MATHEMATICAL FOUNDATIONS OF RISK MEASUREMENT

by

Louis Anthony Cox, Jr.

A.B., Harvard University
(1978)

S.M., Massachusetts Institute of Technology
(1985)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE
DEGREE OF
DOCTOR OF PHILOSOPHY
IN RISK ANALYSIS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 1986

© Louis Anthony Cox, Jr. 1986

The author hereby grants to M.I.T. 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

Certified by

Alvin W. Drake
Thesis Supervisor

Accepted by

Arthur C. Smith
Chairman, Departmental Committee on Graduate Students
MATHEMATICAL FOUNDATIONS OF RISK MEASUREMENT

by

LOUIS ANTHONY COX, JR.

Submitted to the Department of Electrical Engineering
and Computer Science on May 16, 1986 in partial
fulfillment of the requirements for the Degree of
Doctor of Philosophy in Risk Analysis

ABSTRACT

Economic activities that produce deterministic flows of economic costs
and benefits often also produce flows of human health and safety risks
that may affect both occupational populations and the general public.
Consumers, neighbors, and employees may be exposed to health and
safety threats from the operation of production activities that are
nonetheless considered by policy makers to be socially worthwhile. In
deciding how to manage and whether or how to restrict such activities,
it would be helpful to have unambiguous quantitative expressions for
the risks they produce.

This study investigates the theoretical problems of defining and
representing mathematically the (possibly uncertain) health and safety
risks of activities. We define the problem of mathematical risk
measurement as that of assigning numbers to activities in such a way
that (i) higher numbers are assigned to riskier activities; and (ii)
all individuals know how to interpret the resulting numbers in terms
of "equivalent" standardized risks, so that unambiguous communication
about risks by means of such numbers is possible. We also examine
stronger requirements, such as that the numbers be assigned in such a
way that twice as great a number corresponds to twice as great a risk,
and consider conditions under which such requirements are meaningful.
Finally, we consider various non-numerical representations of risk.

Our primary focus is on clarifying the mathematical scope and
boundaries of the set of risk management decision problems, as opposed
to other decision problems that have been addressed in the decision
analysis literature. In addition, we seek to provide a conceptual
framework in which useful normative decision theories for these
problems can be developed. After critically reviewing existing
decision theories that represent a decision maker's potential acts by
probability measures over a static set of (possibly dated) physical
outcomes, we propose that acts should instead be represented by
stochastic processes over physical outcomes. This reflects the fact
that the outcomes of risk management decisions are often not learned
until long after the decisions have been made, and are sometimes not
learned at all.
Simple stochastic process models for both individual and population fatality risks are identified and novel decision problems that they raise that do not appear in traditional decision models are discussed. Our principal contribution is the development of risk scales (called "canonical" risk scales) for these simple models, on which both certain and uncertain risks can be represented. Risks are measured on a canonical risk scale by comparison with simple standardized risks, rather than by the more traditional approach of combining "frequency" and "severity" components. This comparative approach sheds light on some conceptual difficulties with frequency-severity definitions of risk and in several special cases leads to useful numerical bounds on risks with known "frequencies" and known qualitative properties.

An abstract axiomatic approach to risk measurement is taken throughout this work. Empirical problems of data collection and statistical inference are identified but not treated. By focusing on the theoretical foundations of risk definition, representation, and communication, we seek to provide a foundation on which sound normative theories of risk management decision making can be built.

Thesis Supervisor: Dr. Alvin W. Drake

Title: Professor of Electrical Engineering
ACKNOWLEDGEMENTS

It is a pleasure to acknowledge the help and support of several colleagues and friends who have contributed to this work. My greatest thanks are due to Professor Alvin Drake, who persuaded me to take the mathematical foundations of risk measurement as the topic of my doctoral research and who supervised my research with creativity, patience, and insight from the start. I am deeply grateful to him and to the other members of my Committee -- Professors Arnold Barnett, Gordon Kaufman, and Norman Rasmussen -- for constructive reviews and discussions that have helped me to clarify my thinking and that have in many cases improved the final statements of results.

The Operations Research Section at Arthur D. Little, Inc. provided an exceptionally stimulating applied research setting for thinking about theoretical problems of risk measurement and risk management. Much of the motivation for this study and for the specific topics examined in Chapters 7 and 8 has grown out of case work for clients such as the Health Effects Institute, the American Nuclear Insurers, and the U.S. Air Force, who have confronted various practical risk analysis problems and who have come to Arthur D. Little, Inc. for help in thinking systematically about them. I am grateful to Joel Jensen and to David Boodman of Arthur D. Little for encouraging and supporting my academic research and professional involvement in risk analysis and decision sciences as part of my professional practice area.

Finally, I would like to thank John Hershey and Howard Kunreuther of the Wharton School for stimulating conversations that caused me to think harder about some of the questions in Chapter 8 than I otherwise would have.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. RISK MEASUREMENT AND RISK MANAGEMENT</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The Classical Decision Analysis Paradigm</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Risk Management Decisions: Examples and Characteristics</td>
<td>1</td>
</tr>
<tr>
<td>1.3 Roles for the Risk Analyst</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Scope and Limitations</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Overview</td>
<td>8</td>
</tr>
<tr>
<td>2. TECHNICAL BACKGROUND: INDIVIDUAL KNOWLEDGE AND BELIEFS</td>
<td>12</td>
</tr>
<tr>
<td>2.0 Introduction and Overview</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Individual Knowledge</td>
<td>17</td>
</tr>
<tr>
<td>2.1.1 A Language for Expressing Propositions</td>
<td>17</td>
</tr>
<tr>
<td>2.1.2 An Axiomatic Approach to Individual Knowledge</td>
<td>19</td>
</tr>
<tr>
<td>2.1.3 Private Information and Deterministic Channels</td>
<td>20</td>
</tr>
<tr>
<td>2.1.4 Comparison of Deterministic Information Channels: Terminology and Background</td>
<td>21</td>
</tr>
<tr>
<td>2.1.5 Syntactic Characterization of Individual Knowledge</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Individual Beliefs</td>
<td>25</td>
</tr>
<tr>
<td>2.2.1 Finitely Additive Probability Measures as Representations of Individual Beliefs</td>
<td>26</td>
</tr>
<tr>
<td>2.2.2 Probability Measures with Respect to a Coarse Algebra</td>
<td>31</td>
</tr>
<tr>
<td>2.2.3 Formation of Beliefs</td>
<td>35</td>
</tr>
<tr>
<td>A. The Data-Generating Model</td>
<td>35</td>
</tr>
<tr>
<td>B. Coherence, Calibration, and Sufficient Statistics</td>
<td>40</td>
</tr>
<tr>
<td>2.2.4 Comparison of Noisy Information Channels</td>
<td>45</td>
</tr>
<tr>
<td>2.3 Summary and Prospectus</td>
<td>47</td>
</tr>
</tbody>
</table>
### 3. CHOOSING AMONG PROBABILITY DISTRIBUTIONS OVER CONSEQUENCES

3.1 Decision Model 1: The Expected Utility Model
(Fishburn, 1970; Roberts, 1979)

**A. The Model**

**B. Critique**

3.2 Decision Model 2: Single-Attribute Decision Theory
(Machina, 1982, 1983)

3.3 An Alternative Axiomatic Basis for Single-Attribute Utility Theory

3.4 Decision Model 3: Multiattribute Consequences
(Keeney and Raiffa, 1976)

3.5 Decision Model 4: Uncertain Consequences

3.6 Summary

### 4. THEORIES OF RISK MEASUREMENT FOR CLASSICAL DECISION MODELS

4.0 Introduction

4.1 Risk as a Relation Induced by R

4.2 Measurement of Risk in a Single-Attribute Decision Model (Fishburn, 1984)

4.3 Stronger Forms of Risk Measurement in the Single-Attribute Case

4.3.1 Introduction: Extensive Measurement

4.3.2 Extensive Measurement of Risk

4.3.3 Critique

### 5. NEW THEORIES OF SINGLE-ATTRIBUTE RISK MEASUREMENT

5.0 Introduction and Overview

5.1 Risk Measurement for Dichotomous Consequences

5.2 Preference Differences and a One-Parameter Family of Comparative Risk Relations
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3 One-Way Communication Between D.M.'s with the Same Deterministic Value Functions for Consequences</td>
<td>111</td>
</tr>
<tr>
<td>5.4 Comparative Riskiness and Second-Order Stochastic Dominance</td>
<td>118</td>
</tr>
<tr>
<td>5.5 Special Cases and Limited Precision Objective Risk Measurement</td>
<td>119</td>
</tr>
<tr>
<td>5.5.1 Objective Numerical Risk Measurement for Special Classes of C.d.f.'s</td>
<td>120</td>
</tr>
<tr>
<td>5.5.2 Limited-Precision Objective Risk Measurement and Canonical Risk Scales</td>
<td>125</td>
</tr>
<tr>
<td>5.5.3 Extension to Risk Models of the Form (p,F)</td>
<td>127</td>
</tr>
<tr>
<td>5.5.4 Practical Implementation: The Canonical Risk Curve</td>
<td>129</td>
</tr>
<tr>
<td>6. COMMUNICATION BETWEEN TWO INFORMED PLAYERS</td>
<td>132</td>
</tr>
<tr>
<td>6.0 Introduction</td>
<td>132</td>
</tr>
<tr>
<td>6.1 Combining Expert Judgements of Probabilities</td>
<td>133</td>
</tr>
<tr>
<td>6.2 Iterative Communication and Common Knowledge</td>
<td>139</td>
</tr>
<tr>
<td>6.3 Communication Without Common Priors</td>
<td>142</td>
</tr>
<tr>
<td>6.3.1 Use of Likelihood Ratios as Sufficient Statistics</td>
<td>142</td>
</tr>
<tr>
<td>6.3.2 Communicating Arrival Time Distributions</td>
<td>144</td>
</tr>
<tr>
<td>7. MEASUREMENT OF INDIVIDUAL FATALITY RISKS</td>
<td>148</td>
</tr>
<tr>
<td>7.0 Introduction and Overview</td>
<td>148</td>
</tr>
<tr>
<td>7.1 Precis</td>
<td>151</td>
</tr>
<tr>
<td>7.2 Three Fatality Risk Models</td>
<td>154</td>
</tr>
<tr>
<td>7.2.1 Individual Fatality Risks</td>
<td>155</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>155</td>
</tr>
<tr>
<td>B. Individual Fatality Risk Model 1</td>
<td>156</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Population Fatality Risk Models</td>
</tr>
<tr>
<td>7.3</td>
<td>Ranking Changes in Lifetime Distributions</td>
</tr>
<tr>
<td>7.3.1</td>
<td>A Special Case: Competing Risks</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Choosing Among Hazard Functions</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Uncertain Risks: Examples</td>
</tr>
<tr>
<td>7.4</td>
<td>Elements of a Preference Theory for Uncertain and Stochastic Hazard Functions</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Motivation</td>
</tr>
<tr>
<td>7.4.2</td>
<td>A Dynamic Expected Utility Theory for Uncertain and Stochastic Hazard Rates</td>
</tr>
<tr>
<td></td>
<td>A. Stochastic Hazard Trajectories</td>
</tr>
<tr>
<td></td>
<td>B. Preferences Over Stochastic Hazard Functions</td>
</tr>
<tr>
<td>7.5</td>
<td>Remaining Challenges</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Nonmonotonic Time Preferences</td>
</tr>
<tr>
<td>7.5.2</td>
<td>Deriving Risk Preferences</td>
</tr>
<tr>
<td>7.5.3</td>
<td>Controlled Processes</td>
</tr>
<tr>
<td>7.5.4</td>
<td>Statistical Estimation Issues and Practical Bounds</td>
</tr>
<tr>
<td>8.0</td>
<td>MEASUREMENT OF POPULATION FATALITY RISKS</td>
</tr>
<tr>
<td>8.1</td>
<td>Routine Population Risks</td>
</tr>
<tr>
<td>8.1.1</td>
<td>Population Renewal in a Model of Size One</td>
</tr>
<tr>
<td>8.1.2</td>
<td>Population Renewal in a Heterogeneous Population</td>
</tr>
<tr>
<td>8.1.3</td>
<td>An Alternative Approach to Defining Population Fatality Risks</td>
</tr>
<tr>
<td></td>
<td>A. Identifying the Set of Relevant Consequences</td>
</tr>
<tr>
<td></td>
<td>B. Developing a Mathematical Representation for Preferences Over Consequences</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>C. Developing a Value Function For Routine Fatalities</td>
<td>220</td>
</tr>
<tr>
<td>D. A Statistical Model for the Value of Routine Population Mortalities</td>
<td>224</td>
</tr>
<tr>
<td>8.1.4 The Case of Homogeneous Responses Within Cells and Equal Value Weights</td>
<td>227</td>
</tr>
<tr>
<td>8.1.5 Testing for Homogeneity</td>
<td>230</td>
</tr>
<tr>
<td>8.2 Catastrophic Population Risks</td>
<td>231</td>
</tr>
<tr>
<td>8.2.1 Risk Curves for the Compound Poisson Process</td>
<td>232</td>
</tr>
<tr>
<td>A. The Classical Risk Curve</td>
<td>232</td>
</tr>
<tr>
<td>B. Critique</td>
<td>233</td>
</tr>
<tr>
<td>8.2.2 Risk Equity</td>
<td>235</td>
</tr>
<tr>
<td>A. Some Previous Approaches</td>
<td>235</td>
</tr>
<tr>
<td>B. Critique</td>
<td>235</td>
</tr>
<tr>
<td>8.2.3 Limited-Precision Risk Measurement for Catastrophic Fatality Risks</td>
<td>241</td>
</tr>
<tr>
<td>A. Risk Equity Scales</td>
<td>241</td>
</tr>
<tr>
<td>B. Evaluation</td>
<td>244</td>
</tr>
<tr>
<td>8.3 Summary and Conclusions</td>
<td>245</td>
</tr>
<tr>
<td>9. SUMMARY AND CONCLUSIONS</td>
<td>247</td>
</tr>
<tr>
<td>9.0 Introduction</td>
<td>247</td>
</tr>
<tr>
<td>9.1 Risk Models and Decision Models</td>
<td>248</td>
</tr>
<tr>
<td>9.1.1 Models of Risk</td>
<td>249</td>
</tr>
<tr>
<td>9.1.2 Decision Theories for Risk Management</td>
<td>252</td>
</tr>
<tr>
<td>9.2 Expression of Risk and Interpretation of Consequences</td>
<td>253</td>
</tr>
<tr>
<td>9.2.1 &quot;Frequency&quot; and Time-Varying Hazard Rates</td>
<td>255</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>9.2.2 Temporal Resolution of Uncertainty and Ex Ante vs. Ex Post Perspectives</td>
<td>256</td>
</tr>
<tr>
<td>9.3 Risk Measurement</td>
<td>258</td>
</tr>
<tr>
<td>References</td>
<td>261</td>
</tr>
</tbody>
</table>
FIGURES

2.1 Qualitative Probability Structure for Five-State Example 29
5.1 Canonical Risk Curve 130

TABLES

8.1 An Aggregation Artifact in Population Risk Calculations 217
## LIST OF EXAMPLES

<table>
<thead>
<tr>
<th>Example</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Cigarette Smoker</td>
<td>1</td>
</tr>
<tr>
<td>2. Regulating Automobile Emissions</td>
<td>3</td>
</tr>
<tr>
<td>3. Marketing a New Drug</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Russell’s Paradox</td>
<td>18</td>
</tr>
<tr>
<td>2.2 Qualitative Probability</td>
<td>27</td>
</tr>
<tr>
<td>2.3 Dose-Response Relations and Filtrations</td>
<td>32</td>
</tr>
<tr>
<td>2.4 Marginal Probabilities and Cell Probabilities</td>
<td>34</td>
</tr>
<tr>
<td>2.5 True vs. Known Consequences</td>
<td>39</td>
</tr>
<tr>
<td>3.1 Incoherent Choice</td>
<td>53</td>
</tr>
<tr>
<td>3.2 Differential Shifts in Public Health Risks</td>
<td>60</td>
</tr>
<tr>
<td>3.3 Installing Safeguards</td>
<td>68</td>
</tr>
<tr>
<td>3.4 Delayed Health Effects</td>
<td>73</td>
</tr>
<tr>
<td>5.1 The Perils of Paul</td>
<td>101</td>
</tr>
<tr>
<td>5.2 Russian Roulette</td>
<td>102</td>
</tr>
<tr>
<td>5.3 Limited-Precision Objective Risk Measurement on the Poisson Risk Scale</td>
<td>126</td>
</tr>
<tr>
<td>6.1 Competing Information about Poisson Risks</td>
<td>134</td>
</tr>
<tr>
<td>6.2 Treating Insomnia</td>
<td>135</td>
</tr>
<tr>
<td>6.3 A Two-Out-of-Three Reliability Problem</td>
<td>138</td>
</tr>
<tr>
<td>6.4 The Perils of Joe</td>
<td>144</td>
</tr>
<tr>
<td>7.1 Fatal Accidents</td>
<td>164</td>
</tr>
<tr>
<td>7.2 Excess Lifetime Fatality Risks</td>
<td>165</td>
</tr>
<tr>
<td>7.3 Deciding Whether to Operate a Hazardous Facility</td>
<td>169</td>
</tr>
<tr>
<td>7.4 Schrödinger’s Cat</td>
<td>176</td>
</tr>
<tr>
<td>7.5 Checking for Cancer</td>
<td>176</td>
</tr>
<tr>
<td>LIST OF EXAMPLES (cont.)</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.6 Stochastic Hazard Rates</td>
<td>177</td>
</tr>
<tr>
<td>7.7 Probabilistic Hazard Rates</td>
<td>177</td>
</tr>
<tr>
<td>7.8 Superposition of Periodic Hazard Functions</td>
<td>178</td>
</tr>
<tr>
<td>7.9 Individual Car Accidents</td>
<td>179</td>
</tr>
<tr>
<td>7.10 Buying Insurance</td>
<td>197</td>
</tr>
<tr>
<td>8.1 Aggregation of Individual Hazard Rates Over Time</td>
<td>212</td>
</tr>
<tr>
<td>8.2 An Aggregation Artifact in Comparative Risk Assessment</td>
<td>215</td>
</tr>
<tr>
<td>8.3 Population Risk vs. Number of Fatalities</td>
<td>234</td>
</tr>
<tr>
<td>8.4 Non-Existence of a Utility Function for Numbers of Lives Lost</td>
<td>236</td>
</tr>
<tr>
<td>8.5 Fatality Risk Hypergraphs</td>
<td>237</td>
</tr>
</tbody>
</table>
### Notation and Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Set of Acts</td>
</tr>
<tr>
<td>C</td>
<td>Consequence Set</td>
</tr>
<tr>
<td>C.</td>
<td>Continuity Axiom (p.50)</td>
</tr>
<tr>
<td>c.d.f.</td>
<td>Cumulative Probability Distribution Function</td>
</tr>
<tr>
<td>d.m.</td>
<td>Decision Maker</td>
</tr>
<tr>
<td>FSD</td>
<td>First-order Stochastic Dominance</td>
</tr>
<tr>
<td>I.</td>
<td>Independence (Substitution) Axiom (p.51)</td>
</tr>
<tr>
<td>NDT</td>
<td>Normative Decision Theory</td>
</tr>
<tr>
<td>N-M</td>
<td>Von Neumann-Morgenstern (p.58)</td>
</tr>
<tr>
<td>R</td>
<td>Comparative Risk or Preference Relation (p.79)</td>
</tr>
<tr>
<td>r.a.</td>
<td>Risk Analyst</td>
</tr>
<tr>
<td>S</td>
<td>Set of States</td>
</tr>
<tr>
<td>s.d.m.</td>
<td>Social Decision Maker</td>
</tr>
<tr>
<td>W.O.</td>
<td>Weak Order Axiom</td>
</tr>
<tr>
<td><em>-</em></td>
<td>is Isomorphic to (p.55)</td>
</tr>
<tr>
<td>d(c)</td>
<td>Degenerate c.d.f. that jumps from 0 to 1 at c</td>
</tr>
<tr>
<td>E(U;1)</td>
<td>Expected Value of U(c) when c has Distribution 1 (p.50)</td>
</tr>
<tr>
<td>(a p b)</td>
<td>p-Mixture of a and b (p.51)</td>
</tr>
</tbody>
</table>
The classical form of a decision problem without information collection is as follows. A single decision maker, d.m., must choose an act or action, $a$, from a set $A$ of possible acts. When he has done so, he receives or experiences a consequence $c$ from a set $C$ of possible consequences. Which consequence will be produced by a given act may depend on other, exogenous, factors; the set of possible act-consequence mappings will then be indexed by a variable $s$, ranging over an index set $S$. $s$ is often interpreted as the true "state of the world", and $S$ as the set of possible states. Thus, the mapping from acts in $A$ to consequences in $C$ is thought of as being given by some deterministic function $c(a,s)$, where $s$ may be uncertain and $a$ is to be chosen. A goal of classical decision analysis is to establish axioms that will allow the d.m. to deduce (or a decision analyst to prescribe) his preferences for acts from his preferences for consequences and his beliefs about $s$.

1.2 Risk Management Decisions: Examples and Characteristics

In this investigation, we examine decision problems which differ from the classical one in several ways:

(i) There may be a delay between the time at which an act $a$ is chosen and the later time at which its consequence $c(a,s)$ is learned.

Example 1a (The Cigarette Smoker): You are a young decision maker trying to decide whether to smoke a cigarette. You feel that there is a minute chance that doing so now will cause a cancer or some other illness later in your life, but that even if it does, you won't find out for at least several decades (and you may die in an automobile accident by then anyway.)

(ii) There may be a further delay between the time the consequence $c(a,s)$ is learned and the time it actually occurs.
Example 1b: A new medical study shows that people of your type and age who smoked in their youth are virtually certain to experience the onset of a terminal illness, Y, within two to ten years of your present age (if they live that long.) You feel doomed, though not yet ill; you also blame yourself for smoking that cigarette in your youth.

(iii) c(a,s) may never be learned, or may not be learned fully, or may only be revealed gradually as time passes.

Example 1c: A new, expanded study shows that in fact, people of your type and age tend to develop terminal illness Y within a few years whether or not they ever smoked, and that there is no way of telling whether a given case of Y was caused by previous smoking. You still feel doomed, but no longer feel guilty, and on balance feel pleased that you smoked while you had the chance.

Such limitations on what can be known when are characteristic of decision problems in the various fields of health and safety risk analysis, and profoundly affect the character of risk management decisions. When the possibility of collecting information about s is added, either before a is chosen or after, the lag between choice of an act and revelation of its consequences has even greater implications for normative theories of decision making and risk management.

Example 1d: Several years later, after the onset of Y, you read that a medical breakthrough now allows doctors to determine precisely what caused illness Y in anyone who has it. How much, if any, would you pay to learn (or to avoid learning) whether your case of Y was caused by your smoking?

In addition to such knowledge constraints, risk management decision problems in the domain of human health and safety risk analysis typically share the following characteristics:

(iv) The consequence set C includes possible human deaths, injuries, or illnesses;
The choice of act is not made in isolation by a single d.m. At the very least, the d.m. chooses a only after listening to technical information and perhaps advice from an expert who we shall call the risk analyst (r.a.).

Example 2: Regulating Automobile Emissions

A regulator (the d.m.) is trying to decide whether to require automobile manufacturers to adopt emissions control technologies that will reduce the current level, $x$, of tailpipe emissions of a certain currently unregulated pollutant, $X$. He suspects that a reduction $dx$ in the emissions level of pollutant $X$ may reduce the incidence rate $r(Y,Z)$ (= average number of new cases per person-year at risk) of illness $Y$ in target population $Z$. In fact, there are several populations and illnesses, indexed by $z$ and $y$, respectively, that he feels might possibly be affected. But he is unsure how large the reduction in incidence rate, $dr(y,z)$, of each disease $y$ in each group $z$ is likely to be for a given change $dx$ in tailpipe emissions of $X$. Moreover, if $dx$ is implemented, its health effects are likely to be small and gradual enough -- even though the cumulative number of cases prevented may be quite significant -- so that they will be unobservable through any affordable epidemiological study. On the other hand, the short-run economic costs to manufacturers (and hence indirectly to consumers) of using a new control technology to reduce emissions of $X$ are much more certain. The d.m. in this example may feel that he needs two sorts of help to make a confident decision:

- He needs to learn more about the probable magnitudes of the changes $dr(y,z)$ in incidence rates of illnesses in different groups in response to a change $dx$ in emissions; and
- He needs some help structuring the value judgements needed to combine the different $dr(y,z)$ estimates into an overall summary of probable public health impact from $dx$. 

3
Example 3: Marketing a New Drug

A pharmaceutical manufacturer must decide whether to market a new nonprescription drug. FDA and independent laboratory testing have provided no evidence that the drug has harmful side effects, but the manufacturer is concerned about potential sensitive subpopulations and possible long latency period health effects. If there are effects, they may not materialize for twenty years or more, and may not be identifiable as such when they do appear. The d.m. needs to know whether "enough" risk research has been done on the drug to warrant distributing it, given that there are health benefits to consumers expected from its use, even though the risk is still somewhat uncertain.

1.3 Roles for the Risk Analyst

These examples suggest that the relationship between risk analyst and decision maker differs from the traditional relationship between a decision analyst and his client. Instead of just helping the d.m. to formally structure his beliefs about s and his preferences over C, and then to draw out their logical implications (under an assumed set of axioms) for choice of a, the r.a. must supply substantive knowledge about the probable consequences of acts. He may also tackle the difficult problem of helping the d.m. to structure his preferences over consequences involving deaths and illnesses -- consequences for which, in contrast to the assumptions of classical decision theory, the d.m. may initially have no well-defined preferences and which in any case he may never learn.

The r.a. may thus act

(i) As a substantive expert about s, i.e., about the probable consequences of different choices;

(ii) To help structure the d.m.'s (possibly latent) preferences over (a,s) pairs;
(iii) As synthesizer of his own information about s with the d.m.'s; and

(iv) As a logical computer who deduces implications for the d.m.'s choice of an act from (i) the d.m.'s preferences over (a,s) pairs; (ii) their combined knowledge about s; and (iii) normative axioms of risk management decision making.

These roles are conceptually distinct, and it may be appropriate for the d.m. to reserve some of them for himself or for other policy advisors in some situations, e.g., when the r.a. is an expert witness and the d.m. is a judge.

1.4 Scope and Limitations

The purpose of this study is to take some steps toward the development of a normative theory of communication and decision making for the class of decision problems defined by the preceding characteristics. Informally, we ask two questions:

QUESTION 1: How and under what conditions can the r.a. encode his substantive knowledge about the risks of an activity in a way that the d.m. can decode and use? (Here, the d.m. can be anyone who must decide how to restrict his own or someone else's participation in the activity.)

QUESTION 2: Given the r.a.'s message, what risk management decision "should" the d.m. make, and what are the criteria to be used in deciding this?

We will concentrate on Question 1, although it is necessary to understand something about the answers to both in order to fully answer either. We will look at the problem of risk communication in Question 1 as one of developing an "objective" way of measuring risks. By "objective", we will mean "the same for all observers," or "unambiguously interpretable by all observers." Thus, we seek a way of encoding risks as numbers (or
perhaps as other mathematical objects such as probability distributions) so that

- One person (the r.a.) can apply the encoding to what he knows to obtain a numerical or mathematical expression of risk; and

- Anyone else who understands the risk measurement method used, regardless of his own subjective preferences and beliefs, can then decode the resulting expression to obtain the knowledge that the r.a. encoded in it.

