This operating point can be viewed as a threshold on the ROC curve. Since there is correspondence between the thresholds on the likelihood ratio axis and on the points of the ROC curve (Eqs. (3.7) and (3.8)), we use the terms interchangeably. We try to use 'threshold' for the thresholds on the likelihood ratio axis and 'operating point' for thresholds on the ROC curve.

### 3.4. SOME SPECIAL ROC CURVES

In this section we present several 'families' of ROC curves which will be employed in our subsequent analysis.

#### 3.4.1. Discrete Distributions

The ROC curves of discrete distributions consist of a series of disjoint points (Figure 3.4(a)). Allowing for **randomized decision rules** intermediate values can be achieved; then the ROC curves consist of the straight lines that connect the points (Figure 3.4(b)). Therefore, the ROC curves of discrete distributions are piecewise linear.

We will use the following convention to determine the slope of the tangent at the points where a piecewise linear ROC curve is not differentiable (i.e., the corner points):

**CONVENTION.** We **assume that the slope of the tangent to a ROC curve at a point** \((P_F, P_D)\), **where the curve is not differentiable, exist and its value be equal to the value of one of the subgradients of the ROC curve at** \((P_F, P_D)\).

For example, the slope of the tangent to the ROC curve at \((P_F^*, P_D^*)\) in Figure 3.4(b) is less than or equal to \(m_1\) and greater than or equal to \(m_2\).
Figure 3.4(a). ROC Curve for Discrete Distributions *without* Randomization

Figure 3.4(b). ROC Curve for Discrete Distributions *with* Randomization
3.4.2. Exponential Distributions

Suppose that we compare the rate (and hence mean and variance) of two exponential distributions. Then:

\[ P(y \mid H_0) = \lambda_0 \exp(-\lambda_0 y) \quad (3.12) \]

and:

\[ P(y \mid H_1) = \lambda_1 \exp(-\lambda_1 y) \quad (3.13) \]

Without loss of generality assume \( \lambda_1 > \lambda_0 \). Set: \( \alpha = \lambda_0 / \lambda_1 \). Then the ROC curve will be given by the following closed-form expression:

\[ P_D = (P_F)^\alpha \quad (3.14) \]

The slope \( m \) of the tangent at a point \( (P_F, P_D) \) is given by:

\[ m = \frac{dP_D}{dP_F} = \alpha (P_F)^{\alpha - 1} = \alpha \frac{P_D}{P_F} \quad (3.15) \]

We should note here that there exists another interesting case of probability distributions which results in the same family of ROC curves. This is the case of comparing the variances of two Gaussian distributions with equal means when the DM receives two observations. Then:

\[ P(y_1, y_2 \mid H_0) = \frac{1}{2\pi \sigma_0^2} \exp \left[ -\frac{y_1^2 + y_2^2}{2\sigma_0^2} \right] \quad (3.16) \]

and:

\[ P(y_1, y_2 \mid H_1) = \frac{1}{2\pi \sigma_1^2} \exp \left[ -\frac{y_1^2 + y_2^2}{2\sigma_1^2} \right] \quad (3.17) \]

Without loss of generality assume that: \( \sigma_1 > \sigma_0 \), and set: \( \alpha = (\sigma_1 / \sigma_0)^2 \).
3.4.3. Gaussian Distributions with Equal Variance

This is a classic case, which demonstrates the inherent complexity of the ROC curve. Suppose that we want to compare the means of two Gaussian distributions with equal variance. Then:

\[ P(y | H_0) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left[ -\frac{(y - \mu_0)^2}{2\sigma^2} \right] \]  \hspace{1cm} (3.18)

and:

\[ P(y | H_1) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left[ -\frac{(y - \mu_1)^2}{2\sigma^2} \right] \]  \hspace{1cm} (3.19)

Without loss of generality assume that \( \mu_0 = 0, \mu_1 = d \) and \( \sigma^2 = 1 \). Then, the ROC curve is given by the following two parametric equations:

\[ P_F = \Phi[\ln(\eta)/d + d/2] \]  \hspace{1cm} (3.20)

\[ P_D = \Phi[\ln(\eta)/d - d/2] \]  \hspace{1cm} (3.21)

where:

\[ \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp \left[ -\frac{x^2}{2} \right] \, dx \]  \hspace{1cm} (3.22)

is the complementary error function [V68]. The parametric equations and the improper integrals are too cumbersome to manipulate algebraically; thus unfortunately, for these distributions rigorous proofs will have to be substituted by computer simulations.

3.4.4. Conic ROC Curves

These are curves described by the following quadratic expression:

\[ aP_D^2 + 2bP_DP_F + cP_F^2 + 2dP_D - P_F = 0 \]  \hspace{1cm} (3.23)
The slope \( m \) of the tangent to this curve at a point \( (P_F, P_D) \) is given by:

\[
m = \frac{dP_D}{dP_F} = \frac{-bP_D + cP_F - 1}{aP_D + bP_F + d}
\]  
(3.24)

The second derivative of the curve at \( (P_F, P_D) \) is given by:

\[
\frac{d^2 P_D}{dP_F^2} = -\frac{a + 2bd + cd^2}{(aP_D + bP_F + d)^3}
\]  
(3.25)

The coefficients have to satisfy the certain conditions. Since the ROC curve goes through (1,1):

\[
a + 2b + c + 2d - 2 = 0
\]  
(3.26a)

The slope of the tangent at \( (0,0) \) is positive. Therefore:

\[
d \geq 0
\]  
(3.26b)

Moreover, the second derivative of a concave curve is non-positive. At \( (0,0) \), this implies that:

\[
a + 2bd + cd^2 \geq 0
\]  
(3.26c)

and at \( (1,1) \) that:

\[
a + b + d \geq 0
\]  
(3.26d)

Combining this with the nonnegative first derivative at \( (1,1) \) we obtain that:

\[
b + c - 1 \leq 0
\]  
(3.26e)

Finally, we have also derived and present the parametric equations that result in the conic ROC curve:

\[
P_F(\eta) = \frac{bd + a}{ac - b^2} - \frac{an + b}{ac - b^2} \left[ \frac{a + 2bd + cd^2}{an^2 + 2bn + c} \right]^{1/2}
\]  
(3.27)

\[
P_D(\eta) = -\frac{b - cd}{ac - b^2} + \frac{bn + c}{ac - b^2} \left[ \frac{a + 2bd + cd^2}{an^2 + 2bn + c} \right]^{1/2}
\]  
(3.28)
3.5. THE CONDITIONAL INDEPENDENCE ASSUMPTION

**ASSUMPTION 3.1.** *The random variables \( y_1, \ldots, y_N \) are conditionally independent given any hypothesis.*

All the results derived in this thesis depend on Assumption 3.1. As was demonstrated by a simple example in [T89] if the assumption fails to hold then threshold strategies do not have to be optimal. In fact it was also shown that without the conditional independence assumption problems of this type become \( NP \)-complete. Several algorithms for solving such problems have been proposed, but neither bounds on their performance, nor sufficient computational experience exist. New algorithms with constraint decision rules, which will greatly decrease the computational complexity, need to be developed. These will lead to suboptimal solutions, but may perform reasonably well in practice.

3.6. SUMMARY

In this chapter the distributed hypothesis testing model was defined. The Receiver Operating Characteristic (ROC) curve was introduced because it offers a complete and convenient description of the DMs in the binary hypothesis testing framework.
CHAPTER 4

Small Team Problems

4.1. INTRODUCTION

In this chapter, we examine problems of small teams, that is teams which consist of two or three DMs and perform binary hypothesis testing. We present different architectures for these teams, analyze them and compare their performance. Our objective is to obtain design guidelines and building blocks which will assist in the ultimate research goal, the optimal design of large organizations.

REMARK. This remark describes the types of decision rules employed by the DMs (deterministic, randomized, etc.) and may be skipped without loss of continuity. As will be extensively discussed in chapter 6, since in this chapter only the Bayesian formulation of the problems is considered, all types of decision rules are equivalent. Therefore, we assume without loss of generality that the DMs employ what will be defined as decision rules with dependent randomization. These result in all the individual and the team ROC curves being both continuous and concave. We suggest that the interested reader return to this remark after reading chapter 6, where also the Neyman-Pearson formulation of the problems is considered.

In section 4.2, we compare the performance of the two different architectures for the team which consists of two DMs; the tandem architecture (Figure 4.1) and the parallel architecture (Figure 4.2). In the tandem team, one DM is referred to as the primary DM and the other as the consulting DM. The consulting DM makes a decision based on his own observation and transmits it to the primary DM. The primary DM has the responsibility of
the final team decision, which will be based on his own observation and the communication from the consulting DM. In the parallel team, each DM makes a decision and transmits it to a fusion center; the fusion center makes the final team decision using the maximum a posteriori normalized probability decision rule. It is known that the tandem architecture is better than the parallel architecture. Several comparisons have appeared in the literature demonstrating that the tandem architecture achieves superior performance than the parallel architecture [ET82], [R87], [V88]. We present a formal proof, since we are going to employ this result as a lemma in the proof of subsequent results.

In section 4.3, we further analyze the team which consists of two DMs in tandem. The problem has been extensively studied in the literature and necessary conditions for optimality have been derived [E82]. Given two DMs, we would like to determine the optimal configuration for the tandem team (i.e., determine which DM should be made the primary DM). It is intuitively appealing that, if we are given two DMs, with one being better than the other in the ROC curve sense, the better DM should be the primary; this was also supported with explanations on data compression [E82] and with numerical results [R87]. We analyze the problem and show that the conjecture is false in general; in fact, we show that the optimal configuration depends on the external parameters of the problem like the prior probabilities, the cost assignments, as well as, in a counterintuitive way, on the number of different messages which the consulting DM may transmit to the primary DM. On the other hand, we present particular probability distributions for which the conjecture is true and we obtain bounds on the deterioration of the performance of the team, if the better DM is constrained to be the primary DM.

In section 4.4, we discuss the team which consists of two DMs in parallel. This was the first team which was studied in this framework [TS81]. It was shown that even if the DMs are identical and the cost structure symmetric, the optimal decision rules of the two DMs do not have to be identical or symmetric. This was attributed to the explicit dependence of the cost function on the decisions of the DMs. Subsequently, this result was
also demonstrated with a simple example for case where the cost function depends
implicitly on the decisions of the DMs, through the fusion center [T88]. We obtain the
second order optimality conditions to enhance our understanding of the problem and its
counterintuitive solution. We also obtain that bounds on the deterioration of the team
performance, when both DMs are constrained to employing identical decision rules.

Finally in section 4.5, we deal with teams which consist of three DMs. We try to
determine whether a dominant architecture exists. It was conjectured in the literature [E82]
that the two consultant or V-architecture (Figure 4.19) is better than the three DM tandem
architecture (Figure 4.20). We disprove the conjecture and show that the optimal team
architecture depends on the DMs of the team, on the prior probabilities and on the costs.

REMARK. Throughout this discussion we assume that the cost function is such that it is
more costly for the team to err than to be correct (i.e., $J(0, H_1) > J(1, H_1)$ and $J(1, H_0) >
J(0, H_0)$). This logical assumption is made in order to express the optimal decision rules in
the convenient likelihood ratio form with constant thresholds.

4.2. TANDEM VERSUS PARALLEL ARCHITECTURE

Consider the two different architectures for the team which consists of two DMs; the
tandem architecture (Figure 4.1) and the parallel architecture (Figure 4.2). In the tandem
team, one DM is referred to as the primary DM and the other as the consulting DM. The
consulting DM makes a decision based on his own observation and transmits it to the
primary DM. The primary DM has the responsibility of the final team decision, which will
be based on his own observation and the communication from the consulting DM. In the
parallel team, each DM makes a decision and transmits it to a fusion center; the fusion
center makes the final team decision using the maximum a posteriori normalized probability
decision rule. We would like to determine which architecture achieves superior performance:

**PROBLEM 4.1.** Consider a team which consists of two DMs and performs binary hypothesis testing; determine which of the two possible architectures for the team, the tandem architecture or the parallel architecture, achieves superior performance.

![Diagram of Two DMs in Tandem](image1)

**Figure 4.1.** Two DMs in Tandem

![Diagram of Two DMs in Parallel](image2)

**Figure 4.2.** Two DMs in Parallel

As was mentioned above we present a formal proof of this result, which it will be employed as a lemma in the sequel.

**LEMMA 4.1.** Consider a team which consists of two DMs. Then, the tandem architecture achieves at least as good performance as the parallel architecture.
Proof. The proof is simple and straightforward. Consider a team which consists of two DMs in parallel and denote by $\Gamma^* = \{ \gamma_a, \gamma_b, \gamma_1 \}$ the set of the optimal decision rules for the two DMs and for the fusion center respectively. The optimal decision of each DM $n$ depends exclusively on the observation $y_n$ of the DM, for $n = a, b$. Also, the decision of the fusion center depends on the decisions $u_a$ and $u_b$ of the two DMs.

Now consider the same two DMs in a tandem architecture; without loss of generality assume that DM $b$ is the consulting DM. DM $b$ can employ $\gamma_b$ to make his decision based on his own observation. Moreover, DM $a$ can employ $\gamma_a$ and make a preliminary decision based on his own observation and then also employ $\gamma_1$ to make the team decision based on DM $b$'s message and on his own preliminary decision. The proposed decision rules $\tilde{\gamma}_a$ and $\tilde{\gamma}_b$ for the tandem architecture are thus defined by:

$$\tilde{\gamma}_a(y_a, u_b) = \gamma_a(y_a), u_b$$

and:

$$\tilde{\gamma}_b(y_b) = \gamma_b(y_b)$$

The proposed decision rules, though not optimal in general, enable the tandem team to always duplicate the optimal performance of the parallel team. This implies that the tandem architecture can achieve at least as good performance as the parallel. Q.E.D.

Note that this result does not depend on the DMs involved, does not on which of the two DMs is made the primary DM in the tandem architecture and does not depend on the prior probabilities or the costs. Furthermore, the result can be generalized for any number of messages which can be transmitted within a team, as long as the consulting DM in the tandem configuration is allowed to transmit to the primary DM the same number of messages as he (the consulting DM) is allowed to transmit to the fusion center in the parallel configuration.
4.3. TWO DMs IN TANDEM

4.3.1. Problem Formulation and Solution

Given two DMs, we would like to determine the optimal configuration for the tandem team (i.e., determine which DM should be made the primary DM). If one DM is better than the other, it is intuitively appealing that the better DM be made the primary DM. Given two DMs one would expect to have the better DM make the team decision, independent of the prior probabilities and the cost assignments. If this was the case, then the optimal way of organizing two DMs would not change, say, as the prior probabilities of the underlying hypotheses vary. But, it is not true in general. To see this we begin by presenting the necessary conditions which characterize the optimal decision rules of the two DMs from [ET82].

**PROBLEM 4.2.** A team consisting of two DMs in tandem performs binary hypothesis testing. The costs $J(u, H)$ which are incurred by the team when the team decision is $u_t$ and $H$ is the true hypothesis as well as the prior probabilities ($p_i = P(H_i)$, for $i = 0, 1$) are assumed to be known. Each DM receives a conditional independent observation. The consulting DM makes a decision based on his observation and transmits a binary message ($u_c = 0$ or $u_c = 1$) to the primary DM. The primary has the responsibility of the final team decision ($u_p = 0$ or $u_p = 1$) which will be based on his own observation and the communication from the consulting DM. The decision rules for the two DMs which minimize the expected cost are to be determined.

The optimal decision rules have to satisfy the following necessary conditions [E82].

For the primary DM:
If $u_c = 0$:

$$
\begin{align*}
& u_p = 1 \\
& \Lambda(y_p) \geq \frac{1 - P_F^c}{1 - P_D^c} \eta = \eta_0
\end{align*}
$$

(4.3)

If $u_c = 1$:

$$
\begin{align*}
& u_p = 1 \\
& \Lambda(y_p) \geq \frac{P_F^c}{P_D^c} \eta = \eta_1
\end{align*}
$$

(4.4)

For the consulting DM:

$$
\begin{align*}
& u_c = 1 \\
& \Lambda(y_c) \geq \frac{P_F^i - P_F^0}{P_D^i - P_D^0} \eta = \eta_c
\end{align*}
$$

(4.5)

where $\eta$ was defined in Eq.(3.5) with $P_D^i$ and $P_F^i$ respectively the probability of detection and probability of false alarm for the primary DM when $u_c = i$ was sent by the consulting DM ($i = 0, 1$) and, $P_D^c$ and $P_F^c$ respectively the probability of detection and probability of false alarm for the consulting DM. The operating points for two typical DMs are shown in Figure 4.3(a) and (b).

**REMARK 1.** It should be clear by now that the decision thresholds of the two DMs are given by a set of coupled equations. For example:

$$
P_F^0 = P\left(\Lambda(y_p) \geq \frac{1 - P_F^c}{1 - P_D^c} \eta \middle| H_0\right)
$$

(4.6)

**REMARK 2.** The two messages assigned to the consulting DM do not have to be denoted 0 and 1. For that matter they can be denoted $m_1$ and $m_2$. Without loss of generality assume that:

$$
\frac{P(u_c = m_1 | H_0)}{P(u_c = m_1 | H_1)} \geq \frac{P(u_c = m_2 | H_0)}{P(u_c = m_2 | H_1)}
$$

(4.7)
Figure 4.3(a). Operating Points of Typical Primary DM

\[ \eta_0 = \frac{1 - P_F^c}{1 - P_D^c} \eta \]

\[ \eta_1 = \frac{P_F^c}{P_D^c} \eta \]

Figure 4.3(b). Operating Point of Typical Consulting DM

\[ \eta_c = \frac{P_F^1 - P_F^0}{P_D^1 - P_D^0} \eta \]
Then it can be shown that when the primary DM receives \( u_c = m_1 \) he will always be more likely to decide \( u_p = 0 \) and when he receives \( u_c = m_2 \) he will always be more likely to decide \( u_p = 1 \). Hence the interpretation of \( m_1 \) as 0 and of \( m_2 \) as 1.

**Remark 3.** The ROC curve of the team as a whole can be computed and is given by the following two parametric equations:

\[
P_F^t(\eta) = [1 - P_F^c(\eta)] P_F^0(\eta) + P_F^c(\eta) P_F^1(\eta) \tag{4.8}
\]

\[
P_D^t(\eta) = [1 - P_D^c(\eta)] P_D^0(\eta) + P_D^c(\eta) P_D^1(\eta) \tag{4.9}
\]

Note that the team ROC curve depends not only upon the characteristics ('expertise') of the individual DMs, but also on the particular way that they have been constrained to interact (the team or organization architecture).

### 4.3.2. Architecture Comparisons

Suppose that one of the DMs is *better* than the other, i.e., his ROC curve is higher than the ROC curve of the other DM. There exist two candidate architectures for the team; either make the better DM the primary DM or make the better DM the consulting DM (Figure 4.4). Recall that the primary DM makes the final team decision. Consider the following problem:

**Problem 4.3.** Consider two DMs, one better than the other. Determine whether the optimal configuration is independent of the external parameters of the problem (details of cost function, prior probabilities), which determine the value of \( \eta \); that is, determine which configuration of the team (Figure 4.4) yields superior performance, for all values of \( \eta \).
Figure 4.4. Different Configurations for the Two DM Tandem Team

The architecture with the better DM as the primary DM was conjectured [E82] to be better. It will be shown that the problem of the optimal architecture can be reduced to a simpler problem in which the worse DM will have a three piecewise linear ROC curve and the better DM will also have a piecewise linear ROC curve with at most four line segments\(^1\). This restricted problem will yield an example disproving the conjecture. The discussion below is simple and straightforward. We will compare the performance of four teams for a particular value of the threshold; Figure 4.5 should provide a flowchart summary of the analysis.

Consider two DMs, a DM B better than a DM W in the ROC curve sense and suppose that DM B has been designated as the consulting DM. The solution to the problem for some \(\eta^*\) can be described by the three operating points \((P_{F_0}^0, P_{D_0}^0)\) and \((P_{F_1}^1, P_{D_0}^1)\) of the primary DM W and \((P_{F_0}^0, P_{D_0}^0)\) of the consulting DM B, which have been defined above (Figure 4.3).

Consider also a DM W' whose ROC curve consists of the points \((0,0)\), \((P_{F_0}^0, P_{D_0}^0)\), \((P_{F_1}^1, P_{D_0}^1)\) and \((1,1)\) as well as the line segments joining them. Finally consider a DM B' whose ROC curve consists of the points \((0,0)\), \((P_{F_0}^0, P_{D_0}^0)\) and \((1,1)\), the line segments joining them and whose ROC curve lies above the ROC curve of DM W'. There are six different cases for the ROC curve of DM B' depending on where \((P_{F_0}^0, P_{D_0}^0)\) lies with respect to \((P_{F_0}^0, P_{D_0}^0)\) and \((P_{F_1}^1, P_{D_0}^1)\) as can be seen in Figure 4.6.

---

\(^1\) This fact was formalized and generalized in [T89a], where the optimality of the likelihood ratio quantizers (LRQs) was established for a broad class of problems.
Figure 4.5. Summary of the Analysis

Figure 4.6. The Six Cases for the Consulting DM
Consider the 'restricted' team with DM \( W' \) as the primary DM and DM \( B' \) as the consulting DM. Note that the performance of this team will never be better than the performance of the original team since DM \( W' \) is not better than DM \( W \) and DM \( B' \) is not better than DM \( B \) by construction; but for the particular \( \eta^* \) both teams achieve the same performance since the three optimal operating points, namely \( (P_0^0, P_D^0), (P_1^1, P_D^1) \) and \( (P_\infty^\infty, P_D^\infty) \), are attainable by the respective DMs of both teams.

We would now like to compare the performance of the 'restricted' team with the performance of the 'reverse' restricted team which consists of DM \( B' \) as the primary DM and DM \( W' \) as the consulting DM. Note the the performance of the 'reverse' team is obviously never better than the performance of the team which consists of DM \( B \) as the primary DM and DM \( W \) as the consulting DM.

\textit{NOTE}. Henceforth, for the sake of convenience we will only mention the primary DM of each team since it is clear which DM is his 'complementary' consulting DM.

So we compare the 'restricted' team with DM \( W' \) as the primary DM to the 'reverse' team with DM \( B' \) as the primary DM for all six cases of Figure 4.6. If the 'reverse' team performed better than the 'restricted' team in all six cases, we would conclude that the team with DM \( B \) as its primary DM would perform better than the team with DM \( W \) as its primary DM; hence the conjecture would have been proven true. But, there exist certain cases where the 'restricted' performs better than the 'reverse' team; hence an example disproving the conjecture is obtained.

\subsection{The Analysis}

This section may be skipped without loss of continuity. It is here for the sake of completeness to demonstrate the method of analysis employed which resulted in disproving the conjecture.
Figure 4.7. The 'Restricted' DMs of the Analysis

We need only analyze one case of Figure 4.6 which is presented in more detail in Figure 4.7. This is done to demonstrate the method for the analysis of such problems. From Figure 4.7 we see that in this case the slopes of the operating points have to satisfy the following conditions:

\[
\frac{P_D^0}{P_F^c} \geq \frac{P_D^c}{P_F^c} \geq \frac{P_D^1}{P_F^c} \geq \frac{P_D^1 - P_D^0}{1 - P_F^c} \geq \frac{1 - P_D^0}{1 - P_F^c} \geq \frac{1 - P_D^1}{1 - P_F^c} \quad (4.10)
\]

and:

\[
\frac{P_D^c}{P_F^c} \geq \frac{P_D^c - P_D^0}{P_F^c - P_F^0} \geq \frac{P_D^1 - P_D^0}{P_F^c - P_F^0} \geq \frac{P_D^1 - P_D^c}{P_F^c - P_F^c} \geq \frac{1 - P_D^c}{1 - P_F^c} \quad (4.11)
\]

Moreover recalling that the operating points are by construction optimal for the 'restricted' team for \( \eta^* \), the necessary optimality conditions for the primary DM can be written as:

\[
\frac{P_D^0}{P_F^c} \geq \frac{1 - P_F^c}{1 - P_D^c} \geq \frac{P_D^1 - P_D^0}{P_F^c - P_F^0} \quad (4.12)
\]

and:

\[
\frac{P_D^1 - P_D^0}{P_F^c - P_F^0} \geq \frac{P_F^c}{P_D^c} \eta^* \geq \frac{1 - P_D^1}{1 - P_F^c} \quad (4.13)
\]
and for the consulting DM as:

\[
\frac{P_D^c-P_D^0}{P_F^c-P_F^0} \geq \frac{P_D^1-P_D^0}{P_F^1-P_F^0} \eta^* \geq \frac{P_D^1-P_D^c}{P_F^1-P_F^c} \tag{4.14}
\]

Eqs. (4.12)-(4.14) can be summarized with the help of Eqs.(4.10) and (4.11) as:

\[
\min \left\{ \frac{1-P_D^c}{1-P_F^c}, \frac{P_D^c-P_D^0}{P_F^c-P_F^0}, \frac{P_D^1-P_D^0}{P_F^1-P_D^0} \right\} \geq \eta^* \tag{4.15a}
\]

\[
\eta^* \geq \max \left\{ \frac{P_D^c}{P_F^c}, \frac{1-P_D^1}{1-P_F^1}, \frac{P_D^1-P_D^c}{P_F^1-P_D^0}, \frac{P_D^1-P_D^0}{P_F^1-P_F^0} \right\} \tag{4.15b}
\]

We should note here that the probabilities of false alarm and detection for the 'restricted' team are given by:

\[
P_F^I = (1-P_F^c)P_F^0 + P_F^cP_F^1 \tag{4.16}
\]

and:

\[
P_D^I = (1-P_D^c)P_D^0 + P_D^cP_D^1 \tag{4.17}
\]

We now consider the 'reverse' team. The two possibly optimal operating points for the new consulting DM \(W\) are \((P_F^0, P_D^0)\) and \((P_F^1, P_D^1)\).

**CASE A. Operating Point of the Reversed Consulting DM:** \((P_F^0, P_D^0)\)

That is assume from Eq.(4.5):

\[
\frac{P_D^0}{P_F^0} \geq \eta_c \geq \frac{P_D^1-P_D^0}{P_F^1-P_D^0} \tag{4.18}
\]

Then from Eq.(4.3):

\[
\eta_0 = \frac{1-P_F^0}{1-P_D^0} \eta^* \tag{4.19}
\]

and from Eq.(4.4):

\[
\eta_1 = \frac{P_F^0}{P_D^0} \eta^* \tag{4.20}
\]
Combining Eqs. (4.10), (4.15) and (4.17), we obtain that:

\[
\eta_0 \geq \frac{P_D^1 - P_D^c}{P_F^1 - P_F^c} \tag{4.21}
\]

Thus the optimal operating point for the new primary DM $B'$ when $u_c = 0$ is received is either $(P_F^0, P_D^0)$ or $(P_F^c, P_D^c)$.

Similarly combining Eqs.(4.15) and (4.20) we obtain that:

\[
\frac{1 - P_F^c}{1 - P_F^0} \geq \eta_1 \tag{4.22}
\]

Thus the optimal operating point for the new primary DM $B'$ when $u_c = 1$ is received is either $(P_F^1, P_D^1)$ or (1,1). We have four subcases for the operating points of the primary DM $B'$ which we examine separately.

**SUBCASE A.I. Operating Point of the Reversed Primary DM when $u_c = 0$:** $(P_F^0, P_D^0)$

**Operating Point of the Reversed Primary DM when $u_c = 1$:** $(P_F^1, P_D^1)$

Then the optimality conditions can be written as:

\[
\frac{P_D^c - P_D^0}{P_F^c - P_F^0} \geq \eta_0 \geq \frac{P_D^1 - P_D^c}{P_F^1 - P_F^c}
\]

\[
\Rightarrow \quad \frac{1 - P_D^0}{1 - P_F^0} \frac{P_D^c - P_D^0}{P_F^c - P_F^0} \geq \eta^* \geq \frac{1 - P_D^0}{1 - P_F^0} \frac{P_D^1 - P_D^c}{P_F^1 - P_F^c} \tag{4.23}
\]

and:

\[
\frac{P_D^1 - P_D^c}{P_F^1 - P_F^c} \geq \eta_1 \geq \frac{1 - P_D^1}{1 - P_F^1}
\]

\[
\Rightarrow \quad \frac{P_D^0}{P_F^0} \frac{P_D^1 - P_D^c}{P_F^1 - P_F^c} \geq \eta^* \geq \frac{P_D^0}{P_F^0} \frac{1 - P_D^1}{1 - P_F^1} \tag{4.24}
\]

Also from Eq.(4.5):

\[
\eta_c = \frac{P_F^1 - P_F^c}{P_D^1 - P_D^c} \eta^* \tag{4.25}
\]
In view of this, we check our assumption of Eq.(4.18):

\[
\frac{P_D^0}{P_F^0} \geq \frac{P_F^1 - P_F^c}{P_F^1 - P_F^c} \eta^* \geq \frac{P_D^1 - P_D^0}{P_F^1 - P_F^c}
\]  

(4.26)

This is indeed true from Eqs.(4.11) and (4.15).

In this subcase the probabilities of false alarm and detection for the 'reversed' team are given by:

\[
P_F^\gamma = (1 - P_F^c) P_F^c + P_F^0 P_F^1
\]  

(4.27)

and:

\[
P_D^\gamma = (1 - P_D^0) P_D^c + P_D^0 P_D^1
\]  

(4.28)

Comparing the probabilities of false alarm for the 'restricted' and the 'reversed' teams:

\[
P_F^r = (1 - P_F^c) P_F^0 + P_F^c P_F^1 \leq (1 - P_F^c) P_F^c + P_F^0 P_F^1 = P_F^\gamma
\]  

(4.29)

For the 'reversed' architecture to be better:

\[
\frac{P_D^\gamma - P_D^1}{P_F^\gamma - P_F^1} = \frac{P_D^c - P_D^0}{P_F^c - P_F^0} \frac{1 - P_D^1}{1 - P_F^1} \geq \eta^* \geq \frac{1 - P_D^\gamma}{1 - P_F^\gamma}
\]  

(4.30)

The left inequality of Eq.(4.30) does not hold, as it is in contradiction with the right inequality of Eq.(4.24). Therefore, the 'restricted' team results in superior performance in this subcase.

**SUBCASE A.2. Operating Point of the Reversed Primary DM when u_c = 0:** \((P_F^c, P_D^c)\)

**Operating Point of the Reversed Primary DM when u_c = 1:** \((1, 1)\)

Then the optimality conditions can be written as Eq.(4.23) again and:

\[
\frac{1 - P_D^1}{1 - P_F^1} \geq \eta_1 \Rightarrow \frac{P_D^0}{P_F^0} \frac{1 - P_D^1}{1 - P_F^1} \geq \eta^*
\]  

(4.31)

Also from Eq.(4.5):
\[ \eta_c = \frac{1-P_F^c}{1-P_D^c} \eta^* \]  

(4.32)

In view of this we check our assumption of Eq. (4.21):

\[ \frac{P_D^0}{P_F^0} \geq \frac{1-P_F^c}{1-P_D^c} \eta^* \geq \frac{P_D^1-P_D^0}{P_F^1-P_F^0} \]  

(4.33)

This is indeed true from Eqs. (4.11) and (4.15).

In this subcase the probabilities of false alarm and detection for the 'reversed' team are given by:

\[ P_F^\gamma = (1-P_F^0)P_F^c + P_F^0 \]  

(4.34)

and:

\[ P_D^\gamma = (1-P_D^0)P_D^c + P_D^0 \]  

(4.35)

Comparing the probabilities of false alarm for the 'restricted' and the 'reversed' teams:

\[ P_F^\gamma = (1-P_F^c)P_F^0 + P_F^c P_F^1 \leq (1-P_F^0)P_F^c + P_F^0 = P_F^\gamma \]  

(4.36)

For the 'reversed' architecture to be better:

\[ \frac{P_D^\gamma-P_D^0}{P_F^\gamma-P_F^0} = \frac{P_D^c}{P_F^c} \frac{1-P_D^1}{1-P_F^1} \eta^* \geq \frac{1-P_D^\gamma}{1-P_F^\gamma} \]  

(4.37)

The left inequality of Eq. (4.37) does not hold, as it is in contradiction with the right inequality of Eq. (4.15b). Therefore, the 'restricted' team results in superior performance in this subcase.

**SUBCASE A.3. Operating Point of the Reversed Primary DM when \( u_c = 0 \):** \( (P_F^0, P_D^0) \)

**Operating Point of the Reversed Primary DM when \( u_c = 1 \):** \( (P_F^1, P_D^1) \)

We can immediately state that the 'restricted' team results in superior performance in this subcase. Otherwise the optimality of the operating points \( (P_F^0, P_D^0) \) and \( (P_F^1, P_D^1) \) of DM W
and \((P_F^C, P_D^C)\) of DM\( B' \) would be contradicted, because the operating points \((P_F^0, P_D^0)\) and \\
\((P_F^1, P_D^1)\) can be employed by DM\( W' \) and \((P_F^0, P_D^0)\) can be employed by DM\( B' \).

**SUBCASE A.4. Operating Point of the Reversed Primary DM when \(u_c = 0\):** \((P_F^0, P_D^0)\)

**Operating Point of the Reversed Primary DM when \(u_c = 1\):** \((1,1)\)

We can state that the 'restricted' team results in superior performance in this subcase for the same reasons as in Subcase A.3 above.