This is in rough analogy to the use of measurements to communicate knowledge of the physical properties of physical objects, such as temperature or length; the expressions "Liquid x is at temperature 57 degrees" or "Rod L is of length ten," for example, have unambiguous meanings to anyone who know what measurement scales have been used to encode these properties. The "objects" of our study will be human activities, however, and the "properties" of interest will concern the probabilistic character of the outcomes or consequences of these activities.

In practice, the choice of an act from the set A of collectively feasible acts (or risk management decisions regarding who can participate how in what activities and under what conditions) is often determined by the individual choices of several decision makers having different partial information about s. In this study, we limit ourselves to the case of a single r.a. communicating with a single d.m. We also focus on the potential health and safety consequences of the d.m.'s choices, and not on financial consequences, for example. We thus assume that preferences over health and safety consequences can meaningfully be studied in isolation -- an example of what is called a "separability" assumption about preferences over total consequences. Finally, we are primarily concerned here with risk measurement and communication, rather than with risk management decision making -- although as stressed above, the two are closely related. This emphasis is chosen in the belief that a
well-developed theory of risk measurement and communication is a prerequisite for a sound normative theory of risk management, especially since most risk management decision making is decentralized or distributed among multiple decision makers.

Even with these restrictions, a wealth of important risk management problems remain. For example, the following important cases fall within the scope of this investigation:

A. Decentralized Individual Decisions

The d.m. may be an individual trying to decide whether to buy a certain consumer product (e.g., a lawnmower, diving board, or prescription drug), and the r.a.'s advisory role may be filled by the manufacturer or by a regulatory agency trying to find a succinct way of communicating its substantive knowledge about the product's potential hazards to nontechnical potential consumers. Or the d.m. may be a potential employee in a labor market, trying to inform himself about the magnitude of occupational risks before accepting a certain job at a given wage rate; or a potential home buyer, wondering exactly how large the risks are of living near a former hazardous waste site in a development that the state has classified as "safe" for human inhabitation.

B. Hierarchical Public Decisions

A regulator trying to set a health or safety standard for a manufacturer that the manufacturer can interpret clearly and use as an unambiguous constraint in making his own decisions faces essentially the same communication problems as a r.a. trying to express the health and safety risks of an activity to a d.m. Conversely, a manufacturer trying to decide whether his facilities are in compliance with applicable health and safety standards needs to be able to decipher unambiguously the language in which risk standards are expressed, and thus faces essentially the same communication problems as a d.m. trying to understand the results of a risk analysis. (For example, how has the
boundary between uncertainty about consequences and uncertainty about uncertainty about consequences been drawn in deriving an overall expression of the "risks" from an activity? Chapter 7 will present some examples that clarify such distinctions).

C. Centralized Decisions

The d.m. may be a product manufacturer trying to estimate quantitatively the risks to (i) consumers of the product; (ii) his employees; and (iii) neighbors of his production facilities, e.g., health risks from plant emissions, or safety risks from potential catastrophic industrial releases or fires and explosions. Or the d.m. may be a public decision maker trying to decide whether the public and occupational health and safety risks from a new production process or technology are "greater" than the corresponding risks from current practices.

In each of these cases, being able to express clearly the risk from an activity, perhaps as a number on some appropriate scale, so that someone can make comparisons with other risks or with the deterministic economic costs and benefits of the activity, will clarify the basis for choice and provide a foundation for risk management decisions.

1.5 Overview

The remaining chapters develop and apply a mathematical approach to the problem of objective (unambiguous) numerical risk measurement. Normative theories of decision making are reviewed and used to guide the development of measurement theories for different types of risks, but our emphasis throughout is on measurement, in the sense of mathematical representation, rather than on decision making per se. However, it is our hope that the theoretical approach to risk measurement developed in the following chapters will eventually contribute to sound normative theories of risk management decision making.
The following chapters fall into three major groups. Chapters 2 to 4 review past approaches to risk measurement and decision theory and set forth the technical background for a new approach. Chapters 5 and 6 develop this new approach and discuss its limitations, and chapters 7 and 8 apply the new approach to measurement of individual and population fatality risks, respectively.

In more detail, Chapter 2 introduces key ideas of information representation and updating that are crucial to our concept of risk, fully developed in Chapters 7 and 8, as resulting from hazard experienced over time. Chapter 3 reviews and synthesizes normative decision theories for the "classical" case where a d.m. must, in effect, choose among probability distributions over consequences. Von Neumann-Morgenstern (N-M) utility theory is presented briefly from a measurement theory (i.e., preference relation representation) perspective, and some recent generalizations and extensions are reviewed. Chapter 4 summarizes and critically reviews previous contributions to the risk measurement literature, which have tended to focus on choice among probability distributions over a single-attribute consequence.

Chapter 5 introduces our key technical approach to a new theory of objective numerical risk measurement for this classical case, and observes that precise objective numerical risk measurement is impossible even in principle. Limited-precision and non-numerical representations are therefore introduced instead as practical substitutes for the unreachable goal of precise objective numerical measures of risk. Chapter 6 examines the implications for risk communication between two partially informed observers, drawing on past results and using them to put the theoretical approach of Chapter 5, which is primarily applicable to the case of an informed observer (the r.a.) communicating with an uninformed d.m., into perspective.

Chapters 7 and 8 apply and extend the approach proposed in Chapter 5 to measurement of fatality risks for individuals and populations, respectively. Chapter 7 argues that all risks (with the exception of
those from Poisson arrivals) necessarily evolve over time, and extends
the approach of Chapter 5 to apply to dynamic risks. "Uncertain" and
"stochastic" dynamic risks are also introduced, with the former
reflecting probabilistic uncertainty about which of several risk
processes is operating, and the latter arising when assessed risk depends
on information contained in observations other than simply the passage of
time. A generalization of N-M utility theory to uncertain and stochastic
dynamic risks, derived from the work of two mathematical economists
(Kreps and Porteus, 1979) is presented and its inapplicability to
"controlled" stochastic dynamic risk processes in which actions are
interleaved with observations is noted. However, many risk management
decisions have observation and control structures such that a decision is
essentially made at one point in time and its risk consequences learned
only when and if they occur. The temporal expected utility theory of
Chapter 7 is relevant for these situations.

Chapter 8 explores possible applications of the approach developed in
Chapters 5 and 7 to measurement of population risk. In a sequence of
increasingly complex models, it is shown that a theory of political
choice or public decision making is necessary to define population risks
in cases when risks are time-varying. Standard approaches to definition
of population risk are briefly summarized and some difficulties arising
from aggregation artifacts are noted. While statistical theory for large
populations can be successfully exploited for "routine" population risks
if certain ethical axioms are adopted for representing the values of
lives, "catastrophic" population risks are much more difficult to
address. Chapter 8 attempts to clarify the mathematical and conceptual
problems involved and to suggest a line of attack, but proposes no final
solutions.

Finally, Chapter 9 concludes with a summary of our major contributions.
The three most important ones are:

(i) The introduction of mathematical objects other than probability
distribution functions (or measures) over consequences to represent
elements of the decision maker's choice set. The new objects, which include certain simple stochastic processes, for example, are proposed as being better suited than probability measures for representing human health and safety risks.

(ii) Emphasis on the role of time in risk. This leads to the identification of certain "consequences" of risk management decisions, such as the condition of being put at risk and the duration of uncertainty about eventual physical outcomes, that are not traditionally included in the formulation of decision problems, but that may significantly affect preferences among decision options.

(iii) Systematic separation of the "subjective" components of risk preference (defined as those components that may differ across different individuals) from the "objective" component (defined as that on which all individuals will agree); and numerical representation of the latter.

Like any inquiry into foundations, this one is somewhat removed from direct application to risk management decisions. However, a number of immediately applicable results are developed and cited along the way. Moreover, we hope that by concentrating on fundamental questions about the nature of risks and their formal representations, we have helped to clarify the mathematical scope of the field of risk analysis, to sharpen its boundaries, and to lay a conceptual foundation for new normative theories of decision making that will be more useful than current ones for solving practical risk management decision problems.
2.0 Introduction and Overview

To understand in detail the problems of risk measurement and risk communication, it is useful to have models of what individuals know and believe about uncertain propositions. The propositions of interest in decision making are usually claims about what will happen if particular actions are taken, and are nearly always uncertain. Risk measures in the form of numbers, graphs, or other formal symbols are the expressions used by a (partially) informed risk analyst (r.a.) to communicate to a d.m. r.a.'s knowledge or beliefs about the probable consequences of the d.m.'s potential actions. These measures therefore typically reflect not simply objective intrinsic properties of the alternative acts, but rather the r.a.'s knowledge and beliefs. This gives theories of risk measurement a different flavor from theories of measurement for physical quantities. Observers with access to different information may report different expressions for "the risk" of the same activity, and differences in the information available to different observers can go well beyond the measurement noise or error typical of physical measurements.

This chapter provides a conceptual vocabulary and several technical models for discussing and representing an individual's knowledge and beliefs and for explaining how these depend on and may be updated by his observations. In addition, some background on probability measures and related concepts that will be useful in subsequent chapters is given. The material in this chapter is used primarily as technical background for later chapters and need not be read in its entirety to understand them.

The organization and main results of this chapter are as follows.

Section 2.1 introduces a highly formal approach to the representation of individual knowledge based on recent work in epistemic logic (see e.g.,
Rosenschein, 1985). Knowledge is modelled as a set of propositions in an appropriate formal language and a distinction is drawn between empirical propositions (assertions about the state of the world) and epistemic propositions (assertions about what is known). One purpose of this section is to suggest a possible formal philosophical and axiomatic basis for risk measurement and to relate it to other fundamental epistemological questions. A much more important purpose is to motivate and explain the following three sets of technical concepts, which are used as needed throughout the rest of this work:

- An **observation**, or **empirical observation**, is a signal, observed through a "channel", that gives information about the state of the world. Deterministic (noiseless) channels give knowledge; noisy channels produce beliefs that depend on prior beliefs. These are the two kinds of information that an r.a. may wish to communicate to a d.m.

- An observer's **information partition** divides the set of possible states of the world mediating between the d.m.'s act and the consequence of the act into classes such that the observer "knows" which class the true state belongs to, but "knows" nothing else about the state of the world (although he may have beliefs about it). Every deterministic channel creates a unique information partition.

- Information partitions can be partially ordered by a relation of comparative "informativeness" or "fineness", and can be combined by two operations ("meet" and "join") that arise naturally when observers with different information partitions communicate.

These terms and concepts are used in several places in later chapters. The idea of an information partition is especially important and useful. For one thing, it provides a simple model for interpreting phrases like "Observer i's information increases over time" (meaning that his information partition becomes finer over time) or "Observer i is better
informed than observer j" (meaning that i's information partition is finer than j's). Such phrases appear frequently in advanced modern treatments of stochastic processes and their applications (e.g., Arjas, 1981), usually with a much more complicated and general intended interpretation in terms of event algebras, as explained in Section 2.2. It is useful, and always legitimate, to think of information partitions as a simplified model for interpreting such literature, and we use this interpretation in several places to build an intuitive understanding of the nature of risk and uncertainty over time.

Information partitions are used for several purposes in Chapters 6 to 8.

- In Chapter 6, communication between two observers with access to different information (i.e., to different information partitions or to signals from different deterministic channels) is modelled in terms of iterative refinement of individual information partitions. A developing literature has shown that when observers have common prior beliefs, the process of knowledge acquisition and attainment of "common knowledge" through communication can be understood in terms of the meet and join operations between their information partitions (see Theorem 6.3).

- In Chapter 7, the idea of increasing information over time is used extensively to represent updating of beliefs about uncertain risks as time passes. The passage of time in itself acts as a kind of "signal" that gives the observer information about which of several possible risks he is facing. Chapter 7 also allows observations to arrive at different points in time, leading to successive refinements in the information partition with respect to which individual "risk" is calculated. This is a good example of a case in which the increase in information over time may take a more complicated form than refinement of an information partition, but in which thinking of increasing information in these terms provides a simple concrete
interpretation of what is going on. (The more complicated interpretation allows, among other things, for observations through a sequence of noisy channels.)

In Chapter 8, information partitions are used again to represent population heterogeneity. Individuals are classified into groups or "cells" based on their observed characteristics, but not all characteristics that determine individual risk (e.g., probabilistic health response to the d.m.'s actions) are observable. The set of distinguishable cells, corresponding to different combinations of the levels or values of the observable characteristics, defines an information partition.

Thus, the technical ideas related to information partitions introduced in Section 2.1 are important for understanding the theories of risk measurement developed in later chapters. The formal axiomatic treatment of knowledge, by contrast, is used only briefly in Chapter 6, in discussing the idea of "common knowledge" as the result of communication between observers with common priors.

Section 2.2 is more technical and less important for the rest of the chapters than Section 2.1. It deals with beliefs derived from observations through a noisy channel, rather than with knowledge. In this sense, it is a generalization of Section 2.1, since deterministic channels may be thought of as a special limiting case of noisy channels. The idea of an information partition is generalized to that of a probability algebra. Whereas an information partition consists of several mutually exclusive, collectively exhaustive blocks of states such that the observer knows which block contains the true state, an algebra is generated by a partition in which the observer is only able to assign probabilities to the different blocks to indicate his beliefs about how likely he thinks each is to contain the true state. (Here and throughout, we avoid unnecessary mathematical generality. For example, in introducing probability algebras, we assume a finite set of distinguishable states.)
Key ideas introduced in Section 2.2 that are used in subsequent chapters include the following:

- An event is measurable with respect to a probability algebra only if it consists of a union of blocks of the partition generating the algebra. The probabilities of other events (e.g., subsets of these blocks) can not be calculated from the information contained in the probability algebra.

- Probability algebras may be compared by relative "fineness", just like information partitions; moreover, a partial ordering of noisy information channels by comparative "informativeness" can be defined that extends the comparative informativeness relation based on fineness between information partitions.

- The "probability" of an event, as calculated from a prior probability algebra and a sequence of observations through a noisy channel follows a stochastic process similar to a random walk. The average step size in this process increases with the informativeness of the channel(s) through which observations are made.

Section 2.2 also provides treatments of three topics -- qualitative probability, calibration of probability assessments, and foundations of statistical inference -- that are of independent interest. These topics are not used directly in subsequent chapters. But the treatment of qualitative probability as a representation of subjective beliefs introduces a type of reasoning that will be used extensively in Chapters 3 and 4 in discussing representation of preferences. The discussions of calibration and inference suggest some fundamental problems in the use of subjective probabilities and statistical inference to represent and update beliefs, having to do with the fact that these techniques leave essentially unanswered the question of where initial beliefs come from or of what makes them credible. We document some of these difficulties, but do not try to resolve them.
2.1 Individual Knowledge

We now return to the case where the d.m. must choose an act \( a \) from a set \( A \) and then immediately receives a consequence \( c = c(a,s) \) from a set \( C \) of possible consequences, where \( s \), a member of \( S \), may be uncertain. For the moment, we shall assume that the d.m. has no further opportunities for collecting information about \( s \) before choosing \( a \); he must choose \( a \) with only the information that he currently has. Study of this simple situation provides insights into more general decision problems.

If the d.m. knew \( s \), then the solution to his decision problem would be to choose any act in \( A \) yielding a most-preferred consequence in \( C \). His decision problem becomes nontrivial only when he is uncertain about \( s \). To talk about his uncertainty, we will assume that \( S \) is finite and let \( F \) be an algebra of subsets of \( S \), i.e., a set of subsets that is closed under union and complementation (and hence under intersection). Every event \( E \) in \( F \) consists of a subset of states in \( S \).

2.1.1 A Language for Expressing Propositions

For any subset \( E \) of \( S \), let \( (E) \) denote the proposition "\( s \) is in \( E \)" , where \( s \) denotes the true state. An individual's knowledge (as opposed to his probabilistic beliefs) about \( s \) can be expressed through such propositions. To this end, it is useful to have a formal language in which propositions can be expressed. One potentially useful language is generated by the following two rules:

\[ L_1: \text{For any event } E \text{ in } F, \text{ the proposition } (E) \text{ is in } L. \] (Explicitly, (i) For any state \( x \) in \( S \), the proposition \( (x) \), i.e., "\( s \) is \( x \)", is in \( L \); (ii) For any two propositions \( (a) \) and \( (b) \) in \( L \), the proposition \((a) \cap (b)\) is in \( L \), where \( \cap \) denotes conjunction of propositions; (iii) For any proposition \( (a) \) in \( L \), the proposition \((- (a)) \) is in \( L \), where \( - \) denotes negation. Thus, \( L \) is closed under conjunction and negation (and hence under disjunction) of propositions.)
L2: Let $L'$ be any "regular" subset of propositions in $L$; then for any proposition $(a)$ in $L$, the proposition that $(a)$ is in $L'$, denoted by $(L'(a))$, is also in $L$. Thus, $L(a)$ implies $L(L'(a))$, for any "regular" subset $L'$ of $L$.

Rule L1 allows empirical propositions, i.e., propositions about $s$, to be expressed in $L$. Rule L2 allows an infinite hierarchy of propositions and "metapropositions" (i.e., propositions about propositions) to be included in $L$. This opens the door to logical difficulties. The idea of "regularity", which we will not formally specify here is required to avoid difficulties of the following sort:

Example 2.1: (Russell's Paradox)

Let $L'$ denote the set of propositions $(A)$ in $L$ such that $L'(A)$ if and only if $-(A(A))$, i.e., if and only if the set $A$ (of propositions) is not a subset of itself. Is $L'(L')$, i.e., does $L'$ include itself? Answering either "yes" or "no" to this question leads to an apparent paradox.

There is a large, somewhat specialized, literature in the theory of computation (and mathematical logic) that deals with various equivalent notions of regularity (under names such as "recursively enumerable sets", "effectively computable characteristic functions", "decidable propositions" or "strings recognizable by a finite state automaton in a finite amount of time") that suffice to prevent such logical difficulties. The basic idea is that a set $L'$ of propositions is "regular" if, given any candidate proposition (or string of symbols) $(a)$, it is possible to determine whether $(a)$ is in $L'$ by a finite number of applications of a few "rewriting rules" that implicitly define $L'$. For the purposes of modelling individual knowledge, however, we can get by without careful specification of rule L2 by using only a special case dealing with the subset of "known" propositions, for which the defining "rewriting rules" are easily given. We shall, however, assume that $L$ satisfies rule L1, so that empirical propositions about $s$ can be expressed.
2.1.2 An Axiomatic Approach to Individual Knowledge

We now develop a definition of individual knowledge. Let \( K \) be a subset of the propositions in \( L \), and for any proposition \( (a) \) in \( L \), let \( K(a) \) mean that \( (a) \) is in \( K \). \( K \) is to be interpreted as the set of propositions that are "known" to the individual (the d.m. or r.a.), and \( K(a) \) means that he knows \( (a) \). We shall use the following rules or axioms to specify \( K \) more exactly.

**\( K1 \):** For any event \( E \) in \( F \), \( K(E) \) implies \( E \). That is, if it is known that "s is in \( E \)" then \( s \) is, in fact, in \( E \). Thus, empirical propositions in \( K \) are assumed to be true, and may therefore be referred to as "facts".

**\( K2 \):** For any two propositions \( (a) \) and \( (b) \) in \( L \), \( K(a) \) implies \( K((a)v(b)) \), where "v" here denotes disjunction of propositions. In particular, if \( E \) and \( E' \) are two events in \( F \) with \( E \) a subset of \( E' \) (so that \( (E) \) implies \( (E') \)), then \( K(E) \) implies \( K(E') \).

**\( K3 \):** For any two propositions \( (a) \) and \( (b) \) in \( L \), \( K(a) \& K(b) \) implies \( K((a)^(b)) \), where "^" here denotes conjunction of propositions. Thus, if someone knows that \( (a) \) holds and knows that \( (b) \) holds, then he knows that both \( (a) \) and \( (b) \) hold.

**\( K4 \):** For any proposition \( (a) \) in \( L \), \( K(a) \) implies \( K(K(a)) \). Conversely, \( K(K(a)) \) implies \( K(a) \).

These axioms, similar in spirit to a set proposed by Bacharach, 1985, essentially state that someone can't both know something and be wrong about it (rule \( K1 \)); that he can deduce logical consequences of what he knows (rules \( K2 \) and \( K3 \)); and that he knows something if and only if he knows that he knows it. Additional axioms are possible. For example, Bacharach proposes

**\( K5 \):** For any \( (a) \) in \( L \), \([L-K](a)\) implies \( K([L-K](a)) \), i.e., if someone doesn't know something (meaning that the corresponding proposition is in \( L \) but not in \( K \)), then he knows that he doesn't know it.
However, this raises logical problems (about "soundness" and "completeness") and decidability issues of the sort that we wish to avoid. We shall therefore restrict ourselves as far as possible to $K_1$ through $K_4$. A set of propositions satisfying these axioms (with respect to some state set $S$ and an algebra of events $F$ on $S$) will be called a "knowledge base" (Rosenschein, 1985).

2.1.3 Private Information and Deterministic Channels

Axioms $K_1$ to $K_4$ define a logical structure that any individual knowledge base, $K$, might be expected to have. They do not say anything about how (or whether) empirical knowledge about state $s$ might get into $K$ to begin with. In other words, the problem of how to "load" a knowledge base with some initial empirical fact(s) is not addressed. We will use the following fundamental axiom to bridge this gap between empirical reality, $s$, and individual knowledge:

$K_0$: For each state $x$ in $S$, $s = x$ implies $K(B(x))$, where $B(x)$ is an event in $F$. [To express this rule more formally using the structure developed so far, we could write $K(-(s)^{-(K(B(x))})).$

That is, $B(x)$ is a set of states such that if the true state $s$ is $x$, then the individual will know that $s$ is in $B(x)$.

The correspondence $B(.)$, which maps states in $S$ into subsets of states in $S$, may be interpreted as a (noiseless) "observation", meaning a deterministic function that maps states, which are not themselves directly observable, into "signals" that are. $B(x)$ may then be interpreted either as the set of states producing the same signal as $x$, or as the signal itself. We shall choose the former interpretation, and introduce a separate set $Y = [y_1, ..., y_m]$ of "signals" to index the possible distinguishable outcomes of the individual's observation of $s$.

We define a deterministic information channel $C$ as a many-to-one mapping from $S$ to $Y$, i.e., $C$ maps each state $x$ onto a unique signal, say $y(x)$.
B(x) is the set of states that maps onto signal y(x). A deterministic (or "noiseless") channel can be represented as an m x n Boolean matrix with a "1" in the ith row and jth column if and only if state i maps into signal j, and a zero there otherwise. We will use the symbol C to denote both this channel matrix and the corresponding mapping from S to Y.

The effect of axiom K0 is to assume that the individual has a perfect (noiseless) channel through which he receives some information about s. (Bacharach in effect attempts to prove the existence of such a channel from axioms similar to K1 through K5, but his proof, which fails to distinguish among levels of the language in which propositions are stated, appears to be flawed.) The information that the individual receives under K0, namely, that s is in B(s), is called his "private information". Applying axioms K1 to K5 iteratively to B(s) generates an entire knowledge base consisting of the empirical proposition (B(s)), the logical empirical consequences of this proposition (representing the supersets of B(s)), and an infinite hierarchy of "epistemic" propositions. Later we shall consider communication with someone else as an alternative way of getting knowledge. We shall also be interested in beliefs, as well as knowledge, about s, and hence will consider "noisy" channels in which the mapping from states to signals is probabilistic.

2.1.4 Comparison of Deterministic Information Channels: Terminology and Background

It is easy to see that K0 to K5 imply that the state set S can be partitioned into disjoint, collectively exhaustive blocks, say B(1),...,B(m) where m is the number of blocks, such that the individual will know which block contains s, and will know nothing more about s. Namely, the equivalence relation "is indistinguishable from" between states induces this partition, where states x and x' in S are said to be "indistinguishable from" each other (with respect to a communication channel C) if and only if they map onto the same signal under C. States that are indistinguishable from each other are assigned to the same block of the partition, and the partition itself will be called the "information partition" induced by C.
Let $C$ and $C'$ be two deterministic channels on the same state and signal sets, with corresponding information partitions $I$ and $I'$. We can say that $I$ is at least as informative as $I'$ if every two states that are distinct in $I'$ are also distinct in $I$. $I$ is also said to be "finer than" or "a refinement of" or "strictly more informative than" $I'$ if it distinguishes all the pairs of states that $I'$ does and some others as well. Conversely, $I'$ is said to be "coarser than" or "a coarsening of" or "strictly less informative than" $I$ under these conditions. A deterministic channel $C$ may be called (strictly) "more informative than" a second channel $C'$ if and only if the partition induced by $C$ is finer than the partition induced by $C'$. This definition of comparative informativeness of partitions is widely used in computer science in the study of information flows in finite-state machines. (see e.g., Kohavi, 1970, p. 347.)

Given any two partitions, $I$ and $I'$, it may be that neither is finer than the other, and hence that they are not comparable in terms of relative informativeness as defined here. However, in such a case there will be a unique partition that is the coarsest common refinement of $I$ and $I'$. This partition, called the "meet" of $I$ and $I'$ and denoted by $I \wedge I'$, is defined as follows:

Two states $w$ and $x$ are distinguished in $I \wedge I'$ if and only if they are distinguished in $I$ OR they are distinguished in $I'$. Equivalently, $w$ and $x$ are assigned to the same block of $I \wedge I'$ if and only if they are in the same block in $I$ AND they are in the same block in $I'$.

In notation, we could write $B(w;I \wedge I') = B(x;I \wedge I')$ if and only if $B(w;I) = B(x;I)$ and $B(w;I') = B(x;I')$. It should be clear that indistinguishability in $I \wedge I'$ is symmetric, transitive, and reflexive, and hence that $I \wedge I'$ is in fact a partition; thus, the set of partitions is closed under meets. We will see in Chapter 6 that if $I$ and $I'$ are the information partitions used to load the initial knowledge bases of two individuals, 1 and 2, so that 1 will know $B(s;I)$ and 2 will know $B(s;I')$ as their original private information, then $I \wedge I'$ can be thought of as the
partition used to load their common knowledge base when each can communicate his private information perfectly to the other.

Similarly, given any two partitions I and I', there is a unique finest common coarsening of them, denoted by I \lor I' and called the "join" of I and I'. I \lor I' is the finest partition that is refined by both I and I'. It can be constructed from I and I' as follows:

States w and x are assigned to the same block of I \lor I' if and only if there is a sequence of states s_1, s_2, ..., s_r with w = s_1, x = s_r, and such that any two consecutive states in the sequence are either indistinguishable (belong to the same block) in I or are indistinguishable in I'. Equivalently, w and x belong to the same block in I \lor I' if and only if there is a sequence of blocks, drawn alternately from I and I', such that the first member of the sequence contains w, the last member of the sequence contains x, and no two consecutive members of the sequence are disjoint. In this case, w and x will be called "chain-connected".

It is easy to see that the relation "is chain connected to" between states is reflexive, symmetric, and transitive; I \lor I' is simply the partition induced by this equivalence relation. As a computational matter, the block of I \lor I' to which any state x belongs, which we will denote by B(x; I \lor I'), can be found as follows:

1. Let B(x) = B(x; I) \cup B(x; I')

2. Let R = \{B \text{ such that } B \text{ is a block in } I \text{ or in } I' \text{ that intersects } B(x)\}

3. If B(x) = B(x) \cup R then let B(x; I \lor I') = B(x). Otherwise, let B(x) = B(x) \cup R and return to step 2.

(Here, "\cup" denotes set union, and "=" following a "let" denotes an assignment statement rather than equality.)
The join of two information partitions, and this process for computing it have very natural interpretations in terms of communication between two players and the set of states that are "possible" with respect to what the players know. This interpretation is discussed briefly in Chapter 6 (Theorem 6.3) and more extensively in Bacharach (1985) and Geanakoplos and Polemarchkis (1982).

Technical Note: From an abstract mathematical standpoint, the comparison of deterministic channels or their information partitions may be summarized by saying that the set of all partitions of S forms a lattice with respect to the partial ordering (i.e., reflexive, antisymmetric, transitive) relation "is finer than" and the dual operations of meet (= greatest lower bound) and join (= least upper bound). Hence, for example, \(^\land\) and \(\lor\) satisfy associativity, commutativity, idempotency, and absorption, i.e., \(I \lor I' \lor I = I' \lor (I \lor I') = I\). (see e.g., Hartmanis and Stearns, 1966.)

2.1.5 Syntactic Characterization of Individual Knowledge

So far, we have emphasized the "semantic" interpretation of individual knowledge and private information, i.e., the link between what is known (propositions in K) and what is (events in F that include s.) However, it is useful to have a purely syntactic characterization of individual knowledge, i.e., a characterization in terms of symbol-manipulation rules, and we will therefore conclude this section by restating axioms K0 to K5 from this perspective.

We will let K denote the set of strings representing what an individual "knows". The axioms defining this set are as follow:

K0': The string \((K(B(s)))\) is admissible.

K1': If \((a)\) and \((b)\) are two admissible strings, then \(((a)\lor(b))\) and \(((a)\lor(b))\) are admissible strings.
If (a) is an admissible string, then (K(a)) is an admissible string.

No other strings are admissible.

If we interpret B(s) as "The individual knows that s is in B" and (K(a)) as "The individual knows that (a)", then the set of admissible strings defined by the above rules may be interpreted as the knowledge base of an individual whose private empirical information is B(s). We will also say that this is the knowledge base "generated" by B(s). Strings in the knowledge base represent things that are known.

It is a useful to think of an individual's knowledge as consisting of two parts: a core empirical proposition, (B(s)), and a set of latent epistemic propositions that can be derived by a finite number of applications of rules K1' and K2'. Thus, even though an individual "knows" an infinite number of facts, e.g., (B(s)), (K(B(s))), (K(K(B(s))))..., most of this knowledge is latent. Only the initial string (B(s)) corresponding to the empirical proposition need be explicitly stored in order to generate the rest, i.e., the epistemic portion, of the knowledge base. The set K generated by the above rules is automatically "regular", and there is no need to refer to any larger set, L.

2.2 Individual Beliefs

In addition to his knowledge about states, an individual may also have beliefs. His knowledge is, by assumption, correct. Once a fact B(s) is in a knowledge base it stays there, and can never be driven out by subsequent information. The knowledge that s is in B(s) can be refined -- i.e., it may later be learned that s is in some proper subset of B(s) -- but it can never be reversed. Thus, our model of individual knowledge entails what is sometimes called "monotonic reasoning": new facts can be added to a knowledge base through continued observations, but old facts can not be withdrawn.
Beliefs, by contrast, can be revised or even reversed. An individual may initially believe that $s$ is in some subset of states $E$ more strongly than he believes that it is in some other subset $E'$, but subsequent observations may reverse this belief. We shall assume throughout this chapter that beliefs about $s$ can be represented by a (finitely additive) probability measure $p(.)$ on an algebra $F$ of events in $S$.

2.2.1 Finitely Additive Probability Measures as Representations of Individual Beliefs

For the moment, suppose that $F$ is the set of all subsets of $S$. Then by a finitely additive probability measure $p(.)$ on $(S,F)$ we mean a function assigning nonnegative numbers to the states in $S$ in such a way that the following two rules or axioms hold:

$P_1$: For any two states $w$ and $x$ in $S$, $p(w \text{ or } x) = p(w) + p(x)$; and

$P_2$: If $B(s)$ is the smallest set of states such that the individual knows that $s$ in $B(s)$, then $p[B(s)] = 1$.

Here, the probability $p(E)$ of any event $E$ in $F$ is defined as the sum over all states $x$ in $E$ of the numbers $p(x)$ assigned by probability measure $p(.)$ to the states in $E$.

Equivalently, we assume that an individual's prior beliefs about $s$ before receiving his private information $B(s)$ are representable by a set of nonnegative numbers $p(x)$, satisfying rule $P_1$ and the normalization convention $p(S) = 1$; and that his posterior beliefs after learning $B(s)$ are representable by nonnegative numbers $p[x;B(s)]$ obtained by assigning to each state $x$ in $B(s)$ the number $p(s)/p[B(s)]$, and to each state not in $B(s)$ a zero.

When we say that $p(.)$ "represents" an individual's beliefs about $s$, what we mean is that $p(.)$ assigns numbers to events in such a way that for any two subsets of states $E$ and $E'$ in $F$, $p(E)$ is greater than $p(E')$ if and
only if the individual believes that \( s \) is in \( E \) more strongly than he believes that \( s \) is in \( E' \). Or, assuming that he can order events in terms of relative plausibilities, \( p(E) \) is greater than \( p(E') \) if and only if \( E \) is judged (by the individual) to be more plausible than \( E' \), i.e., if and only if he judges that \( s \) is more likely to be in \( E \) than in \( E' \). By a "finitely additive" representation, we mean that the operation of union between disjoint states (and hence between any two disjoint events) is represented by addition of their respective probabilities. (See, e.g., Shiryayev, 1984, for a modern treatment and definitions of this and other probability concepts.) Thus, for any two disjoint events \( E \) and \( E' \) in \( F \),

\[
p(E \cup E') = p(E) + p(E').
\]

The assumption that an individual's beliefs can be represented by a finitely additive probability measure implies some substantive restrictions on the qualitative beliefs that he may hold.

**Example 2.2: Qualitative Probability**

Assume that the qualitative relation "is-judged-at-least-as-likely-as" between events, which we will denote by \( N \), satisfies the following axioms:

**QP1 (Weak Ordering)** \( N \) defines a weak (complete, transitive) ordering on \( S \), i.e., (i) for any states \( x \) and \( x' \) in \( S \), either \( x \) \( N \) \( x' \) or \( x' \) \( N \) \( x \); and (ii) for any three states \( v \), \( w \), and \( x \) in \( S \), \( v \) \( N \) \( w \) and \( w \) \( N \) \( x \) implies \( v \) \( N \) \( x \). (Here, \( x \) \( N \) \( x' \), for example, is read "\( x \) is judged at least as likely as \( x' \).")

**QP2 (Nontriviality)** For any subset \( E \) of \( S \), \( E \) \( N \) (null set). Also, it is not the case that (null set) \( N \) \( S \), i.e., \( S \) is nontrivial.

**QP3 (Monotonicity)** Let \( E \) and \( E' \) be any two disjoint subsets of \( S \), ad let \( E'' \) be any other subset of \( S \) that is disjoint from \( E \). Then \( E' \) \( N \) \( E'' \) if and only if \( (E' \cup E) \) \( N \) \( (E'' \cup E) \).

These three constraints on judged probabilities seem to be minimal requirements for coherent probability judgements.
THEOREM 2.1: Assumptions QP1 to QP3 imply the existence of

- A numerical function, \( p(.) \), mapping subsets of \( S \) into numbers in the interval \([0,1]\); and

- A binary operation, \(*\), mapping \([0,1] \times [0,1]\) into \([0,1]\)

such that for any two subsets \( E \) and \( E' \) of \( S \),

1. \( p(E) \) is at least as great as \( p(E') \) if and only if \( E \subseteq E' \).
2. If \( E \) and \( E' \) are disjoint, then \( p(E \cup E') = p(E) \ast p(E') \).
3. \( \ast \) is associative and commutative.
4. For any number \( x \), \( x \ast p(\text{null set}) = x \).


Although these properties are useful, they do not imply that the probability measure \( p(.) \) must be additive, i.e., that \( \ast \) is +. Indeed, any monotonically increasing function \( f \) with \( f(0) = 0 \) can be used to generate a \( \ast \) with the above properties by defining \( x \ast y \) as \( f'[f(x) + f(y)] \), where \( f' \) denotes the inverse of \( f \). For example, choosing \( f(x) \) to be \( \frac{\ln(1 + kx)}{k} \) will give \( x \ast y = x + y + kxy \) for \( k \) not equal to zero.

To understand the restrictions implied by the assumption of an additive measure, consider a state set \( S \) with five states, numbered from 1 to 5 in order of increasing judged probability. The "axioms of qualitative probability", QP1 to QP3, allow the 32 events in the algebra of all subsets of \( S \) to be partially ordered into sixteen classes, as shown in the "Hasse diagram" of Figure 2.1. In this diagram, an arc from a higher subset to a lower one indicates that the higher one is at least as likely as the lower one according to ranking \( N \). The sum of the numbers in any subset (i.e., the sum of the position numbers of the corresponding states in the judged likelihood ordering of states) roughly indicates the relative likelihood of the subset. Each subset with a sum of \( n \) is more likely (under QP1 to QP3) than some subset with a sum of \( n-1 \) and is less likely than some subset with sum \( n+1 \). However, QP1 to QP3 do not allow the relative ordering of two subsets with the same sum, or even of all subsets with different sums, to be determined. Only the comparisons
Sum of Position Numbers

S = 12345

15

14

13

12

11

10

9

8

7

6

5

4

3

2

1

0

FIGURE 2.1 QUALITATIVE PROBABILITY STRUCTURE FOR FIVE-STATE EXAMPLE
indicated by the partial ordering in Figure 2.1 can be drawn. To
generalize the qualitative probability structure of Figure 2.1 from the
case of five states to the general case of n states appears to be an
interesting and perhaps useful combinatorial algebraic exercise, but will
not be attempted here.

Thus, knowing the probability ordering of the five basic states is
insufficient to determine the ordering of all other events in the algebra
they generate, so that qualitative probability measures are in this sense
weaker than their corresponding additive measures. By the same token,
they can express a wider range of beliefs than numerical (meaning
finitely additive) probability measures. For example, Fine (1973, p. 22)
presents an example that we can transplant to the context of Figure 2.1
as follows. Suppose that

\[ o(1,5) \text{ is judged more likely than } (3,4), \text{ denoted } (1,5) \succ (3,4); \]
\[ o(2,3) \succ (1,4); \]
\[ o(4) \succ (1,3); \text{ and} \]
\[ o(1,3,4) \succ (2,5). \]

These rankings are evidently consistent with the partial ordering in
Figure 2.1, which is all that the axioms of qualitative probability
imply. However, they are inconsistent with any additive probability
measure. For translating \((1,5) \succ (3,4); \ (2,3) \succ (1,4); \text{ and } (4) \succ (1,3)\)
into numerical inequalities and summing would give that \([p(1) + p(5)] +
[p(2) + p(3)] + p(4)\) is greater than \([p(3) + p(4)] + [p(1) + p(4)] +
[p(1) + p(3)], i.e., \(p(1) + p(2) + p(3) + p(4) + p(5) = 1\) is greater than
\(2[p(1) + p(3) + p(4)], \text{ or } p(2) + p(5)\) is greater than \(p(1) + p(3) +
p(4). \) But this contradicts \((1,3,4) \succ (2,5). \) So the assumption of
additivity, i.e., that \(* = +\) in this example, can not be sustained.

In general, there seems to be no simple, compelling reason for assuming
that subjective beliefs can be represented by additive, as opposed to
more general, probability measures. However, the assumption leads to
great simplifications, for example, by allowing the probabilities of all
subsets of $S$ to be calculated from the probabilities of the individual states in $S$ alone. In addition, it can be justified in some practical applications, e.g., where relative frequencies are appropriate. We shall generally assume additive probability measures, for reasons given in Section 2.2.3.

2.2.2 Probability Measures with Respect to Coarse Algebras

We have assumed that $F$ is the set of all subsets of $S$. The smallest events in $F$ are the individual states; thus, $F$ provides complete resolution, meaning that any two states can be distinguished for the purposes of computing event probabilities.

In many situations, complete resolution is impossible, and the smallest subsets of distinguishable states will contain multiple states. It is then necessary to define probability measures without using rule Pl.

Example 2.3: Dose-Response Relations and Filtrations

Suppose that each individual in a population is in one of eight possible states, generated by three binary factors which we will call $e =$ "exposure", $d =$ "enzyme deficiency" and $r =$ "health effect" (or "response"). Each state may be represented by a triple of binary digits (zeros or ones), where $(e,d,r) = (1,0,0)$, for example, represents no exposure, no deficiency, and no health effect; and $(e,d,r) = (0,1,1)$ represents no exposure, presence of the enzyme deficiency, and occurrence of the health effect. For any specific individual, $e$ might represent prenatal exposure to a certain drug taken by the mother during pregnancy; $d$ might represent a biological characteristic or "type" (e.g., enzyme deficiency) that differs across individuals; and $r$ might represent occurrence of a specified illness in the individual during his life.

For convenience, label the states 0 through 7 so that the binary representation of each number is the state that it labels (i.e., $2 = (0,1,0)$, $7 = (1,1,1)$, etc.) Let $F$ be the set of all 256 subsets of $S = \{0,1,2,3,4,5,6,7\}$, i.e., the set of subsets generated under union and
complementation by the eight distinct states in this set; and let \( F' \) be the algebra of 16 events formed as the closure under union and complementation of the "generating set" of four events \([(0,2),(1,3),(4,6),(5,7)]\). Thus, \( F' = \{(\text{null set}), (0,2),(1,3),(4,6), (5,7),(0,1,2,3),(0,2,4,6),(0,2,5,7),(1,3,4,6),(1,3,5,7),(4,5,6,7),(0,1,2,3,4,6),(0,1,2,3,5,7),(0,2,4,5,6,7),(1,3,4,5,6,7), (0,1,2,3,4,5,6,7) = S\}.

Finally, let \( F'' \) be the algebra generated by \([(0,2,4,6),(1,3,5,7)]\), i.e., \( F'' = \{(\text{null set}), (0,2,4,6),(1,3,5,7),S\} \).

We can interpret \( F'' \) as the set of events of interest to a decision maker (perhaps the individual himself) who can eventually observe \( r \), but to whom \( e \) and \( d \) are forever unobservable. Namely, \((1,3,5,7)\) is the event that the individual develops the health response (e.g., cancer) before dying from some other cause, while \((0,2,4,6)\) is the event of no response. Similarly, \( F' \) is the algebra of events that might be of interest to a d.m. who can (eventually) observe \( e \) and \( r \), but not \( d \).

Finally, \( F \) is the algebra of events for a d.m. who can observe \( e, d, \) and \( r \).

An alternative interpretation is that \( F'', F', \) and \( F \) represent three successive algebras available to a d.m. who becomes able to observe more factors (first \( r \) alone, then \( e \) and \( r \), and finally \( e, r, \) and \( d \)) as time passes. Each algebra is "finer" than the preceding one, in the sense that it contains all of the same events and additional ones besides. The sequence \((F'',F',F)\) is also called an "increasing sequence of algebras" or a "filtration".

Now suppose that a d.m. is interested in the probability \( p(1,5) = p(1) + p(5) = p(0,0,1) + p(1,0,1) \), i.e., the probability that a particular individual, perhaps a plaintiff, will show (or would have shown) the health effect if he has (or had had) no enzyme deficiency. Notice that the event \((1,5)\) is not in \( F' \) or in \( F'' \). This means that there is no way of computing \( p(1,5) \) with respect to either of the algebras \( F' \) or \( F'' \). The event \((1,5)\) is not "measurable" with respect to these algebras. More generally, we say that an event \( E \) is "measurable" with respect to an
algebra \( F \), or "\( F \)-measurable", if and only if \( E \) is one of the events in \( F \).

Let \( F \) be an arbitrary algebra on a set \( S \) of states, and let \( G = [E_1, E_2, \ldots, E_r] \) be a set of subsets of \( S \) such that (i) The closure of \( G \) under union and complementation is \( F \); (ii) No proper nonnull subset of any of the events \( E_1, \ldots, E_r \) in \( G \) is \( F \)-measurable; and (iii) Each of the events \( E_1, \ldots, E_r \) is \( F \)-measurable. Then \( G \) must be a partition of \( S \). For the union of its members is \( S \), and the intersection of any two of them, say \( E \) and \( E' \), is \( E \cap E' = -(E \cup -E') \), which must be in \( F \) (since \( F \) is closed under set unions and complementations), and which can not be a proper nonnull subset of either \( E \) or \( E' \) (since by hypothesis no such subset is \( F \)-measurable); the only remaining possibility is that \( E \cap E' \) is the null set. We will call \( G \) the "generating partition" for \( F \). It is easy to see that any algebra on a finite state set \( S \) has a unique generating partition, which we will denote by \( G(F) \).

We can now define a probability measure with respect to an arbitrary algebra \( F \) on \( S \) as a function \( p(.) \) mapping events in \( F \) into nonnegative numbers in such a way that

\[ P1': \text{ For any two events in the generating partition } G(F) \text{ of } F \text{ (and hence for any two disjoint events in } F \text{), say } E \text{ and } E', p(E \cup E') = p(E) + p(E'); \text{ and} \]

\[ P2': p(B) = 1, \text{ where } B \text{ is the smallest } F \text{-measurable subset of states such that it is known that } s \text{ is in } B. \text{ For any state } x \text{ outside } B, p(x) = 0. \]

This is a direct generalization of the definition for the case of complete resolution, except that the role of the states in \( S \) is now played by the blocks in \( G(F) \). The pair \([F, p(.)] \) is called a probability algebra.

The probability of any arbitrary \( F \)-measurable event \( E \) can be calculated very simply from the probabilities of the blocks in \( G(F) \). Since every nonnull event in \( F \) is a union of blocks of \( G(F) \), \( p(E) \) is the sum of the
probabilities of those blocks in $G(F)$ that are contained in $G$. In this sense, any probability measure $p(.)$ on $S$ is uniquely characterized by the probabilities of the blocks in some partition of $S$, and every such partition acts as the generating partition (under union and complementation) of an algebra $F$ on $S$.

Example 2.4: Marginal Probabilities and Cell Probabilities

Suppose that each individual in a population can be classified according to two characteristics, e.g., as a smoker vs. a nonsmoker and as a drinker vs. a nondrinker. Suppose also that it is known that the probability that a randomly selected individual will be a smoker is 50%, while the probability that he will be a drinker is 40%. Can we conclude anything at all about the probability that a randomly selected subject, $s$, will be both a smoker and a drinker?

The answer is that we can conclude only that this probability is no greater than the smaller of the marginal probabilities that it contributes to, namely $\min(0.4, 0.5) = 0.4$ in this example. The probability of "s is a drinker and smoker" is not measurable with respect to the algebra generated by the events [$s$ is a smoker, $s$ is a drinker], for which we have probabilities. On the other hand, this does not mean that we know nothing about the probability of a drinking smoker: we know that it is between 0 and 0.4. Thus, from probabilities about subsets of $S$ other than the blocks in a partition (e.g., the possibly overlapping subsets of smokers and drinkers) it is possible to infer partial information about the probabilities of other events. To get a complete probability measure for some algebra of events, however, with point-valued rather than interval-valued probabilities, it is necessary to start with the probabilities of each of the blocks in some partition.

In summary, we can define a finite probability space in terms of a triple $(S, G, p)$, where $S$ is a finite set of states, $G$ is a partition of $S$, and $p(.)$ is a function assigning to each block of states in $G$ a number, say $p(E_i)$ for block $E_i$, such that the sum of these numbers is one. An entire probability algebra and associated probability measure that extend $G$ and
p(.) to the set of all subsets consisting of unions of blocks in $G$ can then be built up using rules P1' and P2'. Using $(S,G,p)$ rather than $(S,F,p')$ (where $F$ is the closure of $G$ under union and complementation and $p'$ is the extension of $p$ to $F$) to define a probability space seems to have advantages of simplicity, and we will use this formulation henceforth.

2.2.3 Formation of Beliefs

A. The Data Generating Model

If an individual's beliefs are assumed to be representable by a partition $G$ and an initial assignment of probabilities to the blocks of $G$, the question of where these initial probabilities come from must be confronted. As for knowledge, we will assume that beliefs are based on observations; in Chapter 6, we will allow communication with someone else to be another possible source of beliefs. In addition to observation, however, classification plays an essential role in the formation of beliefs. For example, the individual may classify the situation that he confronts as a member of a class of "similar" situations that he has faced in the past, and develop probabilistic predictions about the consequences of each act $a$ in his choice set $A$ on the basis of this classification. Thus, formation of individual beliefs about the probable consequences of acts may depend on gestalt pattern recognition and perceptions.

Although we are unable to provide a reasonable general model of belief formation, we can take some useful initial steps by concentrating on the role of observation rather than classification. (See, however, Hintikka and Suppes, 1970.) One simple model is the following.

Each member $i$ in some "population" (e.g., of people, industrial facilities, laboratory animals, trials of an experiment, observations in a time series, etc.) is thought of as having a true underlying "type", $t(i)$, which determines the results of any possible observation on member $i$. We will call each member of the population a "case". Two types of observation on case $i$ are of interest: a passive observation, $y(i)$, which
summarizes everything that the d.m. knows about case i at the time that he chooses an act from A; and a subsequent observation r(i;a) summarizing everything the d.m. knows about i after choosing act a (including the fact that a was chosen.) Each of these observations consists of a realization of one among several distinguishable (hence observable) outcomes. Thus, the passive observation y(i) is drawn from a set \( Y = \{Y_1, Y_2, \ldots, Y_m\} \) of possible outcomes or "signals", while the observation r(i;a) is drawn from a set \( R = \{r_1, r_2, \ldots, r_n\} \) of possible outcomes or "responses". The signal y(i) observed for case i at the time that a is chosen may also be called the "observed type" of case i. Given that the true type of case i is t(i), y(i) and r(i;a) are in principle given by deterministic functions, y(i) = g[t(i)] and r(i;a) = c[a,t(i)].

In general, there may be many types that map onto a single observed signal y(i) before act a is chosen, but that map onto different observed responses after act a has been chosen. In this case, the response r(i;a) of a case that is of observed type y(i) will be stochastic for an observer (the d.m.) whose only information is y(i). Let T be the set of all possible true types, and let G be the partition of T induced by the passive observation mapping g. (For simplicity, we assume that the act of observation (which is not an act in the choice set A) leaves the type of the observed case unchanged. However, the key ideas to be discussed can be extended to the case where observations lead to changes in type by allowing current type to index possible relations between past history and future responses.) Similarly, let H(a) be the partition of T induced by c(a,·). Thus, two types are in the same block of G if and only if they map under g onto the same observed type; and they belong to the same block of H(a) if and only if they map under c(a,·) onto the same response after a has been chosen. Ideally, the d.m. would like to use his observation y(i) to predict which block of H(a), identified by r(i;a), he will observe if he chooses act a. From this standpoint, he can think of his choice among acts as a choice among probable future observations.

Perhaps the easiest way to model the inference process is to assume that there is a true (but unknown) distribution of cases over "types". Case
i's type, \( t(i) \), can often be conveniently thought of as a vector with many components: \( t(i) = [t_1(i), t_2(i), \ldots, t_r(i)] \), where each component represents the level of some attribute of case \( i \) that helps distinguish it, at least conceptually, from some other cases in some conceivable situations. (The components of \( t(i) \) do not have to be numerical attributes, but this is the simplest interpretation.) Components of \( t(i) \) may be dated to reflect case \( i \)'s characteristics at different times. They may include information about \( i \)'s exposure history or about anything else that could lead \( i \)'s response in some situations and along some dimensions to differ from the corresponding responses of other cases. If all components of \( t(i) \) were known, then \( i \)'s response to any act \( a \) would be known as well.

The possible types, or \( r \)-tuples, may be thought of as cells in an \( r \)-dimensional contingency table. A fundamental assumption supporting statistical inference is that there is a true, though perhaps unknown, distribution of potential cases over the cells in this table. Even if there are only a small number of actual cases available for observation, in principle there are an arbitrarily large number of cases, distributed among the cells in some fixed proportions. (The proportions are necessarily fixed because all relevant aspects of time passage are captured in the \( r \) dimensions that specify a cell.) This is sometimes called a data generating process or model for whatever real cases and observations may be made. Since cases are in principle distributed among cells according to some frequency count, such a model provides intuitive justification for the assumption of additive probability measures (see Example 2.3.)

At the time he chooses \( a \) from \( A \), the d.m. will have observed some components of some cases. In effect, he will have sampled from multivariate marginal distributions in which sample cases drawn from the full distribution of types have been projected onto the subspace of components observable by him as of the decision date. (Whether observation error is incorporated into the definitions of "types" or treated as part of the observed distribution over types is a
matter of taste.) Based on his sample of observed cases from this projected distribution, on assumptions or models of the full joint distribution, and on the observed projection of \( t(i) \) onto the subspace of components that he has observed for case \( i \) at the decision date, the d.m. will try to estimate the projection of \( t(i) \) onto the larger subspace of components that will be observable to him after his decision has been made.

Equivalently, let \( y(i) \) be the block of \( G \) to which case \( i \) belongs, and let \( y(i)^H(a) \) be the partition consisting of blocks that case \( i \) may turn out to lie in once \( a \) has been chosen and \( r(i;a) \) has been learned. Then the d.m.'s inference problem is to estimate the probabilities of the blocks of \( y(i)^H(a) \), given \( y(i) \) and whatever he has learned or assumed about the data generating process distribution of cases over types (cells) on the basis of his sample of previously observed cases.

Since the full joint distribution of type components is in principle underdetermined by the marginal distributions that he has observed (Example 2.4), it will in general be necessary for him to make some assumptions, e.g., that case distributions across different attributes are statistically independent. A particularly strong but useful assumption is that the true prior distribution of types, i.e., the proportion of the population falling in each cell, is known. Let \( p(t) \) denote the prior probability that a randomly drawn case, \( i \), is of type \( t \). Then the prior probability that the d.m. will observe \( y(i) = y \) is just the sum over all types in \( y \) (i.e., over all cells that project onto \( y \) at the decision date) of the values \( p(t) \). Equivalently,

\[
p(y) = p(y;1)p(1) + \ldots + p(y;n)p(n),
\]

where \( p(t) \) is the theoretical proportion of cases that fall in cell \( t \) ( = the prior probability of type \( t \)) for \( t = 1,2,\ldots,N \); \( p(y;t) \) is the conditional probability (0 or 1) that signal \( y \) will be observed if the true type is \( t \); and it is assumed for mathematical simplicity that there is a finite number, \( N \), of types.
Conversely, if signal \( y(i) = y \) is observed, then the posterior probability that \( t(i) = t \) is given by

\[
p'(t) = p(t; y) = \frac{p(y; t)}{p(y)} p(t) \quad \text{for } t = 1, \ldots, n \text{ and } y = 1, \ldots, m.
\]

Since each \( p(y; t) \) is either a zero or a one, this posterior probability is just \( p'(t) = p(t)/p(y) \) for every type in block \( y \) of \( G \), and is equal to zero for all other types.