**CASE B. Operating Point of the Reversed Consulting DM's :** \((P_F^1, P_D^1)\)

The analysis is omitted since it is exactly analogous to the analysis of Case A above; again we conclude that the 'restricted' team results in superior performance.

Hence we can construct two DMs as the ones in Figure 4.7, knowing that the tandem team with the better DM as the consulting DM will achieve superior performance.

### 4.3.4. An Example

Consider the following *Example 4.1*, which disproves the conjecture. The ROC curves for the two DMs are presented in Figure 4.8 and the associated discrete probability density functions conditioned on the two hypotheses are presented in Figure 4.9. The team ROC curves are presented in Figure 4.10 (and in Table 4.1) for both architectures. It is interesting to note that for \(\eta = 1.0\) performance is maximized by making DM\( B \) the consulting DM, while for \(\eta = 0.40\) performance is maximized by making DM\( B \) the primary DM (Table 4.2). Thus, in this special example, the optimal team architecture depends on the value \(\eta\) (i.e., the numerical values of the prior probabilities and the costs).
Figure 4.8. The DMs for Example 4.1.

(a). Better DM

(b). Worse DM

Figure 4.9. Probability Distributions for DMs of Figure 4.8.
Figure 4.10(a). The Team ROC Curves for Both Configurations and Binary Messages

Figure 4.10(b). Close Up of Team ROC Curves of Figure 4.10(a).
Table 4.1. The Team ROC Curves of Figure 4.10.

\[
\begin{align*}
B & \rightarrow W \\
(0, 0) & \rightarrow (0, 0) \\
(0.01, 0.255) & \rightarrow (0.01, 0.255) \\
(0.02, 0.35) & \rightarrow (0.02, 0.35) \\
(0.05, 0.459) & \rightarrow (0.05, 0.459) \\
(0.10, 0.63) & \rightarrow (0.10, 0.63) \\
(0.18, 0.78) & \rightarrow (0.14, 0.71) \\
(0.30, 0.864) & \rightarrow (0.23, 0.805) \\
(0.55, 0.955) & \rightarrow (0.30, 0.87) \\
(0.60, 0.97) & \rightarrow (0.35, 0.889) \\
(0.75, 0.991) & \rightarrow (0.55, 0.955) \\
(1, 1) & \rightarrow (1, 1)
\end{align*}
\]

TABLE 4.2. Configuration Comparison for Binary Messages

(i). \( \eta = 0.4 \) \( [P(H_0) = 0.2857] \)

Operating Point of the Consulting DM:
\[(0.5, 0.91) \rightarrow (0.5, 0.9)\]

Operating Point of the Primary DM when \( u_c = 0 \):
\[(0.1, 0.5) \rightarrow (0.1, 0.51)\]

Operating Point of the Primary DM when \( u_c = 1 \):
\[(0.5, 0.9) \rightarrow (0.5, 0.91)\]

Operating Point of the Team:
\[(0.30, 0.864) \rightarrow (0.30, 0.87)\]

Probability of Error:
\[0.1829 \rightarrow 0.1786^*\]

(ii). \( \eta = 1.0 \) \( [P(H_0) = 0.5] \)

Operating Point of the Consulting DM:
\[(0.2, 0.7) \rightarrow (0.1, 0.5)\]

Operating Point of the Primary DM when \( u_c = 0 \):
\[(0.1, 0.5) \rightarrow (0.2, 0.7)\]

Operating Point of the Primary DM when \( u_c = 1 \):
\[(0.5, 0.9) \rightarrow (0.5, 0.91)\]

Operating of the Team Point:
\[(0.18, 0.78) \rightarrow (0.23, 0.805)\]

Probability of Error:
\[0.20^* \rightarrow 0.2125\]

* Optimal
4.3.5. The Number of Messages

Consider again the example of the previous section and suppose that the number of messages that the consulting DM can transmit to the primary DM is increased from two to three. Then, if the worse DM is made the consulting DM, the team achieves the optimal centralized performance because the consulting DM can transmit his observation to the primary DM. The same result can not be achieved when the better DM is made the consulting DM. Hence, the ROC curve of the team with the worse DM as the consulting DM lies over the ROC curve of the team with the better DM as the consulting DM (Figure 4.11 and Table 4.3). Thus, for the case of three messages making the better DM the primary is optimal even for $\eta = 1.0$ (Table 4.4). Therefore, the optimal team architecture depends also on the number of messages.

The fact that the optimal team architecture depends on the number of messages is not surprising, but the way it does is counterintuitive. Intuitively we think that as the number of messages increases it becomes more likely for the better DM to be placed as the consulting DM in the optimal configuration, because as the number of messages increases the loss of information caused by the fusion of the observation of the consulting DM to a message decreases. This is especially obvious in the two limit cases; in the zero message (isolation) case, the better DM should be the primary DM, thus making the team decision and in the infinite message (centralized) case, the better DM can be made the consulting DM without causing any deterioration in the team performance. In our particular example assuming $\eta = 1.0$, increasing the number of messages from two to three makes the better DM change his role in the optimal team architecture from being the consulting DM to being the primary DM; a counterintuitive result.
Figure 4.11(a). Team ROC Curves for Both Configurations and Ternary Messages

Figure 4.11(b). Close Up of Team ROC Curves of Figure 4.11(a).
Table 4.3. The Team ROC Curves of Figure 4.11.

\[
\begin{array}{c}
\text{B} \rightarrow \text{W} \\
(0, 0) \\
(0.01, 0.255) \\
(0.02, 0.35) \\
(0.06, 0.554) \\
(0.09, 0.659) \\
(0.13, 0.735) \\
(0.23, 0.831) \\
(0.35, 0.915) \\
(0.40, 0.934) \\
(0.75, 0.999) \\
(1, 1)
\end{array}
\quad
\begin{array}{c}
\text{W} \rightarrow \text{B} \\
(0, 0) \\
(0.01, 0.255) \\
(0.02, 0.35) \\
(0.06, 0.554) \\
(0.09, 0.659) \\
(0.13, 0.735) \\
(0.18, 0.786) \\
(0.23, 0.831) \\
(0.35, 0.915) \\
(0.40, 0.934) \\
(0.75, 0.999) \\
(1, 1)
\end{array}
\]

Table 4.4. Configuration Comparison for Ternary Messages

\[
\eta = 1.0 \quad [P(H_0) = 0.5]
\]

Operating Point (0 vs. 1) of the Consulting DM: (0.5, 0.91)  (0.5, 0.9)
Operating Point (1 vs. 2) of the Consulting DM: (0.2, 0.7)  (0.1, 0.5)
Operating Point of the Primary DM when \( u_e = 0 \): (0.0, 0.0)  (0.1, 0.51)
Operating Point of the Primary DM when \( u_e = 1 \): (0.1, 0.5)  (0.2, 0.7)
Operating Point of the Primary DM when \( u_e = 2 \): (0.5, 0.9)  (0.5, 0.91)
Operating Point of the Team: (0.13, 0.735)  (0.18, 0.786)
Probability of Error: 0.1975  0.197*

* Optimal
4.3.6. **Performance Bounds**

In section 4.3.4 it was shown that the intuitively appealing suggestion of designating the better DM as the primary DM is not optimal in general. Still, it is optimal for several probability density functions (as it will be seen in section 4.3.7 below) and even in cases where it is not true (like the counterexample presented above) both architectures have very similar performance. Therefore, it is logical to designate the better DM to be the primary DM and try to obtain a bound on the deterioration of the team performance. A meaningful bound is the deterioration of the team performance relatively to the optimal team performance.

But, consider the following *Example 4.2*, for which the optimal team architecture requires that the better DM be the primary DM. The ROC curves for the two DMs are presented in Figure 4.12 and the associated discrete probability density functions conditioned on the two hypotheses are presented in Figure 4.13. Suppose that \( \eta = 1.0 \). The performance of the team will be measured by the normalized probability of error defined in Eqs.(3.9) and (3.10) as usual.

Consider any \( m \) such that:

\[
m > 1
\]  

(4.38)

and any \( \varepsilon \) such that:

\[
\min\{1/2, \ 1/m\} > \varepsilon > 0
\]  

(4.39)

Suppose that the better DM is the consulting DM. The optimal operating points can be found on Table 4.5 and the optimal normalized probability of error is:

\[
P_1^\varepsilon = \frac{\varepsilon^2}{m + 1}
\]  

(4.40)
Figure 4.12. The DMs of Example 4.2.

Figure 4.13. Probability Distributions for the DMs of Figure 4.12.
Table 4.5. Configuration Comparisons for Example 4.2.

\[ \eta = 1.0 \quad [P(H_0) = 0.5] \]

Operating Point of the Consulting DM:
\[ \left( \frac{\varepsilon}{m + 1}, 1 - \frac{\varepsilon}{m + 1} \right) \quad (0, 1- \varepsilon) \]
Operating Point of the Primary DM when \( u_c = 0 \):
\[ (0, 1- \varepsilon) \quad (0, 1- \varepsilon) \]
Operating Point of the Primary DM when \( u_c = 1 \):
\[ (\varepsilon, 1) \quad (\varepsilon, 1) \]
Operating Point of the Team:
\[ \left( \frac{\varepsilon^2}{m + 1}, 1 - \frac{\varepsilon^2}{m + 1} \right) \quad (0, 1- \varepsilon^2) \]
Probability of Error:
\[ \frac{\varepsilon^2}{m + 1} \ast \quad \frac{\varepsilon^2}{2} \]

* Optimal

Now suppose that the better DM is the primary DM. The optimal operating points can also be found on Table 4.5 and the optimal normalized probability of error is:

\[ P_2^e = \frac{\varepsilon^2}{2} \quad (4.41) \]

Then the deterioration of the team performance is:

\[ \Delta P^e = P_2^e - P_1^e = \frac{m - 1}{m + 1} \frac{\varepsilon^2}{2} > 0 \quad (4.42) \]

and the relative deterioration of the team performance is:

\[ \omega = \frac{\Delta P^e}{P_1^e} = \frac{m - 1}{2} \quad (4.43) \]

Since we can choose any \( m > 1 \) we conclude that the relative deterioration of the team performance can not be bounded this way. But also note that, as \( m \to \infty \), the absolute magnitude of the deterioration of the team performance goes to zero; thus as the relative deterioration increases, the absolute magnitude of the deterioration decreases.
4.3.7. Special Probability Distributions

It should be evident that for the conjecture to exist there are many cases where it holds true. As we already saw in section 4.3.4, the conjecture does not hold necessarily true for discrete distributions. But, our numerical analysis suggests that it holds true for the case of comparing means of Gaussian distributions with equal variance. Unfortunately, due to the complexity of the improper integrals involved, no theoretical results were obtained to substantiate our numerical findings. A similar empirical result exists for the case of the conic ROC curves, where in order to get a rigorous proof we need to analytically obtain the roots of a sixteenth order polynomial.

On the other hand, in the case of exponential distributions with different rates (or equivalently of comparing variances of Gaussian distributions with equal means, when each DM receives two observations) not only it is better to designate the better DM as the primary DM, but also we were able to obtain a proof. Thus consider two such DMs; from section 3.4.2 we know that the worse DM will have variance \( \alpha = \sigma_0^2/\sigma_1^2 \) with \( \sigma_1 > \sigma_0 \) and the better DM will have variance \( \alpha /k \) for some \( k > 1 \). Then the ROC curve of the worse DM is given by:

\[
P^*_D = (P_F)^\alpha
\]  

(4.44)

and the ROC curve of the better DM is given by:

\[
P^b_D = (P_F)^{\alpha/k}
\]  

(4.45)

Suppose that the better DM is made the primary. Then from Eqs.(4.3)-(4.5) and the property of the tangent to the ROC curve:

\[
\eta_c = \frac{P^1_F - P^0_F}{P^1_D - P^0_D} \eta = \frac{dP^b_D}{dP_F} \bigg|_{(P^*_F, P^*_D)} = \frac{\alpha}{k} \frac{P^c_D}{P^c_F}
\]  

(4.46)
\[ \eta_0 = \frac{1-P_F^c}{1-P_D^c} \eta = \frac{dP_D^w}{dP_F} \bigg|_{p_D^0, p_D^0} = \alpha \frac{P_D^0}{P_F^0} \quad (4.47) \]

and:

\[ \eta_1 = \frac{P_F^c}{P_D^c} \eta = \frac{dP_D^w}{dP_F} \bigg|_{p_F^1, p_D^1} = \alpha \frac{P_D^1}{P_F^1} \quad (4.48) \]

Solving the system of Eqs.(4.46)-(4.48) and recalling the concavity of the ROC curve, we obtain that:

\[ (P_F^1, P_D^1) = (1,1) \]

which implies that whenever \( u_c = 1 \) is received from the consulting DM, the primary DM decides \( u_p = 1 \) independent of his own observation. Substituting into Eqs.(4.8) and (4.9), we obtain that the team ROC curve in this case is given parametrically by:

\[ P_F^t = P_F^0 + P_F^c - P_F^0 P_F^c \quad (4.49) \]

and:

\[ P_D^t = P_D^0 + P_D^c - P_D^0 P_D^c \quad (4.50) \]

for some operating points \((P_F^0, P_D^0)\) of the worse DM and \((P_F^c, P_D^c)\) of the better DM.

Now suppose that the better DM is made the primary DM. Then, we can arbitrarily assign to the DMs the following operating points:

\[ (P_F^0, P_D^0): \quad \text{to the new consulting (worse) DM} \]

\[ (P_F^c, P_D^c): \quad \text{to the new primary (better) DM when } u_c = 0 \text{ is received} \]

\[ (1,1): \quad \text{to the new primary DM when } u_c = 1 \text{ is received} \]

Substituting into Eqs.(4.8) and (4.9) we obtain Eqs.(4.49) and (4.50) again. Since for this arbitrary assignment of operating points, the architecture with the better DM as the primary DM can achieve performance equal to the optimal performance of the other architecture, the better DM should always be the primary DM.
4.3.8. Propagation of ROC Curves

Suppose that the better DM is made the consulting DM in the example of section 4.3.7 above. Then from the system of Eqs.(4.44)-(4.48), we can solve for $P^\xi_F$ to obtain:

$$P^\xi_F = \left[ k \sigma^2 (1 - P^\xi_D) \eta \frac{\left[ \sigma_0^2 (1 - P^\xi_D) \right]^{-\gamma} - \left[ \sigma_1^2 (1 - P^\xi_D) \eta \right]^{-\gamma} - \delta}{\left[ \sigma_0^2 (1 - P^\xi_D) \right]^{-\beta} - \left[ \sigma_1^2 (1 - P^\xi_D) \eta \right]^{-\beta}} \right]$$  \hspace{1cm} (4.51)

where:

$$\beta = \frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2}$$  \hspace{1cm} (4.52a)

$$\gamma = \frac{\sigma_1^2}{\sigma_1^2 - \sigma_0^2}$$  \hspace{1cm} (4.52b)

and:

$$\delta = \frac{k \sigma_1^2}{k \sigma_1^2 - \sigma_0^2}$$  \hspace{1cm} (4.52c)

This is an equation of just $P^\xi_F$. We could have substituted for $P^\xi_D$ from Eq.(4.45), but did not do it because of space limitations. If the equation is solved $P^\xi_F$ is obtained. Moreover:

$$P^\xi_D = \left[ \alpha \frac{1 - P^\xi_F}{1 - P^\xi_D} \eta \right]^{-\gamma}$$  \hspace{1cm} (4.53)

By substituting into Eq.(4.45), $P^\xi_D$ is obtained. Finally by substituting for all the probabilities into Eqs.(4.49) and (4.50), the team ROC curve is obtained as a function of $\eta$, the variances and $k$.

It should be clear that the team ROC curve will not be of the same form as the ROC curves of the individual DMs. In fact, it can not be given in a closed form expression. Thus, the results cannot be easily extended to the case of three DMs in tandem even for these particular probability distributions.
4.3.9. Updating the Conjecture

From the above discussions one may conclude that the failure of the conjecture is a result of discrete distributions. It is not, because we can construct continuous distributions which have piecewise linear ROC curves (for example, a series of uniform distributions). Moreover, strictly concave ROC curves will not make the conjecture true, since strictly concave curves which approximate the piecewise linear ROC curves within \( \epsilon \) can be constructed. If more work is to be devoted in proving the validity of a similar conjecture we suggest that particular probability distributions and ROC curve shapes be examined.

4.3.10. The Geometric Approaches

For the sake of completeness, we will mention two geometric approaches for the solution of this problem. The first method was originally presented in [E82]. Suppose that the operating points \((P_F^0, P_D^0)\) and \((P_F^1, P_D^1)\) of the primary DM have been determined somehow and that the optimum threshold of the consulting DM remains to be determined.

The operating points of the primary DM define a rectangle. Reduce the ROC curve of the consulting DM so that it fits exactly the rectangle (i.e., multiply its width by \((P_F^1 - P_F^0)\) and its height by \((P_D^1 - P_D^0)\) ) and place it in that rectangle (Figure 4.14). Then the optimal operating point of the consulting DM is the point of the reduced ROC curve which has tangent with slope \( \eta^* \). To prove this, consider any point \((P_F, P_D)\) on the (original) ROC curve of the consulting DM; this point will have one-to-one correspondence with the point:

\[
\left( (1-P_F^0) P_F^0 + P_F^1 P_F^1, (1-P_D^0) P_D^0 + P_D^1 P_D^1 \right)
\]

of the placed ROC curve. Moreover we know that as in the usual cases the point with tangent slope \( \eta^* \) will be optimal. This method is especially useful for solving the problem when the ROC curves are piecewise linear and for obtaining bounds.
In the second method consider the effects of the movement of the consulting DM's operating point along the ROC curve of the consulting DM (Figure 4.15). When the operating point of the consulting DM is at (0,0) the operating point \((P_F^1, P_D^1)\) of the primary DM is at (1,1) and the other operating point of the primary DM \((P_F^0, P_D^0)\) is at \((P_F^*, P_D^*)\), where \((P_F^*, P_D^*)\) is the maximum likelihood operating point of the primary DM if he was deciding in isolation for the same value of \(\eta^*\). Similarly, when the operating point of the consulting DM is at (1,1) the operating point \((P_F^1, P_D^1)\) of the primary DM is at \((P_F^*, P_D^*)\) and the other operating point of the primary DM \((P_F^0, P_D^0)\) is at (0,0). The optimality condition is that:

\[
m_c = \frac{P_F^1 - P_F^0}{P_D^1 - P_D^0} \eta
\]  

(4.54)
(a). The Operating Point of the Consulting DM

(b). The Operating Points of the Primary DM when $u_c = 0$ and when $u_c = 1$

Figure 4.15. The Second Geometric Approach

where $m_c$ is the slope of the tangent at the consulting DM's operating point. As the operating point moves along the consulting DM's ROC curve, the slope of its tangent decreases from infinity to zero in a continuous manner (by definition, of 3.4.1). At the same time, the right part of Eq.(4.54) also decreases continuously from a finite positive value to another (smaller) finite positive value. Thus the existence of optimal decision rules can also be established this way. Moreover, the fact that the sufficient optimality conditions are not necessary can be obtained since Eq.(4.54) can have more than one solutions.
4.4. TWO DMs IN PARALLEL

4.4.1. Problem Formulation and Solution

**PROBLEM 4.4.** A team consisting of two DMs, DM a and DM b, in parallel performs binary hypothesis testing (Figure 4.2). The costs \( J(u_i, H) \) which are incurred by the team when the decision of the team is \( u_i \) and \( H \) is the true hypothesis as well as the prior probabilities \( (p_i = P(H_i), \text{for } i = 0, 1) \) are assumed to be known. Each DM receives a conditional independent observation, makes a decision based on his observation and transmits a binary message to a fusion center. The fusion center makes its decision so as to minimize the team normalized probability of error. The decision rules for the two DMs which minimize the expected cost are to be determined.

The optimal decision rules of the DMs depend on the optimal decision rule of the fusion center. The fusion center tries to minimize the normalized probability of error of its decision and has two candidate decision rules; the AND decision rule (i.e., decide \( u_i = 1 \) if both DMs send \( u_a = 1 \) and \( u_b = 1 \)) and the OR decision rule (i.e., decide \( u_i = 1 \) if either \( u_a = 1 \) or \( u_b = 1 \)). If the AND decision rule is the optimal decision rule of the fusion center, the optimal decision rules of the DMs have to satisfy the following necessary conditions [E82].

For DM a:

\[
\Lambda(y_a) \begin{array}{c}
\leq \\
\geq
\end{array} \frac{p_i}{p_D} \eta \equiv \eta_a
\]

(4.55a)

and for DM b:

\[
\Lambda(y_b) \begin{array}{c}
\leq \\
\geq
\end{array} \frac{p_i}{p_D} \eta \equiv \eta_b
\]

(4.56a)
If the OR decision rule is the optimal decision rule of the fusion center, the optimal decision rules of the DMs have to satisfy the following necessary conditions [E82].

For DM $a$:

$$\Lambda(y_a) \begin{array}{c} u_a = 1 \\ u_a = 0 \end{array} \begin{array}{c} \frac{1-P^b_E}{1-P^b_D} \eta = \eta_a \\ \frac{1-P^a_E}{1-P^a_D} \eta = \eta_b \end{array} \quad (4.55b)$$

and for DM $b$:

$$\Lambda(y_b) \begin{array}{c} u_b = 1 \\ u_b = 0 \end{array} \begin{array}{c} \frac{1-P^a_E}{1-P^a_D} \eta = \eta_b \end{array} \quad (4.56b)$$

where $\eta$ was defined in Eq.(3.5) with $P^a_D$ and $P^a_F$ respectively the probability of detection and of false alarm of DM $n$, for $n = a, b$.

**Remark 1.** The decision thresholds of the two DMs are given by a set of coupled equations. For example, if the AND decision rule is employed by the fusion center:

$$P^b_D = P\left(\Lambda(y_b) \geq \frac{P^a_E}{P^a_D} \eta \bigg| H_1\right) \quad (4.57)$$

**Remark 2.** The ROC curve of the team as a whole can be computed; for example, if the AND decision rule is employed by the fusion center, the ROC curve of the team is given by the following two parametric equations:

$$P^b_F(\eta) = P^a_F(\eta) P^b_E(\eta) \quad (4.58)$$

$$P^b_D(\eta) = P^a_D(\eta) P^b_F(\eta) \quad (4.59)$$

**4.4.2. Second Order Optimality Conditions**

As was mentioned above the optimality conditions of the previous section are necessary conditions. We also derive the second order necessary conditions for optimality.
in order to enhance our understanding of and intuition about the team behavior. These conditions depend, as the first order conditions do, on whether the AND or the OR decision rule is the optimal rule employed by the fusion center.

**PROPOSITION 4.1.** Consider the team which consists of two DMs in parallel and a fusion center, and performs binary hypothesis testing (Figure 4.2).

(i). If the fusion center employs the AND rule as its optimal decision rule, the second order necessary optimality conditions are given by the following inequalities:

\[-\eta + \eta_a \eta_b \geq 0\]  
(4.60a)

or:

\[\alpha_a \alpha_b P_D^a P_D^b \geq (-\eta + \eta_a \eta_b)^2\]  
(4.61a)

where \(\alpha_n\) is the second derivative of the ROC curve at the operating point \((P_F^n, P_D^n)\), for

\(n = a, b\).

(ii). If the fusion center employs the OR rule as its optimal decision rule, the second order necessary optimality conditions are given by the following inequalities:

\[\eta - \eta_a \eta_b \geq 0\]  
(4.60b)

or:

\[\alpha_a \alpha_b (1 - P_D^a) (1 - P_D^b) \geq (\eta - \eta_a \eta_b)^2\]  
(4.61b)

**Proof.** Only Eqs.(4.60a) and (4.61a) need to be proved because the proof of Eqs. (4.60b) and (4.61b) follow by symmetry. To prove Eqs.(4.60a) and (4.61a), consider the two operating points \((P_F^a, P_D^a)\) and \((P_F^b, P_D^b)\) which satisfy the first order conditions and without loss of generality assume that \(P_F^a \leq P_F^b\). Perturb \(P_F^a\) to \(P_F^a + \varepsilon\), where \(\varepsilon\) is a real number of very small magnitude; then, by Taylor's theorem, the perturbed probability of detection will be \(P_D^a + \varepsilon \eta_a + 0.5\varepsilon^2 \alpha_a\). Similarly, perturb the operating point of the other DM to \((P_F^b - \delta, P_D^b - \delta \eta_b + 0.5\delta^2 \alpha_b)\), where \(\delta\) is a real number of very small magnitude with \(\varepsilon \delta > 0\); note that the two operating points should be perturbed in opposite directions so that they continue to satisfy the first order conditions.
For \((P_F^a, P_D^a)\) and \((P_F^b, P_D^b)\) to be globally optimal, they need to satisfy the following necessary condition:

\[
\frac{\eta}{\eta + 1} P_F^a P_F^b + \frac{1}{\eta + 1} (1 - P_D^a P_D^b) \leq \frac{\eta}{\eta + 1} (P_F^a + \varepsilon)(P_F^b - \delta) + \\
+ \frac{1}{\eta + 1} \left[1 - (P_D^a + \varepsilon \eta_a + 0.5 \varepsilon^2 \alpha_a)(P_D^b - \delta \eta_b + 0.5 \delta^2 \alpha_b)\right] \Rightarrow \\
0 \leq \varepsilon (\eta P_F^b - \eta_a P_D^b) + \delta (-\eta P_F^a + \eta_b P_D^a) + \varepsilon \delta (-\eta + \eta_a \eta_b) + \\
+ \varepsilon^2 (-0.5 \alpha_a P_D^b) + \delta^2 (-0.5 \alpha_b P_D^a) + H.O.T. \ (4.62)
\]

For global optimality, Eq.(4.62) should hold for both positive and negative perturbations \(\varepsilon\) and \(\delta\) which implies that the coefficients of \(\varepsilon\) and of \(\delta\) should be zero; these, as expected, are the first order conditions of Eqs.(4.55a) and (4.56a) respectively. Moreover, the coefficients of \(\varepsilon^2\) and of \(\delta^2\) are non-negative, because the ROC curve is concave and thus has a non-positive second derivative. Hence, if the coefficient of \(\varepsilon \delta\) is non-negative, that is if Eq.(4.60a) holds, then Eq.(4.62) holds as well.

Assume that Eq.(4.60a) does not hold; Eq.(4.62) may hold even in this case as long as:

\[
\phi(\varepsilon, \delta) \equiv \varepsilon^2 (-0.5 \alpha_a P_D^b) + \varepsilon \delta (-\eta + \eta_a \eta_b) + \delta^2 (-0.5 \alpha_b P_D^a) \geq 0 \quad (4.63)
\]

for \(\varepsilon\) and \(\delta\) sufficiently small with \(\varepsilon \delta > 0\).

We minimize \(\phi(\varepsilon, \delta)\) with respect to \(\varepsilon\) and to \(\delta\). Using simple calculus, the function is minimized with respect to \(\varepsilon\) if and only if:

\[
\varepsilon = \delta \frac{(-\eta + \eta_a \eta_b)}{\alpha_a P_D^b} \quad (4.64)
\]

Observe that the coefficient of \(\delta\) in Eq.(4.64) is positive (because we assumed that Eq. (4.60a) does not hold), so that \(\varepsilon \delta > 0\) as required. Substituting for \(\varepsilon\) from Eq.(4.64) into Eq.(4.63), we obtain that Eq.(4.62) is true if and only if Eq.(4.61a) is true. \textbf{Q.E.D.}
COROLLARY 4.1. Consider again the team of the previous proposition. Then, the second order necessary conditions for optimality can be written in the following equivalent form:

(i). If the fusion center employs the AND decision rule as its optimal decision rule:

\[-\eta + \eta_a \eta_b \geq - (\alpha_a \alpha_b P_D^a P_D^b)^{0.5}\]  (4.65a)

(ii). If the fusion center employs the OR decision rule as its optimal decision rule:

\[\eta - \eta_a \eta_b \geq - [\alpha_a \alpha_b (1 - P_D^a) (1 - P_D^b)]^{0.5}\]  (4.65b)

Proof. Again only Eq.(4.65a) needs to be shown, since the proof of Eq.(4.66b) follows by symmetry. We show first that if Eqs.(4.60a) and (4.61a) hold, then Eq.(4.65a) holds as well; if Eq.(4.60a) is true, then Eq.(4.65a) is obviously also true because the right hand side of Eq.(4.65a) is non-positive. Suppose that Eq.(4.60a) does not hold and that Eq.(4.61a) holds; then by taking the square root of each side of Eq.(4.61a), we obtain Eq.(4.65a).

To complete the proof we have to demonstrate that if Eq.(4.65a) is true, then Eqs. (4.60a) and (4.61a) are true as well; so suppose that Eq.(4.65a) holds and distinguish between two cases depending on whether \(-\eta + \eta_a \eta_b \geq 0\) or \(-\eta + \eta_a \eta_b < 0\). If the former is true, then Eq.(4.60a) is obviously true; if the latter is true, take the square of each side of Eq.(4.65a) to obtain Eq.(4.61a). Q.E.D.

4.4.3. Identical DMs

We now focus on the special case in which both DMs are identical. Because of the symmetry of the problem, it seems intuitive that the optimal decision rules of the two DMs will be identical or symmetric at least for the case where \(\eta = 1\) (i.e., perfect symmetry of the variables external to the team). It is known that the optimal decision rules neither have
to be identical nor have to be symmetric even if $\eta = 1$; a simple example with discrete probability density functions was presented in [T88]. We should note that such an example can be constructed for strictly concave ROC curves as well. These examples indicate that it is optimal in general for the team to split the risk unevenly among its members and suggest the need for different levels of command.

From an organizational designer's point of view it is desirable to assign identical decision rules to similar DMs having the same position in the organization; this not only significantly reduces the complexity of the problem, but it is also asymptotically optimal [T88]. So, we suppose that both DMs are restricted to employing identical decision rules and derive the first and second order optimality conditions; these follow directly from Eqs. (4.55a) and (4.56a), and from Corollary 4.1 and thus proofs are omitted.

**COROLLARY 4.2.** Consider the team which consists of two identical DMs in parallel and a fusion center, and performs binary hypothesis testing. Suppose the two DMs are restricted to employing identical decision rules. Then the optimal decision operating point $(P_F^n, P_D^n)$ of the ROC curve satisfies the following conditions:

(i) If the fusion center employs the AND decision rule as its optimal decision rule:

$$\frac{dP_D}{dP_F} \bigg|_{(P_F^n, P_D^n)} = \frac{P_F^n}{P_D^n} \eta$$  \hspace{1cm} (4.66a)

(ii) If the fusion center employs the OR decision rule as its optimal decision rule:

$$\frac{dP_D}{dP_F} \bigg|_{(P_F^n, P_D^n)} = \frac{1 - P_F^n}{1 - P_D^n} \eta$$  \hspace{1cm} (4.66b)

**REMARK:** Because of the concavity of the ROC curve there exists one and only one point on every ROC curve which satisfies each of Eq. (4.66a) and of Eq. (4.66b).
COROLLARY 4.3. Consider again the team of Corollary 4.2. Then, the second order necessary conditions for optimality are:

(i). If the fusion center employs the AND decision rule as its optimal decision rule:

\[- \eta + (\eta_n)^2 \geq \alpha_n P_D^a\]  \hspace{1cm} (4.67a)

(ii). If the fusion center employs the AND decision rule as its optimal decision rule:

\[\eta - (\eta_n)^2 \geq \alpha_n (1 - P_D^b)\] \hspace{1cm} (4.67b)

The above conditions suggest that DMs whose associated ROC curves have steep changes, implying large absolute values for the second derivative, are likely to have identical optimal decision rules (for example, ROC curves for nearly uniform noise).

4.4.4. Performance Bounds

As was mentioned above, it was shown in [T88] that as the number of DMs in a parallel team, which performs binary hypothesis testing, increase to infinity, the decision rules of the DMs may be restricted to be identical without any deterioration in the team performance. Since this is not true for the parallel team which consists of two DMs, we would like to determine whether bounds exist for the deterioration of its performance.

(i). Absolute Bound

Consider two identical DMs whose optimal decision rules are not identical for some given \(\eta^*\), as the ones in Figure 4.16(a); suppose also, without loss of generality, that the optimal decision rule for the fusion center is the AND decision rule. The operating points of the DMs have to satisfy the necessary optimality conditions of Eqs.(4.55a) and (4.56a); without loss of generality assume \(P_F^a < P_F^b\). Also consider two new identical DMs with the
three-piecewise linear ROC curve of Figure 4.16(b), whose underlying probability distributions are presented in Table 4.6. The new DMs are of course worse than the original DMs; still, for the particular $\eta^*$, the new (worse) team can achieve performance equal to the optimal performance of the original (better) team, since the operating points $(P_F^a, P_D^a)$ and $(P_F^b, P_D^b)$ can also be employed by the new (worse) DMs. Therefore, if the DMs are restricted to employing identical decision rules, an upper bound for the deterioration in the performance of the original (better) team, for the particular $\eta^*$, is given by the deterioration in the performance of the new (worse) team.

Thus, the problem is reduced to obtaining bounds for teams which consist of DMs with three-piecewise linear ROC curves\(^2\). Consider again the DMs of Figure 4.16(b) and Table 4.6. The set of possibly optimal decision rules may be found by exhaustive enumeration. Since each DM has to perform a likelihood ratio test, there are only two candidate decision rules for each DM $n$, $n = a, b$:

\[
\begin{align*}
&\text{(i) } u_n = 1 \text{ if and only if } y_n \in \{2\} \\
&\text{(ii) } u_n = 1 \text{ if and only if } y_n \in \{1, 2\}
\end{align*}
\]

Thus, we need to consider six cases:

1. **Both DMs employ (i) and the fusion center employs AND**
2. **Both DMs employ (ii) and the fusion center employs AND**
3. **DM a employs (i), DM b employs (ii) and the fusion center employs AND**
4. **Both DMs employ (i) and the fusion center employs OR**
5. **Both DMs employ (ii) and the fusion center employs OR**
6. **DM a employs (i), DM b employs (ii) and the fusion center employs OR**

For the given $\eta^*$, denote by $P_f^*\text{e}$ the team probability of error when the decision rules

---

\(^2\) The optimality of the Likelihood Ratio Quantizers (LRQs) was formalized and generalized in [T89a].
Figure 4.16(a). Identical DMs with Non-Identical Optimal Decision Rules

Figure 4.16(b). The 'Worse' Identical DMs with Non-Identical Optimal Decision Rules

Table 4.6. Probability Distributions for 'Worse' DMs of Figure 4.16(b).

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$1-P_F$</td>
<td>$P_F - P_D$</td>
<td>$P_D$</td>
</tr>
<tr>
<td>$H_1$</td>
<td>$1-P_D$</td>
<td>$P_D - P_F$</td>
<td>$P_F$</td>
</tr>
<tr>
<td>$y$</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
of case \([j], j = 1, \ldots, 6\), are employed and by \(\Delta P^e\) the minimum deterioration in the team performance, if both DMs are restricted to employing identical decision rules. Since case [3] was assumed to be the optimal set of decision rules:

\[
\Delta P^e \leq P_1^e - P_3^e = \frac{\eta^*}{\eta^* + 1} \left[ -P_F^a (P_F^b - P_F^a) \right] + \frac{1}{\eta^* + 1} \left[ P_D^b (P_D^b - P_D^a) \right] \equiv \Delta P_1^e \quad (4.69a)
\]

\[
\Leftrightarrow \frac{P_D^a (P_D^b - P_D^a)}{P_F^a (P_F^b - P_F^a)} \geq \eta^* \quad (4.69b)
\]

\[
\Delta P^e \leq P_2^e - P_3^e = \frac{\eta^*}{\eta^* + 1} \left[ P_F^b (P_F^b - P_F^a) \right] - \frac{1}{\eta^* + 1} \left[ P_D^b (P_D^b - P_D^a) \right] \equiv \Delta P_2^e \quad (4.70a)
\]

\[
\Leftrightarrow \eta^* \geq \frac{P_D^b (P_D^b - P_D^a)}{P_F^b (P_F^b - P_F^a)} \quad (4.70b)
\]

\[
\Delta P^e \leq P_4^e - P_3^e = \frac{\eta^*}{\eta^* + 1} \left[ P_F^a (1 - P_F^b) + P_F^a (1 - P_F^a) \right] - \frac{1}{\eta^* + 1} \left[ P_D^a (1 - P_D^b) + P_D^a (1 - P_D^a) \right] \equiv \Delta P_4^e \quad (4.71a)
\]

\[
\Leftrightarrow \eta^* \geq \frac{P_D^a (1 - P_D^b) + P_D^a (1 - P_D^a)}{P_F^a (1 - P_F^b) + P_F^a (1 - P_F^a)} \quad (4.71b)
\]

\[
\Delta P^e \leq P_5^e - P_3^e = \frac{\eta^*}{\eta^* + 1} \left[ P_F^b (1 - P_F^b) + P_F^b (1 - P_F^a) \right] - \frac{1}{\eta^* + 1} \left[ P_D^b (1 - P_D^b) + P_D^b (1 - P_D^a) \right] \equiv \Delta P_5^e \quad (4.72a)
\]

\[
\Leftrightarrow \eta^* \geq \frac{P_D^b (1 - P_D^b) + P_D^b (1 - P_D^a)}{P_F^b (1 - P_F^b) + P_F^b (1 - P_F^a)} \quad (4.72b)
\]

\[
0 \leq P_6^e - P_3^e = \frac{\eta^*}{\eta^* + 1} \left[ P_F^a (1 - P_F^b) + P_F^a (1 - P_F^a) \right] - \frac{1}{\eta^* + 1} \left[ P_D^a (1 - P_D^b) + P_D^a (1 - P_D^a) \right]
\]

\[
\Leftrightarrow \eta^* \geq \frac{P_D^a (1 - P_D^b) + P_D^a (1 - P_D^a)}{P_F^a (1 - P_F^b) + P_F^a (1 - P_F^a)} \quad (4.73)
\]
REMARK. Eq.(4.73) does not provide a bound on the deterioration of the team performance incurred by the restriction that both DMs employ identical decision rules, because, according to case [6], the DMs of the team do \textit{not} employ identical decision rules. Nevertheless, it provides a condition for \(\eta^*\) so that case [3] be the optimal set of decision rules employed by the team.

Using simple algebra and since \(P_D^a/P_F^a > P_D^b/P_F^b\) (Figure 4.19(b)), we obtain that, as long as Eq.(4.71b) holds true, Eq.(4.72b) holds true as well; this implies that the bound of Eq.(4.71a) is tighter than the bound of Eq.(4.72a). So, we need only to examine Eqs. (4.69a), (4.70a) and (4.71a) in order to obtain the maximum least upper bound for the deterioration of the team performance.

Furthermore, note that the bound of Eq.(4.69a) is decreasing with \(\eta^*\), while the bounds of Eqs.(4.70a) and (4.71a) are both increasing with \(\eta^*\). Therefore, the maximum least upper bound of Eqs.(4.69a) and (4.70a) occurs when the bounds are equal. Thus:

\[
P_1^e - P_3^e = P_2^e - P_3^e \iff \eta^* = \frac{P_D^b + P_D^a}{P_F^b + P_F^a} \frac{P_D^b - P_D^a}{P_F^b - P_F^a} \tag{4.74}
\]

Similarly, the maximum least upper bound of Eqs.(4.69a) and (4.71a) occurs when the bounds are equal. Thus:

\[
P_1^e - P_3^e = P_4^e - P_3^e \iff \eta^* = \frac{P_D^b}{P_F^b} \frac{1 - P_D^a}{1 - P_F^a} \tag{4.75}
\]

REMARK. Using simple algebra, we verify that the \(\eta^*\) of both Eqs.(4.74) and (4.75) satisfy Eq.(4.73) because:

\[
\frac{P_D^b + P_D^a}{P_F^b + P_F^a} \geq \frac{P_D^b - P_D^a}{P_F^b - P_F^a} = \frac{P_D^b}{P_F^b} \left( \frac{1 - P_D^a}{1 - P_F^a} \right)
\]

\[
= \frac{P_D^b (P_D^b - P_D^a) + (P_D^b + P_D^a) (1 - P_D^b)}{P_F^b (P_F^b - P_F^a) + (P_F^b + P_F^a) (1 - P_F^b)} = \frac{P_D^a (1 - P_D^b) + P_D^b (1 - P_D^a)}{P_F^a (1 - P_F^b) + P_F^b (1 - P_F^a)}
\]

and:
\[
\frac{P_D^a 1 - P_D^b}{P_F^a 1 - P_F^a} \geq \max \left\{ \frac{P_D^a (1 - P_D^b)}{P_F^a (1 - P_F^b)}, \frac{P_D^b (1 - P_D^a)}{P_F^b (1 - P_F^a)} \right\} \geq \frac{P_D^a (1 - P_D^b) + P_D^b (1 - P_D^a)}{P_F^a (1 - P_F^b) + P_F^b (1 - P_F^a)}
\]

Substituting for \( \eta^* \) from Eq.(4.74) into Eq.(4.69a) (or into Eq.(4.70a) ), we obtain the maximum least upper bound of Eqs.(4.69a) and (4.70a). Substituting for \( \eta^* \) from Eq.(4.75) into Eq.(4.69a) (or into Eq.(4.70a) ), we obtain the maximum least upper bound of Eqs.(4.69a) and (4.71a). Thus, the globally maximum least upper bound is obtained when both the above bounds are equal, that is when:

\[
\frac{P_D^b + P_D^a}{P_F^b + P_F^a} \frac{P_D^b - P_D^a}{P_F^b - P_F^a} = \frac{P_D^a 1 - P_D^b}{P_F^a 1 - P_F^b}
\]

\( \Leftrightarrow (P_D^b)^2 = \frac{P_D^a 1 - P_D^b}{P_F^a 1 - P_F^b} (P_F^b)^2 + \frac{P_D^a (P_D^b - P_F^b)}{1 - P_D^a} \)

(4.76a)

(4.76b)

To maximize the least upper bound, the two operating points of the DMs have to satisfy Eq.(4.76b); it is not hard to obtain (see the Appendix to this chapter) that in order to maximize the least upper bound for a fixed \((P_F^a, P_D^a)\), subject to Eq.(4.76b), the operating point \((P_F^b, P_D^b)\) has to satisfy:

\[
(P_F^b, P_D^b) = \left( \left[ \frac{(P_F^b)^2 + (1 + P_D^a) (1 - P_F^b) P_F^a}{P_D^a} \right]^{0.5}, 1 \right)
\]

\[
= \left( \left[ \frac{P_F^a}{P_D^a} (1 + P_D^a - P_F^a) \right]^{0.5}, 1 \right)
\]

(4.77)

Moreover, substituting from Eqs.(4.75) and (4.77) into Eq.(4.69a), we obtain that the maximum least upper bound of the deterioration in the team performance as a function of \((P_F^a, P_D^a)\) is given by:

\[
\Delta P^e \leq \frac{P_D^a 1 - P_D^b}{1 - P_F^a} \left[ 1 - \left[ \frac{P_D^a}{P_D^a} (1 + P_D^a - P_F^a) \right]^{0.5} \right]
\]

(4.78)
Figure 4.17. Probability Distributions that Maximize the Absolute Bound

<table>
<thead>
<tr>
<th>Decision Rules of DMs</th>
<th>Fusion Center Decision Rule</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[i] and [i]</td>
<td>AND 0.22617857†</td>
<td>OR 0.22617857†</td>
<td></td>
</tr>
<tr>
<td>[ii] and [ii]</td>
<td>0.22617857†</td>
<td>0.51834035</td>
<td></td>
</tr>
<tr>
<td>[i] and [ii]</td>
<td>0.18640251*</td>
<td>0.41202552</td>
<td></td>
</tr>
</tbody>
</table>

\( \Delta P^e = 0.03977607 \)
* Globally Optimal

\( \eta = 1.58190398 \quad [P(H_0) = 0.61268893] \)
† Optimal for Identical Decision rules

In order to obtain the largest absolute deviation from optimality, we have to maximize the bound of Eq.(4.78) with respect to \( P_F^a \) and \( P_D^a \), subject to:

\[
0 \leq P_F^a \leq P_D^a \leq 1
\]  

(4.79)

Using calculus, the bound is maximized\(^3\) for \((P_F^a, P_D^a) = (0.16543756, 0.67773498)\), giving \( P_F^b = 0.60758314 \) and:

\[
\Delta P^e \leq 0.03977607
\]

(4.80)

\(^3\) To be more specific, we expressed \( P_F^a \) and \( P_D^a \) in terms of the slopes \( M = P_D^a/P_F^a \) and \( m = (1 - P_D^a)/(1 - P_F^a) \), and then maximized with respect to \( M \) and \( m \) subject to: \( M \geq 1 \geq m \).
Note that this is a tight bound; it is achieved by a parallel team which consists of two DMs, whose underlying probability distributions are presented in Figure 4.17, for $\eta^* = 1.58190398$ obtained from Eq.(4.74) (Table 4.7). Thus, assuming minimum probability of error cost structure, the absolute deviation is not maximized for equal prior probabilities, but for: $P(H_0) = 0.61268893$.

(ii). Relative Bound

Consider the exact same team used in the analysis of the absolute bound, the same given $\eta^*$ and again denote by $P_3^e$ the optimal probability of error of the two DM parallel team. Also, denote by $P_3^s$ the probability of error when the DMs are employing some other decision rules; then define the relative deterioration of the team performance as:

$$\omega^e \equiv \frac{P_s^e - P_3^e}{P_3^e} \quad (4.81)$$

Proceeding with analysis similar to the above, in order to obtain the least upper bound on the relative deterioration, only the following three relative bounds, derived from Eqs.(4.70a), (4.71a) and (4.72a) respectively, need to be considered:

$$\omega^e \leq \frac{P_1^e - P_3^e}{P_3^e} = \frac{\eta^*(P_F^b)^2 + \left[1 - (P_D^b)^2\right]}{\eta^*P_F^aP_D^b + \left[1 - P_D^aP_D^b\right]} - 1 \equiv \omega_1^e \quad (4.82)$$

$$\omega^e \leq \frac{P_2^e - P_3^e}{P_3^e} = \frac{\eta^*(P_F^b)^2 + \left[1 - (P_D^b)^2\right]}{\eta^*P_F^aP_D^b + \left[1 - P_D^aP_D^b\right]} - 1 \equiv \omega_2^e \quad (4.83)$$

$$\omega^e \leq \frac{P_4^e - P_3^e}{P_3^e} = \frac{\eta^*\left[1 - (1 - P_D^b)^2\right] + (1 - P_D^b)^2}{\eta^*P_F^aP_D^b + \left[1 - P_D^aP_D^b\right]} - 1 \equiv \omega_4^e \quad (4.84)$$

To maximize the least upper bound of Eqs.(4.82)-(4.84), Eqs.(4.76) and (4.77) have to hold again (see the Appendix); substituting from Eqs.(4.76) and (4.77) into Eq.(4.82):
\[
\omega^e \leq \frac{1 + P_D^a - P_F^a}{1 - P_F^a + P_D^a \left[ \frac{P_F^a}{P_D^a} (1 + P_D^a - P_F^a) \right]^{0.5}} - 1 \quad (4.85)
\]

In order to obtain the largest absolute deviation from optimality, we have to maximize the bound of Eq.(4.85) with respect to \(P_F^a\) and \(P_D^a\), subject to Eq.(4.79). Using calculus, the bound is maximized for \((P_F^a, P_D^a) \to (0, 1)\) giving:

\[
\omega^e \leq 1 \quad (4.86)
\]

The bound of Eq.(4.86) is a tight bound, since it can be achieved by a team which consists of two DMs, whose underlying probability distributions are presented in Figure 4.18, as \(\varepsilon \to 0\). It is achieved for \(\eta^* = 1.0\) obtained from Eq.(4.75), as can be seen in Table 4.8. Note that this 100% bound is obtained as the team probability of error is going to zero. As the optimal probability of error increases the relative deterioration in the team performance is considerably smaller; in fact, the relative deterioration for the example of Table 4.7, in which the absolute deterioration in performance is maximized, is only 21.3388%.

These results suggest, that it could be worthwhile for the designer of the team to incur some additional error and in the same time considerably simplify the complexity of the problem by restricting similar DMs to employ the same decision rules.

![Figure 4.18](image-url)

**Figure 4.18.** Probability Distributions that Maximize the Relative Bound
TABLE 4.8. Probability of Error of Team of Identical DMs of Figure 4.18.

<table>
<thead>
<tr>
<th>Decision Rules of DMs</th>
<th>Fusion Center Decision Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i] and [i]</td>
<td>$\varepsilon$ †</td>
</tr>
<tr>
<td>[ii] and [ii]</td>
<td>$\varepsilon$ †</td>
</tr>
<tr>
<td>[i] and [ii]</td>
<td>$0.5\varepsilon + 0.705\varepsilon^2$ *</td>
</tr>
</tbody>
</table>

$\omega^* = 1.0$

* Globally Optimal

$\eta = 1.0$ [$P(H_0) = 0.5$]

† Optimal for Identical Decision rules

4.5. TEAMS OF THREE DMs

There exist four different acyclic architectures for a team which consists of three DMs; the two consultant or $V$-architecture (Figure 4.19), the three DM tandem architecture (Figure 4.20), the three DM parallel architecture (Figure 4.21) and the asymmetrical architecture (Figure 4.22). We compare the performance of the different architectures and try to determine whether a dominant architecture exists.

It is worthwhile to understand the changes in the complexity of the decision rules when the number of team members increases from two to three. Three thresholds describe the decision rules of both the two DM tandem architecture (one for the consulting DM and two for the primary DM) and for the two DM parallel architecture (one for each DM and one for the fusion center). Six thresholds describe the decision rules for the $V$-architecture (one for each consulting DM and four for the primary DM), five thresholds describe the decision rules for both the three DM tandem architecture (one for the second consulting DM, two for the first consulting DM and two for the primary DM) and the three DM asymmetrical architecture (two for DM B and one for DM C, DM A and the fusion center), and four thresholds describe the decision rules for the three DM parallel architecture (one for each DM and one for the fusion center). It should be clear that the complexity of the
Figure 4.19. The Two Consultant or V-Architecture

Figure 4.20. The Three DM Tandem Architecture

Figure 4.21. The Three DM Parallel Architecture
Figure 4.22. The Three DM Asymmetrical Architecture

decision rules depends not only on the number of the DMs in the team, but also on the particular architecture of the team. Therefore, we will compare the performance of the different architectures for the teams of three DMs, keeping in mind that the complexity and communication requirements are not the same for all architectures. We should also note that the equations which define the optimal decision rules for the members of the three DM teams are of the same form as, though more complex than, the equations which define the optimal decision rules for the members of the two DM teams⁴.

As was shown in Lemma 4.1, the two DM tandem architecture is superior to the two DM parallel architecture; it will be interesting to determine whether a dominant architecture also exists for the teams which consist of three DMs. It is not hard to construct proofs, similar to the proof of Lemma 4.1, to show that the parallel and asymmetrical architectures cannot perform better than the V-architecture. Therefore, to determine the dominant three DM architecture, if such exists, we have to compare the V-architecture and the tandem architecture.

In several places in the literature it is conjectured and supported by numerical results that the V-architecture is better than the tandem architecture [E82], [R87], [RN87a]. But,

⁴ The solutions for the optimal decision rules of three DM tandem architecture and of the V-architecture have appeared in [E82], [R87] and [RN87a].
consider these two architectures when the primary DM is a very bad DM, that is when the primary DM's observation is extremely unreliable; the $V$-architecture for all practical purposes reduces to the two DM parallel architecture (Figure 4.2), while the three DM tandem architecture reduces to the two DM tandem architecture (Figure 4.1). Then, according to Lemma 4.1, in that particular case the tandem architecture is superior to the $V$-architecture. Thus, the comparison of the three DM tandem and the $V$-architecture depends on the particular DMs involved.

Subsequently, it was suggested that the 'best' $V$-architecture achieves superior performance to the 'best' tandem architecture, where 'best' refers to the optimal configuration of the DMs in a particular architecture. This suggestion can not be tested in general because, as was shown in section 4.3, the 'best' optimal configuration of the DMs in a particular architecture depends not only on the DMs of the team, but also on variables external to the team (i.e., costs, prior probabilities). But, in the special case where all three DMs of the team are identical, there exists (obviously) only one configuration for each architecture.

We would thus like to compare the $V$-architecture and the three DM tandem architecture for the case where all three DMs are identical. Then the problem can be reduced to comparing the performance of these two architectures for DMs whose ROC curve is piecewise linear with at most six line segments (since the tandem architecture has at most five distinct thresholds). Proceeding with analysis similar to section 4.3.3, we obtain the following examples.

If the identical DMs receive binary observations (Figures 4.23 and 4.24), the $V$-architecture achieves centralized performance, while the tandem team does not; thus in this case the $V$-architecture is superior to the tandem architecture (Figure 4.25 and Table 4.9).

But, the tandem architecture which consists of three identical DMs, whose ROC curve and underlying probability distributions are presented in Figure 4.26 and 4.27 respectively, achieves better performance to the $V$-architecture for $\eta = 2.2$, and achieves worse
Figure 4.23. ROC Curve of DM with Binary Observations

Figure 4.24. Probability Distributions for DM of Figure 4.23.
Figure 4.25. Team ROC Curves for DMs of Figure 4.23; Dominant V-Architecture

TABLE 4.9. The Team ROC Curves of Figure 4.25.

<table>
<thead>
<tr>
<th>TANDEM</th>
<th>TWO CONSULTANT (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0, 0.0)</td>
<td>(0.0, 0.0)</td>
</tr>
<tr>
<td>(0.064, 0.216)</td>
<td>(0.064, 0.216)</td>
</tr>
<tr>
<td>(0.256, 0.504)</td>
<td>(0.352, 0.648)</td>
</tr>
<tr>
<td>(0.496, 0.744)</td>
<td>(0.784, 0.936)</td>
</tr>
<tr>
<td>(0.784, 0.936)</td>
<td>(1.0, 1.0)</td>
</tr>
<tr>
<td>(1.0, 1.0)</td>
<td></td>
</tr>
</tbody>
</table>
performance for $\eta = 1.0$ (Table 4.11). Thus, a dominant architecture for a three DM team may not exist and the optimal architecture may depend on $\eta$, even if all the DMs are identical. The team ROC curves for both the $V$-architecture and the tandem architecture can be seen in Figure 4.28 (and Table 4.10).

Furthermore, the tandem architecture which consists of the DMs, whose ROC curve and underlying probability distributions are presented in Figures 4.29 and 4.30 respectively, is superior to the $V$-architecture for all values of the threshold $\eta$, as can be seen from the team ROC curves of Figure 4.31 (and Table 4.12). Thus, the tandem architecture may be superior to the $V$-architecture.

Therefore, a 'globally' dominant architecture for the teams which consist of three DMs does not exist, even if all the DMs are identical; the optimal architecture depends on the DMs involved and on parameters external to the team (i.e., prior probabilities and costs).

4.6. SUMMARY

In this chapter we analyzed the architectures of teams which consist of two or three DMs. We showed that for teams of two DMs the tandem architecture dominates the parallel architecture and that for teams of three or more DMs the optimal architecture depends not only on the characteristics of the team (i.e., the DMs and communications), but also on the variables external to the team (i.e., prior probabilities and misclassification costs).

We disproved two conjectures, one claiming that in the two DM tandem architecture the better DM should be the primary DM, and the other claiming that the optimal architecture for teams of three DM is the two consultant architecture. We also obtained bounds on the team performance for some simple decision schemes, in order to establish the trade off between the complexity of the decision rules and the team performance.
Figure 4.26. DM who Does Not Result in a Dominant Architecture

Figure 4.27. Probability Distributions for DM of Figure 4.26.
Figure 4.28(a). Team ROC Curves for DMs of Figure 4.26; No Dominant Architecture

Figure 4.28(b). Close Up of Team ROC Curves of Figure 4.28(a).
<table>
<thead>
<tr>
<th>TANDEM</th>
<th>TWO CONSULTANT (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0, 0.0)</td>
<td>(0.0, 0.0)</td>
</tr>
<tr>
<td>(0.0, 0.271)</td>
<td>(0.027, 0.271)</td>
</tr>
<tr>
<td>(0.027, 0.412)</td>
<td>(0.027, 0.394)</td>
</tr>
<tr>
<td>(0.081, 0.568)</td>
<td>(0.081, 0.541)</td>
</tr>
<tr>
<td>(0.144, 0.646)</td>
<td></td>
</tr>
<tr>
<td>(0.243, 0.757)</td>
<td>(0.207, 0.793)</td>
</tr>
<tr>
<td>(0.354, 0.856)</td>
<td></td>
</tr>
<tr>
<td>(0.432, 0.919)</td>
<td>(0.459, 0.919)</td>
</tr>
<tr>
<td>(0.588, 0.973)</td>
<td>(0.606, 0.973)</td>
</tr>
<tr>
<td>(0.729, 1.0)</td>
<td>(0.729, 1.0)</td>
</tr>
<tr>
<td>(1.0, 1.0)</td>
<td>(1.0, 1.0)</td>
</tr>
</tbody>
</table>
### Table 4.11. Architecture Comparisons: *Optimal* Architecture Depends on $\eta$

(i). $\eta = 2.2$ [$P(H_0) = 0.3125$]

<table>
<thead>
<tr>
<th></th>
<th>Tandem</th>
<th>Two Consultant (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Point of DM C2:</td>
<td>(0.3, 0.7)</td>
<td>(0.3, 0.7)</td>
</tr>
<tr>
<td>Operating Point(s) of DM C1:</td>
<td>(0.0, 0.1) if $u_{c2} = 0$</td>
<td>(0.3, 0.7)</td>
</tr>
<tr>
<td></td>
<td>(0.3, 0.7) if $u_{c2} = 1$</td>
<td></td>
</tr>
<tr>
<td>Operating Points of DM P:</td>
<td>(0.0, 0.1) if $u_{c1} = 0$</td>
<td>(0.0, 0.1) if $u_{c1} + u_{c2} = 0$</td>
</tr>
<tr>
<td></td>
<td>(0.9, 1.0) if $u_{c1} = 1$</td>
<td>(0.0, 0.1) if $u_{c1} + u_{c2} = 1$</td>
</tr>
<tr>
<td></td>
<td>(0.9, 1.0) if $u_{c1} + u_{c2} = 2$</td>
<td></td>
</tr>
<tr>
<td>Team Operating Point:</td>
<td>(0.081, 0.568)</td>
<td>(0.081, 0.541)</td>
</tr>
<tr>
<td>Probability of Error:</td>
<td>0.1906875*</td>
<td>0.199125</td>
</tr>
</tbody>
</table>

(ii). $\eta = 1.0$ [$P(H_0) = 0.5$]

<table>
<thead>
<tr>
<th></th>
<th>Tandem</th>
<th>Two Consultant (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Point of DM C2:</td>
<td>(0.3, 0.7)</td>
<td>(0.3, 0.7)</td>
</tr>
<tr>
<td>Operating Point(s) of DM C1:</td>
<td>(0.0, 0.1) if $u_{c2} = 0$</td>
<td>(0.3, 0.7)</td>
</tr>
<tr>
<td></td>
<td>(0.9, 1.0) if $u_{c2} = 1$</td>
<td></td>
</tr>
<tr>
<td>Operating Points of DM P:</td>
<td>(0.0, 0.1) if $u_{c1} = 0$</td>
<td>(0.0, 0.1) if $u_{c1} + u_{c2} = 0$</td>
</tr>
<tr>
<td></td>
<td>(0.9, 1.0) if $u_{c1} = 1$</td>
<td>(0.3, 0.7) if $u_{c1} + u_{c2} = 1$</td>
</tr>
<tr>
<td></td>
<td>(0.9, 1.0) if $u_{c1} + u_{c2} = 2$</td>
<td></td>
</tr>
<tr>
<td>Team Operating Point:</td>
<td>(0.243, 0.757)</td>
<td>(0.207, 0.793)</td>
</tr>
<tr>
<td>Probability of Error:</td>
<td>0.243</td>
<td>0.207*</td>
</tr>
</tbody>
</table>

* Optimal
Figure 4.29. ROC Curve for DM Resulting in Dominant Tandem Architecture

Figure 4.30. Probability Distributions for DM of Figure 4.29.
Figure 4.31. Team ROC Curves for Figure 4.29; Dominant *Tandem* Architecture

**TABLE 4.12.** The Team ROC Curves of Figure 4.31.

<table>
<thead>
<tr>
<th>TANDEM</th>
<th>TWO CONSULTANT (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0, 0.0)</td>
<td>(0.0, 0.0)</td>
</tr>
<tr>
<td>(0.0, 0.271)</td>
<td>(0.0, 0.271)</td>
</tr>
<tr>
<td>(0.144, 0.46)</td>
<td>(0.16, 0.46)</td>
</tr>
<tr>
<td>(0.324, 0.676)</td>
<td>(0.336, 0.664)</td>
</tr>
<tr>
<td>(0.54, 0.856)</td>
<td>(0.54, 0.84)</td>
</tr>
<tr>
<td>(0.729, 1.0)</td>
<td>(0.729, 1.0)</td>
</tr>
<tr>
<td>(1.0, 1.0)</td>
<td>(1.0, 1.0)</td>
</tr>
</tbody>
</table>
APPENDIX. Obtaining the Bounds for the Two DM Parallel Team

(i). Absolute Bound

Suppose that \((P^a_F, P^a_D)\) is given; then the problem of maximizing the least upper bound of Eqs.(4.69a)-(4.71a) can be formulated as follows:

\[
\max_{(P^b_F, P^b_D) \in A} \Delta P^e_1 = \frac{\eta^*}{\eta^* + 1} [-P^a_F (P^b_F - P^b_D)] + \frac{1}{\eta^* + 1} \left[ P^a_D (P^b_D - P^b_D) \right] \tag{A.1}
\]

where \(A\) is the set that is defined by all the pairs \((P^b_F, P^b_D)\) satisfying the following four constraints:

\[
(P^b_D)^2 = \frac{P^a_D}{P^a_F} \frac{1-P^a_D}{1-P^a_F} (P^b_F)^2 + \frac{P^a_D}{P^a_D} (P^b_D - P^b_F) \frac{1-P^a_D}{1-P^a_F} \tag{A.2}
\]

\[
\frac{P^b_F}{P^a_F} \leq \frac{P^a_D}{P^a_F} \tag{A.3}
\]

\[
\frac{1-P^b_D}{1-P^b_F} \leq \frac{1-P^a_D}{1-P^a_F} \tag{A.4}
\]

\[
0 \leq P^b_F \leq P^b_D \leq 1 \tag{A.5}
\]

with:

\[
\eta^* = \frac{P^a_D}{P^a_F} \frac{1-P^a_D}{1-P^a_F} \tag{A.6}
\]

Eqs.(A.3)-(A.5) are constraints dictated by the concavity of the ROC curve. We solve Eq.(A.2) for \(P^b_D \geq 0\) and substitute into Eq.(A.1); we then take the derivative of \(\Delta P^e_1\) with respect to \(P^b_F\):

\[
\frac{d\Delta P^e_1}{dP^b_F} = \frac{\eta^*}{\eta^* + 1} \left[ -P^a_F + P^a_F P^b_F \left[ \eta^* (P^b_F)^2 + \frac{P^a_D (P^a_D - P^b_F)}{1-P^a_F} \right]^{-0.5} \right] \tag{A.7}
\]
The derivative in Eq.(A.7) is zero only if \( P_F^b = P_F^\alpha \), which results in a minimum \((\Delta P^b_1 = 0)\). Differentiating one more time both sides of Eq.(A.7), we obtain:

\[
\frac{d^2 \Delta P^b_1}{d(P_F^b)^2} = \frac{\eta^*}{\eta^* + 1} \left[ \eta^* (P_F^b)^2 + \frac{P_D^\alpha (P_D^b - P_F^b)}{1 - P_F^F} \right]^{-1.5} \frac{P_D^\alpha (P_D^b - P_F^b)}{1 - P_F^F} > 0 \quad (A.8)
\]

Thus, for all \( P_F^b \), such that \( P_F^b \geq P_F^\alpha \), the derivative of \( \Delta P^b_1 \) with respect to \( P_F^b \) is positive and increasing; therefore, it is maximized with respect to \( (P_F^b, P_D^b) \) at the boundary point given by Eq.(4.77).

(ii). Relative Bound

Suppose again that \((P_F^\alpha, P_D^b)\) is given; then the problem of maximizing the least upper bound of Eqs.(4.82)-(4.84) can be formulated as follows:

\[
\text{PROBLEM A.2.} \quad \max_{(P_F^b, P_D^b) \in A} \omega^b_{i} = \frac{\eta^* (P_F^b)^2 + 1 - (P_D^b)^2}{\eta^* P_F^b P_D^b + 1 - P_D^\alpha P_D^b} - 1 \quad (A.9)
\]

where \( A \) is the same feasible set as in Problem A.1.

Proceeding in a similar manner as in Problem A.1 above, we show that the relative bound is maximized with respect to \((P_F^b, P_D^b)\) again at the boundary point given by Eq.(4.77). To see this note that the maximization in Eq.(A.9) above is obviously equivalent to:

\[
\max_{(P_F^b, P_D^b) \in A} \psi(P_F^b, P_D^b) = \frac{\eta^* (P_F^b)^2 + 1 - (P_D^b)^2}{\eta^* P_F^b P_D^b + 1 - P_D^\alpha P_D^b} \quad (A.10)
\]

The function \( \psi \) is well defined and non-zero for all the pairs \((P_F^b, P_D^b)\), since both the numerator and the denominator are always strictly positive. Then, the maximization of Eq.(A.10) is equivalent to:

\[
\min_{(P_F^b, P_D^b) \in A} \xi(P_F^b, P_D^b) = \frac{\eta^* P_F^b P_D^b + 1 - P_D^\alpha P_D^b}{\eta^* (P_F^b)^2 + 1 - (P_D^b)^2} \quad (A.11)
\]
This in turn is equivalent to:

$$\max_{(P_F^b, P_D^b) \in A} -\xi(P_F^b, P_D^b) = \frac{-\eta^* P_F^b P_D^b - 1 + P_D^a P_D^b}{\eta^* (P_F^a)^2 + 1 - (P_D^a)^2}$$

(A.12)

But barring constant terms, the maximization of Eq.(A.12) is the exact same maximization of Eq.(A.1). Since Eq.(A.1) is maximized for $(P_F^b, P_D^b)$ given by Eq.(4.77), both Eq.(A.12) and consequently Eq.(A.9) are also maximized for $(P_F^b, P_D^b)$ given by Eq.(4.77).
CHAPTER 5

Large Team Problems

5.1. INTRODUCTION

In this chapter we examine problems of large teams, that is teams which consist in the limit of an infinite number of DMs and perform hypothesis testing. We present and analyze the infinite tandem team and our results complement the results of [T88] for the infinite parallel team. Our objective is to obtain design guidelines for large organizations.