When he chooses \( a \), the d.m. believes that the distribution over types is \( p'(t) \). Hence, he believes that the probability of observing response \( r(i;a) = r \), a block of \( y(i)^r \) when \( y(i)^H(a) \), if he chooses act \( a \) is given by

\[
p(r; y(i), a) = \frac{p(r, y(i); a)}{p(y(i))} = p'(r, y(i); a)
\]

i.e., the sum over all types \( t \) that project onto (or are in) \( y(i)^r \) when act \( a \) is chosen, of \( p'(t) \). The blocks \( y(i)^r \) in \( y(i)^H(a) \) may be interpreted as the states of nature that the d.m. knows are possible when he chooses \( a \), and \( y(i)^r(i;a) \) is the state of nature that turns out to be true after his choice is made. In any particular decision problem involving only one case, we will use \( S \) to denote the set of possible states, \( x \) to denote a generic member of \( S \), and \( s \) to denote the true state. Thus, "states" in this terminology correspond to subsets of (at least potentially) distinguishable types at the time a decision must be made.

In this setting, a d.m. who chooses act \( a \) will be forever unable to distinguish among types that belong to the same block of \( G^H(a) \), i.e., the same state. This raises the possibility of a tradeoff between the consequences of a decision and knowledge of its consequences.

Example 2.5: True vs. Known Consequences

Which of the two following risk control measures would you prefer?
al: A measure that is equally likely to save zero or 1000 lives (depending on whether the underlying hypothesis is correct) but whose effect you will never learn; or

a2: A measure that is equally likely to save zero or 900 lives, and whose effect (actual number of lives saved) you will learn within a year of implementation?

Such choices can arise in public health contexts, where even substantial public health effects may be undetectable or may not be detectable for many years. Tradeoffs between the desirability of probability distributions over consequences and the desirability of the timing of uncertainty resolution are considered more systematically in Chapter 7.

What is important to notice here is that a d.m. may care not only about the probabilities of different future observations, given his current choice, but also about unobservable components of the true type, or about what he would observe if he had perfect information.

B. Coherence, Calibration, and Sufficient Statistics

Let $s_1, s_2, \ldots, s_n$ be the set of possible states that are consistent with $y$, the d.m.'s observations as of the decision date. From $y$ and $p(x)$, his prior probability measure over states $x$ in $S$, the d.m. can compute his posterior probability for any $x$ in $y$ as $p'(x) = \frac{p(y;x)p(x)}{p(y)}$, as described above. Doing so makes his beliefs "coherent", in the sense that they are represented and updated in accord with the assumptions of probability theory. However, it does not make them correct. In this section, we examine the concept of "correctness" of probabilities.

In discussing individual knowledge, one of our defining assumptions was that $K(E)$ implies $E$, i.e., if something is known to be true, then it is in fact true. An analogous requirement for individual beliefs might seem to be that if the d.m believes that the probability of event $E$ is $p(E)$, then the true probability that $s$ is in $E$ is $p(E)$. But the true probability (for someone with perfect information) is either 0 or 1, so this formulation is inadequate.
To clarify the concept of "correct" probabilistic beliefs, we must go beyond the setting of a single one-shot decision and accept the convenient fiction of indefinitely repeated trials and multiple observations. Our approach will be to examine the acquisition of individual beliefs by observations through a noisy channel, in rough analogy to our model of acquisition of individual knowledge by observations through a deterministic channel.

Let $C$ be a matrix with $C(y,x) = p(y;x)$, the conditional probability of observing $y$ if the true state is $x$, in row $y$ and column $x$, $x = 1,\ldots,n$; $y = 1,2,\ldots,m$. (If $i$ is a case selected at random according to the data generating process, then $C(y,x)$ is short for $p[y(i) = y; s(i) = x]$, where $s(i)$ is the true state in case $i$.) We will call $C$ the "channel" mapping true states into probability distributions over observed signals. When the true state is $x$, the probability distribution over signals $y$ in $Y$ is given by column $x$ of $C$. Let $D$ be the set of all $n$-component probability distribution vectors and let $p$, the prior probability vector, be a point in $D$. (Each point $p$ in $D$ extends to a unique corresponding probability measure on the algebra of all subsets of states, as previously explained; it is convenient, however, to deal with the generating vector rather than the whole measure.) If the d.m. observes signal $y$ through channel $C$ and is a coherent Bayesian, then he revises his previous beliefs according to the rules

$$p'(x) = C(y,x)p(x)/p(y),$$

$$p(y) = p(y;1)p(1) + \ldots + p(y;n)p(n),$$

$$q = Cp$$

in matrix notation, where $q$ is the $m$-dimensional column vector of signal probabilities, having components $p(y)$; $p$ is the $n$-dimensional column vector of state probabilities, having components $p(x)$; and $C$ is the channel matrix.
Now imagine the following sequence of experiments or trials. At the beginning of each trial, s is drawn at random so that s = x with probability p(x), for x = 1,...,n. (The vector p specifies a data generating model.) Then the d.m. observes s repeatedly through channel C, updating his beliefs according to the difference equations

\[ p(x; t+1) = \frac{C(y(t), x)/C(y(t))p(x; t)}{q(t)}p(x; t) \]

\[ q(t) = Cp(t), \]

where y(t) is the signal observed on observation t; C(y(t)) is the y(t)th row of C; q(t) is the d.m.'s probabilistic forecast of what signal he expects to see on his next observation; and it is assumed that the conditional probabilities in C remain fixed. If the columns of C are distinct, and hence linearly independent, then q(t) will converge to one of the columns of C as t goes to infinity, and p(t) will converge to the corresponding vertex of the simplex D. Thus, the d.m. will (with probability 1) eventually learn the true state, s. However, this may take many hundreds of trials, especially if two columns of C are close to each other. We shall suppose that a trial ends when p(t) gets "sufficiently close" to a vertex of D.

Let y*(t) = (y(0), y(1),..., y(t)) denote the whole history of the d.m.'s observations up through observation t, and let p(t+1) = p'(t) be his beliefs about s after the t-th observation and before the t+1st observation. According to the above difference equations, p(t+1), and hence q(t+1), depend on the history y*(t) only through p(t) and y(t): the rest of the trajectory is irrelevant. Thus, p(t+1), the probability vector forecast for period t+1 on the basis of observations made through period t, is a "sufficient statistic" for the d.m.'s observations up through period t. (Technically, if we look at the inference problem in terms of trying to identify which column of C is generating the observed sequence of signals, then the likelihood of any sequence of observations y* given s = x, which we will denote by L(y*; x), is a multinomial term depending only on the numbers of each signal in y*.
ratios \( L(y^*;x)/L(y^*;x') \) for each pair of states \( x \) and \( x' \) depend only on these numbers, which are thus "sufficient" statistics for inference. And the corresponding odds ratios, \( p(x;y^*(t))/p(x';y^*(t)) = L[y^*(t);x]p(x)/L[y^*(t);x']p(x') \), which determine the probability vector \( p(t+1) \), indeed depend only on \( y(t) \) and the \( L[y^*(t-1);x]p(x)/L[y^*(t-1);x']p(x') \), and hence on \( y(t) \) and \( p(t) \).

We can now define a more satisfactory concept of "correctness" for probability judgements. Define \( Y^*(p') \) as the set of all trajectories that pass through point \( p' \) in \( D \) in an infinite sequence of repeated trials of the sort just described. A d.m. will be called sequentially well calibrated if the fraction of trajectories in \( Y^*(p') \) that eventually end at vertex \( x \) is exactly \( p'(x) \), for every probability vector \( p' \) in \( D \) that is reachable (with nonzero probability) from \( p \) by some sequence of observations. If a d.m. is not sequentially well-calibrated, then some of his probability judgements will not be accurate.

This concept is a strengthening of the usual concept of "calibration" found in the current statistics literature (Dawid, 1982, 1985; Oakes, 1985; Schervish, 1985) in which the true state (vertex of \( D \)) is disclosed after a single observation in each of an infinite number of trials. Even without a formal analysis, several points seem clear:

- In contrast to the simple axiom \( K(E) \) implies \( E \) for individual knowledge, there is no simple, finitely testable, criterion for assessing the "correctness" of subjective probabilities.

- After any sequence \( y^* \) of observations, a coherent Bayesian d.m. expects his own beliefs to be sequentially well-calibrated. Also, \( E[p(t+k);y^*(t)] = p(t+1) \) for all \( k \) (the martingale property for probabilistic beliefs).

- If the probabilities \( p(t) \) used to describe a d.m.'s beliefs are rounded off to a finite number (e.g., three) of significant digits, as must surely be the case in practice, then with finite probability the d.m. will not be sequentially well-calibrated.
Although careful study of these and related properties of sequential calibration would take us too far afield, the type of analysis involved can be quickly outlined for the special case where the number of possible states is \( n = 2 \) as is traditionally assumed in the calibration literature. Let the two possible states be indexed by \( x_1 \) and \( x_2 \), and let there be two possible signals, \( y_1 \) and \( y_2 \), in \( Y \). The channel \( C \) is specified by two numbers, \( a = \text{pr}[y(t) = y_1; s = x_1] \) and \( b = \text{pr}[y(t) = y_2; s = x_2] \). Let \( p[s = x_i; y^*(t)] \) be denoted by \( p_i(t) \). It turns out to be convenient to use \( w(t) = \ln[p_1(t)/(1-p_1(t))] \), i.e., the logarithmic odds in favor of \( x_1 \), in place of the probability of \( x_1 \), to characterize the d.m.'s beliefs after \( t \) observations. Then beliefs evolve according to the simple stochastic difference equation

\[
w(t+1) = w(t) + v(t)
\]

where \( v(t) = \ln[a/(1-b)] = c \) if \( y(t) = y_1 \); \( v(t) = \ln[(1-a)/b] = d \) if \( y(t) = y_2 \).

Interestingly, the d.m. will necessarily be "wrong" in his beliefs about the p.d.f. of \( v(t) \), since his subjective \( q(t) \) is a convex combination of the columns in \( C \) (with weights, given by the components of \( p(t) \), that depend only on the current \( w(t) \)), while the true \( q(t) \) is one of the columns of \( C \). Assuming linearly independent columns, therefore, his subjective \( q(t) = Cp(t) \) (which he expects to be well calibrated) must differ from the true \( q(t) \), which determines the true p.d.f. of the increment \( v(t) \). Thus, the process \( \{w(t)\} \) is modelled by the d.m. as evolving according to \( w(t+1) = w(t) + v(t) \) where the distribution of \( v(t) \), namely \( q(w(t)) = Cp(t) = C[p_1(t),p_2(t)]' = C[w(t),1]'/(1+w(t)) \) depends only on \( w(t) \); while in fact \( v(t) \) is i.i.d. independent of \( t \) or \( w(t) \), and is determined only by \( s \). The subjective probability vector \( q(t) = Cp(t) \), evolves so that \( E[w(t+1);y^*(t)] = w(t) \), i.e., \( E[V(t)] = 0 \) at any time until \( v(t) \) has been revealed, but the "true" average value for \( v(t) \) is either \( a[\ln(a/(1-b))] + (1-a)[\ln((1-a)/b)] \) or \( (1-b)[\ln(a/(1-b))] + b[\ln((1-a)/b)] \). Similar remarks apply to the evolution of \( q(t) \) and \( p(t) \)
even when \( n \) is greater than 2; for example, the expected displacement in the next period, based on all information available now, is 
\[
\mathbb{E}[p(t+1) - p(t)] = 0.
\]

The process
\[
w(t+1) = w(t) + v(t)
\]

where the \( v(t) \) are i.i.d. increments taking values \( c \) and \( d \) with probabilities \( C(1,s) \) and \( C(2,s) \), respectively, is a one-dimensional random walk with unequal step lengths to left and right. Exact analysis of its behavior is not easy. However, asymptotically, the sum of the increments will be approximately normally distributed with mean and variance proportional to the number of observations, \( t \), taken. Calculating the mean and variance of the step size \( v(t) \) for each column of \( C \) and transforming the \( w(t) \) scale back to a probability scale via the transformation \( p(t) = w(t) / (1 + w(t)) \) allows us to estimate the minimum probable and maximum probable number of observations through \( C \) that would be required to drive \( p(t) \) to within a specified (very small) interval of 0 or 1.

2.2.4 Comparison of Noisy Information Channels

In the special case of deterministic channels, we were able to compare the "informativeness" of channels in terms of the fineness of the induced partitions of \( S \). A clever generalization of this idea due to Blackwell (1951), applies to noisy channels. If \( C \) and \( C' \) are two channels and \( C' = CM \) where \( M \) is a nonnegative matrix with columns summing to one (i.e., a "stochastic matrix"), then it can be shown that \( C \) dominates \( C' \) in several strong senses. (For example, every state-contingent expected payoff vector achievable with an observation through \( C' \) followed by a choice of act from \( A \) is also achievable with an observation through \( C \) followed by a choice of act from \( A \); see McGuire, 1971, Marschak and Miyasawa, 1968.) If \( C \) and \( C' \) are deterministic and \( C' \) is a coarsening of \( C \), then premultiplying the deterministic state vector (with a 1 in row \( s \)) by \( M \).
has the effect of erasing some distinctions between blocks of the partition induced on $S$ by $C$ by mapping all states in two or more blocks into the same signal. If $M$ is stochastic, then $CM$ can be thought of as a combined channel in which the true state is first mapped stochastically into a signal in the signal set of $M$, and then this "garbled" signal is fed through the original channel $C$. Clearly, then $CM$ will be no more informative than $C$ about $s$. However, it also turns out that the converse is true. For example, it is the case that every classical d.m., regardless of his preferences for consequences or his beliefs about $s$, will prefer an observation of $s$ through $C$ to an observation of $s$ through $C'$ if and only if $C' = CM$ for some stochastic matrix $M$. (See e.g., McGuire, op cit.)

A final, novel interpretation of the comparative informativeness partial ordering among channels is in terms of the average absolute step size in the $p(t)$ trajectory when $s$ is repeatedly observed through the channel. Any $m \times n$ channel $C$ may be interpreted as a stochastic correspondence mapping points (probability $n$-vectors) in $D$ into probability distributions over $m$-tuples of points in $D$. The $m$ points in the range of a point $p'$, corresponding to the $m$ possible signals that might be observed next, define a subset of $D$ by their convex closure. Roughly speaking, the size of this set reflects the informativeness of the channel. The effect of postmultiplying $C$ by $M$ is to move the columns of $C$ (each of which is a probability $m$-vector) closer together by averaging them, making them more difficult to distinguish and reducing the size of the set spanned by the $m$ points in $D$ in the range of $p'$. When $n = 2$, the process

$$w(t+1) = w(t) + v(t)$$

where $v(t) = \ln[a/(1-b)]$ or $\ln[(1-a)/b]$ adjusts more slowly because the new values of $a$ and $1-b$, say $a'$ and $1-b'$, in $CM$ will be between the original values of $a$ and $1-b$ in $C$, implying that their ratio $a'/(1-b')$ is closer to 1 than was $a/(1-b)$, and hence
that both possible values of v(t) are closer to zero than they were. Intuitively, pregarbling the state signal by passing it through M before it gets to C reduces the "power" of the signal finally received, e.g., as measured by the variance of v(t), and makes it less informative.

2.3 Summary and Prospectus

This chapter has surveyed a number of technical models for formally representing and updating individual knowledge and beliefs in light of deterministic and noisy observations. Of course, there are still a number of unsolved foundational problems, such as where knowledge of channel structures or initial beliefs come from. However, these types of questions suggest the sort of infinite epistemological regress that typically frustrates foundational work, and are in any case not particular to risk measurement. A somewhat more severe difficulty for risk assessment has to do with the fact that assessed probabilities, which are basic to all aspects of risk analysis, may be expected to follow miscalibrated stochastic trajectories driven by successive observations, as in the simple model discussed above. However, there seems to be no easy solution to the problem of miscalibrated probabilities. In any cases, the models of individual knowledge and beliefs presented in this chapter, namely, information partitions and probability algebras updated by observations taken through channels with known characteristics, are sufficiently simple and useful to serve as a basis for our subsequent analysis of risk measurement and communication among observers who are able to use deterministic or noisy channels to gain information (knowledge or beliefs) about the probable consequences of actions.

As discussed in Section 2.1, the models of this chapter find their primary application in chapters 6 to 8. In Chapter 6, information partitions are used to describe the knowledge conveyed by probability messages exchanged among observers with a common prior probability algebra. In addition, the problem of separating knowledge (represented by a sequence of observations) from beliefs (which depend on the prior probability algebra) for communication among observers with different
priors is considered. Chapters 7 and 8 apply the ideas of information partitions and probability algebras to measurement of individual fatality risks over time and to measurement of population risk, respectively.

The technical material in this chapter has dealt with one of the two central components of normative decision making, namely, the information on which decisions are based. Chapters 3 to 5 consider the other component, namely the d.m.'s preferences for acts, given fixed information about their probable consequences. Chapters 6 to 8 attempt to synthesize these components by presenting normative theories and representations (via risk numbers) of preferences in contexts where information changes over time.
The models of individual knowledge and beliefs discussed in Chapter 2 provide the essential conceptual background for studying communication between a risk analyst (r.a.) and a decision maker (d.m.). We are interested in communication for the purpose of improving risk management decision making, however. It is therefore necessary to understand how the d.m. will use whatever information he has to make his decision. In this chapter, we outline normative models of decision making in the "classical" case where (i) No further opportunities for information collection are available before the decision must be made (perhaps because the d.m. expects the loss from waiting longer to exceed the loss from acting without additional information); and (ii) All relevant uncertainty will be resolved as soon as the d.m. has made his choice. In this chapter, we do not distinguish between types and states: everything the d.m. cares about is assumed to be revealed in the consequence after his choice of a from A is made.

Because the models in this chapter are known in the literature, the presentation here will be brief and no proofs will be given. The results collected here will be used in Chapter 4 as a basis for risk measurement; they also provide a model for the development in Chapters 5, 7, and 8 of normative decision theories in nonclassical settings (e.g., when the consequence of an act is not measurable with respect to the generating partition of the d.m.'s probability measure.) Thus, this chapter is primarily a review and synthesis of current theory in normative decision analysis, and may be skimmed quickly by readers familiar with this literature.

3.1 Decision Model 1: The Expected Utility Model (Fishburn, 1970; Roberts, 1979)

A. The Model

Let A be the set of acts, S the set of states, p(x) the generating set of probabilities over the atoms (states) in S, and C the set of possible consequences or outcomes of the decision. A, S, and hence C are assumed
for mathematical simplicity to be finite. The consequence when the d.m. chooses act a from A will be c(a,s); the d.m. treats it as a random variable or "lottery", which we will denote by l(a), taking the value c(a,x) in C with probability p(x) for each x in S. Hence, choosing among acts amounts to choosing among random variables l(a).

Let L = \{l(a): a is in A\} be the set of all lotteries (or probability measures over C) that the d.m. must choose among. Assume that he has well-defined (complete, transitive) preferences over the consequences in C. For any two consequences c and c' in C, the notation c \geq c' will here mean that the d.m. prefers c at least as much as c'. (Later, this notation will be used to denote a strict preference relation on L.) To choose from L, however, the d.m. needs to deduce or derive preferences for the members of L from these primitive preferences for the members of C. To this end, embed L in the larger set D of all probability distributions (or measures) over C. Assume that D includes each c in C as a special "degenerate" lottery. We seek to construct or derive a preference ordering R over the members of D that will agree with P on C, i.e., an ordering R such that c \geq c' if and only if c \geq c', where c and c' are considered as members of D in the first comparison and as members of C in the second. Then applying R to D will order the elements l(a) of L, and allow a most-preferred act to be selected.

Accordingly, let R denote a binary ordering on D, to be interpreted as the derived preference ordering. The following axioms are traditionally assumed for R.

R1. Weak Ordering: (D,R) is a weak ordering, i.e., any two distributions in D are comparable by R, and R is transitive on D.

R2. Continuity: Let Q be the relation of strict preference defined by a \geq b if and only if a \geq b and not b \geq a. If x, y, and z are any three distributions in D such that x \geq y \geq z, then there are probabilities p and q in (0,1) such that (x p z) \geq y \geq (y q z).
Note on Mixture Notation: Here and elsewhere we use the notation \((1 \ p \ 1')\) to denote the lottery that gives 1 with probability \(p\) and 1' with probability \(1-p\). This notation takes some getting used to, but proves to be very convenient to work with. We call \((1 \ p \ 1')\) a **p-MIXTURE** of 1 and 1'. Formally, we can abandon the interpretation of \((1 \ p \ 1')\) as a lottery, and axiomatize "mixture spaces" directly in terms of the properties

\[
(a \ p \ b) = (b, 1-p, a);
\]

\[
(a \ p \ a) = a; \text{ and}
\]

\[
[(a \ p \ b) \ q \ b] = (a \ pq \ b).
\]

However, there is no advantage for our purposes in proceeding with this more abstract approach (Roberts, 1979.)

R3. (Substitution): Let \(x, y,\) and \(z\) be any three distributions in \(D\). Then \(x \ Q \ y\) if and only if \((x \ p \ z) \ Q \ (y \ p \ z),\) for any \(p\) in \((0,1)\). This is also sometimes called the "monotonicity" or the "independence" axiom in the mathematical psychology and economics literatures, respectively.

**THEOREM 3.1:** The above axioms of weak ordering (W.O.), continuity (C) and independence (I) hold on the set \(D\) of all lotteries over \(C\) if and only if there is an assignment of numbers \(u(c)\) to consequences \(c\) in \(C\) such that for any two lotteries \(l\) and \(l'\) in \(D,\)

\(l \ Q \ l'\) if and only if \(E(u;l)\) is greater than \(E(u;l'),\)

where \(E(u,l)\) denotes the expected value of \(u(c)\) when \(c\) is chosen from \(C\) at random according to the distribution \(l\).

**THEOREM 3.2** (Raiffa, 1968): Given W.O., C, and I, suppose also that

- There are "best" and "worst" consequences, \(c'\) and \(c''\) in \(C\) (i.e., for any \(c\) in \(C, c' \ P \ c \ P \ c''.\)
For each \( c \) in \( C \), there is some \( p \) in \([0,1]\) such that \( c \) is indifferent to \((c' p c'')\). Call this value of \( p \) \( u^*(c) \).

Then

(i) For any two consequences \( a \) and \( b \) in \( C \), \( a \succ b \) and not \((b \succ a)\), i.e., \( a \) is strictly preferred to \( b \) in the primitive preference ordering, if and only if \( u^*(a) \) is greater than \( u^*(b) \).

(ii) For any two lotteries \( l \) and \( l' \) in \( D \), \( l \gtrless l' \) if and only if \( E(u^*; l) \) is greater than \( E(u^*; l') \), and \( l \sim l' \) if and only if \( E(u^*; l) \) is at least as great as \( E(u^*; l') \).

(iii) If \( u \) is any other function assigning numbers to the consequences in \( C \) in such a way that (ii) holds, then \( u(.) = au^*(.) + b \) for some positive \( a \). Conversely, any such \( u(.) \) will satisfy (i) and (ii).

Thus, we have a way of assigning numbers (interpretable as "indifference probabilities") \( u^*(c) \) to consequences in \( c \) in such a way that both primitive preferences over \( C \) and derived preferences over \( L \), as characterized in Theorem 3.1, are represented by the expectation of the resulting numerical random variable. The indifference probabilities \( u^*(c) \) for \( c \) in \( C \) are called (canonical) "utilities", and the representation of \( R \) in Theorem 3.2 is known as the "expected utility" representation of preferences over lotteries.

A d.m. for whom these axioms hold can choose from \( A \) by ordering the lotteries \( l(a) \) in \( L \) according to \( R \), i.e., according to expected utility, and then choosing an act, say \( a^* \), corresponding to a most preferred member of \( L \), i.e., maximizing expected utility.

B. Critique

Many empirical experiments have shown that the expected utility model does not adequately describe actual individual choices.
Example 3.1: Incoherent Choice (Roberts, 1979)

We define three lotteries over monetary consequences as follows:

\[ L_1 = (-100, .5, 0), \text{ i.e., a 50-50 chance at losing $100 or nothing;} \]

\[ L_2 = (-45) = (-45 \ p \ -45), \text{ i.e., a sure loss of $45;} \]

\[ L_3 = (-45)^{-50,.5,45}, \text{ i.e., a sure loss of $45, followed by an unfavorable lottery giving a 50-50 chance at winning $45 or losing $50.} \]

Many people express the preferences

\[ L_1 \succeq L_2, \text{ given a choice between } L_1 \text{ and } L_2; \]
\[ L_2 \succeq L_3, \text{ given a choice between } L_2 \text{ and } L_3. \]

That is, they prefer the 50-50 chance of losing $100 or nothing to a sure loss of $45, and prefer a sure loss of $45 to a compound lottery in which a sure loss of $45 is followed by a 50-50 chance of losing an additional $50 or of winning back the lost $45. But the compound lottery is equivalent, in traditional theory, to \((0,.5,-95)\), since in traditional theory only the probability distribution over final (monetary) consequences and not the way in which it is resolved is considered. (This is sometimes called the "axiom of complexity" for lotteries over consequences. We shall argue in Chapter 7 that it has more to do with the definition of the relevant consequence set, \(C\), than with the processing of lotteries over consequences.) And \((0,.5,-95)\) is clearly preferable to \(L_1 = (0,.5,-100)\), since it gives a higher probability of the more preferred outcome. (That this implies the stated preference is sometimes called the "unequal probability" axiom. We prefer the term "stochastic dominance." Thus, transitivity of \(\succeq\) appears to be violated. (Roberts, 1979, attributes this example to Raiffa.)
More generally, all of the axioms proposed above have been shown to be suspect in at least some potentially important choice situations. For example, the preference relation $P$ over consequences may not be complete (i.e., some consequences may be difficult or impossible to compare) or may violate transitivity. Preferences tend to exhibit discontinuities at $p = 1$ (or $p = 0$). And the independence axiom appears to be violated in many examples. (See e.g., Machina, 1983 for a more systematic survey of such findings.)

One response to these findings is to stress that the expected utility model is prescriptive, rather than descriptive. But the theory can not be applied unless its substantive restrictions, e.g., that any two consequences in $C$ are comparable by the weak preference relation, are met. Such assumptions are necessarily descriptive rather than normative: it makes little sense to tell a d.m. that he "must" or "should" be able to make comparisons or perform other cognitive tasks which in fact he is unable to do. A more constructive response is to attempt to isolate useful classes of decision problems for which the axioms are satisfied. For example, including anxiety, disappointment, and regret along with monetary payoffs in the description of the possible consequences of a compound lottery may make axioms work that failed when only monetary payoffs were included. Learning how to describe consequences fully enough so that the axioms apply allows some empirical contradictions of Model 1 to be removed by showing them to be the results of inadequate description rather than true violations of the axioms. (For example, if each point in $C$ can be represented by a vector in an $n$-dimensional vector space of descriptive attributes, observed violations may occur not among the members of $C$, but among the images of members of $C$ projected onto a lower-dimensional subspace of descriptive attributes.) Finally, alternative decision theories can be developed that relax one or more of the axioms of weak ordering, continuity, and independence, or that use a weaker (e.g., interval-valued rather than number-valued) representation of preferences. In the next subsections, we summarize some alternatives to Model 1 for special types of consequence sets, $C$. 
3.2 Decision Model 2: Single-Attribute Decision Theory (Machina, 1982, 1983)

Suppose that $C$ has a natural numerical structure, i.e., that its members are naturally ordered according to a single continuous numerical attribute (also called a "factor" in psychology and multivariate statistics), such as "magnitude" or "severity" of outcome. We will denote this natural ordering by $P$, and write $c \preceq c'$, for any $c$ and $c'$ in $C$, if and only if $c$ weakly outranks $c'$ in the ordering $P$. (Later, $P$ will be interpreted as strict preference.) Equivalently, suppose that the members of $C$ can be indexed by the numbers in a closed interval $[0,M]$ in such a way that (i) There is a one-to-one correspondence between members of $C$ and numbers in $[0,M]$; and (ii) For any two consequences $c$ and $c'$ in $C$, $f(c)$ is at least as great as $f(c')$ if and only if $c \preceq c'$. Here, $f(.)$ is the function from $C$ to $[0,M]$ mapping each consequence into its corresponding numerical index. $f(.)$ preserves $P$, in the sense that the relation $P$ between two consequences is represented by the relation of weak inequality between the two corresponding numbers assigned to the consequences by $f$. The process of assigning numbers to objects in such a way that an ordinal relationship $P$ between objects is preserved by the numerical ordering of the corresponding numbers is called ordinal measurement of $P$.