Despite the considerable research interest on distributed decision making in a hypothesis testing environment, almost all the results focus on small teams. The explosive combinatorial complexity of the problems in this framework suggests that only simple and very restricted architectures of large organizations can be analyzed. The infinite parallel organization was considered in [T88]. In this chapter the infinite tandem architecture, the other limit case, is analyzed. There exist two main reasons which motivated our research. First, it is desirable to determine under what conditions the probability of error of the infinite tandem team (organization) goes to zero. Then, it is also interesting to examine the behavior of the members of a team when a change in the architecture of the team occurs. It is known that if a new DM is added to a team, all the DMs have to modify their decision rules to adapt to the new architecture; this requires that the problem be solved again from the beginning for the new architecture. For this reason, we introduce a suboptimal decision scheme which is considerably more simple to implement, under which, each DM tries selfishly to maximize the performance of its own personal decision instead of the global team performance. One can make an argument that this decision scheme is more descriptive of human organizations than the optimal decision scheme, since human DMs are reluctant
to sacrifice their individual performance for the good of the organization. Furthermore, we present a second suboptimal decision scheme, under which all the DMs (except the first one) employ identical decision rules; thus, by restricting identical DMs to employ identical decision rules only one decision rule has to be determined.

The suboptimal decision schemes, and in particular the selfish one, can be easily adapted to account for changes in the team architecture; they will not yield optimal performance, but will provide a descriptive flavor of the team dynamics. Moreover, they provide upper bounds on the optimal team probability of error, which is also bounded from below by the probability of error of the centralized team. For all the above reasons, the proposed suboptimal decision schemes are extensively analyzed; conditions for the probability of error of the infinite team to go to zero, when these schemes are implemented, are derived and their performance is compared to the optimal team performance.

Many results have focused on similar problems with small teams and small numbers of messages [ET82], [TV87], [R87], [TP89]. As mentioned above, in [T88] asymptotic results were obtained for the parallel team, where all DMs are identical and transmit a message to a fusion center which makes the final team decision; it was shown that all DMs may be restricted to employing identical decision rules without any degradation in the team performance. A similar result for the tandem team will be established in section 5.4.

For the tandem team, we can obtain that, given the conditional independence assumption for the observations, the optimal decision rule of each DM (except the first one) is given in the form of two likelihood ratios with constant thresholds. These thresholds can be obtained by examining all the solutions of a set of coupled algebraic equations which are usually very hard to solve. Comparison of the performances of the infinite parallel and the infinite tandem teams indicates that the parallel architecture asymptotically performs better than the tandem architecture, contradicting the claims in [VT88] where each sensor employed the Neyman-Pearson test. Note also that if the conditional independence assumption fails, the optimal decision rules do not have to be given by likelihood ratio tests
and the problems are computationally intractable (NP-hard), even for small number of DMs and of communication messages [TA85].

**REMARK.** This remark describes the types of decision rules employed by the DMs (deterministic, randomized, etc.) and may be skipped without loss of continuity. As will be extensively discussed in chapter 6, since in this chapter only the Bayesian formulation of the problems is considered, all types of decision rules are equivalent. Therefore, we assume without loss of generality that the DMs employ what will be defined as decision rules with dependent randomization. These result in all the individual and the team ROC curves being both continuous and concave. We suggest that the interested reader return to this remark after reading chapter 6, where also the Neyman-Pearson formulation of the problems is considered.

In section 5.2, we examine the team which performs binary hypothesis testing and which consists of an infinite number of identical DMs in tandem; each DM communicates a binary messages to his immediate successor. We derive necessary and sufficient conditions on the ROC curve of the individual DM, so that the team normalized probability of error of the infinite team is zero. We also generalize our results for an arbitrary number of hypotheses and an arbitrary number of communication messages.

In section 5.3, we present a suboptimal selfish decision scheme for the infinite tandem team which is computationally simple to implement, and in section 5.4 we present another suboptimal decision scheme under which all DMs (except the first one) are restricted to employing identical decision rules. We again derive necessary and sufficient conditions on the ROC curve of the individual DM, so that when the infinite team employs each of the suboptimal decision schemes, the normalized probability of error is zero. The results are generalized for an arbitrary number of hypotheses and an arbitrary number of messages.

Finally in section 5.5, we present and analyze the results of our computer simulations, comparing the performance of the infinite team under the optimal and the suboptimal
decision schemes. We also try to establish and discuss the trade off between the probability of error and the complexity of the decision rules.

5.2. THE OPTIMAL INFINITE TANDEM TEAM

5.2.1. Problem Formulation

The sequential (tandem) decentralized problem is defined as follows. The team consists of $N$ DMs (Figure 5.1) and there are $M$ hypotheses $H_0, ..., H_{M-1}$ with known prior probabilities $P(H_i) > 0$. Let $y_n$ the observation of the $n$th DM, $n = 1, ..., N$, be a random variable taking values from a set $Y$. The $y_n$'s are conditionally independent and identically distributed given any hypothesis, with a known probability distribution $P(y | H_i)$, $i = 1, ..., M$. Let $D$ be a positive integer. The first DM evaluates a $D$-valued message $u_1$ based on its own observation and transmits it to its successor; that is, $u_1 = \gamma_1(y_1)$, where $\gamma_1: Y \rightarrow \{1, ..., D\}$ is a measurable function. Each subsequent DM evaluates a $D$-valued message $u_n$ based on its own observation $y_n$ and on the message $u_{n-1}$ from its predecessor and transmits it to its successor; that is, $u_n = \gamma_n(y_n, u_{n-1})$ where the measurable function $\gamma_n: Y \times \{1, ..., D\} \rightarrow \{1, ..., D\}$ is called the decision rule of DM $n$. The decision $u_N$ of the final DM $N$ is the team decision and declares one of the hypotheses to be true. The objective is to choose the decision rules $\gamma_1, ..., \gamma_N$ which minimize the normalized probability of error, $P_e(N)$, of the decision of DM $N$.

Consider the case of binary hypothesis testing ($M = 2$) and binary communication messages ($D = 2$). Given some prior probabilities (hence also some $\eta = P(H_0)/P(H_1)$), we generalize the results of [E82], as in [R87], and obtain the necessary conditions for the optimal decision rules of the DMs as a set of likelihood ratio tests with constant thresholds:
Figure 5.1. The Team Consisting of \( N \) DMs in Tandem

\[
\lambda(y_1) \begin{cases} 
  u_1 = 1 & \frac{P(u_N = 1 \mid u_1 = 1, H_0) - P(u_N = 1 \mid u_1 = 0, H_0)}{P(u_N = 1 \mid u_1 = 1, H_1) - P(u_N = 1 \mid u_1 = 0, H_1)} \eta = \eta_1 \\
  u_1 = 0 &
\end{cases}
\]

\[
\lambda(y_n) \begin{cases} 
  u_n = 1 & \frac{P(u_{n-1} = i \mid H_0) - P(u_{n-1} = i \mid u_n = 0, H_0)}{P(u_{n-1} = i \mid u_n = 1, H_1) - P(u_{n-1} = i \mid u_n = 0, H_1)} \eta = \eta_i^0 \\
  u_n = 0 &
\end{cases}
\]

for \( n = 2, 3, ..., N-1 \) and \( i = 0, 1 \) (5.2)

\[
\lambda(y_N) \begin{cases} 
  u_N = 1 & \frac{P(u_{N-1} = i \mid H_0)}{P(u_{N-1} = i \mid H_1)} \eta = \eta_i^N \\
  u_N = 0 &
\end{cases}
\]

for \( i = 0, 1 \) (5.3)

**Remark 1.** The observations of the DMs are conditionally independent given the true hypothesis.

**Remark 2.** The two messages which each DM is able to transmit can be denoted \( m_1 \) and \( m_2 \). Without loss of generality assume that:

\[
\frac{P(u_n = m_1 \mid H_0)}{P(u_n = m_1 \mid H_1)} \geq \frac{P(u_n = m_2 \mid H_0)}{P(u_n = m_2 \mid H_1)}
\]

Suppose that DM \( n \) receives an observation \( y_n \); if DM \( n \)'s optimal decision is \( u_n = m_i \) when he receives \( u_{n-1} = m_j \), then it can be shown that his optimal decision is again \( u_n = m_i \) when he receives \( u_{n-1} = m_i \) (for \( n = 2, 3, ..., N \)). Moreover, if DM \( N \)'s optimal decision is \( u_N = 1 \) (0) when he receives \( u_{N-1} = m_1 \) (\( m_2 \)), then it can be shown that his optimal decision is
again $u_N = 1$ (0) when he receives $u_{N-1} = m_2$ ($m_1$). Then, backtracking it can be seen that $m_1$ is interpreted as 0 and $m_2$ is interpreted as 1.

For any logical selection of decision rules for the DMs, the normalized probability of error for both the centralized team and the parallel team goes very quickly to zero with increasing $N$. It is interesting to analyze the tandem case as well; thus consider the following problem:

**PROBLEM 5.1.** Consider a team consisting of $N$ identical DMs in tandem. Does the normalized probability of error go to zero as $N$ goes to infinity?

We demonstrate that the optimal probability of error of the infinite tandem team does not necessarily have to be zero. For this, we present a special ROC curve for the individual DM and calculate inductively the ROC curve of the team consisting of $N$ such DMs in tandem; this in general can not be done analytically. We then take the limit of the team ROC curve as $N \to \infty$ and show that the team probability of error is bounded away from zero.

Therefore, consider the DM whose ROC curve is presented in Figure 5.2, for some $\alpha > 1$ (the underlying probability distributions are presented in Table 5.1). An example of a DM with such a ROC curve is a person, who observes the outcomes of the tosses of a coin known to be $\alpha$ to 1 biased in favor of the most likely outcome, and tries to decide whether the coin is biased in favor of 'heads' or in favor of 'tails'. We start by obtaining the ROC curve of a team which consists of $N$ such DMs.

### 5.2.2. A Special Example

**LEMMA 5.1.** Consider the tandem team which consists of $N \geq 2$ identical DMs; suppose that the ROC curve of each DM is given in Figure 5.2, for some $\alpha > 1$ (the
Figure 5.2. A Special ROC Curve

<table>
<thead>
<tr>
<th>y</th>
<th>H</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H₀</td>
<td>( \frac{\alpha}{\alpha + 1} )</td>
<td>( \frac{1}{\alpha + 1} )</td>
</tr>
<tr>
<td></td>
<td>H₁</td>
<td>( \frac{1}{\alpha + 1} )</td>
<td>( \frac{\alpha}{\alpha + 1} )</td>
</tr>
</tbody>
</table>

Table 5.1. Probability Distributions for the DM of Figure 5.2
underlying probability mass functions are presented in Table 5.1). Then the team ROC curve consists of the following $2N$ points:

$$(P_{F}^0, P_{D}^0) = (0,0) \quad (5.4a)$$

$$(P_{F}^{2j-1}, P_{D}^{2j-1}) = \left[ \left( \frac{1}{\alpha + 1} \right)^{N+2-2j} \sum_{i=0}^{j-1} \left( \frac{\alpha}{\alpha + 1} \right)^i, \left( \frac{\alpha}{\alpha + 1} \right)^{N+2-2j} \sum_{i=0}^{j-1} \left( \frac{\alpha}{\alpha + 1} \right)^i \right]$$

for $j = 1, 2, ..., \left\lceil \frac{N}{2} \right\rceil \quad (5.4b)$

$$(P_{F}^{2j}, P_{D}^{2j}) = \left[ \left( \frac{1}{\alpha + 1} \right)^{N-1-2j} \left( 1 - \frac{\alpha}{\alpha + 1} \sum_{i=0}^{j} \left( \frac{\alpha}{\alpha + 1} \right)^i \right), \left( \frac{\alpha}{\alpha + 1} \right)^{N-1-2j} \left( 1 - \frac{1}{\alpha + 1} \sum_{i=0}^{j} \left( \frac{\alpha}{\alpha + 1} \right)^i \right) \right]$$

for $j = 1, 2, ..., \left\lceil \frac{N}{2} \right\rceil - 1 \quad (5.4c)$

$$(P_{F}^{N-1+j}, P_{D}^{N-1+j}) = (1 - P_{D}^{N-j}, 1 - P_{F}^{N-j}) \quad \text{for } j = 1, 2, ..., N \quad (5.4d)$$

**Proof.** By induction on $N$. It is trivial to confirm that it is true for the case of $N = 2$. To prove the general case first observe that since the problem is symmetric with respect to the two hypotheses, the team ROC curve will be symmetric with respect to the line $P_D = -P_F + 1$. Thus, only Eqs.(5.4a), (5.4b) and (5.4c) need to be proved. For this consider whether $N$ is odd or even; then Eq.(5.4) can be written accordingly:

**CASE A.** $N = 2k \quad \text{for } k = 1, 2, 3, ...$

$$(P_{F}^0, P_{D}^0) = (0,0) \quad (5.5a)$$

$$(P_{F}^{2j-1}, P_{D}^{2j-1}) = \left[ \left( \frac{1}{\alpha + 1} \right)^{2k-2(j-1)} \sum_{i=0}^{j-1} \left( \frac{\alpha}{\alpha + 1} \right)^i, \left( \frac{\alpha}{\alpha + 1} \right)^{2k-2(j-1)} \sum_{i=0}^{j-1} \left( \frac{\alpha}{\alpha + 1} \right)^i \right]$$

for $j = 1, 2, ..., k \quad (5.5b)$
\[(P_F^{2j}, P_D^{2j}) = \left( \frac{1}{\alpha + 1} \right)^{2k - 2j - 1} \left[ 1 - \frac{\alpha}{\alpha + 1} \sum_{i=0}^{j} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right], \]
\[(\frac{\alpha}{\alpha + 1})^{2k - 2j - 1} \left[ 1 - \frac{1}{\alpha + 1} \sum_{i=0}^{j} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right] \text{ for } j = 1, 2, ..., k - 1 \quad (5.5c)\]
\[(P_F^{2k-1+j}, P_D^{2k-1+j}) = (1 - P_D^{2k-j}, 1 - P_F^{2k-j}) \quad \text{for } j = 1, 2, ..., 2k \quad (5.5d)\]

**CASE B.** \(N = 2k + 1\) for \(k = 1, 2, 3, ...\)

\[(P_F^0, P_D^0) = (0, 0) \quad (5.6a)\]

\[(P_F^{2j-1}, P_D^{2j-1}) = \left( \frac{1}{\alpha + 1} \right)^{2k - 2j + 3} \sum_{i=0}^{j-1} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \left( \frac{\alpha}{\alpha + 1} \right)^{2k - 2j + 3} \sum_{i=0}^{j-1} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \quad \text{for } j = 1, 2, ..., k \quad (5.6b)\]

\[(P_F^{2j}, P_D^{2j}) = \left( \frac{1}{\alpha + 1} \right)^{2k - 2j} \left[ 1 - \frac{\alpha}{\alpha + 1} \sum_{i=0}^{j} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right], \]
\[(\frac{\alpha}{\alpha + 1})^{2k - 2j} \left[ 1 - \frac{1}{\alpha + 1} \sum_{i=0}^{j} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right] \text{ for } j = 1, 2, ..., k \quad (5.6c)\]
\[(P_F^{2k+j}, P_D^{2k+j}) = (1 - P_D^{2k+1-j}, 1 - P_F^{2k+1-j}) \quad \text{for } j = 1, 2, ..., 2k + 1 \quad (5.6d)\]

Suppose that a team of \(N\) identical DMs has a ROC curve defined by Eq.(5.5) (Eq.(5.6)). Then it is a matter of tedious but straightforward algebra to show that the addition of a new DM to the team yields a new team whose ROC curve is given by Eq.(5.6) (Eq.(5.5) respectively).

Consider only the case where \(N = 2k\), since the proof for the case where \(N = 2k + 1\) is analogous; the ROC curve of the team consists of the \((P_F^n, P_D^n)\) points defined by Eq.(5.5), for \(n = 0, 1, ..., 2N - 1\). The new DM, DM \(N+1\), has two decision rules to choose from:
(i). \( u_{N+1} = \begin{cases} 
1 & \text{if } u_N = 1 \text{ and } y_{N+1} = 1 \\
0 & \text{otherwise} 
\end{cases} \)

(ii). \( u_{N+1} = \begin{cases} 
0 & \text{if } u_N = 0 \text{ and } y_{N+1} = 0 \\
1 & \text{otherwise} 
\end{cases} \)

Then, if decision rule (i). is employed:

\[
(P_F^n, P_D^n) = \left( \frac{1}{\alpha + 1} P_F^n, \frac{\alpha}{\alpha + 1} P_D^n \right)
\]  

(5.7)

and if decision rule (ii). is employed:

\[
(P_F^n, P_D^n) = \left( \frac{\alpha}{\alpha + 1} P_F^n + \frac{1}{\alpha + 1}, \frac{1}{\alpha + 1} P_D^n + \frac{\alpha}{\alpha + 1} \right)
\]  

(5.8)

When decision rule (i). is employed the point \((0,0)\) yields again \((0,0)\), the points of Eq.(5.5b) yield the points of Eq.(5.6b), the points of Eq.(5.6c) yield the first \((k-1)\) points of Eq.(5.6c) and the first point of Eq.(5.5c) (i.e., the point corresponding to \(j = 1\)) yields the last point of Eq.(5.6c) (i.e., the point corresponding to \(j = k\)). Similarly, the points of Eq.(5.6d) are obtained by symmetry. Thus all the points of Eq.(5.6) can be obtained by the team of \(N+1\) DMs.

It is not hard to verify analytically that all the other points, that can be obtained from the points of Eq.(5.5) by applying the two decision rules, never lie above the ROC curve defined by Eq.(5.6). This completes the proof. \textbf{Q.E.D.}

\textbf{COROLLARY 5.1.} Consider the ROC curve of the tandem team of \(N \geq 2\) DMs described by Eq.(5.4). Then the slopes of the linear segments of the ROC curve are as follows:

(a). from \((0,0)\) to \((P_F^1, P_D^1)\) \hspace{1cm} \text{slope: } \alpha^N 

(b). from \((P_F^n, P_D^n)\) to \((P_F^{n+1}, P_D^{n+1})\), for \(n = 1, ..., 2N - 3\) \hspace{1cm} \text{slope: } \alpha^{N-1} - n 

(c). from \((P_F^{2N-2}, P_D^{2N-2})\) to \((1,1)\) \hspace{1cm} \text{slope: } \alpha^{-N}
The proof of the corollary can be obtained by inspection of Eq.(5.1) and thus is omitted. Note that, if the first and last linear segments are excluded, the slopes of neighboring segments change by a factor of $\alpha$. Moreover, the ROC curve always contains a linear segment of slope 1; in fact the linear segment of slope 1 is the one joining $(P_{F}^{N-1}, P_{D}^{N-1})$ and $(P_{F}^{N}, P_{D}^{N}) = (1-P_{D}^{N-1}, 1-P_{F}^{N-1})$.

**Lemma 5.2.** The limit of $(P_{F}^{N-1}, P_{D}^{N-1})$ as $N$ goes to infinity is:

$$
\lim_{N \to \infty} (P_{F}^{N-1}, P_{D}^{N-1}) = \left( \frac{1}{\alpha^2 + \alpha + 1}, \frac{\alpha^2}{\alpha^2 + \alpha + 1} \right)
$$

(5.9)

**Proof.** Again we distinguish between $N$ odd and even.

**Case A.** $N = 2k$ for $k = 1, 2, 3,$ ...

Note two facts; first, that $(P_{F}^{N-1}, P_{D}^{N-1})$ is given by Eq.(5.4b) for $j = k$ and second, that instead of taking the limit as $N$ goes to infinity, we can equivalently take the limit as $k = N/2$ goes to infinity. Thus:

$$
(P_{F}^{2k-1}, P_{D}^{2k-1}) = \left( \left(\frac{1}{\alpha+1}\right)^2 \sum_{i=0}^{k-1} \left(\frac{\alpha}{(\alpha+1)^2}\right)^i, \left(\frac{\alpha}{\alpha+1}\right)^2 \sum_{i=0}^{k-1} \left(\frac{\alpha}{(\alpha+1)^2}\right)^i \right)
$$

(5.10)

and taking the limit as $k$ goes to infinity:

$$
\lim_{k \to \infty} (P_{F}^{2k-1}, P_{D}^{2k-1}) = \left[ \left(\frac{1}{\alpha+1}\right)^2 \frac{(\alpha+1)^2}{\alpha^2 + \alpha + 1}, \left(\frac{\alpha}{\alpha+1}\right)^2 \frac{(\alpha+1)^2}{\alpha^2 + \alpha + 1} \right] =
$$

$$
= \left( \frac{1}{\alpha^2 + \alpha + 1}, \frac{\alpha^2}{\alpha^2 + \alpha + 1} \right)
$$

(5.11)

**Case B.** $N = 2k + 1$ for $k = 1, 2, 3,$ ...

This time $(P_{F}^{N-1}, P_{D}^{N-1})$ is given by Eq.(5.4c) for $j = k$:
\[
(P_{F}^{2k}, P_{D}^{2k}) = \left[ 1 - \frac{\alpha}{\alpha + 1} \sum_{i=0}^{k} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i, 1 - \frac{1}{\alpha + 1} \sum_{i=0}^{k} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right]
\]

Taking the limit as \( k \) goes to infinity:

\[
\lim_{k \to \infty} (P_{F}^{2k}, P_{D}^{2k}) = \left[ 1 - \frac{\alpha}{\alpha + 1} \frac{(\alpha + 1)^2}{\alpha^2 + \alpha + 1}, 1 - \frac{1}{\alpha + 1} \frac{(\alpha + 1)^2}{\alpha^2 + \alpha + 1} \right] =
\]

\[
= \left( \frac{1}{\alpha^2 + \alpha + 1}, \frac{\alpha^2}{\alpha^2 + \alpha + 1} \right)
\]

From Eq.(5.11) and Eq.(5.13) we conclude that the Lemma be true. Q.E.D.

A direct consequence of this is the following corollary:

**COROLLARY 5.2.** The limit of \((P_{F}^{N}, P_{D}^{N})\) as \(N \to \infty\) is:

\[
\lim_{N \to \infty} (P_{F}^{N}, P_{D}^{N}) = \left( \frac{\alpha + 1}{\alpha^2 + \alpha + 1}, \frac{\alpha^2 + \alpha}{\alpha^2 + \alpha + 1} \right)
\]

It is easy to verify again from Eqs.(5.9) and (5.14) that indeed the line segment between \((P_{F}^{N-1}, P_{D}^{N-1})\) and \((P_{F}^{N}, P_{D}^{N})\) has slope 1; thus the ROC curve of the infinite tandem team contains this line segment of slope 1. Then, two more corollaries follow:

**COROLLARY 5.3.** Consider an infinite tandem team of identical DMs whose ROC curve is given in Figure 5.2. Suppose that the team performs binary detection and that the threshold is \(\eta = 1\). Then the optimal normalized probability of error for the team is:

\[
P_{e(\infty)} = \frac{0.5\alpha + 1}{\alpha^2 + \alpha + 1}
\]

Proof. From basic detection theory [V68] it is known that the optimal operating point is any point of the ROC curve at which the tangent to the ROC curve is \(\eta = 1\). According to the convention of section 3.4.1, \((P_{F}^{N}, P_{D}^{N})\) is such a point. Then substitute for the limit of
\((P_F^N, P_B^N)\) from Eq.(5.9) and for \(\eta = 1\) into Eq.(3.9), the definition of the normalized probability of error, and the corollary is proved.  Q.E.D.

**REMARK.** If \(\eta = 1\), the above infinite tandem team improves on the performance of an *individual* DM by a factor of:

\[
\frac{P^e(\infty)}{P^e(1)} = \frac{(0.5\alpha + 1)(\alpha + 1)}{\alpha^2 + \alpha + 1}
\]

Thus the normalized probability of error of the infinite tandem team will be between 50% and 100% of the normalized probability of error of an individual DM. Furthermore as \(|\log(\eta)|\) increases the improvement on the performance is even less significant. These facts should suggest the *inefficiencies* of the tandem architecture.

**COROLLARY 5.4.** The normalized probability of error of an infinite tandem team does not have to be zero.

For the sake of completeness the rest of the ROC curve of the infinite tandem team of Lemma 5.1 is derived.

**LEMMA 5.3.** The ROC curve of an infinite tandem team of identical DMs, whose individual ROC curve is given in Figure 5.2, consists of the following points:

\[
\left( \left( \frac{1}{\alpha + 1} \right)^i \frac{1}{\alpha^2 + \alpha + 1}, \left( \frac{\alpha}{\alpha + 1} \right)^i \frac{\alpha^2}{\alpha^2 + \alpha + 1} \right) \quad (5.15a)
\]

\[
\left( 1 - \left( \frac{\alpha}{\alpha + 1} \right)^i \frac{\alpha^2}{\alpha^2 + \alpha + 1}, 1 - \left( \frac{1}{\alpha + 1} \right)^i \frac{1}{\alpha^2 + \alpha + 1} \right) \quad (5.15b)
\]

for \(i = 0, 1, 2, \ldots\)
Proof. Since the points of Eq.(5.15b) are symmetric to the points of Eq.(5.15a), it suffices to prove Eq.(5.15a). Denote: $m = -\left\lfloor \frac{N}{2} \right\rfloor - 1 + j$. Then the points of Eq.(5.4b) can be written as:

$$
\left[ \frac{1}{(\alpha + 1)^{N-2\left\lfloor \frac{N}{2} \right\rfloor - 2m}} \sum_{i=0}^{\left\lfloor \frac{N}{2} \right\rfloor - m} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i, \frac{\alpha}{(\alpha + 1)^{N-2\left\lfloor \frac{N}{2} \right\rfloor - 2m}} \sum_{i=0}^{\left\lfloor \frac{N}{2} \right\rfloor - m} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right]
$$

$$
m = -1, -2, ..., -\left\lfloor \frac{N}{2} \right\rfloor (5.16)
$$

Similarly, denote: $m = -\left\lfloor \frac{N}{2} \right\rfloor + j$. Then the points of Eq.(5.4c) can be written as:

$$
\left[ \frac{1}{(\alpha + 1)^{N-2\left\lfloor \frac{N}{2} \right\rfloor - 2m-1}} \left\{ \sum_{i=0}^{\left\lfloor \frac{N}{2} \right\rfloor + m} \left( \frac{\alpha}{\alpha + 1} \right)^i \right\}, \right.

\left. \frac{\alpha}{(\alpha + 1)^{N-2\left\lfloor \frac{N}{2} \right\rfloor - 2m-1}} \left\{ \sum_{i=0}^{\left\lfloor \frac{N}{2} \right\rfloor + m} \left( \frac{\alpha}{(\alpha + 1)^2} \right)^i \right\} \right]
$$

$$
m = -1, -2, ..., -\left\lfloor \frac{N}{2} \right\rfloor + 1 (5.17)
$$

When the limit is taken as $N \rightarrow \infty$ the summations of the above equations are constant for all $m$; in fact they were calculated in the proof of Lemma 5.2. Then, by examining separately the cases of $N$ odd and $N$ even, it can be shown that Eqs.(5.16) and (5.17) can be written in the form of (5.15a), completing the proof. Q.E.D.

5.2.3. The Main Result

**PROPOSITION 5.1.** Consider a tandem team which consists of $N$ identical DMs and performs binary hypothesis testing. Then, as $N \rightarrow \infty$, the team will achieve zero normalized probability of error, for all thresholds $\eta$ (i.e., independent of the external parameters like
costs and priors), if and only if either the initial slope of the ROC curve of the individual DM is infinite or its final slope is zero.

Proof. To prove the only if part, suppose that the initial slope of the ROC curve of the individual DM is \( m_0 < \infty \) and that the its final slope is \( m_1 > 0 \). Denote: \( \alpha = \max\{ m_0, 1/m_1 \} \). Then, because of the concavity of the ROC curve of the individual DM, each DM of the team is worse than a DM having a two piecewise ROC curve whose first piece has slope \( \alpha \) and whose second piece has slope \( 1/\alpha \) (Figure 5.2). By Lemma 5.2, an infinite tandem team consisting of such (better) DMs will have a non-zero normalized probability of error. Therefore, the infinite tandem team consisting of the original (worse) DMs will also have non-zero normalized probability of error.

To prove the if part, suppose without loss of generality (since the names of the two hypotheses can be interchanged) that \( m_0 = \infty \). The proof consists of proposing decision rules (not necessarily optimal) for the DMs of the team which will result in a zero asymptotic normalized probability of error; that is, it will be shown that given any \( \delta > 0 \), if the proposed decision rules are employed, the normalized probability of error can be made less than \( \delta \) as \( N \to \infty \).

**STEP 1.** Choose some large number \( \eta^* \). Then consider the point \((P_F(\eta^*), P_D(\eta^*))\) of the ROC curve of the individual DM which has slope \( \eta^* \) and denote by \( \varepsilon_0(\eta^*) > 0 \) the probability of detection at that point (Figure 5.3). Then from the concavity of the ROC curve:

\[
\eta^* P_F(\eta^*) < \varepsilon_0(\eta^*) \Rightarrow P_F(\eta^*) < \frac{\varepsilon_0(\eta^*)}{\eta^*}
\]

(5.18)

Note that such a point can always be obtained, regardless of how large \( \eta^* \) is, because of the assumption that the initial slope of the ROC curve is: \( m_0 = \infty \).

**STEP 2.** For some integer \( N^* \), define the following decision rules for the DMs of the team:
Figure 5.3. ROC Curve with Infinite Initial Slope \( (m_0 = \infty) \)

For DM 1:

\[
    u_1 = \chi(y_1) = \begin{cases} 
0 & ; \ \Lambda(y_1) \leq \eta^* \\
1 & ; \ \Lambda(y_1) > \eta^* 
\end{cases} \tag{5.19a}
\]

For DM \( n; \ n = 2, 3, \ldots, N^* \):

\[
    u_n = \chi(y_n, u_{n-1}) = \begin{cases} 
0 & ; \ \Lambda(y_n) \leq \eta^* \text{ and } u_{n-1} = 0 \\
1 & ; \ \text{otherwise}
\end{cases} \tag{5.19b}
\]

For DM \( n; \ n = N^*+1, N^*+2, N^*+3, \ldots \):

\[
    u_n = \chi(y_n, u_{n-1}) = u_{n-1} \tag{5.19c}
\]

Let us elaborate on the decision rules of Eq.(5.16). According to these only the first \( N^* \) DMs influence the decision of the team. A decision \( u_i = 0 \) is propagated through the first \( N^* \) DMs of the team, until a particular DM \( n^* \) \( (n^* \leq N^*) \) receives an observation which makes him decide \( u_{n^*} = 1 \). In that case, the team decision will be \( u_N = 1 \); otherwise,
the team decision will be \( u_N = 0 \). Note that, since \( \eta^* \) was chosen to be large, there is a very small probability for a DM \( n^* \) who receives \( u_{n^*-1} = 0 \) to decide \( u_{n^*} = 1 \) under both hypotheses, but if he does his observation \( y_{n^*} \) suggests that there is an overwhelming probability that \( H_1 \) be the correct hypothesis.

**STEP 3.** If the decision rules of the previous step are implemented the probability of detection of the team is:

\[
P_D^t = 1 - (1 - \varepsilon_0(\eta^*))^{N^*}
\]

(5.20)

In order to make the team probability of detection greater than \( 1 - \delta \):

\[
P_D^t = 1 - (1 - \varepsilon_0(\eta^*))^{N^*} > 1 - \delta \quad \Rightarrow \quad N^* > \frac{\log(\delta)}{\log(1 - \varepsilon_0(\eta^*))}
\]

(5.21)

**STEP 4.** Similarly, the team probability of false alarm is:

\[
P_F^t = 1 - (1 - P_F(\eta^*))^{N^*}
\]

(5.22)

Then from Eqs.(5.22) and (5.18):

\[
P_F^t < 1 - \left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)^{N^*}
\]

(5.23)

In order to make the team probability of false alarm smaller than \( \delta \):

\[
P_F^t < 1 - \left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)^{N^*} < \delta \quad \Rightarrow \quad \frac{\log(1 - \delta)}{\log\left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)} > N^*
\]

(5.24)

**STEP 5.** To show that an \( N^* \) satisfying both Eqs.(5.21) and (5.24) indeed exists, recall that for \( x \approx 0 \):

\[
\log(1 + x) = x
\]

(5.25)

If \( \eta^* \) is selected sufficiently large, \( \varepsilon_0(\eta^*) \) is very close to zero (Figure 5.3); then from Eqs. (5.21), (5.24) and (5.25), the integer \( N^* \) has to satisfy:
Thus by choosing $\eta^*$ sufficiently large, which can be done because of the assumption $m_0 = \infty$, an integer $N^*$ satisfying both Eqs.(5.21) and (5.24) can always be obtained.

**STEP 6.** If the decision rules of Step 2 are employed by the team, for any given threshold $\eta$ (which is not to be confused with our personal choice of large number $\eta^*$), the team normalized probability of error is:

$$P^{e(\infty)} = \frac{\eta}{\eta + 1} P^t_F + \frac{1}{\eta + 1} (1 - P^t_B) < \frac{\eta}{\eta + 1} \delta + \frac{1}{\eta + 1} \delta = \delta$$

(5.27)

Thus for any given $\delta > 0$, the infinite team normalized probability of error can be made less than $\delta$. The proof of the theorem is now complete. Q.E.D.

This is a useful result because it offers a convenient test involving only the individual team member, in order to determine whether the infinite team probability of error is bounded away from zero or not. Moreover, after the proof was completed, it was communicated to us by J.N. Tsitsiklis that similar results had been established in the context of automata with finite memory [C69], [HC70]. There, a finite Markov chain was considered, with the additional restriction that the transition probabilities be time invariant; each state of this chain corresponded to a discrete hypothesis and conditions for the convergence of the chain to a unique (true) state were obtained.
5.2.4. The Necessary and Sufficient Conditions

Denote by $A(\cdot)$ the likelihood ratio of $y \in Y$ the observation of each individual DM and by $A(Y)$ the range of its possible values. Then, Proposition 5.1 can also be stated in the following equivalent form:

**PROPOSITION 5.1a.** The infinite tandem team normalized probability of error is bounded away from zero, for any threshold $\eta$, if and only if there exists some $B > 0$ such that:

$$\|\log[A(Y)]\| < B.$$ 

We would like to elaborate on these necessary and sufficient conditions. For this consider a single DM who receives measurements serially and who has limited memory, so that he can only recall the last one received. Then the conditions imply that if this DM is willing to wait long enough before making his decision, he can achieve any desirable level of performance. In fact, the decision rules of the previous section suggest that the team decision will be $u_N = 0$, unless a DM receives a highly improbable measurement which would indicate that $H_1$ is the true hypothesis with an overwhelming probability; in that case the team decision will be $u_N = 0$.

On the other hand, if the conditions do not hold the likelihood ratio of an individual DM is bounded. This implies that there exists some DM $(N^*+1)$ of the team whose measurement does not influence his decision, because the decision $u_{N^*}$ is so reliable that it is propagated as $u_{N^*+1}$, regardless of $y_{N^*+1}$; since all the DMs are identical, $u_{N^*}$ will be more reliable than the measurement of each subsequent DM and, will eventually become the final team decision as well. Thus only the measurements of the first $N^*$ DMs influence the team decision and since a finite number of measurements results in a non-zero probability of error even in the centralized case, the optimal team normalized probability of error will be bounded away from zero.
It is worthwhile to compare the conditions of Proposition 5.1 to the conditions in [T88] for the DMs of the infinite parallel team to have identical optimal decision rules. The conditions are 'complementary' in the sense that the conditions in [T88] require that a single DM is not able to make a perfectly accurate decision, while our conditions require that a single DM is able to make a decision satisfying any given level of accuracy.

5.2.5. Generalizations

Consider the infinite tandem team; suppose that it performs $M$-ary hypothesis testing and that the communication messages between DMs are also $M$-ary. Then define the following likelihood ratios:

$$
\Lambda_{ij}(y) = \frac{P(y | H_i)}{P(y | H_j)} \quad \text{for } i, j = 0, 1, ..., M-1 \text{ and } i > j.
$$

Also denote by $\Lambda_{ij}(Y)$ the range of the possible values of $\Lambda_{ij}(\cdot)$. The generalization of Proposition 5.1 for the case of $M$-ary hypotheses is:

**PROPOSITION 5.2.** Consider the infinite tandem team which performs $M$-ary hypothesis testing and employs $M$-ary messages. The team normalized probability of error is bounded away from zero, for any prior probability and cost assignments, if and only if there exists some $B_{ij} > 0$ such that:

$$
|\log[\Lambda_{ij}(y)]| < B_{ij}
$$

for some $i, j = 0, 1, ..., M-1$ and $i > j$.

It is easy to find an example for which the probability of error is bounded away from zero if the conditions are violated and the binary proof can be adapted and simulated $M(M - 1)/2$ times to show that the conditions are also sufficient.
In [HC70] an ingenious scheme was presented (and generalized in [K75]) demonstrating that if the Markov chain is constructed with one more node than the number of the discrete hypotheses, then there exist sets of transition probabilities that will enable the chain to converge to a unique (true) state even if the necessary and sufficient conditions do not hold. This idea could be appropriately adapted to our hypothesis testing framework so that decision rules for the DMs be derived for the infinite tandem team to always achieve zero probability of error if the number of the communication messages between DMs is one higher than the number of the hypotheses. Our results can also be extended to teams which have more general acyclic (tree) architectures, as long as the team does not include a DM who receives an infinite number of messages from other DMs and to teams whose DMs are not identical. The proofs are variants of the proof of Proposition 5.1.

5.3 THE "SELFISH" SUBOPTIMAL TANDEM TEAM

5.3.1. Problem Formulation

The non-linear coupled equations of Eqs.(5.1)-(5.3) are very hard to solve; their solution does not have to be unique and their computational complexity increases exponentially with $N$. Moreover, as was discussed in the previous section and will be verified with the computer simulations of section 5.5, when the conditions of Proposition 5.1 hold, the convergence of the optimal team probability of error to zero is slow\(^1\). For these reasons it is worthwhile to suggest the following suboptimal decision scheme which

\[ \frac{1}{(\alpha^2 + \alpha + 1)(\alpha + 1)} \left[ \frac{\alpha}{(\alpha + 1)^2} \right]^k \]

for $k = 0, 1, 2, ...$

---

\(^1\) This is not true for the cases where the conditions of Proposition 5.1 do not hold. In those cases the convergence to the optimal probability of error occurs geometrically fast at a rate of $\frac{\alpha}{(\alpha+1)^2}$. If $\eta = 1$, the difference in probability of error between a team of $N = 2k + 1$ DMs and the (asymptotically optimal) infinite team is:
has two important properties: its computational requirements increase \textit{linearly} with $N$ and, under slightly more restrictive conditions than those of Proposition 5.1, the probability of error goes to zero as $N \to \infty$.

Under this suboptimal decision scheme each DM tries to minimize the probability of error of its personal decision; that is it acts as if it was the last DM in the team. Therefore, this is a \textit{selfish} decision rule since each DM effectively ignores all of its successors in the team and instead of optimizing the global team decision, each DM tries to 'optimize' the performance (i.e., minimize the probability of error) of his own (local) decision. Note that the decision rule for the last DM in the team (i.e., DM $N$) is the same as the optimal decision rule, as his local objective is the same as the global objective; it is just that the upstream DMs do not readjust. Also note that this suboptimal decision scheme can be implemented easily and efficiently even if the DMs of the team are not identical. Formally, the suboptimal decision rules for any DMs are defined as the following likelihood ratio tests with constant thresholds $^2$:

\begin{align}
\Lambda(y_1) \begin{array}{c} \geq \ \\ u_1 = 0 \end{array} \quad \eta = \mu_1 \\
\eta = \mu_1
\end{align}

\begin{align}
\Lambda(y_n) \begin{array}{c} \geq \ \\ u_n = 0 \end{array} \quad \frac{P(u_{n-1} = i \mid H_0)}{P(u_{n-1} = i \mid H_1)} \quad \eta = \mu^*_n \\
\eta = \mu^*_n
\end{align}

for $n = 2, \ldots, N$ and $i = 0, 1$

\textit{From now on we refer to the decision rules just introduced as the "selfish" suboptimal decision rules and to the team which employs them the "selfish" suboptimal tandem team.}

---

$^2$ Comparing the optimal decision rules of Eqs.(5.1)-(5.3) to the suboptimal decision rules of Eqs.(5.28) and (5.29), notice that the suboptimal ones are \textit{decoupled}. 
There exists one more attractive attribute of this decision scheme which is in fact the primary reason we analyzed it. One of the more intriguing questions of organizational design deals with the response of the team to a change in the team architecture. How does the team react to the addition or deletion of a DM? If a change occurs in the team when the optimal decision scheme is implemented, the team can not make use of its current status to facilitate the computation of the new optimal decision rules; the problem needs to be solved again from the beginning. This is not true under the selfish suboptimal decision rule, where it is relatively easy for the team to adapt to such a change. The reason for this is that under the selfish decision rule, the DMs behave less than team members and more like individuals; still the analysis will provide some insight and bounds on the optimal team performance.

5.3.2. The Main Result

PROPOSITION 5.3. Consider a tandem team which consists of N identical DMs and performs binary hypothesis testing; suppose that the team employs the selfish suboptimal decision scheme. Then, as \( N \rightarrow \infty \), the team will achieve zero probability of error, for all thresholds \( \eta \), if and only if both the initial slope of the ROC curve of the individual DM is infinite and its final slope is zero.

REMARK. The necessary and sufficient conditions of Proposition 5.3 are stricter than the ones of Proposition 5.1, since they require that both the initial slope of the ROC curve be infinite and the final slope be zero, instead of either the initial slope be infinite or the final slope be zero.

Proof. To prove the only if part recall from Proposition 5.1 that if the ROC curve of the individual DM has both finite initial slope and non-zero final slope, the asymptotic team
probability of error is bounded away from zero even if the optimal decision rules are employed; therefore in that case, it is also bounded away from zero when the selfish decision rules are employed.

![ROC Curve with Infinite Initial Slope and Non-Zero Final Slope](image)

**Figure 5.4.** ROC Curve with Infinite Initial Slope and Non-Zero Final Slope

Consider the case where either the initial slope is infinite or the final slope is zero, but not both. We need to show that the asymptotic team probability of error under the selfish decision rules is bounded away from zero, even though it goes to zero under the optimal decision rules (as was established in Proposition 5.1). For this suppose without loss of generality that the ROC curve of the individual DM has infinite initial slope and non-zero final slope; then consider the DM whose ROC curve is presented in Figure 5.4 along with the underlying probability distributions in Table 5.2. It should be obvious that if the prior probabilities are such that \( \eta < 1 - \varepsilon \) the optimal decision for every DM is \( u = 1 \)

\(^3\) It is also easy to see that if \( \eta \geq 1 + \varepsilon \), then the suboptimal decision rules coincide with the optimal decision rules and are:

\[
u_1(y_1) = \begin{cases} 0 & ; y_1 = 0 \\ 1 & ; y_1 = 1 \end{cases}\]
Table 5.2. Probability Distributions for the DM of Figure 5.4.

\[
P^\varepsilon(n) = \frac{\eta}{\eta + 1} \quad ; \quad \text{for } n = 1, 2, 3, ...
\]

Thus also:

\[
P^\varepsilon(\infty) = \frac{\eta}{\eta + 1} > 0
\]

The proof of the only if part is now complete.

The proof of the if part consists of proposing decision rules for the DMs of the team which will result in a zero asymptotic probability of error; that is, it will be shown that given any \( \delta \) such that \( 1 > \delta > 0 \), if the proposed decision rules are employed, the probability of error can be made less than \( \delta \) as \( N \to \infty \). It will also be shown that the proposed decision rules are indeed the selfish suboptimal decision rules.

Denote by \( P_F^n \) the probability of false alarm and by \( P_D^n \) the probability of detection of the decision of the \( n \)th DM of the team. Then Eq.(5.29) can be written in the following form:

\[
\mu_n(y_n, u_{n-1}) = \begin{cases} 
0 ; & y_n = 0 \text{ and } u_{n-1} = 0 \\
1 ; & \text{otherwise}
\end{cases}
\]

These of course lead to \( P^\varepsilon(\infty) = 0 \) (Proposition 5.1).
Figure 5.5. ROC Curve with Initial Slope: $m_0 = \infty$ and Final Slope: $m_1 = 0$

$$
\Lambda(y_n) = \begin{cases} 
  u_n = 1 & \frac{1 - P_F^{n-1}}{1 - P_D^{n-1}} \eta = \mu_n^0 ; \ u_{n-1} = 0 \\
  u_n = 0 & \frac{P_F^{n-1}}{P_D^{n-1}} \eta = \mu_n^1 ; \ u_{n-1} = 1 
\end{cases}
$$

for $n = 2, 3, 4, \ldots$  \tag{5.30}

**STEP 1.** Pick some large number $\eta^*$. Then consider the point $(P_F(\eta^*), P_D(\eta^*))$ of the ROC curve of the individual DM which has slope $\eta^*$ and denote by $\varepsilon_0(\eta^*) > 0$ the probability of detection at that point (Figure 5.5). Then from the concavity of the ROC curve:

$$
\eta^* P_F(\eta^*) < \varepsilon_0(\eta^*) \Rightarrow P_F(\eta^*) < \frac{\varepsilon_0(\eta^*)}{\eta^*} \tag{5.31}
$$

Note that such a point can always be obtained, regardless of how large $\eta^*$ is, because of the assumption that the initial slope of the ROC curve is: $m_0 = \infty$. 
STEP 2. Similarly, for the same large number $\eta^*$, consider the point $(P_F(1/\eta^*), P_D(1/\eta^*))$ of the ROC curve of the individual DM which has slope $1/\eta^*$ and denote by $1 - \epsilon_1(\eta^*)$ the probability of false alarm at that point (Figure 5.5). Then from the concavity of the ROC curve:

$$\frac{1}{\eta^*}[(1 - \epsilon_1(\eta^*)) - 1] + 1 < P_D(1/\eta^*) \Rightarrow P_D(1/\eta^*) > 1 - \frac{\epsilon_1(\eta^*)}{\eta^*} \quad (5.32)$$

Note that such a point can always be obtained, regardless of how large $\eta^*$ is, because of the assumption that the initial slope of the ROC curve: $m_1 = 0$.

STEP 3. Assume without loss of generality (since the names of the hypotheses can be interchanged) that the decision threshold $\eta$ satisfies:

$$\eta > \frac{1 - \epsilon_0(\eta^*) - \epsilon_1(\eta^*)}{\eta^*} \quad (5.33)$$

Then replace every DM of the team by a worse DM whose ROC curve is presented in Figure 5.6 and underlying probability distributions are presented in Table 5.3. Steps 2–6 of Proposition 5.1 can now be repeated for the infinite tandem team consisting of the newly defined DMs to show that $P^e(\infty) = 0$, but for the sake of completeness they are included here as well.

REMARK. If the assumption of Eq.(5.33) does not hold, replace every DM of the team by a worse DM whose ROC curve is presented in Figure 5.7 and underlying probability distributions are presented in Table 5.4. This case is of course symmetric to the previous one and thus its proof follows by symmetry.
Figure 5.6. ROC Curves of the Worse DM of Step 3 and of the Original DM

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₀</td>
<td>(1 - \frac{c₀(\eta^<em>)}{\eta^</em>})</td>
<td>(\frac{c₀(\eta^<em>)}{\eta^</em>})</td>
</tr>
<tr>
<td>H₁</td>
<td>(1 - c₀(\eta^*))</td>
<td>(c₀(\eta^*))</td>
</tr>
</tbody>
</table>

Table 5.3. Probability Distributions for the DM of Figure 5.6.
Figure 5.7. ROC Curve of the Worst DM when Eq.(5.33) is Not True

Table 5.4. Probability Distributions for the DM of Figure 5.7.
**STEP 4.** For some integer $N^*$, define the following decision rules for the DMs of the team:

For DM 1:

$$u_1 = \gamma(y_1) = \begin{cases} 
0 & ; \ y_1 = 0 \\
1 & ; \ y_1 = 1 
\end{cases} \quad (5.34a)$$

For DM $n$; $n = 2, 3, ..., N^*$:

$$u_n = \gamma_n(y_n, u_{n-1}) = \begin{cases} 
0 & ; \ y_n = 0 \text{ and } u_{n-1} = 0 \\
1 & ; \text{ otherwise} 
\end{cases} \quad (5.34b)$$

For DM $n$; $n = N^*+1, N^*+2, N^*+3, ...$

$$u_n = \gamma_n(y_n, u_{n-1}) = u_{n-1} \quad (5.34c)$$

These may be equivalently written in terms of likelihood ratios as follows:

For DM 1:

$$u_1 = \gamma(y_1) = \begin{cases} 
0 & ; \ A(y_1) \leq \eta^* \\
1 & ; \ A(y_1) > \eta^* 
\end{cases} \quad (5.35a)$$

For DM $n$; $n = 2, 3, ..., N^*$:

$$u_n = \gamma_n(y_n, u_{n-1}) = \begin{cases} 
0 & ; \ A(y_n) \leq \eta^* \text{ and } u_{n-1} = 0 \\
1 & ; \text{ otherwise} 
\end{cases} \quad (5.35b)$$

For DM $n$; $n = N^*+1, N^*+2, N^*+3, ...$

$$u_n = \gamma_n(y_n, u_{n-1}) = u_{n-1} \quad (5.35c)$$

**STEP 5.** It is easy to see that if the decision rules of the previous step are implemented the probability of detection of the team is:

$$P_D^i = 1 - (1 - \epsilon_0(\eta^*))^{W^*} \quad (5.36)$$
In order to make the team probability of detection greater than $1 - \delta$:

$$P_D^t = 1 - (1 - \varepsilon_0(\eta^*))^{W^*} > 1 - \delta \quad \Rightarrow \quad N^* > \frac{\log(\delta)}{\log(1 - \varepsilon_0(\eta^*))} \quad (5.37)$$

**STEP 6.** Similarly, the team probability of false alarm is:

$$P_F^t = 1 - (1 - P_F(\eta^*))^{W^*} \quad (5.38)$$

Then from Eqs.(5.38) and (5.31):

$$P_F^t < 1 - \left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)^{W^*} \quad (5.39)$$

In order to make the team probability of false alarm smaller than $\delta$:

$$P_F^t < 1 - \left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)^{W^*} < \delta \quad \Rightarrow \quad \frac{\log(1 - \delta)}{\log(1 - \varepsilon_0(\eta^*))} > N^* \quad (5.40)$$

**STEP 7.** To show that an $N^*$ satisfying both Eqs.(5.37) and (5.40) indeed exists, recall that if $\eta^*$ is selected sufficiently large, $\varepsilon_0(\eta^*)$ is very close to zero (Figure 5.5); then from Eqs.(5.22), (5.37) and (5.39), the integer $N^*$ has to satisfy:

$$\left[\frac{\log(1 - \delta)}{\log(1 - \varepsilon_0(\eta^*))}\right] = \left[\frac{\log(1 - \delta)}{\varepsilon_0(\eta^*)}\right] \geq N^* \geq \left[\frac{\log(\delta)}{\log(1 - \varepsilon_0(\eta^*))}\right] = \left[\frac{\log(\delta)}{\varepsilon_0(\eta^*)}\right] \quad (5.41)$$

Thus by choosing $\eta^*$ sufficiently large, which can be done because of the assumption $m_0 = \infty$, an integer $N^*$ satisfying both Eqs.(5.37) and (5.40) can always be obtained.
STEP 8. If the decision rules of Step 4 are employed by the team, for any given decision threshold \( \eta \) which satisfies Eq.(5.33) (and which is not to be confused with our personal choice of large number \( \eta^* \)) the team normalized probability of error is:

\[
P_{e(\infty)} = \frac{\eta}{\eta + 1} P_F^I + \frac{1}{\eta + 1} (1 - P_D^I) < \frac{\eta}{\eta + 1} \delta + \frac{1}{\eta + 1} \delta = \delta \quad (5.42)
\]

Thus for any given \( \delta > 0 \), the normalized probability of error of the infinite team which consists of the worse DMs (of Figure 5.6) can be made less than \( \delta \); therefore, the normalized probability of error of the infinite team which consists of the original (better) DMs can be made less than \( \delta \).

STEP 9. What remains to be verified is that the decision rules of Step 4 for the worse DMs (of Figure 5.6) are indeed the selfish decision rules of Eqs.(5.28) and (5.29) for those DMs.

It is not hard to verify that for DM 1 the suboptimal decision rule of Eq.(5.28) and the proposed decision rule of Eq.(5.34a) are identical. Recall that \( \eta \) was assumed to satisfy Eq.(5.33) and that \( \eta^* \) can be selected sufficiently large. Then:

\[
\eta^* > \eta > \frac{1 - \varepsilon_0 - \varepsilon_1}{\eta^* - \varepsilon_1} > \frac{1 - \varepsilon_0}{\eta^*} \quad (5.43)
\]

Then it should be clear from Figure 5.6 that the decision thresholds for both decision rules coincide.

Moreover, as was shown above, under the decision rules of Step 4:

\[
(P_F^n, P_D^n) = \begin{cases} 
\left(1 - \left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)^n, 1 - (1 - \varepsilon_0(\eta^*))^n\right) ; & n = 1, 2, ..., N^* \\
\left(1 - \left(1 - \frac{\varepsilon_0(\eta^*)}{\eta^*}\right)^{N^*}, 1 - (1 - \varepsilon_0(\eta^*))^{N^*}\right) ; & n = N^*+1, N^*+2, N^*+3, ...
\end{cases} \quad (5.44)
\]

Substituting into the thresholds of Eq.(5.41):
\[
\mu^0_n = \begin{cases} 
\frac{(1 - \epsilon_0(\eta^*))^{n-1}}{\eta^*} & ; \ n = 2, 3, ..., N^* \\
\eta^* \left(1 - \epsilon_0(\eta^*)\right)^{n-1} \left(1 - \epsilon_0(\eta^*)^W\right) & ; \ n = N^*+1, N^*+2, N^*+3, ...
\end{cases}
\] (5.45)

\[
\mu^1_n = \begin{cases} 
\frac{1 - \left(1 - \epsilon_0(\eta^*)\right)^{n-1}}{\eta^*} & ; \ n = 2, 3, ..., N^* \\
\frac{1 - \left(1 - \epsilon_0(\eta^*)\right)^{n-1}}{1 - (1 - \epsilon_0(\eta^*)^W)} \left(1 - \epsilon_0(\eta^*)^W\right) & ; \ n = N^*+1, N^*+2, N^*+3, ...
\end{cases}
\] (5.46)

To show the above thresholds are indeed the thresholds of the suboptimal decision rules, we need to show that:

\[
\mu^0_n \begin{cases} \leq \eta^* & ; \ n = 2, 3, ..., N^* \\
> \eta^* & ; \ n = N^*+1, N^*+2, ...
\end{cases}
\] (5.47)

and:

\[
\mu^1_n \leq \frac{1 - \epsilon_0(\eta^*)}{\epsilon_0(\eta^*)} ; \ n = 2, 3, 4, ...
\] (5.48)

For this, we now need to show that there exists some integer \(N^*\) satisfying Eq.(5.47). Since we want the team probability of error to go to zero asymptotically, we also need to verify that \(N^*\) satisfies Eq.(5.41). Recall that \(\eta^*\) is large; then:

\[
\mu^0_n = \mu^0_{n-1} \frac{1 - \epsilon_0(\eta^*)}{\eta^*} > \mu^0_{n-1} ; \ n = 2, 3, ..., N^*
\] (5.49)

Therefore, it suffices to show that:
\[
\mu_{N^*+1} = \eta \left( \frac{1 - \epsilon_0(\eta^*)}{\eta^*} \right)^{N^*} \geq \eta^* > \eta \left( \frac{1 - \epsilon_0(\eta^*)}{\eta^*} \right)^{N^* - 1} = \mu_{N^*} \quad (5.50)
\]

Taking logarithms:
\[
N^* \left[ \log \left( 1 - \frac{\epsilon_0(\eta^*)}{\eta^*} \right) - \log \left( 1 - \epsilon_0(\eta^*) \right) \right] \geq \log \left( \frac{\eta^*}{\eta} \right) > \quad (5.51)
\]
\[
> (N^* - 1) \left[ \log \left( 1 - \frac{\epsilon_0(\eta^*)}{\eta^*} \right) - \log \left( 1 - \epsilon_0(\eta^*) \right) \right]
\]

Using Eq.(5.25) to approximate:
\[
N^* \left[ - \frac{\epsilon_0(\eta^*)}{\eta^*} + \epsilon_0(\eta^*) \right] \geq \log \left( \frac{\eta^*}{\eta} \right) > (N^* - 1) \left[ - \frac{\epsilon_0(\eta^*)}{\eta^*} + \epsilon_0(\eta^*) \right] \quad (5.52)
\]

This can be rewritten as:
\[
\frac{\log \left( \frac{\eta^*}{\eta} \right)}{\epsilon_0(\eta^*) \left( 1 - \frac{1}{\eta^*} \right)} + 1 > N^* \geq \frac{\log \left( \frac{\eta^*}{\eta} \right)}{\epsilon_0(\eta^*) \left( 1 - \frac{1}{\eta^*} \right)} \quad (5.53)
\]

Therefore, an integer \( N^* \) satisfying Eq.(5.47) indeed exists. Moreover, for \( \eta^* \) sufficiently large:
\[
\eta^* \log(1 - \delta) - \epsilon_0(\eta^*) > \frac{\log \left( \frac{\eta^*}{\eta} \right)}{\epsilon_0(\eta^*) \left( 1 - \frac{1}{\eta^*} \right)} + 1 > N^* \geq \frac{\log \left( \frac{\eta^*}{\eta} \right)}{\epsilon_0(\eta^*) \left( 1 - \frac{1}{\eta^*} \right)} > \frac{\log(\delta)}{\epsilon_0(\eta^*)} \quad (5.54)
\]

Thus, this integer \( N^* \) also satisfies Eq.(5.41).

Finally to show that Eq.(5.48) holds, observe that for \( \eta^* \) sufficiently large (which implies that \( \epsilon_0(\eta^*) \) is very small), apply Taylor's theorem to obtain that:
For all \( n = 2, 3, 4, \ldots : \)

\[
\mu_n^1 = \eta \frac{1 - \left(1 - \frac{\epsilon_0(\eta^*)}{\eta^*}\right)^{n-1}}{1 - \left(1 - \frac{\epsilon_0(\eta^*)}{\eta^*}\right)^n} = \eta \frac{(n-1) \frac{\epsilon_0(\eta^*)}{\eta^*}}{(n-1) \epsilon_0(\eta^*)} = \eta \frac{1}{\frac{\epsilon_0(\eta^*)}{\eta^*}} < \frac{1 - \epsilon_0(\eta^*)}{\frac{\epsilon_0(\eta^*)}{\eta^*}}
\]  

(5.55)

Thus, the decision rules of Step 4 correspond indeed the selfish decision rules, for the integer \( N^* \) defined by Eq.(5.53). Q.E.D.

5.3.3. Generalizations

Denote by \( \Lambda(\cdot) \) the likelihood ratio of \( y \) the observation of each individual DM and by \( \Lambda(Y) \) the range of its possible values. Then, Proposition 5.3 can also be stated in the following equivalent form:

**PROPOSITION 5.3a.** Consider the infinite tandem team which employs the selfish suboptimal decision rules and performs binary hypothesis testing. The team normalized probability of error will be bounded away from zero, for any threshold \( \eta \), if and only if there exists some \( B > 0 \) such that either:

\[
\log[\Lambda(Y)] < B
\]

or:

\[
\frac{1}{B} < \log[\Lambda(Y)].
\]

Similarly, for the \( M \)-ary hypothesis testing case, Proposition 5.2 can be modified as follows:

**PROPOSITION 5.4.** Consider the infinite tandem team which employs the selfish suboptimal decision rules and performs \( M \)-ary hypothesis testing. The team normalized probability of error will be bounded away from zero, for any prior probability and cost assignments, if and only if there exists some \( B_{ij} > 0 \) such that either:

\[
\log[\Lambda_{ij}(Y)] < B_{ij}
\]
\[ \alpha: \]

\[ \frac{1}{B_{ij}} < \log[\Lambda_{ij}(Y)] \]

under either hypothesis \( H_i \) or \( H_j \), for some \( i, j = 1, 2, ..., M-1 \) and \( i > j \).

It is easy to find an example for which the probability of error is bounded away from zero if the condition is violated and the binary proof may be adapted to show that the conditions are also sufficient. Since under the selfish decision scheme each DM tries to minimize his personal probability of error, it does not make sense to assign to each DM a number of messages different from the number of the alternative hypotheses. Our results may be extended to teams with more general acyclic (tree) architectures, as long as the team does not include a DM who receives messages from an infinite number of other DMs. Finally, as was mentioned above, they can also be applied to teams whose DMs are not identical. The proofs are variants of the proof of Proposition 5.3.

5.4. THE "IDENTICAL" SUBOPTIMAL DECISION SCHEME

5.4.1. Problem Formulation

In [T88] it was established that for infinite parallel team which consists of identical DMs, all the DMs may be restricted to employing identical decision rules without any degradation in the team performance. We would like to establish a similar result for the infinite tandem team; therefore, we examine the suboptimal decision scheme under which all the DMs of the infinite tandem team (except the first one) are restricted to employing identical decision rules.

The problem can be represented by a Markov chain which includes three transient states and two ergodic classes each consisting of two states (Figure 5.8). The process begins at state \( H \) which represents the environment; the true state of the environment will be either \( H_0 \) or \( H_1 \), with corresponding prior probabilities \( P(H_0) \) and \( P(H_1) \). Then, the
process moves to the true state of the environment. Given the true state of the environment (represented by the states \(H_0\) and \(H_1\)), the first DM will make his decision which becomes the tentative decision of the team and the process moves to the corresponding recurrent state (state 0 (1) if the decision is \(u_1 = 0\) (1)). Then given the tentative team decision, every subsequent DM makes his own decision using the common (identical) decision rule, which in turn results to a new tentative team decision.

Note that the transition probabilities from every state (except state \(H\)) in the Markov chain of Figure 5.10 are controlled states, meaning that the transition probabilities are determined by the decision rules of the DMs. In fact, \((P^E_F, P^E_D)\) denotes the operating point of the first DM, and \((P^0_F, P^0_D)\) and \((P^1_F, P^1_D)\) denote the identical operating points of every other DM \(n\) when he receives from his predecessor \(u_{n-1} = 0\) or \(u_{n-1} = 1\) respectively.

Given the identical decision rules of DMs 2 through \(N\), the optimal decision rule of DM 1 is given by the following likelihood ratio test with constant threshold:

\[
\Lambda(y_1) \quad \begin{cases} 
    u_1 = 1 & \frac{P(u_N = 1 | u_1 = 1, H_0)}{P(u_N = 1 | u_1 = 0, H_0)} - \frac{P(u_N = 1 | u_1 = 0, H_1)}{P(u_N = 1 | u_1 = 1, H_1)} \geq \eta \equiv \eta_1 \\
    u_1 = 0 & 
\end{cases} \tag{5.56}
\]

Using simple induction, the threshold of Eq.(5.56) can be written as follows:
\[
\eta_1 = \left[ \frac{\sum_{k=0}^{N-3} (P_F^1 - P_F^0)^k + P_F^1 (P_F^1 - P_F^0)^{N-2}}{P_D^0 \sum_{k=0}^{N-3} (P_D^1 - P_D^0)^k + P_D^1 (P_D^1 - P_D^0)^{N-2}} \right] - \frac{P_F^0 \sum_{k=0}^{N-2} (P_F^1 - P_F^0)^k}{P_D^0 \sum_{k=0}^{N-2} (P_D^1 - P_D^0)^k} \eta \Rightarrow 
\]

\[
\eta_1 = \left( \frac{P_F^1 - P_F^0}{P_D^1 - P_D^0} \right)^{N-1} \eta 
\]

(5.57)

Observe that for \( N = 2 \) (i.e., the optimal two DM tandem team), the threshold of Eq.(5.57) reduces to the threshold of Eq.(4.5) as expected. On the other hand, the optimal identical decision rule of DMs 2 through \( N \) will not be given by a likelihood ratio test. Denote by \( f(\cdot) \) the ROC curve of the identical DMs. Then given the decision rule of the first DM, to obtain the optimal identical decision rule of the DMs the following non-linear optimization problem has to be solved:

\[
\min P^e(N) = \frac{1}{\eta + 1} P_F(N) + \frac{\eta}{\eta + 1} [1 - P_D(N)] 
\]

(5.58)

subject to:

\[
P_F(N) = P_F^0 \sum_{k=0}^{N-2} (P_F^1 - P_F^0)^k + P_F^1 (P_F^1 - P_F^0)^{N-1} 
\]

(5.59)

\[
P_D(N) = P_D^0 \sum_{k=0}^{N-2} (P_D^1 - P_D^0)^k + P_D^1 (P_D^1 - P_D^0)^{N-1} 
\]

(5.60)

\[
P_D^i = f(P_D^i) ; \quad i = 0, 1. 
\]

(5.61)

**Remark** 1. As the number of DMs increases, the decision of the first DM obviously becomes increasingly less significant. This can be seen from Eq.(5.57) where (unless the fraction is equal to 1) \( \eta_1 \) goes either to zero or to infinity, and the decision of DM 1 becomes trivial. This can also be seen from Eqs.(5.59) and (5.60); as \( N \) increases the influence of DM 1's operating point, \((P_F^1, P_D)\), on the team operating point, \((P_F(\infty), P_D(\infty))\), decreases.
REMARK 2. Taking the limit of Eq.(5.59) as $N \to \infty$, we obtain that:

$$P_F(\infty) = \frac{P_F^0}{1 - (P_F^1 - P_F^0)} \quad (5.62)$$

This result can be easily confirmed. In the limit as $N \to \infty$, the team operating point is $(P_F(\infty), P_D(\infty))$. Suppose that a new identical DM, employing the identical decision rules, is added to the team. Since steady state has been reached:

$$P_F(\infty) = [1 - P_F(\infty)]P_F^0 + P_F(\infty)P_F^1 \quad (5.63)$$

Solving for $P_F(\infty)$, Eq.(5.62) is obtained again.