Technical Note: In this situation, the set $[0,M]$ ordered by numerical inequality is said to be "isomorphic" to the set $C$ ordered by $P$; in mathematical notation, we could say that the system $([0,M], \text{is-at-least-as-great-as}) \ast= (C,P)$, where we use the symbol $\ast= $ to denote isomorphism between two algebraic systems, meaning that there is a one-to-one correspondence between the objects (e.g., consequences), relations among corresponding objects, and operations on corresponding objects of the two systems. Isomorphic systems are distinguished only by the names and interpretations assigned to their elements; they have identical algebraic structures.

Note that $M$ can be normalized to 1 without loss of generality. That is, $([0,M], \text{is-at-least-as-great-as}) \ast= ([0,1], \text{is-at-least-as-great-as})$. 

55
Although we are here abandoning our usual simple mathematical framework of finite sets, many of the concepts already developed still apply or can readily be extended. For example, the set of all probability measures on $C$ becomes more difficult to define when $C$ is continuous; however, we can deal instead with the set $D$ of all cumulative probability distribution functions (c.d.f.'s) over $C$, and take these distribution functions as the objects for which preferences are to be derived. Thus, $L$ will now represent the set of cumulative probability distributions over consequences corresponding to different acts in $A$, and our strategy for solving the problem of choosing from $A$ will again be to embed $L$ in $D$, develop a preference ordering over $D$, and then apply it to $L$ to find a most preferred member $a^*$ in $A$. However, a different set of axioms, or substantive assumptions about preferences, will be used in carrying out this construction.

Since $C$ is isomorphic to $[0,M]$, we can take $D$ to be the set of all c.d.f.'s over either $C$ or $[0,M]$; it does not matter which interpretation is used. (In effect, the numbers in $[0,M]$ are just the names of the consequences in $C$.) We will interpret $D$ as the set of c.d.f.'s over $[0,M]$. Any member $d$ in $D$ can be thought of as a single point in an appropriate (infinite-dimensional Banach) mathematical space, and we can consider functions and relations defined on these points.

In analogy to Model 1, let $R$ denote the binary relation "is-at-least-as-preferred-as" between points in $D$. Originally, the d.m. may not know his own $R$ (or it may not exist). That is, we assume that he knows his preferences for the outcomes in $C$, but that he needs help in deriving his preferences over $D$. To this end, we apply the mathematical framework from Model 1, invoking the following assumptions:

**Axiom CR1:** $R$ is a weak ordering (complete and transitive) on $D$;

**Axiom CR2:** Let $Q$ be the strict relation defined by $d Q d'$ if and only if $d R d'$ and $-(d' R d)$, for any two distribution functions $d$ and $d'$ over $[0,M]$. (Here, "-" means "not".) Let $F$ and $G$ be two members of $D$ such
that \( F \succ Q \succ G \), and let \( \{ F_n(.) \} \) be a sequence of c.d.f.'s such that \( F_n(x) \) approaches \( F(x) \) wherever \( F(x) \) is continuous. (Mathematically, the sequence \( F_n \) converges to \( F \) in the topology of weak convergence.) Then for all sufficiently large \( n \), \( F_n \succ Q \succ G \).

Axiom CR3: Let \( F, G, \) and \( H \) be any three c.d.f.'s in \( D \) such that \( F \succ R \succ G \). Then \( (F \succ p \succ H) \succ (F \succ p \succ G) \) for any \( p \) in \( [0,1] \).

These are just the axioms of weak ordering, continuity, and independence from Model 1 transferred to the context of a continuous numerical consequence set.

**Theorem 3.3:** If weak ordering and continuity (axioms CR1 and CR2) hold, then there is a function \( V(.) \) from \( D \) to the set of real numbers such that for any two c.d.f.'s \( d \) and \( d' \) in \( D \),

\[
V(d) \text{ is at least as great as } V(d') \text{ if and only if } d \succ R \succ d'.
\]

That is weak ordering and continuity suffice for ordinal measurement of preferences over \( D \).

**Note:** In mathematical terminology, \( V(.) \) is properly called a "functional", rather than a function, since its range \( D \) is a subset of an infinite-dimensional space.

**Theorem 3.4:** If weak ordering, continuity, and independence (i.e., axioms CR1 to CR3) hold, then there is a function \( u(c) \) defined on \([0,M]\) such that for any two points \( d \) and \( d' \) in \( D \),

\[
d \succ R \succ d' \text{ if and only if } E[u(c);d] \text{ is at least as great as } E[u(c);d'].
\]

(Here, \( E[u;d] \) denotes the expected value of \( u(c) \) under the probability distribution \( d \) on \( c \).)
**Corollary 1:** Under axioms CR1 to CR3, the functional \( V^*(.) \) assigning the numerical value \( V^*(F) = E[u;F] \) to any c.d.f. \( F \) in \( D \) provides for ordinal measurement of \( R \) on \( D \).

Of course, any other monotonically increasing function of \( V^*(F) \) also defines an ordinal measurement scale for \( R \) on \( D \), i.e., if \( f(.) \) is any monotonically increasing function, then \( f(V^*(d)) \) and \( f(V^*(d')) \) will stand in the same ordinal relationship to each other as \( V^*(d) \) and \( V^*(d') \), and hence both will represent \( R \) equally well. (In other words, an ordinal measurement scale is unique only up to monotonic transformations.) However, the function \( u(.) \) is much more restricted.

**Corollary 2:** Let \( u \) and \( u^* \) be any two numerical functions on \([0,M]\) such that for some \( R \) satisfying CR1 to CR3,

\[
d \ R \ d' \text{ if and only if } E[u;d] \text{ is at least as great as } E[u;d'] \text{ if and only if } E[u^*;d] \text{ is at least as great as } E[u^*;d'],
\]

for any two c.d.f.'s \( d \) and \( d' \) in \( D \). Then \( u(.) = au^*(.) + b \) for some \( a \) greater than zero. (That is, \( u(.) \) is unique only up to positive affine transformations.)

A function \( u(c) \) with the above properties is called a "von Neumann-Morgenstern (N-M) utility function". It represents the d.m.'s primitive preferences on \( C \), as well as his derived preferences on \( D \).

These results, which go back to von Neumann and Morgenstern, are essentially restatements of Model 1 in the changed mathematical setting of a continuous numerical consequence set. The same empirical criticisms that applied to Model 1 apply here also. However, a major conceptual advance due to Mark Machina (1982) allows some of these criticisms to be overcome. The basic idea is to weaken the assumption of independence, replacing it with one of "smoothness" of preferences, and then to show that most of the key qualitative results of expected utility theory can still be made to apply. Formally, we keep axioms CR1 and CR2, which
guarantee the existence of an ordinal function $V(.)$ representing $R$ on $D$, and replace the independence assumption, Axiom CR3, with the following

**Axiom CR4**: The functional $V(.)$ representing $R$ on $D$ is "smooth", in the sense of being differentiable at every point $d$ in $D$.  

This condition is not quite as straightforward as it appears, since we have not defined derivatives in infinite-dimensional spaces. However, regarding each c.d.f. in $D$ as a point leads to notions of differentiation similar to those for finite-dimensional spaces. Technically, we shall assume, following Machina, that $V(.)$ is Frechet differentiable (with respect to the $L_1$ metric defined below) at each point in $D$.

**THEOREM 3.5** (Machina, 1982): Under axioms CR1, CR2, and CR4, there exists an absolutely continuous function $u(.;d)$ on $[0,M]$ such that $V(d') - V(d)$ approaches $E[u(c;d);d'] - E[u(c;d);d]$ as $d'-d$ approaches zero. Moreover, $u(.;d)$ is differentiable on $(0,M]$ (except possibly at a countable number of points) and is unique up to positive affine transformations (see definition in Corollary 2 of Theorem 3.4), as for ordinary N-M utility functions, for any $d$ in $D$.

In this theorem, $E[u(c;d);d]$ denotes the expected value of the "local utility function" $u(c;d)$ with respect to the c.d.f. $d$. The notation $u(c;d)$ indicates that the local utility function $u(.;.)$ for consequences $c$ is a function not only of $c$, but also of the point $d$ in $D$ at which it is being evaluated. $u(c;d)$ is called a (single-attribute) local utility function over the consequence attribute $c$, and represents preferences over this attribute, in the sense that $u(c';d(c'))$ is at least as great as $u(c'';d(c''))$ if and only if $c' \preceq c''$, for any two outcome values $c'$ and $c''$ in $C$. Here, $d(c')$, for example, denotes the degenerate c.d.f. corresponding to outcome $c'$, i.e., it is the c.d.f. that jumps from 0 to 1 at $c'$. Any such function, including $u$, that represents preferences over deterministic consequences in $C$ is called an (ordinal) "value function" over $C$.  

59
Technical Note: Again, mathematical details are important. To say that \( d - d' \) approaches zero in the above theorem, we need a metric for comparing the distance between two points in \( D \). We choose for this purpose the \( L_1 \) metric, which defines the "distance" between two c.d.f.'s \( d \) and \( d' \) in \( D \) as the total area between them, i.e., as the difference between the areas beneath them. (This metric is invariant under ordinal transformations of the consequence axis and induces the topology of weak convergence used in axiom CR2, and so is a natural choice.)

Example 3.2: Differential Shifts in Public Health Risks

Suppose that a change \( dx \) in the level of automobile tailpipe emissions of a pollutant \( X \) will produce an uncertain change \( dr \) in the public health risk of a disease \( Y \), as measured by the average annual incidence rate \( r \) of some health effect \( Y \) in the general population. If the base line emissions level of \( X \) is \( x \), and the base line incidence rate of \( Y \) is \( r = r(x) \), then \( dr = r(x + dx) - r(x) \). Both \( r \) and \( dr \) will typically be uncertain (to a d.m. trying to decide whether to implement \( dx \), or how large a \( dx \) to implement), e.g., because \( r(x) \) may also depend on a wide variety of omitted explanatory variables. Let \( F \) be the c.d.f. of \( r \), and \( F' \) be the c.d.f. of \( r + dr \). Axioms CR1, CR2, and CR4 then imply that the value of any proposed differential (i.e., sufficiently small) shift \( dr \) can be measured according to the corresponding change in local expected utility, \( dV = E[u(r;F);F'] - E[u(r;F);F] \). Given two alternative proposed shifts, say \( dr \) and \( dr' \), with uncertain effects on the risk level \( r \), a choice between them may be made by choosing the one corresponding to the larger of \( dV \) and \( dV' \). In the special case where the base line level of \( r \) is certain, moreover, uncertain differentials \( dr \) can be ranked in terms of their expected values, just as in the traditional (Theorem 3.4) case with smooth (locally linear) utility functions.

The idea of the local utility function is Machina's great contribution to the development of a satisfactory decision theory for choices among c.d.f.'s over a numerical interval. It exploits the continuity of the consequence set in a way that was impossible for discrete consequences.
Intuitively, the smoothness assumption (Axiom CR4) guarantees that local utility functions are locally linear, and hence that expected utility analysis is appropriate for ranking differential changes, thus making the independence axiom CR3 unnecessary. This illustrates a useful principle: that instead of there being a single unified normative theory of decision making for problems specified by the three sets C, A, and S, there can appropriately be multiple theories for different mathematical structures of the consequence set C. A consequence set that is isomorphic to [0,1] will support a different decision theory, using different axioms and achieving different characterizations of rational choice, than a consequence set isomorphic to a subset of the integers. Machina calls his decision theory for the former case "generalized expected utility analysis."

Many empirically observed violations of the conventional theory leading to Theorem 3.4 can be explained using generalized expected utility analysis by a simple hypothesis about the shapes of the local utility functions (essentially, that they "fan out" from the origin; see Machina, 1983.) Specifically, behavior that violates the independence axiom, and which would therefore be judged "irrational" by the standards in Theorem 3.1, may be consistent with the existence of smoothly varying local utility functions.

The main importance of generalized expected utility analysis for our purposes, however, lies in its normative implications for decision making. The major concepts and results, that are used later in our discussion of risk measurement procedures may be summarized as follows:

Definition (FSD): Let F and G be two c.d.f.'s over [0,M] such that F(c) is no greater than G(c) for each c in [0,M]. Assume that preferences are increasing in c. Then we will say that F "preference-dominates G in the sense of first-order stochastic dominance", written F FSD G. Conversely, G "risk-dominates" F by first-order stochastic dominance (FSD).
THEOREM 3.6: Under axioms CR1, CR2, and CR4 (i.e., assuming that preferences over members of D are representable by a Frechet-differentiable ordinal value function V), a necessary and sufficient condition for F FSD G to imply that V(F) is at least as great as V(G), for all F and G in D, is that \( u(c;d) \) is nondecreasing in c for all d in D.

If greater values of c in \([0,M]\) correspond to more severe, less preferred consequences, so that preferences are strictly decreasing in c, then the preceding formulation still applies to the attribute \( c' = M-c \), and all d.m.'s will agree that V(F) is at least as great as V(G) whenever F FSD G. Note, however, that FSD only partially orders the set D of c.d.f.'s over \([0,M]\), since many pairs of c.d.f.'s intersect.

THEOREM 3.7: In the conventional case (axioms CR1 to CR3) where \( u(c;d) = u(c) \) is the same for all d in D, a conclusion analogous to that of Theorem 3.6 still holds: for any two c.d.f.'s F and G in D, V(F) is at least as large as V(G) for every monotonically increasing u(c) if and only if F FSD G. In fact, the following are equivalent:

1. \( V(F) \) is at least as large as \( V(G) \) for every increasing utility function \( u(c) \);
2. F FSD G;
3. There are random variables c(s) and e, where e is nonnegative, such that G is the c.d.f. of c(s) and F is the c.d.f. of c(s) + e.

The significance of this result, for our purposes, is the following. If all we know about a d.m.'s preferences is that he prefers larger c values to smaller ones, then we can conclude he will prefer c.d.f. F to c.d.f. G if and only if F FSD G. If F and G cross, then there will be some sets of preferences that satisfy axioms CR1 to CR3 for which F is preferred to G; while there will be other preferences, also consistent with these
axioms, for which F will not be preferred to G. The FSD partial ordering of c.d.f.'s thus provides a tight bound on what can be deduced about preferences for lotteries over consequences from the knowledge that preferences are monotonic in c.

3.3 An Alternative Axiomatic Basis for Single-Attribute Utility Functions

Machina's work exploits the mathematical structure of the consequence interval \([0,M]\) in one way, by using it to define a concept of "smoothness" for the ordinal preference measurement functional \(V\). A second way to exploit the continuity of the interval is through its "solvability" properties when preferences are monotonic. A theory of preference measurement based on this approach will now be presented.

Suppose again that consequences \(c\) in \(C\) have a natural numerical structure induced by an empirical relation \(S\) of comparative severity. That is, there is a one-to-one correspondence between consequences \(c\) in \(C\) and numbers \(f(c)\) in \([0,M]\); and for any two consequences \(c'\) and \(c''\) in \(C\), \(f(c')\) is greater than \(f(c'')\) if and only if \(c' S c''\), where \(S\) is the binary relation "is-more-severe-than". Thus, \(f\) is empirically defined. For example, if \(c'\) and \(c''\) denote two possible numbers of gallons of oil spilled in a tanker collision, then \(c' S c''\) might denote the empirical relation that \(c'\) is a larger number of gallons than \(c''\).

In addition to the empirical relation \(S\), we will assume that there is a judgemental preference relation \(P\) such that for any two consequences \(c'\) and \(c''\) in \(C\), \(c' S c''\) if and only if \(c' P c''\). (Note that in this context \(P\) represents strict preference.) That is, more severe consequences are less preferred. Thus, \(S\) and \(P\) contain exactly the same information. We now wish, as usual, to extend the primitive preference relation \(P\) over \(C\) to a derived preference relation \(R\) of weak preference over \(D\), where \(D\) is the set of all c.d.f.'s over \([0,M]\) \(*=*= C\), so that choices among distributions in \(D\) can be made. Theorem 3.3 showed that if axioms of weak ordering and continuity hold, then such an \(R\) will exist and can be
represented by an ordinal numerical functional $V(.)$, i.e., for any $F$ and $G$ in $D$, $V(F)$ will be at least as great as $V(G)$ if and only if $F \leq G$. Thus, $V(.)$ is an ordinal measure of derived preferences $R$ for c.d.f.'s in $D$. Theorem 3.4 showed that if the independence axiom also holds, then $V(F)$ takes the very special form of an expectation, i.e., $V(F) = E[u;F]$ for some utility function $u(c)$ over $C$, or is a monotonic transformation of these expectations. Moreover, there is a one-to-one correspondence between the set of relations $R$ satisfying axioms $CR1$ to $CR3$ and the set of utility functions (unique up to change of scale and origin) representing them. Finally, any increasing monotonic transformation of a utility function $u(c)$ is itself an ordinal value function over the set of consequences.

These results show the existence of a value functional $V(.)$ over c.d.f.'s and a cardinal (meaning unique up to choice of origin and scale) utility function $u(.)$ over consequences, but they do not explain how either $u(.)$ or $V(.) = E[u;.]$ might be derived in practice. The following, complementary, approach remedies this deficiency.

Let $F$ be any c.d.f. over $C$ (or, equivalently, over $[0,M]$). Then $0 \leq F \leq M$, where $0$ and $M$ are the degenerate distributions that put probability 1 on the endpoints 0 and $M$ of $[0,M]$, respectively. Hence $V(c)$ decreases continuously from $V(0)$ to $V(M)$ as $c$ increases continuously from 0 to $M$; by the assumption of continuity, therefore, there must be some value of $c$, say $CE(F)$, such that $V(F) = V[CE(F)]$. That is, there exists in principle a deterministic outcome $CE(F)$ that is judged indifferent to the uncertain outcome corresponding to c.d.f. $F$. (This is the key "solvability" property referred to above: it clearly does not hold for discrete attributes, such as numbers of lives lost.) By the assumption that preferences are strictly decreasing in $c$, this outcome will be unique. It is called the "certainty equivalent" of $F$.

In the abstract, $CE(.)$ is a functional assigning numbers in $[0,M]$ to distributions in $D$. It is therefore a possible candidate for another ordinal value functional, different from $V(.) = E[u;.]$, representing $R$ on
D. In fact, we have the following alternative derivation of the expected-utility representation of \( R \):

**THEOREM 3.8:** Let \( CE(.) \) be any functional mapping distributions over \([0,M]\) into numbers in \([0,M]\) and satisfying the following two axioms:

**CE1:** For any deterministic consequence \( c \) in \( C \equiv [0,M] \), \( CE[d(c)] = c \) (where \( d(c) \) is the degenerate distribution putting probability 1 on \( c \)).

**CE2:** If \( F \) and \( F' \) are any two c.d.f.'s in \( D \) such that \( CE(F) = CE(F') \), then for any \( r \) in \([0,1]\) and any other c.d.f. \( H \) in \( D \), \( CE[rF + (1-r)H] = CE[rF' + (1-r)H] \).

These are necessary and sufficient conditions for there to be a continuous monotonic N-M utility function \( u(c) \) over \([0,M]\), unique up to choice of scale and origin (i.e., up to positive affine transformations), such that

\[
u[CE(F)] = \nu(F) = E[u;F] \text{ for all } F \text{ in } D,\]

i.e., such that

\[
CE(.) = u'[E(u;.)],
\]

where \( u' \) is the inverse function of \( u \). Under these conditions,

\( F \) FSD \( G \) implies that \( CE(F) \) is at least as great as \( CE(G) \),

and \( CE[rF + (1-r)H] \) lies between \( CE(F) \) and \( CE(H) \), for any two c.d.f.'s \( F \) and \( H \) in \( D \). Note that \( rF + (1-r)H \) is a c.d.f. that is a weighted average (or convex combination) of the c.d.f.'s \( F \) and \( H \); it is not the probabilistic mixture \( (F,r,H) \).

**Proof:** This is an application and interpretation of Theorem 1 of Chew, 1983.
If we look at CE(.) as a monotonic transformation of V(.) = E[u(.)], namely the transformation defined by u' (whose existence and monotonicity follow from the assumed monotonicity of P), then these results may not appear surprising. The significance of Theorem 3.8 is that it allows the usual expected utility representation of preferences to be derived from only the two axioms CE1 and CE2. In addition, if we assume that the intuitive interpretation of CE(F) as a certainty equivalent, i.e., as a certain consequence that is indifferent to the random consequence corresponding to F, satisfies CE1 and CE2, then Theorem 3.8 provides a way of empirically assessing u(.) from observed indifferences between lotteries and their certainty equivalents. For example, if origin and scale are fixed by the "canonical" normalization u*(0) = 0, u*(M) = 1 for the case of increasing preferences, then the canonical utility function u* is found by solving

\[ u^*(c) = p \text{ such that } CE(0,p,M) = c, \]

for each c in [0,M]. For decreasing preferences, the same results can be applied after replacing c with c' = M - c.

3.4 Decision Model 3: Multiattribute Consequences (Keeney and Raiffa, 1976)

We have argued that a normative decision theory for choice among probability distributions over consequences ought to take advantage of the natural mathematical structure of the consequence set. In other words, we are assuming throughout this section that the goal of a normative decision theory is to enable the d.m. to extend his primitive preferences P over consequences to a set of derived preferences R over lotteries represented by c.d.f.'s over consequences. How this extension can best be made, however, will in general depend on the structure of C and P, as in the example of continuous C and monotonic P given above.

A second, outstanding example of adept exploitation of structure is given by multiattribute utility theory as summarized in Keeney and Raiffa,
1976. Namely, if the consequence set \( C \) has the mathematical structure of a closed rectangular subset of an \( n \)-dimensional Euclidean vector space, formed as the Cartesian product space of \( n \) single-attribute intervals, then the \( N-M \) utility function \( u(c) = u(c_1, c_2, \ldots, c_n) \) can in many cases be decomposed as a function of \( n \) single-attribute utility functions. For a full development of this idea, see Keeney and Raiffa, 1976. As one important application, suppose that \( n = 2 \), so that \( C = C_1 \times C_2 \), where \( C_1 \) and \( C_2 \) are two bounded single-attribute intervals (e.g., representing probability of an accident and severity of an accident if one does occur.) Then if one adds to the assumptions of Theorem 3.2 (Decision Model 1) the further assumption that preferences among lotteries over consequence vectors differing only in one of their two components do not depend on the fixed level of the other component -- the assumption of "mutual utility independence" -- it follows that \( u(c) \) can be decomposed as either a sum or a product of two single-attribute utility functions, \( u_1(.) \) and \( u_2(.) \), over attributes \( C_1 \) and \( C_2 \), respectively. Thus, a decision theory for consequence sets isomorphic to bounded rectangular subsets of Euclidean \( n \)-space can be based on decision theories for consequence sets isomorphic to \([0,1]\) if appropriate independence (or "separability") conditions can be established.

3.5 Decision Model 4: Uncertain Consequences

\( C \) may have a more complicated structure than a finite set, a bounded interval, or a Cartesian product of bounded intervals -- the three cases considered in the preceding subsections. Each of these sets is essentially static, for example, whereas many actual consequences are essentially dynamic. In this final subsection on normative decision models, we introduce a very simple dynamic consequence set. This line of development is continued in Chapter 7, which considers consequences that are represented by simple stochastic processes, rather than by static numbers or vectors.

Let \( A \) be a finite set of acts, \( S \) a finite set of states, and \( C \) a finite set of consequences, as in Model 1. Each consequence \( c \) in \( C \) consists of
two parts, labelled \( s \) and \( x \), where \( s \) is a state in \( S \) and \( x \) is a member of a set \( X \) of "physical outcomes" (e.g., gallons of oil spilled.)

The decision-to-consequence process works as follows:

(i) First, the d.m. chooses an act \( a \) from \( A \).

(ii) Secondly, "Nature" reveals the true state \( s \) to the d.m. If the true state turns out to be \( s \) and the d.m. has chosen act \( a \), then the set of possible physical outcomes will be some subset of \( X \) depending on \( s \), and there will be a probability measure \( f(.;s,a) \) over events in \( X \).

(iii) The physical outcome \( x \) in \( X \) is realized according to the probability measure \( f(x;s,a) \).

A consequence, in this setting, consists of a state \( s \) followed by a physical outcome \( x \) drawn from \( X \); thus, it is a pair \((s,x)\). An act is a partially controllable probability distribution over consequences. The part that is controllable is the conditional distribution over \( X \) given \( s \), namely, \( f(.;s,a) \). The d.m. can choose any set of such conditional distributions (one for each possible value of \( s \)) corresponding to an act \( a \) in \( A \). The marginal distribution over \( s \), however, which we will denote by \( p(s) \), is not affected by the choice of \( a \), and hence is not controllable by the d.m.

**Example 3.3: Installing Safeguards**

The designer of a new business or industrial facility realizes that it will be located in a flood plain, and that there is a chance that it will be exposed to a major flood at some time during its operational life. He must choose from a set \( A \) of precautionary safeguards, including a wide range of drainage and pumping system alternatives having varying costs and capacities to handle flooding. Let \( s = 0 \) if the facility in fact experiences no flood over its operational life, and let \( s = 1 \) otherwise; the marginal distribution \([p(s = 0), p(s = 1)]\) is not controllable by the
d.m. On the other hand, given a flood, i.e., that \( s = 1 \), the conditional distribution of the dollar value, \( x \), of property lost and damaged will depend on which alternative \( a \) in \( A \) the designer has chosen. Here, the set \( X \) of physical outcomes is an interval of possible dollar damage figures. Analogous interpretations of \( A \), \( S \), and \( X \) occur in many situations, from buying insurance to installing burglar alarms to retrofitting nuclear reactors, where the decision about what precautions to take must be made before learning what precautions will be needed.

This example should make clear that the standard formulation of a decision problem in terms of the three sets \((A,S,C)\) is now being replaced with a formulation in terms of \((A,S,X)\), with both \( S \) and \( X \) folded into the "consequence" set, \( C \), which is a subset of \( S \times X \). Explicitly, the problem data \([c(a,s), P, p(s)]\) for the standard problem have been replaced by the richer set \([f(x; s, a), P, p(s)]\), where the primitive preference relation \( P \) is now defined over the pairs \((s, x)\). The d.m.'s problem is still to extend his primitive preference ordering \( P \) over deterministic consequences \((s, x)\) to a derived ordering \( R \) over acts (or a superset of the set of acts.) However, an act is now a set of conditional probability distributions over \( X \), one for each possible \( s \).