REMARK 3. As we are going to show in the following section, this suboptimal decision scheme leads to zero normalized probability of error for the infinite tandem team, because, given any $\delta > 0$, identical decision rules can be derived which result to a team normalized probability of error smaller than $\delta$. But, no 'uniquely optimal' identical decision rules exist for the infinite team; that is, given any decision rule for all the DMs of the infinite team (except the first one), we can always find some $\delta > 0$ such that the team normalized probability of error is larger than $\delta$.

From now on we refer to the decision rules just introduced as the "identical suboptimal decision rules".

5.4.2. The Main Result

PROPOSITION 5.5. Consider a tandem team which consists of $N$ identical DMs and performs binary hypothesis testing; suppose that all the DMs of the team (except the first one) employ the identical decision rules. Then, as $N \to \infty$, the team will achieve zero probability of error, for all thresholds $\eta$, if and only if either the initial slope of the ROC curve of the individual DM is infinite or its final slope is zero.
REMARK. The necessary and sufficient conditions of Proposition 5.5 are the same conditions of Proposition 5.1 for the optimal decision rules. Therefore, successful use of this suboptimal decision scheme does not impose any additional restrictions on the DMs.

Proof. The only if part follows immediately from Proposition 5.1. To prove the if part, we will derive identical decision rules for the DMs which take the normalized probability of error asymptotically to zero; that is given any \( \delta > 0 \), the identical decision rules will result in the team normalized probability of error to be less than \( \delta \). Without loss of generality we assume that the slope of the tangent to the ROC curve at (0,0) is infinite.

Pick some large number \( \eta_0 \). Then consider the point \( (P^0_F(\eta_0), P^0_D(\eta_0)) \) of the ROC curve of the individual which has slope \( \eta_0 \) and denote by \( \varepsilon_0 > 0 \) the probability of detection at that point. From the concavity of the ROC curve:

\[
P^0_F(\eta_0) < \frac{\varepsilon_0}{\eta_0}
\]  

(5.64)

Denote by \( k \) the limit of the slope of the ROC curve as the curve moves towards (1,1); clearly, \( 0 \leq k < 1 \). Pick some number \( \eta_1 \), a little larger than \( k \). Then consider the point \( (P^1_F(\eta_1), P^1_D(\eta_1)) \) of the ROC curve of the individual which has slope \( \eta_1 \) and denote by \( 1 - \varepsilon_1 \) the probability of false alarm at that point. From the concavity of the ROC curve:

\[
P^1_D(\eta_1) \geq 1 - \eta_1 \varepsilon_1
\]

(5.65)

Suppose that the operating points \( (P^0_F(\eta_0), P^0_D(\eta_0)) \) and \( (P^1_F(\eta_1), P^1_D(\eta_1)) \) constitute the identical decision rule for the DMs in the team and denote by \( \pi_{ij} \) the steady state probability of state \( i \) when \( j \) is the true hypothesis \( (i, j = 0, 1) \). Then, depending on the true hypothesis, the two ergodic Markov chains are presented in Figure 5.9. Under the identical decision scheme, the normalized probability of error of the infinite tandem team can be denoted by:

\[
P^e(\infty) = c_{10} \pi_{10} + c_{01} \pi_{01}
\]

(5.66)
where $c_{ij}$ are the normalized prior probabilities defined in Eq.(3.3). It is easy to calculate the steady state probabilities of the chains of Figure 5.9:

\[
\pi_{10} = \frac{P_F^0(\eta_0)}{P_F^0(\eta_0) + 1 - P_F^1(\eta_1)} < \frac{\epsilon_0}{\eta_0} = \frac{\epsilon_0}{\epsilon_0 + \epsilon_1} = \frac{\epsilon_0}{\epsilon_0 + \eta_0 \epsilon_1}
\]  
(5.67)

\[
\pi_{01} = \frac{1 - P_D^1(\eta_1)}{P_D^0(\eta_0) + 1 - P_D^1(\eta_1)} \leq \frac{\eta_1 \epsilon_1}{\epsilon_0 + \eta_1 \epsilon_1}
\]  
(5.68)

From Eqs.(5.66)-(5.68), we conclude that the following condition is sufficient to guarantee that the asymptotic team normalized probability of error be smaller than the given $\delta > 0$:

\[
\max \left\{ \frac{\epsilon_0}{\epsilon_0 + \eta_0 \epsilon_1}, \frac{\eta_1 \epsilon_1}{\epsilon_0 + \eta_1 \epsilon_1} \right\} < \delta
\]  
(5.69)

Suppose that we select the operating points in such a way that always:

\[
\epsilon_0 = \alpha \epsilon_1
\]  
(5.70)

for some $\alpha > 1$. Then, Eq.(5.69) reduces to:

\[
\max \left\{ \frac{\alpha}{\alpha + \eta_0}, \frac{\eta_1}{\alpha + \eta_1} \right\} < \delta
\]  
(5.71)

For Eq.(5.71) to hold:

\[
\eta_0 \frac{\delta}{1 - \delta} > \alpha > \eta_1 \frac{1 - \delta}{\delta}
\]  
(5.72)
Note that by selecting an \( \epsilon_1 \) sufficiently small, an \( \alpha \) satisfying Eq.(5.72) can always be obtained, because we assumed that the slope of the ROC curve at \((0,0)\) is infinite. \textbf{Q.E.D.}

5.4.3. **On the Optimal Identical Decision Rules**

To obtain the optimal identical decision rules for the DMs we have to solve the problem of Eqs.(5.58)-(5.61) as \( N \to \infty \). Taking the limits, the problem may be rewritten in the following form:

\[
P^{e(\infty)} = \frac{1}{1 + \eta} \frac{P_F^0}{P_F^0 + 1 - P_F^1} + \frac{\eta}{1 + \eta} \left( 1 - \frac{1 - P_D^1}{P_D^0 + 1 - P_D^1} \right) \tag{5.73}
\]

subject to (5.61).

Suppose that \((P_F^1, P_D^1)\) has been fixed, and that \(P^{e(\infty)}\) is to be maximized with respect to \((P_F^0, P_D^0)\). For this we differentiate Eq.(5.73) with respect to \(P_F^0\):

\[
\frac{\partial P^{e(\infty)}}{\partial P_F^0} = \frac{1}{1 + \eta} \frac{1 - P_F^1}{(P_F^0 + 1 - P_F^1)^2} + \frac{\eta \eta_0}{1 + \eta} \frac{1 - P_D^1}{(P_D^0 + 1 - P_D^1)^2} > 0 \tag{5.74}
\]

where \(\eta_0\) is the slope of the tangent to the ROC curve at \((P_F^0, P_D^0)\). For any \((P_F^1, P_D^1)\), the partial derivative is positive. Moreover for the region \(0 \leq P_F^0 \leq P_F^1\), the partial derivative is decreasing with \(P_F^0\); this indicates that, given \((P_F^1, P_D^1)\), \(P^{e(\infty)}\) is maximized for \((P_F^0, P_D^0) = (0,0)\).

Similarly suppose that \((P_F^0, P_D^0)\) has been fixed, and that \(P^{e(\infty)}\) is to be maximized with respect to \((P_F^1, P_D^1)\); differentiating Eq.(5.73) with respect to \(P_F^1\):

\[
\frac{\partial P^{e(\infty)}}{\partial P_F^1} = \frac{1}{1 + \eta} \frac{P_F^0}{(P_F^0 + 1 - P_F^1)^2} + \frac{\eta \eta_1}{1 + \eta} \frac{P_D^0}{(P_D^0 + 1 - P_D^1)^2} > 0 \tag{5.74}
\]

where \(\eta_1\) is the slope of the tangent to the ROC curve at \((P_F^1, P_D^1)\). For any \((P_F^0, P_D^0)\), the partial derivative is positive. Moreover for the region \(P_F^0 \leq P_F^1 \leq 1\), the partial derivative is
increasing with $P^1_F$; this indicates that, given $(P^0_F, P^0_D)$, $P^e(\infty)$ is maximized for $(P^1_F, P^1_D) = (1,1)$.

But if $(P^0_F, P^0_D) = (0,0)$ and $(P^1_F, P^1_D) = (1,1)$, the ergodicity assumptions are violated. In fact, these operating points will result in four (degenerate) ergodic Markov chains, each consisting of a single state. The resulting normalized probability of error will be equal to the one of the first DM, namely:

$$P^e(\infty) = \frac{1}{1+\eta} P^c_F + \frac{\eta}{1+\eta} (1-P^c_D) \quad (5.75)$$

Therefore under the identical decision scheme, the limit of the normalized probability of error is zero, but no decision rules exist to make it equal to zero.

5.4.4. Generalizations

Denote by $\Lambda(\cdot)$ the likelihood ratio of $y \in Y$ the observation of each individual DM and by $\Lambda(Y)$ the range of its possible values. Then, Proposition 5.5 can also be stated in the following equivalent form:

PROPOSITION 5.5a. Consider the infinite tandem team which employs identical suboptimal decision rules and performs $M$-ary hypothesis testing. The team normalized probability of error is bounded away from zero, for threshold $\eta$, if and only if there exists some $B > 0$ such that:

$$|\log[\Lambda(Y)]| < B .$$

Similarly for the $M$-ary hypothesis testing case:

PROPOSITION 5.6. Consider the infinite tandem team which employs identical suboptimal decision rules and performs $M$-ary hypothesis testing. The team normalized
probability of error is bounded away from zero, for any prior probability and cost assignment, if and only if there exists some $B_{ij} > 0$ such that:

$$\|\log[\Lambda_{ij}(\Theta)]\| < B_{ij}$$

for some $i, j = 0, 1, ... , M-1$ and $i > j$.

It is easy to find an example for which the probability of error is bounded away from zero if the conditions are violated. The binary proof can be adapted to show that the conditions are also sufficient; the process will include $M$ ergodic Markov chains, each containing $M$ recurrent states.

When the necessary and sufficient conditions do not hold, even if more (but finite) messages than the number of hypotheses are allowed and the DMs (except the first and the last) are restricted to employing identical decision rules, the team normalized probability of error will be bounded away from zero. The reason is that when the optimal decision rules are implemented, the extra messages are used to 'create an unbounded likelihood ratio'; this can not happen if all DMs use identical decision rules.

Our results can also be extended to teams which have more general acyclic (tree) architectures, as long as the team does not include a DM who receives an infinite number of messages from other DMs and to teams whose DMs are not identical. The proofs are variants of the proof of Proposition 5.5.

5.5. NUMERICAL STUDIES

In this section we perform numerical studies on the optimal and the 'selfish' suboptimal decision rules; the reason for which results for the 'identical' decision rules are not presented is that for a finite number of DMs the optimal identical decision rules are not given in the form of a likelihood ratio test with constant thresholds (but are given as the
solution of the problem of Eqs.(5.58)-(5.61)), and that for an infinite number of DMs the optimal identical decision rules depend on the level of performance that the team desires to achieve (as was explained in subsection 5.4.3).

The probability distributions of the observations are Gaussian with variance \( \sigma^2 = 100 \); under \( H_0 \) the mean is \( \mu_0 = 0 \) and under \( H_1 \) the mean is \( \mu_1 = 10 \). Note that in this case the initial slope of the ROC curve of the individual DM is \( m_0 = \infty \) and that the final slope is \( m_1 = 0 \); thus the team probability of error should go to zero asymptotically under both the optimal and the suboptimal decision scheme, according to the theoretical results presented above.

In Figure 5.10, we present the probability of error for the centralized case, the optimal case and the suboptimal case for a team consisting of \( N = 10 \) DMs as a function of the threshold \( \eta \) (namely the prior probabilities). The number \( N = 10 \) was selected because it was shown in [P88] that it is sufficient for a parallel team to achieve performance very close to the asymptotic one, i.e., as \( N \to \infty \), for a special class of problems. It is interesting to note that as the prior probabilities become very unequal (the minimum prior probability is less than 0.333), the performance of both decision schemes becomes roughly equal (though never exactly equal). Also, note that for \( \eta = 1 \) (i.e., equal priors) the optimal and the suboptimal decision rules achieve identical performance; in fact the suboptimal decision rules become optimal in this case.

The above conclusions can be derived from Figures 5.11, 5.12 and 5.13 as well, where the ROC curves for the optimal, suboptimal and centralized teams are presented for three different teams \((N = 2, 6, 10)\); the ROC curves of the optimal and of the suboptimal team are tangent to each other when \( \eta = 1 \). Note that as \( N \), the number of DMs in the team increases, the performance of the centralized team becomes increasingly superior to the performance of the other two teams; but, in the same time, the performance of the suboptimal team is \textit{very close} to the performance of the optimal team in all three cases.
Figure 5.10. $P^e(10)$: Error Probability for a Ten DM Team ($N = 10$) vs. $\eta$ (log. scale)

Figure 5.11. ROC Curves for the Three Teams of $N = 2$ DMs
Figure 5.12. ROC Curves for the Three Teams of $N = 6$ DMs

Figure 5.13. ROC Curves for the Three Teams of $N = 10$ DMs
In Figure 5.14, we compare the performance of the centralized team and the optimal team, by presenting the percentage deterioration in performance between the two teams. Note that as the prior uncertainty increases (i.e., $\log(\eta) \to 0$), the deterioration of performance increases; this occurs because the inferior performance of the (worse) optimal team is more evident when the prior uncertainty is maximized. Moreover, note that the deterioration of performance increases with $N$, the number of DMs in the team, since the effect of the addition of a new DM to the optimal tandem team decreases with $N$.

In Figure 5.15, we compare the performance of the optimal team and of the suboptimal team, by presenting the percentage deterioration in performance between the two teams. As was already stated, both teams achieve identical performance for $\eta = 1$ (i.e., $\log(\eta) \to 0$); their performance also becomes roughly equal as $|\log(\eta)|$ increases. Moreover, the deterioration of performance increases with $N$, since the number of the suboptimal decisions made by the team increases with $N$. It is worthwhile to note that maximum percentage deterioration of performance in this (optimal vs. suboptimal) case increases more slowly than in the previous (centralized vs. optimal) case.

![Figure 5.14. Performance Comparison of Centralized and Optimal Team](image-url)
Figure 5.15. Performance Comparison of Optimal and Suboptimal Team

Figure 5.16. Optimal Team ROC Curves for $N = 1$ (A), 2 (B), 6 (C) and 10 (D) DMs
In Figure 5.16, we present the ROC curves of several optimal teams. As expected, the more the identical DMs in the team, the better the team performance. Also, the effect of the addition of a DM to the team performance decreases with the number of DMs in the team.

In Figure 5.17, we compare the performance of the centralized, the optimal and the suboptimal team as a function of \( N \), the number of DMs in the team, for \( \eta = 3 \). The normalized probability of error of the centralized team decreases exponentially to zero; on the other hand, for both the other two teams, it decreases to zero more slowly.

In Figure 5.18, we compare the performance of the centralized and the optimal team, by presenting the percentage deterioration in performance between the two teams, for \( \eta = 1 \) \((P(H_0) = 0.5)\), \( \eta = 3 \) \((P(H_0) = 0.75)\) and \( \eta = 9 \) \((P(H_0) = 0.5)\). The deterioration in performance increases exponentially fast with \( N \), at a rate that is roughly independent of \( \eta \).

Finally, consider the comparison of the performance of the optimal and the suboptimal team, presented in Figure 5.19 as the percentage deterioration of performance between the two teams, for different values of the threshold \( \eta \). This percentage deterioration in performance reaches a steady-state level, as the number of DMs increases. Note that, as mentioned above, for \( \eta = 1 \) the performance of both teams is identical; as \( \eta \) increases the deterioration in performance between the two teams also increases.

We would like to repeat that the implementation of the suboptimal decision rules is much easier than the one of the optimal decision rules. Each DM of the team needs to know only very little about the rest of the team in order to implement the suboptimal decision rules; for example, it does not need to know anything about its successors. The suboptimal decision rules are much less rigid and can be easily adapted to accommodate for changes in the team; also, for the optimal decision rules the computational requirements grow exponentially with time, while for the suboptimal decision rule they just grow linearly with \( N \). Thus, depending on the particular application of the team and on the available resources, the team designer has to evaluate the trade offs and decide which decision rules should be implemented.
Figure 5.17. $P_e(N)$: Probability of Error for $N$ DM Team ($\eta = 3$)

Figure 5.18. Performance Comparison of Centralized and Optimal Team
Finally, consider the comparison of the performance of the optimal and the suboptimal team, presented in Figure 5.19 as the percentage deterioration of performance between the two teams, for different values of the threshold $\eta$. This percentage deterioration in performance reaches a steady-state level, as the number of DMs increases. Note that, as mentioned above, for $\eta = 1$ the performance of both teams is identical; as $\eta$ increases the deterioration in performance between the two teams also increases.

We would like to repeat that the implementation of the suboptimal decision rules is much easier than the one of the optimal decision rules. Each DM of the team needs to know only very little about the rest of the team in order to implement the suboptimal decision rules; for example, it does not need to know anything about its successors. The suboptimal decision rules are much less rigid and can be easily adapted to accommodate for changes in the team; also, for the optimal decision rules the computational requirements grow exponentially with time, while for the suboptimal decision rule they just grow linearly with
$N$. Thus, depending on the particular application of the team and on the available resources, the team designer has to evaluate the trade offs and decide which decision rules should be implemented.

5.6. SUMMARY

The tandem team which consists of $N$ DMs was analyzed. The optimal decision rules were derived. Moreover, a computationally simple suboptimal decision scheme was presented. Necessary and sufficient conditions for the team probability to go to zero as the number of DMs grows to infinity were obtained for both decision schemes and numerical studies were performed to compare their performance.
CHAPTER 6

The Team ROC Curve
and Randomized Decision Rules

6.1. INTRODUCTION

In chapter 3, the Receiver Operating Characteristic (ROC) curve was introduced as a very useful and convenient tool for describing the individual DM in the binary hypothesis testing environment. By exploring its properties we were able to obtain the results that were presented in chapters 4 and 5. It is now time to take a closer look at the ROC curve of a team of DMs (team ROC curve), in order to determine whether the 'nice' properties are preserved under different operating conditions. It would be very helpful to establish that the team ROC curves be indeed a direct generalization of the individual ROC curves and that the properties be maintained. Then, the results for teams consisting of individual DMs could be immediately aggregated for teams consisting of teams of DMs, thus providing building blocks for higher organizations.

But, as it has been repeatedly demonstrated in this thesis, problems in the hypothesis testing framework are much more intricate than they look. In [R87] it was shown that the ROC curve for the two DM parallel team does not have to be concave. This should not be surprising. One 'team ROC curve' is obtained when the fusion center is restricted to employing the AND decision rule; similarly, a second 'team ROC curve' is obtained when the fusion center is restricted to employing the OR decision rule. The (global) team ROC curve consists of the upper envelope of these two curves, which does not have to be concave.
The issues dealing with the concavity of the team ROC curve are addressed in this chapter. Is the non-concavity a consequence of 'discontinuous decisions'; that is, is it caused because the fusion center has to choose between the AND and the OR decision rules with nothing in between? Does the two DM parallel team ROC curve become concave when the fusion center also receives an observation (i.e., for the $V$-architecture), so that its decision become 'continuous', since the decision thresholds can be changed in a continuous manner? Does the team ROC curve become concave when all the individual ROC curves are smooth, so that all the decisions are 'perfectly continuous'? Is the non-concavity a characteristic of parallel teams, or is it encountered in tandem teams as well? Claims have appeared in the literature [TP89a] stating that if the ROC curves of the individual DMs are concave, then the ROC curve of the two DM tandem team is also concave; a counterexample to it will be provided in section 6.3. Moreover, it will be shown that the non-concavity of the team ROC curve is a general result which transcends both the team architecture and the smoothness of the ROC curves of the members of the team.

This lack of concavity indicates that randomized (as opposed to deterministic) decision rules may be in order. Recall from chapter 3 that a randomized decision rule is characterized by a set of deterministic decision rules and by an associated discrete random variable; the realization of the random variable selects from the set the deterministic decision rule to be employed. In fact, by using a particular definition for the randomization process, the team ROC curve may be made concave. Of course nothing comes for free; as we are going to show in the sequel, in order to achieve concave team ROC curves in this manner, some degree of joint synchronization between the DMs of the team needs to be introduced which implies somewhat of a compromise on the decentralization aspects of the team.

Given the above discussion, a very important question arises: how are the results of the previous two chapters affected? The results of the previous two chapters are not affected at all. The reason for this being that the problems were addressed in a Bayesian framework; the non-concave team ROC curve and the issues of randomization
arise in the Neyman-Pearson formulation of these problems. We will return to this point in section 6.4.

In section 6.2, we present a simple example of a non-concave ROC curve for a two DM tandem team. We use it to introduce the various randomization strategies and to analyze the different issues involved with the Bayesian and Neyman-Pearson approach to the problems. Then in section 6.3, we present several examples to demonstrate that the non-concavity of the team ROC curve is independent of the smoothness of the ROC curves of the members of the team. Finally in section 6.4, we summarize the results and discuss the significance of the non-concave team ROC curve and its consequence on the rest of this thesis.

6.2. RANDOMIZED DECISION RULES: AN EXAMPLE

6.2.1. The Example

Consider the two DM tandem team (Figure 4.1) that performs binary hypothesis testing. The ROC curves for the two DMs presented in Figure 6.1 and the associated discrete probability density functions conditioned on the two hypotheses are presented in Figure 6.2.

REMARK 1. The primary DM is better than the consulting DM, since the ROC curve of the primary DM includes all the points of the ROC curve of the consulting DM and one extra point (namely, (0.2, 0.7)).

REMARK 2. For the tandem team of two DMs to have a non-concave ROC curve, the ROC curves of the individual DMs do not have to be of the form of Figure 6.1; in fact, one DM does not have to be better than the other. The DMs for this example were selected because their team ROC curve was easy to calculate.
Figure 6.1. ROC Curves for the DMs of the Example in 6.2.1.

Figure 6.2. Probability Distributions for the DMs of Figure 6.1.
6.2.2. Deterministic Decision Rules

Suppose that both DMs are restricted to employing deterministic decision rules. Then, each DM can operate only in the marked points of its ROC curve in Figure 6.1. In particular, the consulting DM can operate at (0.0, 0.0), (0.1, 0.5), (0.5, 0.9) and (1.0, 1.0), and the primary DM can also operate at them plus the additional point (0.2, 0.7). Then, the team ROC curve will also consist of a (finite) set of points. These points are presented in Table 6.1, together with the associated operating points of the DMs. The notation is consistent with the notation introduced in Chapter 4. Note that the team can only perform at the levels of probability of false alarm given in the leftmost column of Table 6.1. Therefore, if a Neyman-Pearson requirement of the form $P_F^T \leq \alpha$ is imposed on the team, the team will be constrained to operate at the (marked) point with (the largest) probability of false alarm less than or equal to $\alpha$.

<table>
<thead>
<tr>
<th>$(P_F^T, P_D^T)$</th>
<th>$(P_F^0, P_D^0)$</th>
<th>$(P_F^1, P_D^1)$</th>
<th>$(P_F^2, P_D^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.00, 0.00)</td>
<td>(0.0, 0.0)</td>
<td>(0.0, 0.0)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(0.01, 0.25)</td>
<td>(0.0, 0.0)</td>
<td>(0.1, 0.5)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(0.02, 0.35)</td>
<td>(0.0, 0.0)</td>
<td>(0.2, 0.7)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(0.10, 0.63)</td>
<td>(0.0, 0.0)</td>
<td>(0.2, 0.7)</td>
<td>(0.5, 0.9)</td>
</tr>
<tr>
<td>(0.14, 0.70)</td>
<td>(0.1, 0.5)</td>
<td>(0.5, 0.9)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(0.18, 0.80)</td>
<td>(0.2, 0.7)</td>
<td>(0.5, 0.9)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(0.28, 0.85)</td>
<td>(0.2, 0.7)</td>
<td>(1.0, 1.0)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(0.30, 0.86)</td>
<td>(0.1, 0.5)</td>
<td>(0.5, 0.9)</td>
<td>(0.5, 0.9)</td>
</tr>
<tr>
<td>(0.35, 0.88)</td>
<td>(0.2, 0.7)</td>
<td>(0.5, 0.9)</td>
<td>(0.5, 0.9)</td>
</tr>
<tr>
<td>(0.60, 0.97)</td>
<td>(0.2, 0.7)</td>
<td>(1.0, 1.0)</td>
<td>(0.5, 0.9)</td>
</tr>
<tr>
<td>(0.75, 0.99)</td>
<td>(0.5, 0.9)</td>
<td>(1.0, 1.0)</td>
<td>(0.5, 0.9)</td>
</tr>
<tr>
<td>(1.00, 1.00)</td>
<td>(1.0, 1.0)</td>
<td>(1.0, 1.0)</td>
<td>(0.5, 0.9)</td>
</tr>
</tbody>
</table>
6.2.3. Independent Joint Randomization

Suppose that each DM is allowed to *independently* randomize his decision rule. By independently we mean that each of the DMs has an "appropriately biased coin" which he flips (prior to making his decision) in order to determine which decision rule should be implemented for his decision. In this case the ROC curves of the individual DMs (Figure 6.1) will also include the line segments joining the marked points. For example, each of the DMs can operate at (0.02, 0.10) by using a coin that selects (0.0, 0.0) with probability 0.8 and selects (0.1, 0.5) with probability 0.2.

Turning our attention to the team ROC curve, suppose that the team wishes to operate at a point on the line segment joining (0.14, 0.70) and (0.18, 0.80), say at (0.16, 0.75); this requires the team to operate half the time at (0.14, 0.70) and half the time at (0.18, 0.80). This can be achieved through independent randomization by having the two DMs independently choose between their two corresponding choices of operating points with probability 0.5. Effectively, the consulting DM does not have to randomize; he should always operate at (0.1, 0.5). Moreover, when the primary DM receives \( u_c = 0 \), he should always operate at (0.2, 0.7), and when the primary DM receives \( u_c = 1 \), he should operate at (0.5, 0.9) with probability 0.5 and at (1.0, 1.0) with probability 0.5.

*Remark.* Recall that given the operating points of the DMs, the team operating point is given by Eqs.(4.16) and (4.17).

But, suppose that the team wishes to operate at a point on the line segment joining (0.10, 0.63) and (0.14, 0.70), say at (0.12, 0.665); this requires the team to operate half the time at (0.10, 0.63) and half the time at (0.14, 0.70). This can not be achieved through independent randomization. If, just like in the previous case, each of the two DMs independently chooses between his two corresponding choices of operating points with probability 0.5, from Table 6.1 it can be seen that the team will be operating with
probability 0.25 at each of (0.02, 0.35), (0.10, 0.63), (0.14, 0.70) and (0.30, 0.86); this implies that the actual team operating point will be (0.14, 0.635) (i.e., the weighted sum of the previous four points), and not (0.12, 0.665) as intended. The reason for this is that since both DMs are randomizing independently and since each DM is not informed of the outcome of the other DM's randomization, the operating point of the consulting DM and the two operating points of the primary DM may be (randomly) combined in an undesirable way (for example, (0.5, 0.9) for the consulting DM with (0.1, 0.5) and (0.5, 0.9) for the primary DM).

The moral of the story is more general. Suppose that a team wants to operate at a point on the line segment joining two consecutive points of its ROC curve. If more than one DMs are required to randomize their decision rules, the points on the line segment can not be achieved; consequently, the team ROC curve is not concave in that region.

Returning to the team ROC curve of our particular example, we obtain from Table 6.1 that only the line segments joining (0.10, 0.63) and (0.14, 0.70), and joining (0.28, 0.85) and (0.30, 0.86) can not be achieved through independent randomization. We explicitly calculated the team ROC when independent randomization is allowed; it consists of all the points of Table 6.1 plus the points (179/1420, 939/1420) and (0.296, 0.856), and includes of course all the line segments joining any two consecutive points (solid line in Figure 6.3).

It is interesting to note in this case that despite both DMs being allowed to randomize, only one has to each time; hence, the team ROC curve remains piecewise linear. Moreover, the randomizations used to obtain the points on the newly added line segments are presented in Table 6.2.
Figure 6.3(a). Team ROC Curve With Independent and Dependent Randomization

Figure 6.3(b). Close Up of the Team ROC Curve of Figure 6.3(a).
(straight line: independent, dotted line: dependent)
Figure 6.3(c). Close Up of the Team ROC Curve of Figure 6.3(b).
(straight line: independent, dotted line: dependent)
Table 6.2. The Non-Concave Portions of the Team ROC Curve: Randomization Probabilities and Operating Points

(i). Line Segment From: (0.10, 0.63) To: (179/1420, 939/1420)
Randomization Probability: \( p_p \leq 37/213 \) [only the primary DM is randomizing]

<table>
<thead>
<tr>
<th></th>
<th>((P_F^p, P_D^p))</th>
<th>((P_F^0, P_D^0))</th>
<th>((P_F^1, P_D^1))</th>
<th>((P_F^E, P_D^E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_p)</td>
<td>(0.10, 0.63)</td>
<td>(0.0, 0.0)</td>
<td>(0.2, 0.7)</td>
<td>(0.5, 0.9)</td>
</tr>
<tr>
<td>(1-p_p)</td>
<td>(0.25, 0.81)</td>
<td>(0.0, 0.0)</td>
<td>(0.5, 0.9)</td>
<td>(0.5, 0.9)</td>
</tr>
</tbody>
</table>

(ii). Line Segment From: (179/1420, 939/1420) To: (0.14, 0.70)
Randomization Probability: \( p_p \leq 11/71 \) [only the primary DM is randomizing]

<table>
<thead>
<tr>
<th></th>
<th>((P_F^p, P_D^p))</th>
<th>((P_F^0, P_D^0))</th>
<th>((P_F^1, P_D^1))</th>
<th>((P_F^E, P_D^E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1-p_p)</td>
<td>(0.05, 0.45)</td>
<td>(0.0, 0.0)</td>
<td>(0.5, 0.9)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(p_p)</td>
<td>(0.14, 0.70)</td>
<td>(0.1, 0.5)</td>
<td>(0.5, 0.9)</td>
<td>(0.1, 0.5)</td>
</tr>
</tbody>
</table>

(iii). Line Segment From: (0.28, 0.85) To: (0.296, 0.856)
Randomization Probability: \( p_c \leq 0.05 \) [only the consulting DM is randomizing]

<table>
<thead>
<tr>
<th></th>
<th>((P_F^p, P_D^p))</th>
<th>((P_F^0, P_D^0))</th>
<th>((P_F^1, P_D^1))</th>
<th>((P_F^E, P_D^E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_c)</td>
<td>(0.28, 0.85)</td>
<td>(0.2, 0.7)</td>
<td>(1.0, 1.0)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(1-p_c)</td>
<td>(0.60, 0.97)</td>
<td>(0.2, 0.7)</td>
<td>(1.0, 1.0)</td>
<td>(0.5, 0.9)</td>
</tr>
</tbody>
</table>

(iv). Line Segment From: (0.296, 0.856) To: (0.30, 0.86)
Randomization Probability: \( p_c \leq 0.05 \) [only the consulting DM is randomizing]

<table>
<thead>
<tr>
<th></th>
<th>((P_F^p, P_D^p))</th>
<th>((P_F^0, P_D^0))</th>
<th>((P_F^1, P_D^1))</th>
<th>((P_F^E, P_D^E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1-p_c)</td>
<td>(0.14, 0.70)</td>
<td>(0.1, 0.5)</td>
<td>(0.5, 0.9)</td>
<td>(0.1, 0.5)</td>
</tr>
<tr>
<td>(p_c)</td>
<td>(0.30, 0.86)</td>
<td>(0.1, 0.5)</td>
<td>(0.5, 0.9)</td>
<td>(0.5, 0.9)</td>
</tr>
</tbody>
</table>
6.2.4. Dependent Joint Randomization

Consider again the problem of the previous section and suppose that the team wants to operate at a point on the line segment joining (0.10, 0.63) and (0.14, 0.70), say (0.12, 0.665) as before. Further suppose that instead of each DM flipping his own personal coin, the team designer flips a 'global team' coin, the outcome of which is somehow transmitted to the DMs prior to making their decision. Thus, the DMs may use the outcome of this coin to make their decisions. But under this protocol, by knowing the outcome of the 'global team' coin, each DM also knows the decision rule that the other DM is going to implement; hence, the term dependent randomization. Then, the DMs may combine their decisions in such a way that only desirable combinations of their decision rules take place and the points in any line segment joining two points of the team ROC curve become feasible.

In our example, the DMs may agree off-line the if the outcome of the (assumed unbiased) global coin is heads [tails] the consulting DM will employ (0.5, 0.9) [(0.1, 0.5)] as his operating point, and the primary DM will employ (0.0, 0.0) and (0.2, 0.7) [(0.2, 0.5) and (0.5, 0.9)] as his operating points. This results in the team operating at the desired (0.12, 0.665) point. The team ROC curve under dependent randomization is presented in Figure 6.3 with a dotted line where it differs from the team ROC curve under independent randomization.

Again this corresponds to a more general result. Suppose that every point of team ROC curve under the deterministic decision rules (or even under independent randomization for that matter) is represented by the corresponding vector starting at the origin and ending at that point. Then the team ROC curve under dependent randomization will be given by the convex hull of all the vectors. Hence, any team ROC curve under dependent randomization will always be concave.

Why then don't we invoke the dependent randomization assumption from the beginning and put to rest all the intricacies which arise when it is missing? The reason for
not doing this, without any reservations, is that this assumption imposes indirectly, if not directly, restrictions on the decentralized nature of the team. Since communications between the members of the team are limited, we can only reluctantly accept that the outcome of the 'global' coin will be transmitted off-line to all the DMs prior to making their decision. Moreover, even if we are able to endow the team with this extra communication capacity, it may be more beneficial for the team to have information different from the outcome of the 'global' coin transmitted.