Let \( A^* \) denote the closure of \( A \) under mixture, i.e., the set such that if \( a \) and \( b \) are any two acts in \( A \) and \( p \) is any number in \([0,1]\), then the mixture \((a,p,b)\) is in \( A^* \). \( A^* \) may be thought of as the set of all randomized acts. Embedding \( A \) in \( A^* \) allows an ordering \( R \) over \( A^* \) to be developed that applies to the acts in \( A \) as a special case. Similarly, let the set of all probability distributions (or measures) over the set \( C \) of \((s, x)\) pairs be denoted by \( D \). We will let \( P \) denote the weak preference relation between \((s, x)\) pairs, and will use \( R^* \) to denote the weak preference relation between probability distributions in \( D \). Note that these probability distributions no longer correspond to acts, which is why we have introduced the new notation \( R^* \) for the preference relation between them, instead of continuing to use \( R \). We will assume that \( R^* \) is a primitive preference relation, i.e., that the d.m. knows his own preferences for lotteries over consequences, or state-physical outcome
pairs. This knowledge, in turn, may result from a previous round of N-M
utility analysis extending P to R*, as discussed in preceding models.

Let R denote the derived preference relation on A*, the set of randomized
acts. Suppose that R on A* and and R* on D satisfy the following axioms:

SR1 (Weak Ordering): R and R* are complete, transitive binary relations
on A* and D, respectively.

SR2 (Continuity): Let Q and Q* be the strong orderings corresponding to R
and R*, respectively (e.g., a Q b means a R b and not b R a.) Then Q
and Q* satisfy the continuity axiom of Model 1 (see Axiom R2 of Model 1
for definition) on A* and D, respectively.

SR3 (Independence): Q and Q* satisfy the substitution, or "independence",
axiom of Model 1 (see Axiom R3 of Model 1 for definition) on A* and D,
respectively. For example, x Q y if and only if (x p z) Q (y p z), for
any randomized acts x, y, and z in A*.

We will call these three axioms W.O., C, and I, respectively.

THEOREM 3.9: W.O., C, and I imply the existence of a weighting function
w(s,x) such that for any two acts a and b in A*, a R b if and only if
E[w;a] is at least as great as E[w;b]. Moreover, if w'(s,x) is any other
weighting function with this property, then w'(s,x) = aw(s,x) + b(s),
where a is nonnegative.


In this theorem, E[w;a] denotes the sum over all pairs (s,x) in C of
w(s,x)f(x;s,a). Note that p(s) does not enter this expression
explicitly; its effects are incorporated into w(s,x). Thus, E[w;a] is a
generalization of the usual expectation operator. The use of the E[.;;] notation is justified by the following result.
THEOREM 3.10: Suppose that in addition to W.O., C, and I, (axioms SR1 to SR3), the derived preference relation R also satisfies the following two axioms:

SR4 (Nontriviality for Outcomes): There are at least two nonindifferent physical outcomes in X, say \( x' \) and \( x'' \), that have positive probabilities of occurrence \( f(x';s,a) \) and \( f(x'';s,a) \), at each state-act pair \( (s,a) \).

SR5 (State Independent Preferences): Preferences R among any two acts \( a \) and \( b \) in \( A^* \) depend only on their implied corresponding marginal distributions \( f(x;a) \) and \( f(x;b) \) over physical outcomes \( x \) in \( X \), and do not otherwise depend on \( s \).

Then there exists a unique probability measure \( p(.) \) on \( S \) and a unique (up to choice of origin and scale) N-M utility function \( u(x) \) such that \( w(s,x) = p(s)u(x) \) in Theorem 2.10.


Thus, under SR1 to SR5, preferences among acts will have an expected utility representation. Theorem 3.10 establishes the existence of a unique subjective probability measure \( p(s) \) over states \( s \) in \( S \), and we shall assume that this subjective measure is "correct", i.e., that it agrees with some "objective" probability measure, despite the difficulties with this concept (see Section 2.2.)

Axiom SR5 has often been implicitly or explicitly assumed in practical applications of both single-attribute and multi-attribute decision theories. For example, the axiom of "reduction of compound lotteries" for gambles over monetary outcomes asserts that all that matters in choosing among acts with monetary physical outcomes is the respective marginal probability distributions over monetary outcomes induced by the acts: the way in which this distribution is resolved does not affect preferences. However, there is no logical necessity for this assumption. And in fact, it is often violated by observed behavior (see e.g., Example 3.1 and Keller, 1985.)
Recently, Bell (e.g., Bell, 1985) has argued for the importance of "disappointment", "regret", and similar phenomena in the formation of both prospective preferences for and retrospective evaluations of acts, and has shown that such concepts can help explain observed departures from the reduction of compound lotteries axiom. For example, in Example 3.3 (Installing Safeguards), a d.m. might feel differently about an outcome of $100,000 worth of property damage if it resulted from a minor flood after he had decided not to install expensive safeguards than he would if it resulted from a major flood after he had installed the best available safeguards. Thus, his utility function for consequences might well be of the form $u(s,x)$, or even $u(a,s,x)$, rather than simply $u(x)$ as axiom SR5 would imply. We will consider utility functions of the form $u(s,x)$, i.e., defined on full consequences $(s,x)$. The extension to utility functions of the form $u(a,s,x)$, where $a$ presumably enters $u$ only through $f(.;s,a)$, requires (i) a distinction between prospective (choice oriented) and retrospective (evaluation or regret oriented) preferences; and (ii) development of "local" utility functions taking the probability measure $f(.;s,a)$ as an argument. Machina's theory, outlined in Section 3.3, illustrates the latter development. Preferences for the same outcomes evaluated at different times will be considered in Chapter 7.

**THEOREM 3.11** (Karni, 1985): If $R$ on $A^*$ and $R^*$ on $D$ satisfy SR1 to SR3, i.e., W.O., C, and I, then there exists an N-M utility function $u(s,x)$ on $C$ such that for any two consequence lotteries $d$ and $d'$ in $D$, $d R^* d'$ if and only if $E[u;d]$ is at least as great as $E[u;d']$.

Here, $E[u;d]$ is the sum over all $(s,x)$ pairs in $C$ of $u(s,x)d(s,x)$, where $d(s,x)$ is the probability of $(s,x)$ under lottery $d$.

Suppose now that in addition to SR1 to SR3, $R$ and $R^*$ satisfy the following two axioms:

**SR6 (Nontriviality for Acts):** There are at least two nonindifferent acts in $A^*$.
SR7 (Strong Consistency): Suppose that for two acts \(a\) and \(b\) in \(A^*\),
\[f(x;s,a) = f(x;s,b)\] for all \(x\) in \(X\) and for each \(s\) in \(S\) except \(s = s'\).
Then \(R\) and \(R^*\) agree on the ranking of \(a\) and \(b\), where \(a\) and \(b\) are considered as acts for \(R\) and as consequence lotteries for \(R^*\). (See Karni, 1985, for a more formal statement of this condition.)

Then there exists a probability measure \(p(.)\) on \(S\) such that for any two acts \(a\) and \(b\) in \(A^*\),
\[a R b \text{ if and only if } E[u;a] \text{ is at least as great as } E[u;b],\]
where \(E[u;a]\) is the sum over all \((s,x)\) pairs in \(C\) of \(p(s)u(s,x)f(x;s,a)\), and \(u\) is the N-M utility function from above.

Proof: Karni, 1985, Chapter 1.

The significance of this theorem is that it links derived preferences \(R\) over acts (i.e., \(n\)-tuples of consequence lotteries, one for each state \(s\) in \(S\)) to the assumed primitive preferences \(R^*\) over consequence lotteries. The link is that both orderings are represented by the same utility function \(u(s,x)\) on \(C\).

Example 3.4: Delayed Health Effects

Consider a young worker's decision to take a risk (e.g., to work in a hazardous occupation in exchange for hazard pay) that might have long-latency health effects, e.g., some form of cancer. Let \(x = 1\) if the effect eventually occurs before the worker dies of other causes, and let \(x = 0\) otherwise. Similarly, let \(s = 1\) if the worker eventually gets married before dying or contracting the disease, and let \(s = 0\) otherwise. (In a realistic application, \(s\) could be refined to include other information, e.g., \(s = 0\) for no marriage, \(s = 1\) for married but with no living children, \(s = n\) for married with \(n\)-1 children for \(n = 2,3,...\). Or \(s\) could be a vector indicating family size, health status, economic status, and so forth.) Then preferences over pairs \((s,x)\) might
very plausibly depend on s as well as x, making the representation of preferences in Theorem 3.11 appropriate.

3.6 Summary

We have introduced the idea of a normative decision theory (NDT) as a set of rules for extending primitive (known) preferences over a set such as the set C of consequences to derived preferences over a set A of acts, i.e., the objects of choice. Within an NDT, preferences over A can be derived or deduced from primitive preferences with the help of an explicitly stated set of assumptions or axioms. There are many normative decision theories, appropriate for different mathematical structures of the choice and consequence sets.

In the "classical" case, A is a subset of the set of all probability distributions (or measures) over C, primitive preferences are over C, and the extension of primitive to derived preferences can be made by means of an expected utility representation. Special cases include the following:

- C is isomorphic to a finite subset of the integers (Section 3.1);
- C is isomorphic to the unit interval [0,1] ("single-attribute" decision theory, Sections 3.2 and 3.3);
- C is isomorphic to the n-fold Cartesian product of [0,1], i.e., to the unit n-cube, for some integer n ("multiattribute" decision theory, Section 3.4);
- C is isomorphic to the Cartesian product of n finite subsets of the integers ("discrete-attribute" decision theory, not covered.) For example, C might be a subset of n-dimensional binary vectors, indicating the status (e.g., alive or dead) of each of n members of a population; this formulation is used in Chapter 8.
C is a "mixed" set with both continuous and discrete components, or a Cartesian product of such sets.

In each of these classical cases, the set $C$ is static and completely determines the set $D$ in which $A$ is embedded. The elements of $C$ are also embedded in $D$ as special, degenerate cases. By extending the primitive ordering $P$ on $C$ to a derived ordering $R$ on $D$, where $R$ agrees with $P$ on $C$, the problem of choice from $A$ is implicitly solved: an optimal choice is to choose the highest-ranking member of $A$.

Alternatives to this classical analysis are obtained by making $C$ dynamic. What this means intuitively is that the d.m. only gradually learns the consequence $c$ that is eventually realized: his information about $c$ increases over time. A simple example was given in Section 3.5, for a model of "state-dependent" preferences. In this model, each consequence $c$ consisted of a pair $(s,x)$, where both members were learned after the act was chosen, but with $s$ learned before $x$ and giving information about the remaining probability distribution $f(x;s,a)$ over $x$. More generally, $c$ may consist of the realization of a controlled stochastic process in which choices of acts are interleaved with and affect the gradual resolution of uncertainty about $c$. The definition of primitive preferences over $C$ and the formulation of an act set linked to partial realizations of $C$ then require care. Chapter 7 will develop a normative decision theory for a class of cases in which $C$ is not isomorphic to any static set, but must be described in terms of stochastic processes. This is a natural setting for NDT's that will be applicable to realistic risk management problems.
CHAPTER 4  THEORIES OF RISK MEASUREMENT FOR CLASSICAL DECISION MODELS

4.0 Introduction

The preceding chapter has surveyed several normative decision models for deriving preferences over acts from preferences for consequences. In this chapter and the next, we examine the problems of numerically "measuring", i.e., representing by numbers, that component of preferences that has to do with "risk".

This chapter reviews and criticizes past work on mathematical risk measurement. Most past work in this area has been done by mathematical psychologists interested in representing preferences for static risks, meaning probability measures over consequences. We will continue to work in this framework for the next three chapters; in Chapter 7, however, we turn to more realistic "dynamic" risks, represented by stochastic processes over consequences.

Section 4.1 introduces our basic framework for thinking about risk. We explicitly define the risk of an act as a component of preference, and later use the normative theories of preferences for acts surveyed in Chapter 3 to derive inferences about the mathematical structures of risk measures. Section 4.2 summarizes theories of ordinal and stronger risk measurement, based on some slight modifications of a review, synthesis, and extension of the state of the art by Fishburn (1984). The purposes of this section are

- To exhibit some alternative axiom systems for risk measurement that provide for ordinal and for expected utility representations without using the usual axioms of expected utility theory; and

- To show the sensitivity of the expected utility representation to different axioms by exhibiting several less restricted forms implied by less restrictive axioms.
Section 4.4 introduces the very important idea of "extensive measurement", in which an empirical operation as well as an ordinal comparative relation (e.g., the relation "is at least as risky as") is represented by the numbers on the measurement scale. This adds a much tighter mathematical structure to the measurement scale, typically making it cardinal instead of ordinal. After stating the relevant mathematical result (Theorem 4.4), we critically review some ingenious efforts in mathematical psychology to apply cardinal measurement scales to the objective numerical measurement of the risk of probability distribution functions over (continuous, single-attribute) consequences. We conclude that any such effort must ultimately fail: the set of internally consistent preferences over distributions (e.g., preferences that are representable by expected utility) is richer than the set of numbers that would be used to represent them, and any deterministic relation between distributions and a numerical index of their desirabilities will be over-determined. This observation sets the stage for Chapter 5's analysis of limited-precision objective numerical risk measurement.

4.1 Risk as a Relation Induced by R

Intuitively, a risk may be defined by four elements:

- Its SOURCE (risk from what?)
- Its TARGET (risk to what or to whom?)
- Its EFFECT (risk of what?)
- Its CAUSAL MECHANISM (the path, route, or method by which the source produces its effect on the target.)

For our purposes, the source of a risk will always be a human decision, act, or activity taken by or participated in by the d.m. The target may be the d.m. himself, as in the case of the risk of heart attack to a d.m. who chooses to participate in the activity of daily coffee drinking. Or
it may be a human population, as in the risk to neighbors of an industrial facility from the decision to operate the facility in a certain way. In reliability applications, the target might be a machine or a population of machines. The defining effect for a risk will usually be a change in the health status of one or more people, although a change in the operational status of a machine could be considered just as easily. We do not consider risk mechanisms.

Formally, then, a risk is specified by a triple \((A, B, C)\), where \(A\) is a set of acts; \(B\) is a set of individuals (usually interpreted as people, but possibly interpretable as a set of machines or facilities) affected by the choice of an act; and \(C\) is a set used to describe the effects of acts in \(A\) on the members of \(B\). For example, \(C\) might be a set of binary vectors showing which members of the population \(B\) live and which die as a consequence of the choice of an act \(a\) in \(A\). We will call \(B\) the target population for the risk. Given a triple \((A, B, C)\), let \(a\) be an act in \(A\) and let \(c\) be a consequence in \(C\) defined with respect to target population \(B\). Then the "risk" of \(c\) to \(B\) from \(a\) will be denoted by \(r(c; a)\), and the total "risk" associated with act \(a\) in this context will be denoted by \(r(a)\). We will give conditions under which \(r(a)\) is numerical and can be interpreted as an ordinal (or stronger) measure of an empirical comparative risk relation between acts.

Throughout, we interpret risk as a component of preference. For example, each act \(a\) in \(A\) may produce some deterministic costs and benefits, in addition to whatever (initially uncertain) effects in the risk effect set \(C\) it produces. Thus, the full set of consequences from an act may belong to the Cartesian product of several sets, e.g., a set of benefits, a set of economic costs, and so forth, of which the set \(C\) of risk effects is only one. However, we shall assume that preferences among outcomes (and lotteries over consequences, etc.) that differ only in the component corresponding to \(C\) depend only on the value of \(C\) in each outcome, and not on the common levels at which the other consequence components are fixed. Thus, the relation "is-at-least-as-risky-as" between acts is a subrelation on component \(C\) induced by the relation \(R = \)
"is-at-least-as-preferred-as" on the full set of consequences of act A. It is not necessary for the full relation $R$ to have been derived in order to study the risk relation, however: we assume that risk can be studied in isolation.

Our treatment of risk as a separable component of preference departs slightly from a tradition in mathematical psychology (Coombs and Huang, 1970; Krantz et al, 1971; Luce, 1980) that defines the "risk" of a monetary random variable as a number that, taken together with the mean of the random variable, completely determines preferences among such monetary random variables. Thus, risk is still a component of preference (among random variables), but in a slightly different sense than ours. However, there are mathematical difficulties with this definition, as will be discussed below, and we will stay with the interpretation of risk as a subrelation induced by an (actual or potential) preference relation on the full set of consequences when all consequence components outside $C$ are the same across the consequences being compared.

4.2 Measurement of Risk in a Single-Attribute Decision Model (Fishburn, 1984)

We shall now develop a theory of risk measurement for risk in the following simple decision model. The act-to-consequence mapping is assumed to work as follows:

(i) First, the d.m. chooses an act $a$ from a finite or infinite set $A$ of acts.

(ii) Secondly, an accident either occurs or doesn't occur. The probability of occurrence, denoted by $p(a)$ or simply by $p$, depends only on which act $a$ from $A$ has been chosen. Note that $p(a)$ denotes the probability of an accident given that act $a$ has been chosen, and not the probability of act $a$.

(iii) If an accident does occur, then its magnitude (or severity) will be
described by a random variable taking values in \( C = (0, M]\). The c.d.f. of its severity if act \( a \) has been chosen and the accident actually occurs will be denoted by \( F(a) \).

In this setting, a choice among acts is equivalent to a choice among pairs \([p(a), F(a)]\). Indeed, if we augment \( F(a) \) by extending it from the positive interval \((0, M]\) to the nonnegative interval \([0, M]\) with a point mass of \( 1 - p(a) \) at \( c = 0 \), then the choice among acts could be thought of as a choice among c.d.f.'s in the set \([F(a): a \in A]\). However, the formulation in terms of pairs \([p(a), F(a)]\), with \( F(a) \) conditional on the occurrence of an accident, is more convenient since it is compatible, for example, with the state-dependent preferences formulation of Example 3.3 (installing safeguards) with \( x = c \) and \( p(a) = p(s = 1; a) \). In fact, since \( p \) as well as \( F \) is allowed to be conditioned on \( a \), so that the likelihood of an accident as well as its consequences (if one occurs) might depend on the precautions taken, for example, this is a particularly useful model.

The mathematical problem of risk measurement in this framework is as follows. Let \( D \) denote the set of all c.d.f.'s over the interval \([0, M]\), where \( 0 \) is interpreted as zero loss from an accident and \( M \) is interpreted as the worst (or most severe) loss. Let \( L = [0, 1] \times D \) be the set of all pairs \((p, F)\) where \( p \) is in \([0, 1]\) and \( F \) is in \( D \). The set of acts \( A \) is isomorphic to a subset of \( L \), and may be thought of as being embedded in \( L \). Thus, the problem is to derive a preference ordering \( R \) over \( L \); the restriction of \( R \) to \( A \) then solves the choice problem. We will interpret the relation \((p, F) R (p', F')\) between two members \((p, F)\) and \((p', F')\) in \( L \) as saying that \((p, F)\) is "at-least-as-risky-as" \((p', F')\). Theorem 4.1 below provides for ordinal measurement of the "is-at-least-as-risky-as" relation on \( L \) by a numerical function \( r(p, F) \). We will let \( Q \) denote the corresponding strong relation "is strictly riskier than".

A priori, what might we be willing to say about the comparative risk relation \( R \)? Three axioms that seem intuitively reasonable in defining comparative risk in the above decision framework are as follows:
**F1 (Weak Ordering):** R is a weak order (connected and transitive) on L.

**F2 (Continuity):** If \((p,F) \preceq (q,G) \preceq (h,F)\) for any \(p, q, h \in [0,1]\) and any \(F, G \in D\), then there is some \(w\) in \([0,1]\) such that \((wp + (1-w)h,F) \sim (q,G)\), where \(\sim\) denotes the indifference relation corresponding to \(R\), i.e., \(a \sim b\) iff \(a R b\) and \(b R a\), for any \(a, b \in L\). In mixture notation, for some \(w\) in \([0,1]\) (interpreted as either a mixture probability whose resolution is unobservable or as a nonprobabilistic weight in a deterministic convex combination) \([\{(p,F),w,(h,F)\}] \sim (q,G)\). Similarly, if \((p,F) \preceq (q,G) \preceq (p,H)\), then for some \(w\) in \([0,1]\), \([\{(p,F),w,(p,H)\}] \sim (q,G)\).

**F3 (Monotonic Preferences):** If \(p\) and \(q\) are two numbers in \([0,1]\) with \(p\) greater than \(q\), then \((p,F) \preceq (q,F)\) for any \(F \in D\). Here, as usual, \((p,F) \preceq (q,F)\) means \((p,F) R (q,F)\) and not \((q,F) R (p,F)\). Similarly, if \(c\) and \(c'\) are any two outcomes in \(C *= [0,M]\) such that \(c\) is greater than \(c'\), then \(c R c'\). Here, \(c\) and \(c'\) are considered as degenerate members of \(L\), i.e., \(c = (1,d(c))\), \(c' = (1,d(c'))\), where \(d(c)\), for example, is the cumulative distribution function that jumps from 0 to 1 at \(c\).

Axiom F3 can be summarized by saying that preferences are monotonically decreasing in both probabilities and consequences.

**Theorem 4.1 (Fishburn, 1984):** Axioms F1 to F3 imply that \(R\) on \(L\) can be represented by a continuous functional \(r(.,.)\), monotonically increasing in its first argument, assigning numbers \(r(p,F)\) to points \((p,F)\) in \(L\) in such a way that for any two points \((p,F)\) and \((q,G)\) in \(L\), \(r(p,F)\) is at least as great as \(r(q,G)\) if and only if \((p,F) R (q,G)\). Any such \(r(.,.)\) is unique only up to monotonic increasing transformations.

Proof: Fishburn, 1984, p. 399 provides a simple constructive proof.

Theorem 4.1 shows that the assumptions of weak ordering, continuity, and monotonicity suffice for ordinal measurement of the preference relation \(R\) by a continuous monotonic functional \(r(.,.)\). Note that the \(r(.,.)\) of this construction plays essentially the same role as the \(V(.)\) functional.
in single-attribute theory, namely, to provide an ordinal indication of preference for c.d.f.'s over \([0,M]\). Recall from Theorem 3.3 that weak ordering and continuity axioms such as F1 and F2 generally suffice for ordinal measurement; here, they have been augmented by a monotonicity axiom, F3, that guarantees that \(r(p,F)\) is increasing in \(p\).

To go further, it is necessary to introduce some additional qualitative assumptions about the comparative risk relationship \(R\). Since \(r(.,.)\) is a deterministic ordinal value function on the two-component product space \(L = [0,1] \times D\), it is natural to look at the axioms used in the construction of two-attribute value functions (see e.g., Keeney and Raiffa, 1976.) Even though \(D\) is a set of c.d.f.'s over an interval, and therefore has a different mathematical structure than a numerical attribute, the theory of ordinal value functions for attributes suggests the following axiom:

\[
F4 \text{ (Preferential Independence): For any two accident probabilities } p \text{ and } q \text{ in } [0,1] \text{ and any } F \text{ and } G \text{ in } D, \ (p,F) R (p,G) \text{ if and only if } (q,F) R (q,G).
\]

Interpretively, preferences among the conditional c.d.f.'s over losses given an accident are independent of the probability of an accident, i.e., of the level at which the first attribute (accident probability) is fixed. To obtain the standard result from two-attribute value theory of multiplicative separability (see e.g., Keeney and Raiffa, 1976), it is also necessary to make the following technical assumption, known in the measurement theory literature as "Thomsen's condition" (see e.g., Krantz et al, 1971):

\[
F5 \text{ (Thomsen's Condition): For any } p, q, w \text{ in } [0,1] \text{ and any } F, G, H \text{ in } D, \text{ if } (p,F) I (q,G) \text{ and } (q,H) I (w,F), \text{ then } (p,H) I (w,G).
\]

Unlike the preceding conditions, there is no obvious intuitive interpretation for F5. But it is implied by any representation of the form \(r(p,F) = v(p)w(F)\), to which we now turn.
THEOREM 4.2 (Fishburn, 1984): If the risk function \( r(p,F) \) satisfies axioms F1 to F5, then it can be decomposed as

\[
r(p,F) = v(p)w(F),
\]

where \( v(p) \) is continuous and increasing in \( p \), \( v(0) = 0 \), and \( w(d(c)) \) is continuous and increasing in \( c \) over the interval from 0 to \( M \) (where \( d(c) \) is the c.d.f that jumps from 0 to 1 at \( c \)). Moreover, if we normalize \( v(.) \) and \( w(.) \) by setting \( u(1) = v(1) = 1 \), then the resulting "canonical" risk components, denoted by \( v^*(.) \) and \( w^*(.) \), respectively, representing any given preference ordering \( R \) are unique up to a power transformation (i.e., any \( v \) and \( w \) obtained by raising \( v^* \) and \( w^* \) to a positive power will also represent \( R \) on \( L \), and only such pairs of functions will represent \( R \) on \( L \).)

Proof: This is a restatement of Theorem 2 of Fishburn, 1984, which in turn derives from conjoint measurement theory (Krantz et al, 1971), as does the usual two-attribute value theory for separable value functions presented in Keeney and Raiffa, 1976. Note that if \( r(p,F) = v(p)w(F) \) is an ordinal measure of preferences for \( (p,F) \) pairs on \( L \), then so is \( \ln[r(p,F)] = \ln[v(p)] + \ln[w(F)] \); hence, we can present the result in terms of either additive or multiplicative separability. In either case, the Thomsen condition and the other assumptions are necessary as well as sufficient. We have chosen the multiplicative form as giving a more intuitive normalization, with \( r(0,F) = 0 \) rather than minus infinity.

To be compatible with an expected utility representation, we would have to have \( v(p) = p \) and \( w(F) = E[u;F] \) for some utility function \( u(c) \) on \([0,M]\). Theorem 4.2 is evidently consistent with such a representation, but allows other forms as well. However a slight strengthening of the continuity axiom F2 together with the usual independence or "substitution" property from Model 1 and a reasonable assumption about stochastic dominance produce an expected utility representation.
THEOREM 4.3 (Fishburn, 1984): Suppose that the comparative risk relation R on L satisfies the following axioms:

F1 (Weak Order): R is a weak order on L.

F2* (Strong Continuity): If a, b, and c are any three members of L such that a \(\leq\) b \(\leq\) c, then there is some w in (0,1) such that (a,w,c) I b, where (a,w,c) stands for wa + (1-w)c. (Note that each member of L is a pair of the form (p,F).)

F3 (Monotonic Preferences): If p and p' are two numbers in [0,1] with p greater than p', and if c and c' are two numbers (representing consequences) in [0,M] with c greater than c', then (p,F) Q (p',F) for any F in D, and (q,d(c)) Q (q,d(c')) for any q in [0,1].

F4* (Substitution Independence): For any a, b, and c in L and any mixture weight w in (0,1], a \(\leq\) b if and only if (a,w,c) Q (b,w,c). Here, (a,w,c) denotes the convex combination aw + (1-w)c.

F5* (Stochastic Dominance): For any two members a and b of L, b FSD a implies a R b. (Recall that (p,F) FSD (q,G) if and only if \((1 - p) + pF(c)\) is at least as great as \((1 - q) + qG(c)\) for every c in (0,M].)

Then R is represented on L by a risk function of the form

\[ r(p,F) = pE[u;F] \]

where u(.) is a nonnegative increasing "disutility" function on [0,M], unique up to scalar multiplication.