Still, if concave ROC curves are desirable, there exist other ways to circumvent this issue. One may assume that the DMs agreed off-line on a particular sequence for employing the different sets of decision rules. This implies that the dependent randomization has taken place off-line. If an outside observer is informed of the sequence, he will know that at each instance of the game that the outcome of the team is biased towards one or the other outcome. But to an observer who is not aware of the sequence, the team performance will be identical to the case of the dependent randomization. Another alternative is to assume the existence of a common clock for all the team members; the clock will be used for selecting the decision rules to be employed. Since teams in our framework are implicitly assumed to operate in a synchronous manner, the imposition of the common clock is not really restrictive.

6.3. SMOOTH ROC CURVES

As was clearly demonstrated in the previous section, the ROC curve of a team does not have to be concave if independent randomization is employed by the DMs. Further, it was shown that for the concavity to be lost both DMs have to 'abruptly jump' from their original operating points to new operating points (for example, refer to Table 6.1 to see how all the operating points of the DMs change as the team ROC curve moves from (0.10, 0.63) to (0.14, 0.70)). Also in the example of the two DM parallel team in [R87], the team
ROC curve becomes non-concave when the decision rule of the fusion center abruptly switches from AND to OR. Therefore, it is logical to hypothesize that the non-concave team ROC curve is a result of individual ROC curves that are not strictly concave and/or smooth. In this section we show that this hypothesis is not true, as the non-concave ROC curve is a more general characteristic encountered even when all the relevant ROC curves are very smooth. In order to demonstrate this we employ the two DM tandem team, thus eliminating the discontinuous ramifications of the fusion center.

We now present examples of two DM tandem teams with binary communications whose ROC curves are not concave. The DMs of the teams are only allowed to employ deterministic decision rules.

6.3.1. Strictly Concave Primary DM and Piecewise Linear Consulting DM

Suppose that the ROC curve of the consulting DM presented in Figure 6.4. It is obviously three segment piecewise linear which implies that he receives four discrete observations. Since the decision rules are deterministic, the primary DM can only operate at the marked points of the ROC curve; the line segments joining the marked points are just presented for convenience.

Also, suppose that the ROC curve of the primary DM is given by Eq.(3.14) for $\alpha = 0.2$ (Figure 6.5). This means that the consulting DM is trying to distinguish between two exponential distributions with different means, one being five times larger than the other. This is not only a strictly concave ROC curve, but it is also very smooth curve as it is infinitely differentiable at every point (except at the origin).

The team ROC curve is not concave. It is presented in Figure 6.7, together with the ROC curve of the primary DM so that the improvement can be duly noted.
Figure 6.4. Three Segment Piecewise Linear ROC Curve

Figure 6.5. Strictly Concave ROC Curve of DM with Exponential Observations (for $\alpha = 0.2$)
Figure 6.6(a). Team ROC Curve with Strictly Concave Primary DM

Figure 6.6(b). Close Up of Team ROC Curve of Figure 6.6(a).
6.3.2. Piecewise Linear Primary DM and Strictly Concave Consulting DM

We now reverse the position of the DMs of the team of the previous section and designate the piecewise linear DM of Figure 6.4 as the primary DM, and the strictly concave DM of Figure 6.5 as the consulting DM. The team ROC curve is not concave in this case as well. It is presented in Figure 6.7, together with the ROC curve of the consulting DM so that the improvement can be noted.

Figure 6.7(a). Team ROC Curve with Strictly Concave Consulting DM
Figure 6.7(b). Close Up of Team ROC Curve of Figure 6.7(a).

Figure 6.7(c). Close Up of Team ROC Curve of Figure 6.7(a).
6.3.3. Both DMs Strictly Concave

Since this is the most interesting case, we are not satisfied to simply present two strictly concave DMs which yield non-concave team ROC curve, but we try to convey some intuition to it. Consider again the team of section 6.3.2. The team consisting of a strictly concave consulting DM and a piecewise linear primary DM has a non-concave ROC curve. If we can approximate the piecewise linear ROC curve with a strictly concave ROC curve (within some \( \varepsilon > 0 \)), then the team consisting of the original consulting DM and the new DM as the primary DM will also have a non-concave team ROC curve. Thus, we concentrate our efforts to approximating arbitrarily close any piecewise linear DM with a strictly concave DMs.

Consider any piecewise linear DM. His ROC curve consists of the following \( N+2 \) points \( (N \geq 1) \):

\[
(P_{F_0}, P_{D_0}) = (0, 0), \ (P_{F_1}, P_{D_1}), \ldots, \ (P_{F_N}, P_{D_N}), \ (P_{F_{N+1}}, P_{D_{N+1}}) = (1, 1)
\]

Also consider a DM who receives two observations: the discrete observation, whose ROC curve was just described, and one exponential observation, whose ROC curve is given by Eq.(3.14) for some \( \alpha < 1 \). Then, it is not hard to obtain (using some tedious but straightforward algebra) that the ROC curve of this DM will be given by:

\[
P_D^F = P_{D_i} + (P_F^1 - P_F)^\alpha \sum_{k=i+1}^{N+1} \left[ \frac{(P_{F_k} - P_{F_k})}{(P_{D_k} - P_{F_k})} \left( \frac{P_{D_k} - P_{D_k}}{P_{F_k} - P_{F_k}} \right)^{1-\alpha} \right] - \alpha ; \ P_F(i) \leq P_F^F \leq P_F(i+1)
\]

for: \( i = 0, 1, \ldots, N \) \hspace{1cm} (6.1)

where:

\[
P_F(0) = 0 \hspace{1cm} (6.2a)
\]

\[
P_F(i) = P_{F_i} + \left( \frac{P_{F_i} - P_{F_{i+1}}}{P_{D_i} - P_{D_{i+1}}} \right)^{1-\alpha} \sum_{k=i+1}^{N+1} \left( P_{F_k} - P_{F_k} \right) \left( \frac{P_{D_k} - P_{D_k}}{P_{F_k} - P_{F_k}} \right)^{1-\alpha} \hspace{1cm} (6.2b)
\]
and:

\[ P_F(N+1) = 1 \]  \hspace{1cm} (6.2c)

**REMARK.** One way to obtain the ROC curve of Eq.(6.1) is to consider a 'dummy' two DM tandem team. In that team, the primary receives the exponential observation, the consulting DM receives the discrete observation and can communicate \( N + 1 \) messages to the primary DM. Clearly, this 'dummy' team can achieve centralized performance. Then as a function of the threshold \( \eta \), the team probability of false alarm and detection will be given as follows:

\[
P_F^i(\eta) = P_{F_i} + \left( \frac{\alpha}{\eta} \right)^{\frac{1}{1-\alpha}} \sum_{k=1}^{N+1} (P_{F_k} - P_{F_{k-1}}) \left( \frac{P_{D_k} - P_{D_{k-1}}}{P_{F_k} - P_{F_{k-1}}} \right)^{\frac{1}{1-\alpha}} ; \quad \eta(i+1) \leq \eta \leq \eta(i) \]

\[
P_D^i(\eta) = P_{D_i} + \left( \frac{\alpha}{\eta} \right)^{\frac{1}{1-\alpha}} \sum_{k=1}^{N+1} (P_{F_k} - P_{F_{k-1}}) \left( \frac{P_{D_k} - P_{D_{k-1}}}{P_{F_k} - P_{F_{k-1}}} \right)^{\frac{1}{1-\alpha}} ; \quad \eta(i+1) \leq \eta \leq \eta(i) \]

\[
\text{for: } i = 0, 1, ..., N + 1
\]

where:

\[
\eta(0) = \infty \]

\[
\eta(i) = \alpha \frac{P_{D_i} - P_{D_{i-1}}}{P_{F_i} - P_{F_{i-1}}} ; \quad i = 1, 2, ..., N + 1
\]

\[
\eta(N+2) = 0
\]  \hspace{1cm} (6.5c)

It is easy to check that the ROC curve of Eq.(6.1) is not only strictly concave, but it is differentiable as well. Moreover, as \( \alpha \to 1 \) (i.e., as the exponential observation becomes increasingly useless) the ROC curve approaches the original piecewise linear ROC curve. Thus, by selecting \( \alpha \) sufficiently close to 1 we can approximate arbitrarily close a piecewise linear ROC curve with a strictly concave and differentiable ROC curve.
REMARK. The ROC curve of a DM who receives an exponential observation is presented in Figure 6.8 for different values of $\alpha$; note that the observation becomes increasingly useless as $\alpha \to 1$.

Figure 6.8. ROC Curves for Exponential Observations for Different Values of $\alpha$

Consider thus a DM who receives the discrete observation of Figure 6.4 and an exponential observation with $\alpha = 0.9$. From Eqs.(6.1) and (6.2) the ROC curve of this DM is given by:

$$P_D = \begin{cases} 
5.5603 \, (P_F)^{0.9} & ; \quad 0.0 \leq P_F \leq 0.1 \\
0.7943 \, (P_F - 0.1)^{0.9} & ; \quad 0.1 \leq P_F \leq 0.200000762 \\
0.2445 \, (P_F - 0.2)^{0.9} & ; \quad 0.200000762 \leq P_F \leq 1.0
\end{cases}$$

(6.6)

The ROC curve of Eq.(6.6) is presented in Figure 6.9. It is also presented in Figure 6.10 together with the ROC curve of the piecewise ROC curve of Figure 6.4, in order to demonstrate that they are indeed very similar.
Figure 6.9. ROC Curve of Eq.(6.6)

Figure 6.10. ROC Curves of DMs of Figures 6.4 and 6.9
Consider the two DM tandem team with the DM of Figure 6.9 as the primary DM, and the DM of Figure 6.5 as the consulting DM. The team ROC curve is presented in Figure 6.11 and it clearly is not concave. It is also presented in Figure 6.12 together with the team ROC curve of Figure 6.7(a) (piecewise linear primary DM, same consulting DM), in order to directly compare them. As intended and expected, the two team ROC curves look identical for all practical purposes.

Therefore, even strictly concave and differentiable individual ROC curves may yield a non-concave ROC curve for the two DM tandem team.

REMARK. This result is also true for infinitely differentiable individual ROC curves, since this 'smoothing' process can be repeated more than once.

6.4. DISCUSSION

6.4.1. Summary

Three different types of decision rules were discussed: the deterministic decision rules, the decision rules with independent randomization and the decision rules with dependent randomization. If a DM employs deterministic decision rule and receives a discrete observation, his ROC curve consists of a countable number of disjoint points.

Under either the deterministic decision rules or under decision rules with independent randomization, the ROC curve of any team does not have to be concave\(^1\). This is so because the ROC curves of the two simplest possible teams (i.e., the two DM tandem team and the two DM parallel team) do not have to be concave even if the individual ROC curves of the DMs are infinitely differentiable (i.e., very smooth). Under the decision rules with

\(^1\) The notion of concavity of a ROC curve under deterministic decision rules does not make sense if all DMs receive discrete observations because we can not define the concavity of a number of isolated points. In this case, the upper bound of the convex hull of the isolated points is concave (also, see section 3.3).
Figure 6.11. Team ROC Curve for Strictly Concave DMs

Figure 6.12. Team ROC Curves of Figures 6.7(a) and 6.11
dependent randomization, the ROC curve of any team is concave, because if the team can operate at a point \((P_F^0, P_D^0)\) and at a point \((P_F^1, P_D^1)\), then it can operate at any point on the line segment joining \((P_F^0, P_D^0)\) and \((P_F^1, P_D^1)\).

When the various problems in the distributed hypothesis testing framework are analyzed from the Bayesian viewpoint, the non-concavity of the team ROC curve does not cause any additional difficulties whatsoever. In order to see this recall that, given a threshold \(\eta\), a necessary condition that the optimal team operating point needs to satisfy is that its tangent has slope \(\eta\) (or that \(\eta\) is its subgradient, if the tangent does not exist). From a geometric point of view, to obtain the optimal operating point, a line of constant slope \(\eta\) should 'slide down' the \(P_D\) axis and the first point at which the line intersects with the ROC curve is the optimal point. Then, it can be easily obtained by contradiction that the optimal operating point will lie in a concave region of the ROC curve (or at a corner point if the ROC curve either is piecewise linear or consists of a set of disjoint points).

A direct consequence of this is that the optimal operating point is the same under any type of decision rule. This is not at all surprising. In fact, it is expected since in the various papers where the Bayesian formulation was considered and the optimal decision rules were obtained, it was shown that the optimal decision rules are deterministic and given by likelihood ratio tests with constant thresholds (for example, see [TS81], [ET82] and [PA86] among others).

This is the reason for not bringing up earlier the concept of the non-concave team ROC curve. Since in the previous two chapters we employed the Bayesian formulation of the problems, we assumed without loss of generality that the DMs may use decision rules with depended randomization. We were thus able to include as part of the ROC curve the line segments joining (corner) points of the ROC curves of discrete DMs (or teams of DMs) and obtain 'nice' looking figures of the ROC curves. Therefore, all the results obtained in the previous two chapters are not affected. For example, the optimal configuration for the two DM tandem team which
consists of the DMs of Figure 4.8 depends on the value of the threshold $\eta$ under any type of decision rules.

**REMARK.** This is also the reason for Definition 3.1a. Since in the Bayesian formulation the optimal operating points lie in a concave region of the ROC curve (or at a corner point if the ROC curve either is piecewise linear or consists of a set of disjoint points), the ROC curve may be replaced by the upper bound of the convex hull of the ROC curve, without loss of generality.

### 6.4.2. The Neyman-Pearson Formulation

On the other hand, the type of the decision rules is very important when the Neyman-Pearson formulation of the problems is considered. It is immediately obvious that if randomization is not allowed, all levels of false alarm are not attainable when, for example, the DMs of the team receive discrete observations. Furthermore, since for a given level of false alarm the team may be required to operate at a point where the team ROC curve is not concave, the particular type of the decision rules employ will determine the level of probability of detection that the team is able to achieve. For example, for a false alarm level of, say, 0.25, the optimal configuration for the two DM tandem team which consists of the DMs of Figure 4.8 may depend on the type of decision rules employed; further work is needed to determine the optimal architecture for each type (by the way, under both types of randomized decision rules, for $P_F^t = 0.25$ the optimal configuration has the worse DM as the primary DM and achieves $P_D^t = 0.829$).

**REMARK.** Now that the three types of decision rules have been defined and analyzed, Definition 3.1b of the better DM for the Neyman-Pearson formulation can be justified. Note that the ranking of two DMs (or teams of DMs) may depend on the type of the decision rules employed.
We would like to again emphasize that, independent of the type of the decision rules employed, all the corner points in the various examples throughout this thesis can be reached by the corresponding teams. An easy way to verify this is to recall that, for piecewise linear ROC curves, the optimal solution to a Bayesian problem will always lie on a corner point. Furthermore, we argued above that, under any type of decision rules, the optimal solution of the Bayesian problems can always be reached. Therefore, under any type of decision rules, all corner points can be reached.

But, even if only corner points are considered, all the results derived in the previous two chapters for the Bayesian framework are also valid in the Neyman-Pearson framework (for example, no optimum configuration exists for the two DM tandem team, no optimum architecture exists for the three DM team, and the infinite tandem team, with binary messages, performing binary hypothesis testing, cannot achieve perfect detection unless the conditions of Proposition 5.1 hold).

6.4.3. The Significance of the Non-Concave Team ROC Curve

What is the significance of the non-concave team ROC curve? Mathematically, it states that a series of concave sets (i.e., individual ROC curves) gets mapped into a non-concave set (i.e., the team ROC). Since this mapping, represented by the likelihood ratio tests with the coupled thresholds, is highly non-linear, the result is not surprising.

From the organizational design point of view, the non-concave team ROC curve supports the chaotic theories for teams which state that teams do not operate in a prescribed and predictable pattern. The team ROC curve becomes concave if dependent randomization is allowed; that is, it becomes concave if the outcome of the 'global team coin' is communicated to all the DMs, or if a common clock exists, or if the DMs have agreed off
line on a (pre)randomized sequence. This implies that the inconsistencies in the behavior of a team may be cured with increased communication and/or supervision (communication of the outcome of the 'global coin'), coordination (existence of common clock) and increased training and cooperation (knowledge of and agreement on pre-randomized sequence). Therefore, this result reinforces the theories about the positive effects of coordination, cooperation and training on the performance of an organization.
CHAPTER 7

A Multiple Hypothesis Testing Paradigm

7.1. INTRODUCTION

As has been mentioned repeatedly throughout this thesis, the combinatorial complexity of the distributed hypothesis testing problems almost always explodes as the number of the team DMs, or of the hypotheses, or of the messages increases. On the other hand, simple, suboptimal and usually intuitive solutions have been shown to perform relatively well. Moreover, in chapter 4 it was shown that the optimal team architecture in general depends on the characteristics of the environment. Therefore, it becomes apparent that it is worthwhile for the designer of the team to: (i) sacrifice some of the team performance, in order to keep the complexity under control and, (ii) take into consideration the specific characteristics of the environment in which the team operates, in order to improve the team performance. To demonstrate this approach, a complex multiple hypothesis testing problem with a team consisting of several DMs is presented.

In section 7.2, the problem is defined and the optimal solution is presented. The team has to select one of \( M \) different hypotheses, taking into account different costs for hypothesis misclassification. The team consists of \( N \) DMs; one of them is referred to as the primary DM and is responsible for the final team decision, and the rest are known as the consulting DMs (Figure 7.1). Each consulting DM receives his observation and then computes and communicates a \( U \)-ary message to the primary DM. Upon receipt of his own observation and the messages from the consulting DMs, the primary DM makes the final team decision declaring one of the \( M \) hypotheses to be true. This version of the problem has been formulated and solved in [TP90]. In general, in order to determine the optimal
team configuration (i.e., the optimal primary DM), the problem needs to be solved $N$ times; each time with a different primary DM.

In section 7.3, a suboptimal decision scheme is presented which has a dual objective. The first objective is to design a protocol that makes an educated selection, among the DMs of the team, for the DM who will be designated as the primary DM. The second objective is to decrease the computational complexity of the original problem by reducing the complex $M$-ary hypothesis testing problem into a series of binary hypothesis testing problems. The proposed decision scheme is a corrected adaptation of the scheme in [D89] for object recognition with a single sensor.

Finally in section 7.4, in order to further demonstrate the proposed problem dependent analysis, we examine a particular instant of the problem of section 7.2 where the multiple hypothesis consists of a number of distinct attributes. The observation of each consulting DM contains information for just one of the attributes, which implies that each consulting DM has his particular domain of expertise. The observation of the primary DM contains information about all the attributes of the problem. We present some thoughts on how the suboptimal decision scheme of the previous section can be modified in order to take maximum advantage of the particular structure of this problem.

### 7.2. PROBLEM DEFINITION AND SOLUTION

**PROBLEM 7.1.** Consider a team which consists of $N$ DMs and performs $M$-ary hypothesis testing. Each hypothesis $H_m$ occurs with a known prior probability $p_m = P(H_m)$, for $m = 0, 1, ..., M-1$. All the DMs receive conditional independent observation. One DM is designated as the primary DM and the others are referred to as the consulting DMs (Figure 7.1). Each consulting DM $n$ transmits to the primary DM a $U$-ary message $u_n \in \{0, 1, ..., U-1\}$ based on his observation $y_n$, $n = 2, 3, ..., N$. The primary DM considers his own observation and the communications from the consulting DMs and makes the final team
Figure 7.1. The Team of Problem 7.1.

decision \( u_1 \in \{0, 1, ..., M-1\} \), declaring one of the hypotheses to be true. There exists a cost \( J(u_1, H) \) associated with the team deciding \( u_1 \) when \( H \) is the true hypothesis. The decision rules which minimize the team normalized probability of error are requested.

Given the communication vector \( u = (u_2, u_3, ..., u_N) \in \{0, 1, ..., U-1\}^{N-1} \) transmitted from the consulting DMs to the primary DM, the optimal decision rule of the primary DM is given by the following likelihood ratio tests with constant thresholds:

\[
\sum_{m=1}^{M-1} \Lambda_m(y_1) p_m[J(i, H_m) - J(j, H_m)] \prod_{n=2}^{N} \frac{P(u_n | H_m)}{P(u_n | H_0)} \begin{cases} \geq & u_1 \neq i \\ < & u_1 = j \end{cases} \begin{cases} \geq & p_0[J(j, H_0) - J(i, H_0)] \end{cases}
\]  \hspace{1cm} (7.1)

for: \( i = 1, 2, ..., M-2 \), and \( j = m+1, m+2, ..., M-1 \), where:

\[
\Lambda_m(y_n) = \frac{P(y_n | H_m)}{P(y_n | H_0)} \hspace{1cm} ; \hspace{1cm} m = 1, 2, ..., M-1; \hspace{1cm} n = 1, 2, ..., N. \hspace{1cm} (7.2)
\]

The optimal decision rules of the consulting DMs are given by:
\[ \sum_{m=1}^{M-1} A_m(y_n) p_m \sum_{u_1=0}^{M-1} \left[ P(u_1 \mid u_n = i, H_m) - P(u_1 \mid u_n = j, H_m) \right] J(u_1, H_m) \geq \] 
\[ u_n \neq i \]
\[ u_n \neq j \]
\[ p_0 \sum_{u_1=0}^{M-1} \left[ P(u_1 \mid u_n = j, H_0) - P(u_1 \mid u_n = i, H_0) \right] J(u_1, H_0) \] 
(7.3)

for: \( i = 0, 1, \ldots, U-2; \ j = i+1, i+2, U-1; \ n = 2, 3, \ldots, N. \)

Suppose that the optimal team configuration is to be determined for some given prior probabilities and some given costs. Problem 7.1 needs to be solved \( N \) different times each time with a different primary DM to obtain the optimal team performance. Recalling the results of section 4.3 for a team with a single consulting DM performing binary hypothesis testing, one concludes that the optimal configuration for the team can only be determined in this manner because it depends both on the prior probabilities and on the associated costs.

REMARK. As usual in this thesis, the \( U \)-ary messages from the consulting DMs to the primary DM are denoted as \( 0, 1, \ldots, U-1 \) for the sake of simplicity. For that matter they can be denoted with any \( U \) distinct symbols or names. Note that consulting DM transmitting a 0 message does not necessarily indicate that \( H_0 \) is the correct hypothesis. The important fact is that each consulting DM has a \( U \)-ary message in his disposition and should try, together with the primary DM, to make optimal use of them so that the team expected cost be minimized.

A Gauss-Seidel algorithm for determining the optimal decision rules was presented in [TP90]. But, the optimality conditions of Eqs.(7.1) and (7.3) are just necessary conditions. If there exists a suboptimal set of decision rules for the team which satisfies these conditions, then the algorithm could converge to a suboptimal cost. Even if we assume that this does not occur for some 'well behaved' probability density functions, implementation of the algorithm for six or more hypotheses is highly non-trivial even if the consulting DM
can only transmit binary messages; it is hard to determine the two five-dimensional decision regions and calculate the corresponding probabilities. In fact, we do not have any feeling for what these regions would look like. It should be an interesting problem for future research to implement an algorithm which solves a (Gaussian) six hypothesis testing problem for a team which consists of six DMs and binary communications. Since the thresholds on the likelihood ratio hyperplane can be translated to thresholds in the observation axis, this should provide some insight on how information should be summarized and fused.

In conclusion, the determination of the optimal decision rules and their associated decision regions is highly non-trivial, even for a small number of alternative hypotheses and communication messages; this suggests that it should be worthwhile to abandon optimality for a suboptimal scheme which is easily implementable.

7.3. A SUBOPTIMAL DECISION SCHEME

7.3.1. The Normalized Prior Probabilities

(i). Elementary Hypotheses

As can be seen from Eq.(3.5), the definition of the decision threshold for binary hypothesis testing, the ratio of the additional costs incurred by the team when it makes the wrong decision influences the decision in the same way that the ratio of the prior probabilities does; for example, doubling the additional cost incurred by the team when it makes the wrong decision under hypothesis $H$ is interpreted by the team, in the optimal decision rules, as doubling the relative frequency of occurrence of $H$. We try to extend this notion to the $M$-ary hypothesis case by introducing the normalized prior probabilities.
We try to summarize all of the information given by the prior probabilities and the misclassification costs into 'relative frequencies' and thus define the (elementary) normalized prior probabilities as follows:

\[
\pi_m = \frac{p_m \sum_{u=0}^{M-1} \frac{p_u}{1-p_m} [J(u, H_m) - J(m, H_m)]}{\sum_{q=0}^{M-1} p_q \sum_{u=0}^{M-1} \frac{p_u}{1-p_q} [J(u, H_q) - J(q, H_q)]}; \quad \text{for } m = 0, 1, ..., M-1 \quad (7.4)
\]

The restrictive assumption, implied in the above definition, is that given the true hypothesis \( H \) the relative frequencies of the different types of error are given by the prior probabilities.

Note that for the binary hypothesis testing (where given the true hypothesis only one type of error may occur), the normalized prior probabilities are 'optimal' in the sense that their ratio is equal to the decision threshold \( \eta \) (i.e., \( \eta = \pi_0/\pi_1 \)). Furthermore, in the case of the minimum error cost function (i.e., \( J(u, H_m) = 1 \) if \( u \neq m \) and \( J(u, H_m) = 0 \) if \( u = m \)) the normalized prior probabilities reduce to the actual prior probabilities; since all errors are equally costly the relative frequencies of the hypotheses are not affected.

(ii). Composite Hypotheses

Consider \( F_0 \) and \( F_1 \), two distinct subsets of \( \{0, 1, ..., M-1\} \), and the binary hypothesis testing problem between the following two composite hypotheses, that is hypotheses which consist of a set of elementary hypotheses:

\[
H_{F_0}: \sum_{m \in F_0} \frac{p_m}{\sum_{q \in F_0} p_q} P(y \mid H_m) \quad \text{vs.} \quad H_{F_1}: \sum_{m \in F_1} \frac{p_m}{\sum_{q \in F_1} p_q} P(y \mid H_m) \quad (7.5)
\]

For this binary problem the hypotheses occur with prior probabilities:
\[
\frac{\sum_{m \in F_i} p_m}{\sum_{q \in F_0 \cup F_1} p_q} \quad ; \quad \text{for } i = 0, 1.
\]  
(7.6)

The normalized prior probability of \( F_i \) for this binary problem is defined as:

\[
\pi_{F_i, F_{1-i}} = \frac{\sum_{m \in F_i} \sum_{u \in F_{1-i}} \frac{p_u}{\sum_{q \in F_{1-i}} p_q} [J(u, H_m) - J(m, H_m)]}{\sum_{j=0}^1 \sum_{m \in F_j} \sum_{u \in F_{1-j}} \frac{p_u}{\sum_{q \in F_{1-j}} p_q} [J(u, H_m) - J(m, H_m)]}
\]  
(7.7)

The two sets \( F_0 \) and \( F_1 \) of the hypotheses have to be mutually exclusive, but do not have to be collectively exhaustive (i.e., \( F_0 \cap F_1 = \emptyset \), \( F_0 \cup F_1 \subseteq \{0, 1, \ldots, M-1\} \)). Note that in the case of the minimum error cost function, the normalized prior probability of \( F_i \) of Eq.(7.7) reduces to the prior probability of \( F_i \) of Eq.(7.6). Moreover, the decision threshold for the composite binary hypothesis testing problem is given by:

\[
\eta(F_0, F_1) = \frac{\sum_{q \in F_0} p_q \sum_{m \in F_0} \sum_{u \in F_1} p_u [J(u, H_m) - J(m, H_m)]}{\sum_{q \in F_1} p_q \sum_{m \in F_1} \sum_{u \in F_0} p_u [J(u, H_m) - J(m, H_m)]}
\]  
(7.8)

Hence just like in the usual binary case, the decision threshold can be broken down as the product of two ratios; the first being a ratio of the prior probabilities and the second being a ratio of the (weighted) additional costs.
7.3.2. The Binary Decision Tree

Consider the following suboptimal decision scheme for a single DM to perform $M$-ary detection. The multiple hypothesis testing problem will be broken into $M-1$ binary hypothesis testing problems. For this an appropriate binary decision tree\(^1\) is constructed (Figure 7.2). Consider any tree with $M$ leaves (i.e., terminal nodes) having the following property: there exists a single non-terminal node that has exactly two edges emanating from it, and the rest of the non-terminal nodes have exactly three edges emanating from them. It is not hard to see inductively that such a tree contains exactly $M-1$ non-terminal nodes. The special non-terminal node with just the two edges is referred to as the *source node* and the other $M-1$ non-terminal nodes are referred to as the *decision nodes*; also every one of the $M$ terminal nodes corresponds to one of the $M$ hypotheses. The hypotheses can be assigned to the terminal nodes in any order.

In the decision tree, there exists a unique (directed) path from the source node to each and every of the terminal nodes; in fact, starting from the source node any hypothesis can be declared (i.e., any terminal node can be reached) with a series of binary decisions. There is a total of $M-1$ binary decisions that can be made; one in each decision node of the tree. Consider any node $t$ (except the source node) of the decision tree. The immediate predecessor of $t$ on the path from the source node to $t$ is called the *parent* node of $t$.\(^2\) Similarly, consider any non-terminal node $t$; the two nodes $t_0$ and $t_1$ adjacent to $t$, which are not the parent of $t$, are called the *children* nodes of $t$. Finally, for any node $t$, the terminal nodes of the subtree with $t$ as the source node is referred to as the *feasible* set of $t$ and is denoted by $F(t)$. Note, that the feasible set of the source node consists of all the

---

1 A *tree* is a connected acyclic graph. A *binary decision tree* is defined as a directed tree which has the following three properties: one *source* node (i.e., a node with no arcs directed into it); there are exactly two directed arcs emanating from each non-terminal node; each terminal node is a destination (i.e., the node's single arc is directed into it). For a complete treatment of graph theory, the reader is referred to [PS82].

2 The source node does not have a parent node.
hypotheses \( H_0, H_1, \ldots, H_{M-1} \), and that the feasible set of any terminal node consists of the one hypothesis that has been assigned to that terminal node.

Consider any non-terminal node \( t \) and denote by \( t_0 \) and \( t_1 \) its two children. Then, the binary decision that is to be made at node \( t \), is conditional on one of the hypotheses on the feasible set of \( t \) being true, and requires the selection between two composite hypotheses: the feasible set of \( t_0 \) and the feasible set of \( t_1 \); that is:

\[
H_{t_0} : \sum_{m \in F(t_0)} \frac{p_m}{\sum_{q \in F(t_0)} p_q} P(y \mid H_m) \quad \text{vs.} \quad H_{t_1} : \sum_{m \in F(t_1)} \frac{p_m}{\sum_{q \in F(t_1)} p_q} P(y \mid H_m) \quad (7.9)
\]

The prior probabilities for the hypotheses were defined in Eq.(7.6), as was the decision threshold in Eq.(7.8). Thus, at every decision node \( t \), the primary DM will be employing a likelihood ratio test like the one in Eq.(3.6), comparing the likelihood ratio of the probability density functions defined in Eq.(7.9) to the decision threshold.

7.3.3. The Suboptimal Algorithm

A suboptimal algorithm is now proposed for solving the Problem 7.1; thus the team architecture with a single primary DM, the consulting DMs and the binary communications is preserved so that the structure of the original problem be preserved. In order to compensate for the extra error incurred by the team, the algorithm needs to be computationally simple and easy to implement.

Suppose for the moment that a specific configuration for the team is given (i.e., the primary DM has been specified); this assumption will be relaxed in the sequel. Moreover, suppose that there exist exactly \( M - 1 \) consulting DMs. This is not a really restrictive assumption. If the actual number of consulting DMs is less than \( M - 1 \), some 'dumb' (totally worthless) DMs can be added to the team; if more than \( M - 1 \) consulting DMs exist,
the primary DM can fuse the messages of some groups of DMs into single binary messages using some threshold rule (for example, see section 4.4).

Given a primary DM and an associated binary decision tree, the consulting DMs will be assigned so that they can assist the primary DM in making the team decision. The suggested way to do this is to assign one consulting DM to each decision node. Then at every decision node \( t \), the primary DM and the corresponding consulting DM will behave like a two DM tandem team in order to make a decision for the binary hypothesis testing problem \( (H_0 \ vs. \ H_t) \). It is important to note that, unfortunately, when composite hypotheses are considered, the conditional independence assumption is lost because, given the true state of the environment, each and every DM can update his 'beliefs' on the probability distributions of the other DMs (to see this, consider the composite hypothesis which consists of gaussian elementary hypotheses with different means and very small variance; then, knowledge of the true hypothesis clearly gives specific information on the probable values of this observation). As we are try to design a suboptimal decision scheme, each two DM tandem team is analyzed as though the conditional independence assumption was still valid.