Proof: This follows from Theorem 5 of Fishburn, 1985. Our axiom F5* implies Fishburn's axiom C1*, and our assumption that severity is bounded from above by a constant M implies his axiom C2* (set y = M.)
Theorem 4.3 gives an expected (dis)utility representation for pairs \((p,F)\) in \(L\). This axiomatization gets rid of the unintuitive Thomsen condition, which is implied by the remaining axioms. If the conditions of the theorem are accepted, then choice among acts in \(A\) can be made by computing \(r(a) = r[p(a),F(a)]\) for each \(a\) in \(A\), and then choosing an act \(a^*\) with the smallest risk \(r(a^*)\). More generally, if the acts in \(A\) have additional consequences in sets other than the risk effect set \(C = [0,M]\), and if preferences are separable with respect to \(L\), then the number \(r[p(a),F(a)] = p(a)E[u;F(a)]\) will be "sufficient" for the pair \([p(a),F(a)]\), in the sense that a d.m. with risk preferences represented by \(R\) on \(L\) will make the same decision knowing only \(r(a)\) as if he knew \([p(a),F(a)]\).

4.3 Stronger Forms of Risk Measurement In the Single-Attribute Case

There have been several attempts in the mathematical psychology literature to restrict the form of the risk function \(r(p,F)\) further in the above decision framework, in which choices among acts are tantamount to choices among \([p(a),F(a)]\) pairs. In this subsection, we briefly review this literature, and then argue that such efforts are in general inappropriately restrictive in the range of risk preferences that they can represent.

4.3.1 Introduction: Extensive Measurement

One way to strengthen ordinal measurement is to look for properties other than order that the assignment of numbers should preserve. For example, in measuring subjective probabilities (Theorem 2.1), we required that if \(E\) and \(E'\) are disjoint events in the algebra \(F\), then \(p(E u E')\) should equal \(p(E) + p(E')\), i.e., the numbers \(p(E)\) assigned to events \(E\) in \(F\) should not only reflect the comparative ordinal relation "is-judged-to-be-at-least-as-likely-as" between events, but should also be made in such a way that the operation of disjoint union between events is represented by the operation of addition between the numbers representing those events. Similar comments apply to the measurement of
physical properties such as length or weight, where concatenation of rods or stacking of weights on a balance pan are natural empirical operations that should be correctly represented by addition of the numbers assigned to the rods or weights, respectively, in the measurement process.

More generally, let \((O,Q,*\)) be an empirical system, where \(O\) is a set of "objects", e.g., rods, weights, or events; \(Q\) is a comparative relation allowing ordinal comparisons among the members of \(O\); and \(*\) is an empirical operation such that \(O\) is closed under \(*\). Let \(r(.)\) be a function mapping the objects in \(O\) into numbers in such a way that for any two objects \(a\) and \(b\) in \(O\),

(i) \(r(a)\) is greater than \(r(b)\) if and only if \(a Q b\). Thus, \(r(.)\) is an ordinal measure of \(R\) on \(O\).

(ii) \(r(a * b) = r(a) + r(b)\).

Then \(r(.)\) is said to be an extensive measure, or to provide for **extensive measurement** of \(R\) on \(O\).

The importance of extensive measurement is indicated by the following result.

Let \((O,Q,*)\) be an empirical system such that

(i) \((O,*)\) is a group (i.e., \(*\) is associative on \(O\), \(O\) has an identity, and each element of \(O\) has an inverse under \(*\));

(ii) \((O,Q)\) is a simple order (i.e., \(Q\) is complete and transitive on \(O\) and asymmetric, so that \(a Q b\) implies not \(b Q a\), for any \(a, b\) in \(O\));

(iii) Let \(e\) denotes the identity element in \(O\), i.e., \(a*e = a\) for any \(a\) in \(O\). Then for any two elements \(a\) and \(b\) in \(O\) such that \(a Q b\), the \(n\)-fold concatenation of \(b\) with itself, denoted by \(nb = b*b*...*b\) (\(n\) times) satisfies \(nb Q a\). That is, no element in \(O\) is "infinitely greater than"
any other in terms of the comparative relation $\mathcal{Q}$. (This is sometimes called an "Archimedean axiom").

(iv) For any $a$, $b$, and $c$ in $0$, $a \mathcal{Q} b$ if and only if $(a * c) \mathcal{Q} (b * c)$.

This is called the "monotonicity" axiom. For example, if $0$ is a set of lotteries, $*$ an operation of $p$-mixture (so that there is actually a continuous family of operations), and $\mathcal{Q}$ represents a strict preference relation, then the monotonicity axiom becomes the substitution axiom for preferences among lotteries.

Then $(0, \mathcal{Q}, *)$ is called an "Archimedean ordered group".

**THEOREM 4.4 (Holder, 1901):** Every Archimedean ordered group is isomorphic to the set of real numbers under the comparative relation of strict inequality and the numerical operation of addition. That is, $(0, \mathcal{Q}, *) \cong (\text{Reals}, \text{Inequality}, +)$ for any Archimedean ordered group $(0, \mathcal{Q}, *)$.

**Proof:** Krantz et al, 1971.

Since ordered Archimedean groups are realistic models for a wide range of empirical relational systems, including distance and weight measurement (but not classical temperature or density measurement, for example), Theorem 4.4, which is the basic theorem of extensive measurement, is widely applicable.

### 4.3.2 Extensive Measurement of Risk

To apply the idea of extensive measurement to risks, it is necessary to introduce an empirical operation $*$ between the objects of risk comparison. We shall continue to assume that these objects are pairs of the form $(p,F)$ in $L$, and shall take the operation of convolution between members of $L$ as the operation $*$. That is, if $(p,F)$ and $(q,G)$ are two members of $L$, then their convolution, denoted by $(p,F)*(q,G)$ and defined by $(p,F)*(q,G) = (1-(1-p)(1-q), F*G)$, where $F*G$ is the convolution...
integral of $F$ and $G$, is also assumed to be in $L$. (Hence, our usual bounded interval $[0,M]$ must be replaced by the unbounded interval $[0,\infty)$.) Interpretively, if $X$ any $Y$ are two random variables taking values in $[0,\infty]$ according to the c.d.f.'s $(1-p) + pF(c)$ and $(1-q) + qG(c)$, respectively, then $(p,F)*(q,G)$ is the member of $L$ corresponding to $X + Y$. The operation $*$ is known to be associative and commutative.

**Theorem 4.5:** Suppose that the comparative risk relation $R$ on $L$ satisfies the following axioms:

**KR1 (Weak Order):** $(L,R)$ is a weak order (complete and transitive);

**KR2 (Monotonicity):** For any $a$, $b$ and $c$ in $L$, $a R b$ if and only if $(a * c) R (b * c)$;

**KR3 (Archimedean):** For any $a$, $b$, $c$, and $d$ in $L$, $a R b$ implies that $(na * c) R (nb * d)$ for sufficiently large $n$.

Then there exists an extensive risk measure $r(p,F)$ on $L$, and $r(\ldots)$ is unique up to choice of scale.


The strength of the extensive measurement structure is made most clear by examining the consequences of the following additional mild "regularity" assumptions.

**KR4 (Monotonic Preferences):** For any $c$ in $[0,\infty]$ and any $a$ in $L = [0,1] \times D$, $a R [a * (1,d(c))]$. Thus, adding a sure loss of $c$ to the probabilistic distribution of losses if an accident occurs never reduces risk.
KR5 (Continuity): If \( a(n) \) is a sequence of points in \( L \) that approaches \( a \) (in the topology of weak convergence for the corresponding c.d.f.'s) as \( n \) increases to infinity, then \( r[a(n)] \) approaches \( r(a) \).

Here, the c.d.f. corresponding to the point \((p,F)\) in \( L \) is the one with value \( p + F(c) \) at point \( c \) in \([0,\infty)\). Actually, axiom KR5 can be weakened to apply only to sequences in which the corresponding c.d.f.'s all have the same mean and variance, and the following result will still hold.

**THEOREM 4.6:** If \((L,R,*)\) satisfies axioms KR1 to KR5, then \( R \) is represented by a risk function of the form

\[
r(a) = wE(a) + (1-w)Var(a)
\]

for some \( w \) in \((0,1)\), where \( E(a) \) and \( Var(a) \) are the mean and variance of the random consequence \( c \) having the c.d.f. over \([0,\infty)\) corresponding to point \( a \) in \( L \). In mixture notation, \( r(a) = [E(a),w,Var(a)] \).

**Proof:** This result follows directly from Theorem 13, p. 127 of Krantz et al., 1971, which is again based on a refined version of the Pollatsek-Tversky (1970) model. We have not had to explicitly assume the axiom that Krantz et al call "scalar monotonicity", which applies to actuarily fair lotteries, since all our consequences lie in \([0,\infty)\).

Theorem 4.6 is amazingly strong. It implies that the judged "risk" associated with any probability distribution over consequences is a weighted average of the mean and variance of the distribution, where the weight depends only on which ranking \( R \) satisfying KR1 to KR5 is being represented. As one consequence, Theorem 4.6 implies that any loss distribution that has both a higher mean and a higher variance than a second distribution will be judged "riskier" than the second distribution by all d.m.'s.
We shall soon argue that this type of result is too strong, in that it rules out a wide variety of perfectly "rational" risk preferences. Before expressing this criticism, however, we shall summarize a few other results of the same type.

Luce (1980), citing research by Coombs and Bowen (1971), notes that the Pollatsek-Tversky model of Theorem 4.6 is empirically incorrect, in that individuals do not in fact consider equally risky all distributions with equal means and variances. He therefore introduces an alternative axiom, which we will call "scale invariance", essentially stating that if two distributions \( d \) and \( d' \) over consequences are considered equally risky, then the corresponding distributions when each consequence is scaled up or down by a constant factor are also considered equally risky. For example, if two normal distributions over dollar payoffs are judged equally risky, then the same distributions over cents instead of dollars will also be considered equally risky. Under this assumption, and assuming also that risk has an expected (dis)utility representation as \( r(a) = E[u;a] \), Luce shows that \( r(a) = E[c^w;a] \), and that this risk measure is unique up to choice of scale. In other words, \( u(c) = kc^w \) for some \( k \) greater than 0.

Finally, in a careful elaboration on this theme that again exploits a scale invariance axiom, Luce and Weber (1984) establish essentially the following form for the risk measurement function:

\[
r(p,F) = (1-p) + p[a + bE(c^w_F)],
\]

where \( a, b, \) and \( w \) are parameters that depend on the ordering \( R \) to be represented.

4.3.3 Critique

The above types of results are strong enough to rule out many apparently reasonable risk preferences. Our view is that they are too strong: that any axiomatization that reduces the set of possible orderings \( R \) over
points in L to a one-parameter or few-parameter family is too restrictive. We shall return to this point shortly.

The above axioms can be criticized more simply on a case-by-case basis. For example, the Pollatsek-Tversky model (Theorem 4.6) is based on the assumption that if X and Y are two random variables taking values in the numerical consequence set \([0, \infty)\), then it is reasonable to require that \(r(X + Y) = r(X) + r(Y)\), where \(r(X)\), for example, denotes the risk assigned to the c.d.f. representing X. But is this really reasonable? Letting \(r(c)\) denote the risk of the degenerate c.d.f. \(d(c)\) putting probability \(1\) on outcome \(c\), we see that the axiom implies that \(r(2c) = r(d(c) \cdot d(c)) = 2r(c)\) for all \(c\), i.e., the d.m.'s value function over consequences must be linear. This means that an additional loss is not valued any more or less when it is added onto a large base line loss than when it is added onto a small one, so that the equivalent of "wealth effects" (i.e., increasing or decreasing marginal disutilities, in the economic sense, in \(c\)) are ruled out a priori. But there is no intuitively obvious reason why the normative theory should not apply to preferences exhibiting such nonlinearities, especially if the payoff is in dollars.

On a more abstract level, the requirement that the convolution operation be represented by addition is disturbing because it is not invariant under transformations of the consequence axis. For example, if the severity axis is changed from \(c\) to \(c' = \ln c\), then the requirement that \(r(a' + b') = r(a') + r(b')\), would imply very different empirical behavior than \(r(a + b) = r(a) + r(b)\). (Here, primes denote values on the logarithmic scale, and \(a\) stands for the c.d.f. \(d(a)\).) The two sets of indifferences among c.d.f.'s implied by the additivity of \(*\) on two different scales are inconsistent. The problem is that one can not simultaneously require additivity on more than one nonlinearly related scale. The convolution additivity axiom of the Pollatsek-Tversky model, however, fails to specify which among all possible scales additivity is to hold on, and thus leaves itself open to inconsistencies.
Luce's scale invariance axiom can be similarly attacked. The assumption that indifferences between lotteries are preserved under arbitrary rescalings of payoffs violates common risk attitudes that are not ordinarily considered undesirable or irrational. For example, it implies that if the certainty equivalent of a 50-50 gamble between -$0.10 and $0.10 is \( \text{CE}(-0.10, 0.5, 0, 10) = 0 \), then \( \text{CE}(-100, 0.5, 100) = 0 \) also -- a proposition that few people would agree with. Similar restrictions apply even when the range of gambles is nonnegative. The scaling invariance axiom also can not hold simultaneously on multiple, nonlinearly related scales, e.g., on both a linear and a logarithmic scale for some variable (such as peak ground acceleration in measuring the severity of an earthquake.) As for the additive convolution axiom in the Pollatsek-Tversky model, one cannot consistently require the axiom to hold on different scales, but no special scale on which it is presumed to hold is specified. (For money, of course, there is a natural scale, but we are interested in more general applications.)

These criticisms are specific to the particular axioms proposed. However, as indicated above, a more general argument can be offered to show that any axiom system, past or future, that reduces the set of possible risk rankings of c.d.f.'s to a one-parameter (or few-parameter) family can be subjected to some such criticism. Let \([R]\) denote the set of all comparative risk relations on \(D\), the set of c.d.f.'s over \([0,M]\).

Under a reasonable set of restrictions, each ordering \(R\) in \([R]\) can be represented by an expected disutility function, \(u(c)\), so that \(F R G\) if and only if \(E[u;F] \geq E[u;G]\), for any two c.d.f.'s \(F\) and \(G\) in \(D\). Sufficient conditions for this representation have been presented in this chapter. We shall assume that preferences are monotonically decreasing in \(c\), so that preferences over \(c\) can be represented by a continuous, strictly increasing disutility function \(u\). (We shall use the terms "utility function" and "disutility" function interchangeably, since we care only about the expected value representation.)
Each distinct ordering $R$ is represented by a unique continuous canonical utility function taking values in $[0,1]$. Let us temporarily use the notation $u(R)$ to denote this function, and let $R(u)$ denote the unique ordering of c.d.f.'s implied by $u$. The cardinality of $[R]$ is evidently the same as the cardinality of the set of all continuous monotonically increasing canonical utility functions representing orderings. But this set is much larger than the set of points in $[0,1]$, i.e., than a one-parameter family. For example, consider the family of canonical functions

$$u(c) = \frac{1 - \exp(-wc/M)}{1 - \exp(-w)},$$

all of which run from 0 at $c = 0$ to 1 at $c = M$. This is a one-parameter family of functions, indexed by $w$, corresponding to a one-parameter set of orderings $R(u(w))$. The members of any other one-parameter family of orderings, such as those in the Pollatsek-Tversky model, can be put into one-to-one correspondence with the members of this family. But the above family of functions, known in the decision analysis literature as "exponential utility functions" displays a qualitative property, known as constant risk aversion (which implies, for example, that the buying and selling price of any lottery will agree) not shared by other utility functions. In fact, it is easy to identify other one-parameter families (e.g., those with constant proportional risk aversion) that imply different qualitative behavior, e.g., that the buying price for any lottery will be less than its selling price.

Thus, any one-parameter family of orderings $R[u(w)]$ must contain an essential restriction. It can contain (exactly) all orderings corresponding to constant risk aversion, or all orderings corresponding to constant proportional risk aversion (for example), but it can not contain both. Thus, if we want a normative theory that will apply to at least these two classes of preferences, we must reject the hypothesis of a one-parameter family of risk orderings.
This argument can be extended to more than one parameter. In general, a useful normative theory must be applicable to a reasonably wide range of preferences, corresponding to a reasonably wide range of utility functions. A theory that reduces risk preference to one or a few parameters must necessarily rule out many families of utility functions corresponding to apparently reasonable preferences that should be included within the scope of a useful normative theory. Indeed, many would argue that expected utility theory, which reduced the cardinality of the set \([R]\) of representable preferences to the cardinality of the set of all continuous monotonically increasing functions on \([0,1]\), is itself too restrictive and needs to be expanded, for example by introducing Machina's "local" utility functions, in order to be fully useful. One-parameter or few-parameter models of risk preferences will then be a fortiori unacceptable.

In summary, the mathematical psychology tradition that attempts to represent preferences among c.d.f.'s by a one-parameter family of risk functions, or that seeks to define risk as the one parameter which, in addition to expectation, explains preferences among c.d.f.'s, is incompatible with the expected utility representation (or its generalizations) combined with the assumption that the scope of the risk measurement theory should include at least several different one-parameter classes of utility functions, corresponding to different qualitative behaviors.
CHAPTER 5  NEW THEORIES OF SINGLE-ATTRIBUTE RISK MEASUREMENT

5.0 Introduction and Overview

Chapters 2 to 4 have provided the technical background on representation of probabilities and preferences needed to develop theories of risk measurement. The principal contribution of this chapter is a new approach, developed in Sections 5.2 to 5.4, to measurement of risk in the classical case where acts are represented by probability distribution functions over a continuous single-attribute consequence set. We assume that the r.a. knows the distributions corresponding to different acts and that he is seeking a measure of risk that will allow him to communicate at least some of this information to a d.m. This measure should ideally be objective, in the sense that it is not necessary for either the r.a. or the d.m. to know the other's subjective risk attitude in order for both to understand what the risk measure announced by the r.a. means.

The basic analytic approach is as follows. We first recognize that the principal challenge in contriving objective numerical measures of risk is overcoming ambiguity. For example, when a r.a. expresses his own expected utilities for different distributions, the resulting numbers confound his subjective risk attitude with objective aspects of the distributions, making the final number impossible to interpret unambiguously for someone who doesn't know the r.a.'s utility function. We therefore focus on finding reference points that all d.m.'s will agree on, and then use these reference points to calibrate objective scales for risk measurement.

This approach is carried out as follows. Recall from Chapter 3 that one consequence cumulative probability distribution function (c.d.f.) preference dominates another in the sense of first-order stochastic dominance (FSD) if and only if all d.m.'s with increasing utility functions prefer the first to the second. The FSD criterion applies equally well to undesirable consequences upon reorientation of the consequence scale (replace the negative consequence attribute c with the
positive attribute $M - c.$) Thus, all d.m.'s will agree on the preference ordering of probability distributions that stand in the FSD relation to each other.

For a numerical risk scale to be "objective" by our definition it is necessary that all d.m.'s agree that higher values are dispreferred to (or more risky than) lower values. This unanimity requirement suggests associating higher risk values with stochastically dominated distributions. More specifically, a continuous numerical objective risk scale that represents an objective ordering of consequence distributions must index a continuous one-parameter family of such distributions that is totally ordered by FSD. This is the necessary and sufficient condition for all expected utility decision makers who agree on the ordering of deterministic consequences (i.e., having monotonic utility functions over the consequence attribute) to agree that risk is unambiguously increasing along the risk scale, given that the scale indexes (or assigns numbers to) consequence distributions. Any continuous one-parameter family of consequence distributions totally ordered by FSD is said to be a CANONICAL RISK SCALE (CRS) calibrated by that family of distributions, where higher parameter values correspond to stochastically dominated distributions within the family.

Any CRS provides for objective ordinal risk measurement within the family of distributions used to calibrate it. Moreover, any CRS that runs from zero to infinity (e.g., the family of exponential distributions indexed by its parameter) exhausts the representational power of the nonnegative real line, in the sense that there is a one-to-one correspondence between the nonnegative numbers and the distributions in this family. However, there are many such CRS scales, and it is impossible to align or scale them so that every d.m. will simultaneously agree that the same values on all the different scales represent the same degree of risk. For example, let CRS' be the family of exponential distributions and CRS" a family of (nonexponential) gamma distributions indexed by the first parameter with the second parameter fixed. Then it is impossible to align the scales so that two or more d.m.'s with different utility functions will be
indifferent between each exponential distribution on CRS' and the gamma
distribution having the same numerical value on CRS".

This is an important point. Another way of making it is as follows: let
CRS' and CRS" be two canonical risk scales, calibrated by two different
families of distributions, and consider the indifference curves in the
quadrant with CRS' as the horizontal axis, CRS" as the vertical axis, and
a common origin representing zero risk. Then no matter how these two
axes are transformed (so long as order is preserved on each), it will be
impossible to make the indifference curves for d.m.'s with different
utility functions coincide. This means that objective numerical risk
measurement as we have defined it is impossible. No numerical risk scale
can be constructed such that if each consequence distribution is assigned
a unique number on it, then (i) All consequence distributions that are
assigned the same number are judged to be indifferent (or equally risky)
by all expected utility d.m.'s; and (ii) Distributions that are
unanimously dispreferred (or judged more risky) by all expected utility
d.m.'s are assigned higher numbers. In fact, objective numerical risk
measurement in this sense is impossible even if the set of d.m.'s and the
set of allowable distributions are sharply restricted, e.g., to d.m.'s
with exponential utility functions and to the set of normal
distributions. This impossibility result is the essential content of
Theorem 5.4.

Although objective numerical risk measurement is impossible, useful
approximate (interval-valued) and nonnumerical risk measures can be
constructed. For example, the FSD relation can be used to establish
upper and lower bounds, measured on a CRS, for distributions not in the
calibrating family for the CRS. This permits "limited-precision"
comparison of arbitrary distributions on a CRS. Section 5.4 develops
this idea and proves that some common parametric families, such as the
Poisson, exponential, gamma (with one fixed parameter) and normal (with
proportional mean and variance) can be used to calibrate canonical risk
scales that are unique up to choice of unit. The CRS construction is
also extended to risks of the form \((p,F)\), studied in Chapter 3.
The chapter concludes with a proposed form of nonnumerical risk measurement in which the risk of a distribution is represented by a "canonical risk curve" plotting risk on a CRS against a measure of risk attitude (the certainty equivalent of a standardized risky prospect) for an exponential utility function. Several advantages of this proposed representation are listed, including the fact that any expected utility d.m. can in principle calculate his own expected utility from it.

In addition to these major themes, this chapter contains several special topics and theoretical developments of independent interest. For example, Section 5.3, which presents the impossibility result, establishes it in the context of risk measurement for normal distributions, and establishes several useful results that apply specifically to the family of normal distributions. In addition, the concepts of second-order stochastic dominance and comparative "riskiness" traditionally used in economics and finance are reviewed.

Section 5.2 presents an important technique, known as difference measurement, that provides the theoretical underpinnings for many of the specific results in this chapter, including those for normal distributions, and that is used in Chapter 8 in developing a theory of measurement for the values of statistically independent fatalities. Difference measurement is the third and last major type of measurement (besides ordinal and extensive measurement) studied in mathematical measurement theory, and provides an alternative to extensive measurement for deriving a cardinal scale. It allows us to show that two expected utility d.m.'s who agree on their preferences for deterministic consequences (including preference intensities as well as preference orders) will under plausible conditions differ from each other only in their degree of "relative risk aversion", which can be represented on a single numerical scale. Thus, the indifference curves for such d.m.'s on any pair of CRS axes will belong to a one-parameter family of indifference curves.
This pins down exactly how much ambiguity must be overcome by a measure of risk. Two d.m.'s can communicate without knowing each others' risk attitudes by sending messages expressed as a function of the single numerical parameter indicating relative risk aversion. Thus, messages more complicated that the canonical risk curve are unnecessary, while less complicated (e.g., numerical) messages will in general leave some ambiguities unresolved.

The use of difference measurement applied to preferences for consequences to establish a cardinal value scale (i.e., a scale unique up to choice of scaling factor) fills a logically necessary role. For a probability distribution over a consequence scale to be well defined, it is necessary that the scale be "sufficiently rigid". For example, on a consequence scale that is purely ordinal, the distinction between any two distributions on the same interval is not meaningful. A monotonic transformation of the consequence scale can turn the cumulative distribution function (c.d.f.) for one into the c.d.f. for the other. If consequences are not structured by any natural "concatenation" operation, as in extensive measurement, but are only distinguished by the d.m.'s preferences, then difference measurement allows enough structure to be imposed on the consequence scale so that probability distributions can be meaningfully defined.

While most of this chapter is concerned with the conventional case of a continuous single-attribute consequence, for which considerations of the sort just mentioned are relevant, there are other mathematical settings that are at least as important for modelling risks. One is the case of dichotomous consequences, which are useful in modeling a variety of "catastrophic" risks in which the quantity of interest is the time until the (first) arrival of a "marker" event (e.g., the death of an individual, the failure of a safety system, and so forth.) This setting is treated in detail in Chapter 7. Section 5.1 below introduces a simple theory of risk measurement for dichotomous consequences based on extensive measurement, where the empirical operation of concatenation of risk exposures is represented by addition of risk numbers. This leads to
a definition of a proposed unit of measurement -- the "mort" -- for catastrophic risks. A probability of \( p \) for a dichotomous marker event corresponds to \(-\ln(1 - p)\) "morts" of risk for the event. Morts (and "micromorts", defined as millionths of a mort) provide an intuitively convenient alternative to probability for measuring catastrophic risk. Although there is nothing profound about this transformation, morts have the advantage over probability that they are measured on a cardinal scale that is closed under addition; thus, ideas such as "twice as much risk" make sense for morts, while the phrase "twice as probable as" are uninterpretable (as is obvious for probabilities greater than 0.5.) For small risks, however, the magnitude of a catastrophic risk will be virtually the same whether it is expressed in morts or in probability. Chapter 7 uses flows of morts over time as a convenient intuitive representation of dynamic risks.

5.1 Risk Measurement for Dichotomous Consequences

Let \( p \) represent the probability of a personally or socially catastrophic accident (e.g., the d.m.'s death) whose consequence, if it occurs, has only one distinguishable level, or whose range of potential consequence severities is small enough to be uninteresting by comparison with the fact of occurrence. Let \( r(p,d) \) denote a numerical measure of "risk" for such an accident, where \( p \) is its probability of occurrence and \( d \) is its magnitude if it does occur. \( d \) is assumed to be the same for all risks being compared, so \( r(p,d) \) effectively depends only on \( p \).

Given any two pairs, \((p,d)\) and \((p',d)\), with \( p \) larger than \( p' \), \( r(.,..) \) will accurately represent preferences between them if and only if \( r(p,d) \) exceeds \( r(p',d) \); thus, \( r(p,d) = p \) is itself a satisfactory ordinal measure of risk for pairs \((p,d)\). However, any monotonically increasing transformation of \( p \) will also be a satisfactory ordinal risk measure, and some may be more intuitively useful or informative than others or than \( p \) itself. In particular, it is known that most people have difficulty thinking about or intuitively understanding very small probabilities, and transformations of the probability scale (e.g., to an odds or log-odds
scale) have been shown to be useful in making quantitative expressions of likelihood more understandable.

Borrowing the idea of extensive measurement, we seek an empirically meaningful operation that might reasonably be thought to "double" risk, and then use this operation to narrow the range of allowable representations for risks. To this end, suppose that risks are generated by "arrivals" in a stochastic process. Specifically, assume that the occurrence of a catastrophe corresponds to the first arrival in a counting process \( N(t) \), where \( t \) is an index of cumulative exposure (most conveniently, but not necessarily, though of as time.) When \( N(t) \) jumps from 0 to 1, an accident has occurred.