Since the two DM tandem team has been completely analyzed (section 4.3), the three decision thresholds (two for the primary DM and one for the consulting DM) of Eqs.(4.3), (4.4) and (4.5) respectively, which completely define the optimal team decision making process can be written by inspection:

For the primary DM:

\[
\begin{align*}
\text{If } u_{c,t} &= 0 : \quad \Lambda(y_p) \begin{cases} 
1 & \quad 1 - P_{E_t}^c \\
0 & \quad 1 - P_{D_t}^c
\end{cases} \quad \eta_t = \eta_0, t
\end{align*}
\]
If $u_{c,t} = 1$:

$$\Lambda(y_p) \quad \begin{cases} u_{p,t} = 1 \\ u_{p,t} = 0 \end{cases} \quad \frac{p^c_{c,t}}{p^c_{D,t}} \eta_t = \eta_{1,t} \quad (7.11)$$

For the consulting DM:

$$\Lambda(y_c,t) \quad \begin{cases} u_{c,t} = 1 \\ u_{c,t} = 0 \end{cases} \quad \frac{p^c_{1,t} - p^c_{0,t}}{p^c_{D,t} - p^c_{0,t}} \eta_t = \eta_{c,t} \quad (7.12)$$

where the notation has the usual meaning and the subscript $t$ indicates that the thresholds are associated with decision node $t$ (i.e., $H_{s_0}$ vs. $H_{s_1}$).

7.3.4. Discussion

The suboptimal decision scheme will perform especially well as compared to the optimal decision scheme, when there exist particular hypotheses (different for each DM) which a DM can detect better than others. Such DMs will be assigned at the decision node of their 'expertise', thus reducing the loss of information caused by the processing of their observations into messages, and consequently reducing the degradation of the team performance caused by the suboptimal decision rules.

Reviewing the decision scheme of the previous subsection, one DM is designated as the primary DM (DM 1 is the primary DM for the team of Figure 7.1), just like in the optimal decision scheme. Then the multiple hypothesis testing problem is broken down into a sequence of $M-1$ binary hypothesis testing problems with composite hypotheses; this sequence is represented by a decision tree (Figure 7.2). The consulting DM who is associated with every such binary problem and the primary DM operate as a two DM tandem team. Every time the primary DM makes a decision, the process effectively moves into a new node in the decision tree and a new consulting DM comes to assist the primary DM. The new two DM tandem team will make its decision based only on the observations.
of its two DMs. Thus, it is assumed that the primary DM does not take explicitly into consideration the messages that he may have previously received from other consulting DMs; these messages are taken implicitly into consideration by the arrival of the decision making process at that particular decision node. When the process reaches a terminal node of the tree, the hypothesis corresponding to that terminal node is declared to be true as the final team decision.

According to the suboptimal decision scheme, all of the consulting DMs are not taken into consideration in every decision. Therefore, this scheme ignores some of the available information, and hence sacrifices some performance, but in the same time offers two main advantages. First, as was already mentioned, the computational complexity is considerably reduced; in fact by choosing the decision tree appropriately, the on-line complexity can be reduced to \( O(\log_2 M) \) and thus the decision process is sped up. Second, each of the consulting DMs improves his performance by adjusting his binary message to the particular binary detection problem; that is, each consulting DM concentrates on a particular binary problem and gives a better 'expert' (more specific) opinion for that problem.

The decision scheme can be characterized as quasi-sequential. It is not sequential in the traditional sense which wants DMs to have the options to stop and decide in favor of some hypothesis, or to receive more information. But, it is sequential in the sense that the final team decision is reached with a sequence of preliminary decisions. At every step the primary DM receives a controlled new observation (decision) from the appropriately selected consulting DM, but does not receive a new (personal) observation; moreover, at every sequential step the primary DM is faced with a different binary hypothesis testing problem.

The suboptimal decision scheme has been presented and the optimal decision rules for the members of the team have been derived. Then three important questions, which deal with the optimal team architecture and configuration, naturally arise from the above discussion:
(1). *Which of the DMs of the team should be designated as the primary DM?*

(2). *How can the optimal binary decision tree be constructed?*

(3). *Which consulting DM should be assigned to each decision node of the tree?*

The following three sections attempt to address the above questions.

### 7.3.5. Selecting the Primary DM

As was seen in section 4.3, the optimal configuration for the two DM tandem team which performs binary hypothesis testing, even in the special case where one DM is better than the other, depends on parameters external to the team, namely the prior probabilities and the misclassification costs. Therefore, it should be evident that no 'globally optimal' primary DM exists for all possible prior probabilities and costs. Consequently, an intuitive and logical suboptimal solution is to employed. For this consider the DM with the minimum individual normalized probability of error (or an approximation to it, if the calculation of actual error probability is too demanding computationally), for the given priors and costs, and designate him to be the primary DM. This choice for the primary DM is 'robust' in the sense that it leads to good performance, as compared to the optimal, even for a bad choice of the decision tree and consulting DMs.

### 7.3.6. Constructing the Binary Decision Tree

There exists an exponential number of binary decision trees and there are several heuristic ways of selecting one for the problem. We will use a corrected variant of the algorithm presented in [D89].

Define the following set which consists of the \( \frac{M(M - 1)}{2} \) possible permutations of pairs of hypotheses (pairs with identical elements are not included):
\[ H_2 = \{(H_m, H_q) \mid m = 0, 1, ..., M - 2, \text{ and } q = m + 1, m + 2, ..., M - 1\} \quad (7.13) \]

Consider the DM who has been selected to be the primary DM; the observation \( y_1 \) of the primary is distributed with density \( P(y_1 \mid H_m) \) under \( H_m \) \((m = 0, 1, ..., M - 1)\). Then assign to each pair \((H_m, H_q)\) the distance \( d(H_m, H_q, \pi_m, \pi_q) \), where \( d(\cdot) \) is some selected stochastic distance measure (for example, the variational distance or the \( J \)-divergence; see [KP90]) that is weighted by the \( \pi_m \)'s, the normalized prior probabilities of \( H_m \) defined in Eq.(7.5). Note that a small stochastic distance measure indicates similarity (i.e., closeness) between two hypotheses; hence, a small distance measure implies a difficult decision (i.e., a decision with large error probability). Suppose that \( M > 2 \); the decision tree is created as follows:

**STEP k** \((k = 1, 2, ..., \lfloor M(M-3)+4 \rfloor / 2)\). Consider the pair of hypotheses with the \( k \)-th smallest distance. Then, there exist three possible cases:

**CASE I.** If neither hypothesis of the pair belongs to the feasible set of a previously defined decision node, then define two terminal nodes at the lowest level of the tree and assign one of the two hypotheses to each. Connect the two terminal nodes by arcs to a new decision node at the second level of the tree; direct the arcs towards the terminal nodes. By definition, the feasible set of the newly generated decision node consists of the union of the feasible sets of the two terminal nodes; that is, it consists of the two hypotheses of the pair with the \( k \)-th smallest distance.

**CASE II.** If only one hypothesis \( H^* \) of the pair belongs to the feasible set of a previously defined decision node, denote by \( r^* \) the node whose feasible set \( F(r^*) \) contains \( H^* \) and has the maximum cardinality. Then, define a new terminal node at the same level as \( r^* \) and

\[ \text{[Footnote]}^3 \text{ If } M=2 \text{ the construction of the tree is obviously trivial.} \]

\[ \text{[Footnote]}^4 \text{ Note that the algorithm described in [D89] is not correct; it needs some correction so as to avoid a deadlock.} \]
assign $H^*$ to it. Connect the newly defined terminal node and $t^*$ by arcs to a new decision node at the next higher level of the tree; direct the arcs towards the new terminal node and towards $t^*$. By definition, the feasible set of the new decision node is $F(t^*) \cup \{H^*\}$; if this set has cardinality $M$ then stop.

**CASE III.** If both hypotheses of the pair belong to the feasible set of some previously constructed decision node(s), then there exist two possibilities. If both hypotheses of the pair belong to the *same* feasible set of some previously defined decision node, then go to the next step $k+1$. Otherwise, if both hypotheses of the pair do not belong together to any feasible set of a previously defined node, denote by $t_0$ ($t_1$) the node of the feasible set with the maximum cardinality that contains the first (second) hypothesis of the pair; connect these two nodes with an arc to a new decision node $t$ at a level that is one higher than the maximum level of $t_0$ and $t_1$; direct the arcs towards $t_0$ and towards $t_1$. Again, the feasible set of $t$ consists of the union of the feasible sets of $t_0$ and $t_1$. If the feasible set has cardinality $M$, then stop.

The last node generated by the above algorithm is the source node. If a new decision node is defined at the $k$-th step of the algorithm, then the feasible set of the newly generated decision node contains the two hypotheses having the $k$-th smallest distance. The decision tree is generated in such a way so as to postpone the most difficult decisions (i.e., decisions between pairs of hypotheses with small distances). We correctly assume that a 'close' error will be less costly, since in the determination of the distance between the hypotheses not only the probability density functions, but also the prior probabilities and the costs were considered (through the normalized prior probabilities).

**Example 7.1.** A seven hypothesis example is presented to demonstrate the construction of a binary decision tree. Before considering the symmetric distance matrix of Table 7.1, note
Table 7.1. Distance Matrix for Example 7.1.

<table>
<thead>
<tr>
<th></th>
<th>$H_0$</th>
<th>$H_2$</th>
<th>$H_3$</th>
<th>$H_4$</th>
<th>$H_5$</th>
<th>$H_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>-</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$H_1$</td>
<td>1</td>
<td>-</td>
<td>10</td>
<td>4</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$H_2$</td>
<td>8</td>
<td>10</td>
<td>-</td>
<td>2</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>$H_3$</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>-</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$H_4$</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>$H_5$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>$H_6$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 7.2. The Binary Decision Tree of Example 7.1.

that it does not have to necessarily be symmetric because the distance measure may not be symmetric (for example, $I$-divergence).

The algorithm proceeds as follows to generate the decision tree of Figure 7.2:

**STEP 1.** $d(H_0, H_1) = 1$; $D_1$ is created; $F(D_1) = \{H_0, H_1\}$.

**STEP 2.** $d(H_2, H_3) = 2$; $D_2$ is created; $F(D_2) = \{H_2, H_3\}$.

**STEP 3.** $d(H_5, H_6) = 3$; $D_3$ is created; $F(D_3) = \{H_5, H_6\}$.
STEP 4. \(d(H_1, H_3) = 4; D_4\) is created; \(F(D_4) = \{H_0, H_1, H_2, H_3\}\).

STEP 5. \(d(H_0, H_3) = 5\); go to Step 6.

STEP 6. \(d(H_4, H_5) = 6; D_5\) is created; \(F(D_5) = \{H_4, H_5, H_6\}\).

STEP 7. \(d(H_4, H_6) = 7\); go to Step 8.

STEP 8. \(d(H_0, H_2) = 8\); go to Step 9.

STEP 9. \(d(H_2, H_4) = 9\); \(S\) is created; \(F(S) = \{H_0, H_1, H_2, H_3, H_4, H_5, H_6\}\): STOP.

In Figure 7.2, the terminal nodes are white, the decision nodes are grey and the source node is black. As expected six (6) terminal nodes generate five (5 = 6–1) decision nodes. The interested reader should note that the algorithm of [D89] would break down because at Step 4 node \(D_4\) would not be constructed and the algorithm would not recover.

7.3.7. Assigning the Consulting DMs

In the previous two sections, an algorithm for selecting a primary DM and, given a primary DM, an algorithm for reducing a multiple hypothesis testing problem into a sequence of binary hypothesis testing problems have been presented. Given a primary DM and a binary decision tree, it remains to be seen how should the consulting DMs be assigned to the decision nodes of the tree. The concept of the expert DM which was discussed in Chapter 2 comes now into consideration.

The assignment problem is formulated as a *maximum weight bipartite matching problem*. The two sets of nodes of the bipartite matching problem consist of the \(M-1\) consulting DMs and of the \(M-1\) non-terminal nodes of the tree. Each consulting DM \(n\) is connected to every decision node \(i\) with an arc which has an associated weight \(w_{n,i} \geq 0\); thus the maximum weight matching will also be a *maximum* matching (i.e., all the nodes

---

5 For an in depth treatment of combinatorial optimization the reader is referred to [PS82].
will be matched), since each consulting DM will be matched with a non-terminal node and vice-versa. How should the weights $w_{n,t}$ be selected so that the solution to this matching problem yield a good solution of the original $M$-ary hypothesis testing problem?

Consider some non-terminal node $t$ of the decision tree. There are three factors that need to be considered for a successful assignment of a consulting DM to the binary problems of node $t$. The first factor is the conditional normalized probability of error that the consulting DM incurs when he solves the problem represented by $t$, conditional on the true hypothesis belonging to the feasible set of $t$. It is conditional because by the definition of the suboptimal decision scheme, the consulting DM at decision node $t$ makes his decision conditional on the true hypothesis being on the feasible set of $t$ (Eq.(7.12)). Furthermore, if an error has already occurred in the decision process, the message of the consulting DM, who is supposed to address the wrong problem, is not going to significantly alter the performance of the team (recall that as the decision process moves down the decision tree, the hypotheses become 'more similar' both in probability and in cost). Clearly, the smaller the normalized probability of error of a consulting DM the more suitable the DM is for that binary problem.

The second factor that needs to be considered is the individual performance of the primary DM at each decision node $t$ (i.e., the primary DM's normalized probability of error conditional on one of the hypotheses in the feasible set of $t$ being true). It is obvious that the better the performance of the primary DM, the less the need for a consulting DM and vice-versa. Denote by $P_{s,t}^e$, the normalized probability of error (or its approximation) of the primary DM 1 at the decision node $t$ (i.e., $H_0$ vs. $H_q$), conditional on the true hypothesis being on the feasible set of $t$.

The third factor is the prior probabilities and the detection costs. Consider some DM $n$ who can be associated with either of two decision nodes; decision node $t$ or decision node $s$. Furthermore, assume that DM $n$ has the exact same performance (i.e., normalized probability of error) at both decision nodes $t$ and $s$, conditional on the true hypothesis being
in the feasible set of \( t \) and \( s \) respectively. Other things being equal, on which of the two
decision nodes should DM \( n \) be assigned?

DM \( n \) should be associated with the decision node whose feasible set has the larger
sum of elementary normalized prior probabilities. This is easier to comprehend in the case
of the minimum error cost function; in this case, DM \( n \) should be associated with the
decision node whose feasible set has the larger sum of prior probabilities. This should be
intuitive as, by being associated with the decision node whose feasible set has the larger
sum of prior probabilities, DM \( n \)'s decision will be taken into account by the team more
often.

Thus, the weight \( w_{n,t} \) of assigning the consulting DM \( n \) to the binary hypothesis
problem of a non-terminal node \( t \) of the decision tree is defined as follows:

\[
w_{n,t} = (1 - P^d_{n,t}) P^e_{1,t} \left[ -\log \left( 1 - \sum_{m \in F(t)} \pi_m \right) \right]
\]

(7.14)

**REMARK 1.** Clearly: \( w_{n,t} \geq 0 \).

**REMARK 2.** The logarithm in Eq.(7.14) is included to emphasize the assignment of the
better DMs to the ('important') non-terminal nodes which are high on the tree (because
these nodes have large normalized prior probability \( \pi_t \)). In fact, note that since for the
source node \( \pi_s = 1 \), the consulting DM with the smallest normalized probability of error for
the source node's binary problem (i.e., \( H_{s_0} \) vs. \( H_{s_1} \)) will always be assigned there.

**REMARK 3.** There exist several algorithms to solve the maximum weight bipartite
matching problem which require \( O(M^3) \) time [PS82].

**REMARK 4.** As has already been mentioned, the (restrictive) implicit assumption in the
definitions of the normalized prior probabilities for both elementary and composite
hypotheses (Eqs.(7.4) and (7.7)) is that, given the true hypothesis, the relative frequencies
of the different types of error are given by the prior probabilities. In the decision scheme
that was presented, the $M$-ary hypothesis testing problem was broken down into a sequence of composite binary hypothesis testing problems; the composite hypotheses consist of similar hypotheses. Thus even though the assumption is not true in general, it is quite reasonable for the proposed decision scheme especially if the probability density functions are smooth.

**Remark 5.** To compute the decision thresholds a Gauss-Seidel algorithm similar to the one described in [TP90] should be employed. As stated earlier, the algorithm is guaranteed to converge to decision rules which satisfy the necessary optimality conditions, but there are no guarantees that these will actually be the optimal decision rules. Still, in the binary hypothesis testing case only one likelihood ratio exists; therefore, we can perform one dimensional search and guarantee that the algorithm converges to the optimal decision rules. This is an additional computational advantage of the suboptimal decision scheme.

### 7.3.8. Improving on the Suboptimal Decision Scheme

The suboptimal decision scheme described above employs common sense and intuition, together with some straightforward mathematics, to yield a hopefully acceptable solution for the very complicated problem 7.1. Because of the several subjective and arbitrary choices that are made (selection of the primary DM, construction of the binary decision tree and the matching problem), it is virtually impossible to obtain analytically any meaningful bounds for the performance of the decision scheme; this could be an interesting topic for future research.

We do not claim to have devised the 'best' suboptimal decision scheme. On the contrary there exist several ways to improve it, but our objective was just to demonstrate that relatively simple procedures may be designed for successfully tackling complicated problems. Indeed, the performance of the decision scheme can be improved, without
forbidding increases in the computational complexity being induced if, for example, one or more of the following factors are taken into consideration:

1. A better algorithm for the generation of the binary decision tree should be derived. This should take into account its particular application in our problem. For example, in the proposed algorithm the pair of hypotheses with the $k$th smallest distance is considered at each step $k$. Once two elementary hypotheses are to be combined into a composite hypothesis, the probability distribution of the composite hypothesis consists of an appropriate convolution of the distributions of the elementary hypotheses. Therefore, it should be preferable at every step of the algorithm to calculate a new distance matrix by taking explicitly into consideration the newly generated composite hypothesis, instead of always using the initial distance matrix (i.e., the distance matrix between the elementary hypotheses).

2. A more sophisticated selection of the weights for the matching problem.

3. Relevant results from information theory should be employed in this framework.

4. For the decision of a particular consulting DM to be considered by the primary DM, a particular sequence of decisions must be made first. For example, for the decision of the consulting DM corresponding to decision node $D_2$ ($H_2$ vs. $H_3$) to be considered, the team must decide 0 at the source node $S$ and decide 1 at decision node $D_4$. This provides some additional information to the consulting DM corresponding to $D_2$; he can therefore update his beliefs of the distribution of the observation of the primary DM, in order to produce an even better decision.

Furthermore, the most important advice for improving the performance of the suboptimal decision scheme is to try to take maximum advantage of the particular structure of the problem. We should carefully examine the characteristics of the problem (i.e., DMs,
hypotheses, costs, etc.) and try to assist the 'mathematical solutions' by making educated choices, which the (simple, suboptimal) analysis may overlook. Keeping this in mind, we present a specific example in the following section.

7.4. A MULTIATTRIBUTE HYPOTHESIS TESTING PROBLEM

Consider a particular instance of the problem 7.1, in which each of the multiple hypotheses consists of a set of independent or loosely dependent attributes. For example, suppose that the multiple hypotheses describe the characteristics of a refrigerator; the size, the color, the weight, the temperature and the motor of the refrigerator are almost independent attributes (the size of a refrigerator contains almost no information about its color and temperature), but knowledge of all of them can provide considerable information on determining the its exact model and make.

Moreover, suppose that the observation of each of the consulting DMs contain information on just one of the attributes. In the refrigerator example, one consulting DM has a tape measure and can measure its size, another has a chart and can determine its exact color, another has a scale and can weight it, another has a thermometer and another has a voltmeter and an ampmeter. On the other hand, the observation of the primary DM contains some information about all the attributes. The objective is to design a decision protocol that will optimize the team performance.

The problem just described fits the framework of the more general problem 7.1 since the team performs multiple hypothesis testing and consists of a single primary DM, and several consulting DMs. We are going to argue that the additional structure imposed on the original problem leads to very good and considerably simpler solutions.

The observation of only one of the DMs contains information on all the attributes of the hypotheses; this DM presents us with a clearcut choice for the primary DM of the team. Furthermore, since each of the consulting DMs is an 'expert' in just one of the attributes of
the hypotheses, he is immediately 'matched' with a particular subproblem, and can use his messages to transmit to the primary DM information about just this particular attribute without any degradation in the team performance. Therefore, the optimal configuration of the team is determined by inspection. Finally since the various attributes of the hypotheses are independent or loosely dependent, the primary DM and the appropriate consulting DM can indeed perform as a two DM tandem team because the conditional independence assumption will not be violated.

Hence, the only remaining issue that needs to be resolved involves the sequence with which the decisions are made; that is, it involves the construction of the decision tree (Figure 7.3). The decision tree will be a little different than the one described in the previous section. First, it does not have to be binary; for example, if the refrigerator can have one of three different colors, it should be worthwhile to endow the 'color expert' with three messages, one for each color.

Second, since the attributes are independent with each other, the depth of the decision tree will be constant (since knowledge of one attribute does not contain any information about another, all the attribute combinations are possible).

Third, note that in this type of problem every consulting DM is associated with an attribute of the multiple hypothesis, and is not associated with a composite hypothesis testing problem. Consequently, each consulting DM will be associated with all the decision nodes at a given level of the decision tree and will not be associated with just one decision node (for the decision tree of Figure 7.3, the consulting DM for the first attribute is associated with node A, the consulting DM for the second (third) attribute is associated with all the B (C) nodes). To see this, consider again the refrigerator example and suppose that its attributes are examined in the given order: size, color, weight, etc. In the third stage of the game when the weight of the refrigerator is examined (the C nodes in Figure 7.3), independent of the team decisions up to that point (i.e., the decisions on size and color), the two DM tandem team consisting of the primary DM and the 'weight expert' consulting DM
will always have to decide on the weight of the refrigerator. Thus, the 'weight expert' is associated with all the decisions that take place in the third level (from above) of the decision tree.

Therefore, to construct the decision tree only a testing sequence for the attributes of the multiple hypotheses has to be generated. To do this, we employ some stochastic distance measure and arrange the sequence of the decisions in such a way, so that the more difficult decisions are postponed for later (i.e., the lower levels of the tree).

7.5. SUMMARY

As the number of the DMs, or of the hypotheses, or of the messages increases, the combinatorial complexity of problems in this framework increases exponentially. In order to keep the complexity under control suboptimal decision schemes need to be derived. Since we also desire that the suboptimal decision schemes achieve good performance, the particular characteristics of the problem should be taken into account. This leads to problem dependent analysis which, we believe, should be the focus of future research. To
demonstrate this approach, we discussed the issues involved with the reduction of a multiple hypothesis testing problem into a sequence of simpler hypothesis testing problems, under different operating conditions.
CHAPTER 8

Concluding Remarks

8.1. SUMMARY AND CONCLUSIONS

In this section we present a summary of our results and conclusions. We not only try to summarize the results, but also try to demonstrate in a coherent manner that the various chapters of this thesis constitute integral parts of the same global problem. This global problem is the normative design of decision making organizations and the understanding of the decision making process of both human and non-human decision making agents.

Since modeling decision making is an interdisciplinary research field, there exist several widely different approaches for it. The need for the development of a normative decision theory emanates from these approaches; this theory should generate models and hypotheses which will be tested in practice in order to generate better and more accurate normative/descriptive models. It should be viewed as a tool to assist decision makers in improving the quality of their decisions.

The decentralized hypothesis testing model was selected as the framework for our research. Decentralized, because the decisions in any organization usually occur in a decentralized manner. The hypothesis testing framework, because it provides for some very interesting paradigms in an environment which is easy to describe, so that scientists from diverse disciplines can appreciate and incorporate the results into their own research. Furthermore, it provides for some very convenient modeling tools (for example, the Bayesian cost function for the team objective and the discrete messages for the communications) which enable us to capture the various important aspects of organizational
decision making without having to address the usual stumbling blocks (for example, the utility function, bounded rationality, etc.).

Problems in the decentralized hypothesis testing framework have been shown to be very hard computationally (NP-hard). On the other hand, these problems have centralized counterparts with trivial solutions which imply that the difficulties arise because of the decentralization. We initially concentrated into 'small' problems, that is problems of teams which consist of a small number of decision makers endowed with a limited number of alternative decisions, in order to isolate the effects of the decentralization from the effects of the combinatorial complexity.

Starting with the simplest possible teams which consist of two decision makers, we determined that the tandem architecture is better than the parallel. We disproved the conjecture which stated that in the tandem architecture the better decision maker should be made the primary decision maker (i.e., make the final team decision) and we further analyzed the parallel architecture. Even if both decision makers are identical, they do not have to employ identical or even symmetric decision rules; tight bounds were obtained on the deterioration of the team performance if both decision makers are restricted to employing identical decision rules. This constituted our first attempt to establish the trade off between the complexity of the decision rules (i.e., the suboptimal identical decision rules versus the truly optimal decision rules).

Also, we showed that the optimal architecture and configuration for teams consisting of three decision makers depend on the particular decision makers of the team, on the communication protocols and on parameters external to the team like the prior probabilities and the misclassification costs. Thus, we disproved a second conjecture which claimed that the two consultant architecture must be better than the architecture of three decision makers in tandem.

Therefore even in our restricted hypothesis testing framework, generalizations can not be made and results which are robust with respect to the particular characteristics of the
team or of the environment can not be obtained. The inherent complexity that these problems exhibit gives rise to several counterintuitive results. Hence, one of the things we learned, and learned very well indeed, is to never accept results to be true, however 'obvious' and 'intuitive' they may seem, until a formal proof has been established. Thus we realized that organizational problems should always be approached with a very cautious and critical eye.

Since there do not exist generalizations which can be established from small teams, we then decided to focus our attention on very large teams, namely the infinite tandem team. We obtained necessary and sufficient conditions on the individual decision maker for the team probability of error to go to zero. We developed an asymptotically optimal 'selfish' decision scheme for the same team and again obtained necessary and sufficient conditions. We performed numerical studies to establish the trade off between the team performance and the complexity of the decision rules and concluded that the suboptimal scheme performs reasonably well. Furthermore, the inefficiencies of the tandem architecture in terms of performance versus number of decision makers were exposed; this, together with the fact that the performance of the tandem architecture is very susceptible to decision maker disabilities and communication link failures, suggests that the tandem architecture is to be preferred only for very small organizations.

We then described the different types of decision rules which DMs employ and discussed their implications on the performance of the team. It was shown that in order for teams of DMs to maintain nice properties of individual DMs (and thus allow our results to be generalized for even larger teams), additional assumptions on the team coordination and/or training need to be introduced.

As was mentioned in chapter 2, Simon claims that because of the complexity of the environment, scientists striving to model decision making have but two alternatives. The first is to build optimal models by making simplifying assumptions. We undertook this approach; we used perfectly cooperating decision makers, each receiving a conditionally
independent observation, the environment consists of a discrete number of hypotheses, a Bayesian cost function as the team objective, no inconsistencies in the prior beliefs, no indecisiveness, no biases, no individual objectives, no conflicts or regrets, no deadlines, very limited and instantaneous communications. Even in this ideal and very restricted environment, the models exhibited great inherent complexity, presented counterintuitive results and did not offer any generalizations. Since real life decision making situations are certainly more complex, we have to focus on Simon's second alternative which is to build heuristic models that maintain greater environmental realism.

To demonstrate this approach, in the last part of this research a multiple hypothesis testing problem was formulated. The optimal decision rules for the problem are computationally very demanding. For this reason, a suboptimal decision scheme was discussed which attempts to reduce the complexity and, in the same time, take advantage of the particular characteristics of the decision makers and of the hypotheses, in order to minimize the deterioration of the team performance from the optimal.

The simplicity of the distributed hypothesis testing model and the fact that good solutions for the problems can be readily obtained suggest that the "dirty word optimality" is the cause of the severe difficulties. Therefore, it may be advisable to forgo optimality in favor of a simple suboptimal solutions which perform 'well enough.' But, this is exactly what human decision makers do according to the theory of satisficing of Simon. Humans do not strive for optimality; rather, they try to make a decision which will result to a satisficing outcome.

We also arrived to an important conclusion which seems to apparently be in complete disagreement with Simon's beliefs about the human decision maker. Simon in [S81, p. 65] states that: "A man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself." But, for the very simple environment of chapter 4 (binary hypothesis testing, two or three perfectly cooperating decision makers receiving
conditionally independent observations, Bayesian cost objective function, binary or ternary communication messages) counterintuitive results were obtained and the inherent complexity of the decision rules were demonstrated. Therefore unless we claim that either the environment described above is not simple enough, or that the decision rules (given in the form of likelihood ratio tests with constant thresholds) are simple enough, or that man, viewed as a behaving system, does not strive to optimize, our results are in contradiction with Simon's claim. But, we also have to keep in mind that the disagreement may be a consequence to Simon talking about humans, while we talk about mathematical models.

In summary, we tried to demonstrate the difficulties of modeling decision making, even under very restrictive assumptions. In order to overcome the difficulties, we should develop heuristics which should take into account the particular characteristics of the organization and of the environment in which it operates.

8.2. Suggestions for Future Research and Concluding Remarks

We distinguish future research directions into two categories. The first category includes some immediate research goals that follow directly from the results in this thesis. The second category includes some longer range research goals which require a complete rethinking of the approach employed.

Beginning with the first category, it would be interesting to obtain an absolute bound, in terms of the normalized probability of error, on the deterioration of the two decision maker tandem team performance when the better decision maker is designated as the primary decision maker (section 4.3). It is not hard to obtain a bound of one ninth (1/9), but it does not seem close to being a tight bound. Using our usual methods, we were able to reduce the problem to a few hundred cases; still, due to the complexity of the problem, we were not able to obtain meaningful bounds for many of these.
Numerical studies need to be performed in order to be able to explicitly compare the optimal team performance of the multiple hypothesis testing problem (chapter 7) with the performance of the suboptimal decision scheme. Moreover, ways to improve the suboptimal decision scheme have to be developed in the context of problem dependent methods of analysis.

In the long run, we believe that further research should be devoted to modeling decentralized decision making and decision making in general. We need to spend some time to rethink the problems and the models. All kinds of decision makers are continuously faced with real time decisions. Good decisions are being made even by unsophisticated agents without resorting to a series of complicated analyses. Therefore, the models, or at least part of the models, need to be modified, so that simple protocols can be formed for a decision system to make intelligent decisions. The decision making process may not be an optimization after all. It will not be the first time that the wrong mathematical tools have been employed for modeling the decision making process. Recall that in the fifties and early sixties, engineers influenced by the advances in information theory tried to model the decision maker as a channel; after several elegant mathematical results were obtained, it became evident that a simple channel could not provide for a rich enough framework for modeling decision making and the information theory approach was abandoned.

Similarly, the utility theory developed by von Neumann and Morgenstern in their second edition of Theory of Games and Economic Behavior, published in 1947, under certain rationality properties, associates each and every decision maker with his individual utility function. Then, every decision that is made is viewed as the solution of the optimization problem of maximizing the utility function; the shape of this artificial function is assumed to be such that all the decisions are justified (again under certain rationality properties). For example, consider a person who is commuting to and from his work every day by driving through the (fixed) shortest distance path. Suppose, and isn't this usually the case, that he does not adjust his path with respect to the traffic; that is, suppose that he
does not try to determine daily the (variable) shortest time path. Although his behavior is
suboptimal, because nobody enjoys spending time commuting, utility theory tries to
justify his decisions by assuming that the driver maximizes his utility by not having to think
and driving for a longer time, instead of having to think and drive for a shorter time. Thus,
by associating to the driver an appropriate utility function, his suboptimal decisions become
optimal.

Note that while the decisions made are considered by the theory as the solutions to the
problem of utility maximization, first the decisions are actually made and then a utility
function that renders the decisions optimal is constructed. Moreover, although the utility
type theory provided for great innovative thinking and for a useful mathematical quantization
tool, it is so artificial, abstract and dependent on 'well behaved' utility functions for the
decision makers that it can not be used for realistic model of decision making with any kind
of practical applications, especially since human decision makers do not have the wits to
maximize.

Subsequently, the theory of satisficing tried to somehow separate the decision making
process from the optimization process by modeling the decision making process as the
search for a feasible solution to a more constrained optimization problem. Although this
type is very appealing because it sounds natural and closer to the actual behavior of
decision makers, it can not still be quantified in any rigorous form. By constructing
appropriate satisficing conditions for each individual decision maker, we are revisiting in a
way the utility functions. On the other hand, the theory of satisficing is a powerful theory
which will continue to interest decision scientists because of its simplicity and solid grasp
of (sometimes sophisticated) human behavior. But we also feel that the decision making
process may have to be further separated from an optimization process; maybe, as several
of the results of this research indicated, the actual optimization is the source of the modeling
difficulties.
We definitely believe that a normative decision theory needs to be developed. But, it seems evident that mathematics alone can not give the solution to real life decision making problems. Mathematics can be used to test, confirm or disprove aspects of decision making theories. But, we have to keep in mind that for the time being a strict mathematical environment is too sterile for capturing even the simplest functions which take place in the human brain and result to good decision making; otherwise, we may drown in a drop of water by spending too much time and effort to formulate 'convenient' assumptions which yield desired results. For example, recall the team which consists of two identical decision makers in parallel (section 4.4); if we had assumed that the two decision makers employ identical decision rules, we could have obtained the then optimal (simple) decision rules and proceeded to draw conclusions which would not be true in general, since they would have been based on the added assumption.

Finally, we would like to state clearly and unequivocally that we do not advocate in any way the abandonment of the effort to mathematically model the decision making process; we simply state that in order to make best use of the models, we need to keep in mind that mathematics can not do it alone. Other decision theories, less mathematically oriented, need to be developed in parallel. The interaction between these theories will result in better understanding and improved models. Modeling decision making is an interdisciplinary research effort and as long as decision makers of different disciplines keep making decisions it will remain one.
REFERENCES