With little loss of generality, we can assume that arrivals are generated by a homogeneous Poisson process, possibly after transformations of the \( t \)-axis to make it linear in the hazard rate (or instantaneous arrival intensity) \( h(t) \). (Here, \( h(t)dt \) may be loosely interpreted as the conditional probability that the first arrival occurs in the interval \( (t,t+dt] \), given that it has not occurred in \( [0,t] \), for sufficiently small \( dt \).) Given a total exposure duration \( T \), the probability of an arrival in \( [0,T] \) is \( 1 - \exp(-hT) \), where \( h \) is the arrival rate in the Poisson process.

If the risk of a catastrophic accident is generated by exposure to some hazard, then it may be reasonable to define total risk as proportional to total exposure. If time is the exposure variable, for example, then it might seem intuitively correct to define risk to a target as doubling when its time at risk is doubled.

Example 5.1: The Perils of Paul

Paul is tied to a railroad track, waiting for his young lady to rescue him. He knows that it will take her exactly one hour to get to him. Meanwhile, trains arrive according to a Poisson process with an average rate of \( h \) trains per hour. The probability that a train will arrive before he is rescued is \( 1 - \exp(-h) \). Let \( r[1 - \exp(-h)] \) denote the magnitude of the risk to Paul associated with this probability of death.
A week later, Paul (who was successfully rescued the first time) is again tied to the track; this time, however, he knows that rescue will not arrive for two hours. (Equivalently, he may still be at risk for only an hour, but train traffic intensity may have doubled.) Then we can define his risk of death to be twice as great as on the previous occasion: \( r[1 - \exp(-2h)] = 2r[1 - \exp(-h)] \).

**Example 5.2: Russian Roulette**

Similar examples can be given for a discrete exposure variable. Let \( r(p) \) be the risk of death to a man who is given a revolver with \( 1/p \) chambers and forced to play one round of Russian roulette with it. Compare this to the risk facing a man who will be forced to play a round of \( (1/p) \)-chamber Russian roulette, and then -- if he survives -- to play another round with a second identical, freshly loaded, \( (1/p) \)-chamber revolver. Then the risk in the second case, namely, \( r[1 - (1-p)(1-p)] \), may be defined as twice the risk in the first case.

If the operation of concatenation of risk exposures is to be represented by addition of corresponding risk numbers, then the risk measure \( r(p) \) must satisfy the functional equations (see e.g., Aczel, 1966)

\[
r[1 - \exp(-hT)] = Tr[1 - \exp(-h)]
\]

in continuous time (or for continuous exposure variables), and

\[
r[1 - (1 - p)^k] = kr(p)
\]

in discrete time (or for discrete exposure variables.) Both equations yield the solution

\[
r(p) = -\ln(1 - p),
\]

unique up to positive scalar multiplication, i.e., up to choice of scale.
Note that this risk scale runs from \( r(0) = 0 \) to \( r(1) = \infty \), so that \( p = 1 \) is "infinitely greater" than any other \( p \). Thus, risks of certain events are not comparable to risks of uncertain events according to this risk measure. However, the measure

\[
r(p) = -\ln(1 - p)
\]

has some attractive properties, e.g., linearity for small probabilities and additivity in total exposure. If it is applied to probabilities of death, and if we choose the scale so that a \( p \) of 1 in a million = 1 on this scale -- one "micromort", in Ron Howard's terminology (Howard, 1984) -- then we can say that a death probability of \( p \) is equivalent to

\[
1,000,000[-\ln(1 - p)] \text{ micromorts.}
\]

We will also define a mort of risk for the arrival of an event as a million micromorts. For values of \( p \) less than 0.1, the departure from linearity is less than 10% of the value of \( p \). For values of 0.2 and greater, however, the nonlinearity of \( r(p) \) in \( p \) becomes increasingly significant. For example, the difference in risk between death probabilities of 0.9 and 0.8 is 693,141 micromorts, while the difference between \( r(0.3) \) and \( r(0.2) \) is only 133,532 micromorts. Thus, an increase of 0.1 in the probability of death is more than five times more risky, according to this risk scale, when the increase is made from a baseline of 0.8 than when it is made from a baseline of 0.2.

This analysis shows how effective the assumption of additivity for representing an empirical operation (such as concatenation of risk exposures) can be in restricting the possible form of the risk function \( r(p) \).

5.2 Preference Differences and a One-Parameter Family of Comparative Risk Relations

A second, powerful, result can be obtained for preferences among c.d.f.'s over a cardinal consequence scale. Essentially, it states a reasonable
set of conditions under which the set of preference orderings that are consistent with an expected utility representation constitute a one-parameter family. Our development of this theory will make use of the third and final major approach to measurement (in addition to ordinal measurement and extensive measurement), namely, difference measurement.

As usual, we take the interval \([0, M]\) as the consequence set, and assume that preferences are strictly monotonically decreasing in \(c\) as \(c\) ranges from 0 to \(M\). We are specifically interested not just in a single d.m.'s preference ordering over the set \(L\) of c.d.f.'s on \([0, M]\), but in the extent to which different d.m.'s may have different preference orderings. We shall therefore assume that the consequence severity axis is ordinarily "objective", in the sense that all d.m.'s agree on the ordinal severity ranking of consequences \(c\) in \([0, M]\). Thus, if \(v_1(c)\) and \(v_2(c)\) are two ordinal value functions representing the preferences of d.m.'s 1 and 2 for consequences in \(c\), then \(v_1(c') > v_1(c'')\) if and only if \(v_2(c') > v_2(c'')\), for any two consequences \(c'\) and \(c''\) in \([0, M]\). For example, if \(c'\) and \(c''\) represent two possible numbers of people killed in a plane crash, then all d.m.'s may agree that \(c'\) is worse (more severe) than \(c''\) if, and only if, \(c'\) is larger than \(c''\).

D.m.'s may agree on more than the ordinal ranking of consequences, however. They may also agree on strengths of preference between different outcomes. This concept can be made precise using the notion of preference differences. Suppose that \((c', 0.5, c'')\) is a lottery that has a 50-50 chance of producing consequences \(c'\) or \(c''\), and consider a choice between the following two acts, which may be thought of as alternative insurance policies:

- act \(a'\) reduces the consequence from \(c'\) to \(c^*\) if \(c'\) would have occurred; otherwise (i.e., if \(c''\) occurs) it has no effect;

- act \(a''\) reduces the consequence from \(c''\) to \(c^{**}\) if \(c''\) would have occurred; otherwise (i.e., if \(c'\) occurs) it has no effect.
The ordinal ranking of consequences by decreasing severity (increasing desirability) is \(c', c^*, c'', c^{**}\). A choice between \(a'\) and \(a''\) requires more than such ordinal information for consequences, however. It requires the d.m. to consider which of the following two alternatives he finds preferable: (i) A reduction in consequence from \(c'\) to \(c^*\); or (ii) A reduction in consequence from \(c''\) to \(c^{**}\). If he prefers \(a'\), then we could say that he prefers \(c'\) to \(c^*\) **MORE STRONGLY** than he prefers to \(c''\) to \(c^{**}\). Alternatively, the **PREFERENCE DIFFERENCE** between \(c'\) and \(c^*\) is larger than the preference difference between \(c''\) and \(c^{**}\). This is the intuitive notion that we want to exploit. If all d.m.'s agree on the relative rankings of preference differences between consequences, then it follows under weak conditions that the consequence scale is unique up to choice of origin and scale, i.e., there will be an "objective" (unanimously agreed-on) cardinal scale for consequence severities. One set of sufficient conditions for this result is now presented.

For any four consequences \(a, b, c,\) and \(d\) in \(C \ast = (0, M]\), let \(ab \sim cd\) mean that \(a\) is preferred to \(b\) at least as strongly as \(c\) is preferred to \(d\). Alternatively, \(ab \sim cd\) means that "\(a\) is preferred to \(b\) at least as much as \(c\) is preferred to \(d\)." An empirical interpretation of this relation has just been suggested. Five reasonable axioms for \(\sim\) follow:

**M1 (Weak Order):** \(\sim\) is a weak order (connected and transitive) on \(C \times C\).

**M2 (Signed Differences):** If \(ab \sim cd\), then \(dc \sim ba\).

**M3 (Monotonicity):** If \(ab \sim cd\) and \(bf \sim dg\), then \(af \sim cg\). (Intuitively, think of \(ab\) as the interval from \(a\) to \(b\) and \(bf\) as the interval from \(b\) to \(f\); then \(af\) is the interval from \(a\) to \(f\). Think of \(\sim\) as the relation "is at least as long as" between these intervals.)

**M4 (Solvability):** If \(ab \sim cd\) and \(aa\) ("nonnegative" differences) then there are consequences \(c'\) and \(c''\) in \(C\) such that \(ac' \leq E cd E c''b\), where \(E\) is defined from \(M\) by \(wx E yz\) if and only if \(wx \leq M yz\) and \(yz \leq M wx\). (Intuitively, if interval \(ab\) is longer than \(cd\), then there are points \(c'\) and \(c''\) in \(ab\) such that \(ac'\) is as long as \(cd\) and \(c''b\) is as long as \(cd\).)
M5 (Archimedean): Let $c_1', c_2', \ldots$ be a sequence of consequences in $C$ such that $c_2c_1 E c_3c_2 E \ldots$ and not $c_2c_1 E c_1c_1$. Assume that there are consequences $c^{**}$ and $c^*$ in $C$ such that $c^{**}c^* M c_1c_1 M c^*c^{**}$ for all $c_i$ in the sequence. (Interpretively, we have a sequence of consequences whose successive members are equally far apart in terms of severity. The severity distance between any two consecutive members of the sequence is nonzero. Any pair of consequences in the sequence are less far apart in terms of severity than are the consequences $c^*$ and $c^{**}$.) Then the sequence is necessarily finite.

Axiom M5 is an "Archimedean" axiom because it asserts the impossibility of having an infinite difference in severity (composed of an infinite number of finite equi-different segments) between any two consequences $c^*$ and $c^{**}$ in $C$. The remaining axioms should be self-explanatory.

THEOREM 5.1: Suppose that $C$ is any consequence set (possibly finite) such that the preference difference relation $M$ satisfies axioms M1 to M5 on $C \times C$. Then there exists a cardinal value function $v(.)$ on $C$, unique up to choice of origin and scale, such that $ab M cd$ if and only if $v(a) - v(b)$ is greater than or equal to $v(c) - v(d)$, for any four consequences $a$, $b$, $c$, and $d$ in $C$.

Proof: Axioms M1 to M5 define a structure known in the measurement literature as an "algebraic difference structure" (see e.g., Krantz et al, 1971 or Roberts, 1979) for the preference difference relation $M$. Theorem 5.1 is proven for general algebraic difference structures in Chapter 4 of Krantz et al, 1971.

Corollary: When $C *= [0,M]$ and preferences are strictly decreasing in $c$, axioms M1 to M5 imply that there exists a unique canonical value function, $v^*$, strictly increasing on $[0,M]$, such that $v^*(0) = 0$ and $v^*(M) = 1$ and such that $v^*$ represents preferences and preference differences on $C$, i.e.,

$v^*(a) - v^*(b)$ is at least as great as $v^*(c) - v^*(d)$ if and only if $ab M cd$, for any four consequences $a$, $b$, $c$, and $d$ in $C$; and
\( v^*(c') \) is at least as great as \( v^*(c'') \) if and only if \( c' \succ c'' \),

where we define \( c' \succ c'' \) (here interpreted as \( c' \) is at least as severe as \( c'' \)) in terms of \( M \) by \( c' \succ c'' \) if and only if \( c',0 \succ M \succ c'',0 \).

for any \( c' \) and \( c'' \) in \( C \). Of course, the same representation holds if preferences are strictly increasing in \( c \), and \( \succ \) is interpreted as weak preference.

**Technical Comments:** A similar result has recently been introduced into the decision analysis literature using a slightly different set of axioms, under the name "measurable value functions" (Dyer and Sarin, 1979.) Measurable value functions enjoy the same intuitive interpretations and cardinal properties as in Theorem 5.1, but are derived from axioms for "positive" rather than "algebraic" difference structures (as defined in Krantz et al, 1971). An alternative axiom system for finite consequence sets, due to Scott and again leading to a cardinal difference measurement scale, is reported by Roberts, 1979. In all three axiom systems the representation and uniqueness properties are the same; in all three also, some of the axioms are in principle empirically unverifiable (since, for example, there are an infinite number of sequences that the Archimedean axiom must apply to.) Only Scott's axioms are necessary and sufficient. In algebraic difference measurement, all axioms except solvability (M4) are necessary, i.e., they are implied by the representation (namely, \( ab \succ M \succ cd \) if and only if \( v(a) - v(b) \) is at least as great as \( v(c) - v(d) \).)

We shall henceforth accept the Dyer and Sarin axiomatization of preference differences and refer to value functions satisfying the representation in Theorem 5.1 and its corollary as "measurable" value functions, in keeping with the recent decision analysis literature, even though the term "measurable" is burdened with other meanings.

N-M utility theory and measurable value functions provide two different ways of obtaining a cardinal value function over the consequence axis
That is, the canonical utility function $u^*(c)$ of single-attribute utility theory and the canonical measurable value function $v^*(c)$ of the corollary to Theorem 5.1 are each unique, each represent ordinal preferences, and each reflect something additional about preferences for different consequences. However, the shape of $u^*(.)$, e.g., its concavity or convexity, is generally interpreted in terms of "risk attitude" (see e.g., Keeney and Raiffa, 1976, Chapter 4) while the shape of $v^*(.)$ is usually interpreted in terms of strength of preference for different consequences.

Given the two cardinal value functions $u^*(c)$ and $v^*(c)$ on $[0, M]$, it is natural to ask whether there is any necessary connection between them. Empirically, it turns out that for positive (desirable) attributes $u^*(c)$ generally lies above $v^*(c)$ throughout $[0, M]$ except at the endpoints, where $u^*(0) = v^*(0) = 0$ and $u^*(1) = v^*(1) = 1$ (Krzysztofowicz, 1983.) This implies that the local "coefficient of risk aversion", which is defined as

$$n(c) = -\frac{u''(c)}{u'(c)}$$

for $u(.)$ at $c$, where primes denote derivatives, is greater than the analogous "coefficient of value satiation",

$$m(c) = -\frac{v''(c)}{v'(c)}$$

for the deterministic value function, over the entire range of consequences $[0, M]$.

Technical Note 1: Both $m(c)$ and $n(c)$ are independent of which versions of $u(c)$ and $v(c)$ are used within the equivalence classes of positive affine transformations of $u^*(c)$ and $v^*(c)$. There is therefore no need to specify the canonical forms in defining $m(c)$ and $n(c)$. See Keeney and Raiffa, 1976, Chapter 4 for a proof of this assertion. The term "coefficient of value satiation" is due to Dyer and Sarin, 1982.
Technical Note 2: Both \( u(c) \) and \( v(c) \) will be assumed to be smooth enough to be at least twice differentiable throughout \([0,M]\). Moreover, for simplicity, we will orient the consequence scale \([0,M]\) so that preferences are increasing over this interval. For undesirable consequences \( c \), this entails replacing attribute \( c \) with the new attribute \( M-c \).

We can now define a coefficient of RELATIVE RISK AVERSION relating \( u^*(c) \) and \( v^*(c) \). Since \( u^*(c) \) and \( v^*(c) \) are both monotonically increasing in \( c \) there will be a unique function \( f \) mapping values \( v \) in \([0,1]\) into values \( u \) into \([0,1]\) such that \( u(c) = f(v(c)) \). Interpretively, \( f \) is the single-attribute utility function defined over the attribute "value", as given by the measurable value function \( v^*(\cdot) \). Then \( k(x) = -f''(x)/f'(x) \) is defined as the coefficient of relative risk aversion at \( x \), for any value \( x \) in \([0,1]\).

Bell and Raiffa, in a 1979 Harvard Business School manuscript ("Marginal Value and Intrinsic Risk Aversion", cited in Dyer and Sarin, 1982) argue on normative grounds that the coefficient of relative risk aversion for a rational d.m. should be constant. Dyer and Sarin (1982) conjecture that \( k(x) \) is constant and independent of the attribute on which a d.m.'s preferences are assessed. Krzysztofowicz (1983) presents empirical evidence that \( k(x) \) is indeed constant for any single attribute, but that the value of the constant may differ for different attributes.

If we assume that constant relative risk aversion holds, then the following very strong link between utility and measurable value functions can be established.

**THEOREM 5.2:** Suppose that a d.m.'s preferences \( R \) among c.d.f.'s on \([0,M]\) have an expected utility representation with utility function \( u \), i.e., \( F R G \) if and only if \( E[u;F] \) is at least as great as \( E[u;G] \) for any two c.d.f.'s \( F \) and \( G \) over \([0,M]\). Suppose also that the d.m. has a measurable value function \( v \) on \([0,M]\), and let \( k \) be the constant coefficient of relative risk aversion linking \( u \) and \( v \). Then exactly one of the following cases holds:
(i) $u^*(c) = v^*(c)$ and $k = 0$.

(ii) $u^*(c) = \frac{1 - \exp(-kv(c))}{1 - \exp(-k)}$ and $k$ is positive. $n(c)$ is greater than $m(c)$ and $u(c)$ lies above and is more concave than $v(c)$ throughout $[0,M]$.

(iii) $u^*(c) = \frac{\exp(kv(c)) - 1}{\exp(k) - 1}$ and $k$ is negative. $n(c)$ is less than $m(c)$ and $u(c)$ lies below and is less concave than $v(c)$ throughout $[0,M]$.

Proof: This is Theorem 1 of Krzysztofowicz, 1983; see also Dyer and Sarin, 1982.

Note that this result does not assert that d.m.'s have or should have exponential utility (or disutility) functions for consequences, which would certainly be false; in fact, it says very little about the shape of $u(c)$. Rather, it says that utility functions for the special attribute "value", which will in general be some nonlinear transformation of the consequence scale, should be exponential (or linear). Krzysztofowicz's evidence suggests that case (ii) is usual for desirable attributes and we shall generally assume this henceforth.

The significance of this result for our understanding of risk measurement is this: that under plausible conditions, the large cardinality of $[R]$, the set of admissible risk orderings over the set $D$ of all c.d.f.'s on $[0,M]$, is not due to potential differences in risk attitudes, but to potential differences in value functions over $[0,M]$. The conditions of Theorem 5.2 imply that if all d.m.'s agree in their deterministic preferences among consequences in $C$, then the set $[R]$ of their possible risk orderings over the c.d.f.'s in $D$ is only a one-parameter family, indexed by the parameter $k$. This is the situation that we explore in the next section.
5.3 One-Way Communication Between D.M.'s With The Same Deterministic Values for Consequences

Consider a group $G$ of d.m.'s who wish to communicate with each other about risks. Each risk will be modelled as a c.d.f. in the set $D$ of all c.d.f.'s over a single-attribute interval consequence set $C = [0,M]$. The members of the group share the following common knowledge:

- Each member of group $G$ has the same deterministic preferences over consequences $c$ in $C$, and these preferences are representable by a common measurable value function $v^*(c)$, strictly increasing in $c$, with greater values of $c$ strictly preferred (possibly after reorientation of the consequence axis.)

- Each member of group $G$ has ordinal preferences over $D$ that have an expected utility representation. However, different members of $G$ may have different N-M utility functions over $c$, reflecting differences in risk attitude. We will let $R(k)$ denote the risk ordering over c.d.f.'s in $D$ held by individuals with relative risk coefficient $k$.

- In addition to these pieces of common knowledge, each member of $G$ knows his own coefficient of relative risk aversion, say $k(i)$ for member $i$. It is convenient to think of $k(i)$ as $i$'s "type", which initially is known only to himself.

It follows from Theorem 5.2 that the members of group $G$ have preference (and hence risk) orderings $R(k)$ over $D$ that belong to a one-parameter family of orderings, represented by the one-parameter family of canonical utility functions

$$u^*(c;k) = \frac{[1 - \exp(-kv(c))]}{[1 - \exp(-k)]}$$

and indexed by $k$. The risk ordering for any $k$ is the reverse of the preference ordering $R(k)$. We shall make the further assumption that $k$ is positive.
The simplest case of communication between members of G involves one-way communication between an informed member i who we will call the risk analyst (r.a.) and an uninformed member j who we will call the decision maker (d.m.). The d.m. must choose an act a from a set of acts A, and knows nothing about what consequences in the risk consequence set C will result from his choice. (He may, however, know a great deal about the deterministic cost and benefit flows associated with each act a.) The r.a. knows the c.d.f. over consequences in C corresponding to each act in A. The r.a.'s problem is to communicate this knowledge to the d.m., or at least enough of it to help the d.m. make a choice.

Suppose the d.m. is trying to choose between only two acts, say a' and a". What can the r.a. say to help him? Most of the obvious suggestions are obviously unsatisfactory. For example,

- The r.a. could warn him that a" is riskier than a', meaning that \( F(a') R(i) F(a") \), where \( F(a) = F(c;a) \) denotes the conditional c.d.f. over consequences in C given act a and \( R(i) \) is an abbreviation for \( R(k(i)) \). But this leaves unclear whether the d.m. would agree with the r.a.'s risk judgement, i.e., whether \( F(a') R(j) F(a") \).

- The r.a. could announce the numerical values of \( 1-E[u^*(i);F(a')] \) and \( 1-E[u^*(i);F(a")] \), calling these the numerical "risks" of a' and a", respectively. (Here, \( u^*(i) \) is an abbreviation of \( u^*(.;k(i)) \).) This would be satisfactory if r.a. and d.m. had the same utility function. But it again leaves ambiguous how much of the resulting risk index, \( r(a) = 1-E[u^*(i);F(a)] \), is due to \( F(a) \), which is what the d.m. cares about, and how much is due to his own personal relative risk attitude, \( k(i) \).

- The r.a. could announce \( r(a) = 1-E[u^*(i);F(a)] \) for a = a', a", and also announce his own "type", \( k(i) \). But in general he will be indifferent between some c.d.f.'s, all of which map onto his announced value \( r(a) \), that the d.m. is not indifferent between.
Thus, the d.m. can not unambiguously translate the announced risk \( r(a) \) for an act \( a \) into a single corresponding value on his own canonical utility scale. Knowing only \( r(a) \), he will remain uncertain about his preferences for \( a \).

The r.a. could tell the d.m. the whole c.d.f. \( F(a) \) corresponding to each \( a \). However, for arbitrary distributions this requires a lot of information to be transferred; moreover, the d.m. may not be able to easily understand or interpret the entire c.d.f. even if it can be transmitted to him.

The essential challenge of risk communication in this context is for the r.a. to find a way of encoding his knowledge of \( F(a) \) into a message, say \( m[F(a)] \), that will be useful to the d.m. who will receive it. Let \( M(i) \) be the set of possible messages that the r.a. can send to the d.m. In many applications, \( M(i) \) may be restricted to having a very simple structure, e.g., it may be necessary for the r.a. to communicate his knowledge numerically using messages in \( M(i) = [0,1] \).

The possibilities for effective communication depend on

(i) The mathematical structure of the set of c.d.f.'s corresponding to acts in \( A \);

(ii) The mathematical structure of the message set \( M(i) \) available to the r.a.; and

(iii) What each player knows before communication begins.

We now present several results for increasingly restricted message sets. Let \( D^* \) denote a subset of \( D \) that includes all c.d.f.'s corresponding to acts in \( A \), as perceived by the r.a. We assume that the members of \( D^* \) are "well-behaved" (e.g., bounded with at most a finite number of discontinuities.)
THEOREM 5.3: Let $M(i)$ include all well-behaved functions on the real line. If the r.a. sends as his message the function $r(k;a) = 1 - E[u^*(k);F(a)]$, considered as a function of $k$, then any d.m. will in principle be able to compute his own risk index $r(a;k(j)) = 1 - E[u^*(k(j));F(a)]$ for act $a$ from this message.

Proof: The risk function $r(k;a)$, which is the expected disutility of $a$ considered as a function of $k$, is just a translated, scaled version of the Laplace transform of $F(a)$ with $k$ as the frequency domain variable. We assume that $F$ is well behaved so that this transform exists for positive real $k$. It is known that the Laplace transform of a distribution uniquely determines the distribution (see e.g., Ross, 1970, p. 5), from which any d.m.'s expected utility can in principle be computed, so Theorem 5.3 is proved.

THEOREM 5.4: It is impossible to assign numbers to arbitrary c.d.f.'s in $D^*$ in such a way that for any two c.d.f.'s $F$ and $G$ in $D^*$, the number $r(F)$ assigned to $F$ is at least as great as the number $r(G)$ assigned to $G$ if and only if every d.m. agrees that $F \preceq R G$, i.e., that $F$ is at least as risky (or undesirable) as $G$.

Proof: The proof is in a sense trivial, and essentially consists of the observation that the first-order stochastic dominance relation FSD on the set $D$ of c.d.f.'s is a partial ordering, so that $(D,\text{FSD})$ can not be isomorphic to $(\text{Reals}, \text{Inequality})$. Sufficient conditions for proving the impossibility of the representation are that $D$ contains at least three distributions, say $F_1$, $F_2$, and $F_3$, such that $F_1 \text{ FSD } F_3$ and no other FSD relations hold among any other pair drawn from this set of three c.d.f.'s.

It is useful to establish the impossibility of numerical representation of objective (unanimous) comparative risk by using a special construction for normal distributions that yields additional results. Accordingly, assume that $D^*$ contains the c.d.f.'s corresponding to truncated normal distributions (truncated at the endpoints of the interval $[0,M]$) with positive means and with standard deviations that are sufficiently small.
compared to their means so that the truncation can be ignored. If the consequence of act \( a \) is a normally distributed random variable with mean \( m \) and variance \( v \), then

\[
r(a;k) = 1 - (1 - \exp[-k(m - (vk)/2)])/[1 - \exp(k)]
\]

is the risk, i.e., the canonical expected disutility, of act \( a \) to a d.m. of type \( k \). Hence, \( r(a';k) \) is less than \( r(a'';k) \) if and only if \( m' - v'k/2 \) exceeds \( m'' - v''k/2 \). As \( k \) increases from 0, these two expressions both decrease linearly in \( k \); hence, they can cross at most once. Thus, the nonnegative real line can be divided into two parts by a single point, say \( k^*(a',a'') = \min[0, 2(m' - m'')/(v' - v'')] \), such that

1. All d.m.'s of type less than \( k^*(a',a'') \), if any, will prefer one act, say \( a' \), to the other; while

2. All players of type greater than \( k^*(a',a'') \) will prefer the other act, say \( a'' \), to \( a' \).

More generally, there is a one-to-one correspondence between the c.d.f.'s of normal distributions \( N(m,v) \) and the set of straight lines with nonnegative slopes \( v \) and intercepts \( m \).

To complete the demonstration, let \( L_1 \) and \( L_2 \) be any two intersecting lines with positive slopes and intercepts, e.g., representing the normal distributions \( N(100,1) \) and \( N(50,2) \), respectively. If the representation posited in the theorem held, we would have to have that \( r(L_1) = r(L_2) \), since neither line is preferred by all d.m.'s to the other (i.e., neither lies beneath the other for all \( k \).) Let \( L_3 \) be parallel to \( L_1 \) and have a greater intercept, e.g., \( L_3 \) might represent the normal distribution \( N(110,1) \). Then \( r(L_3) \) must be greater than \( r(L_1) \), since \( L_3 \) lies above \( L_1 \) everywhere, i.e., for all types \( k \). But we also must have \( r(L_3) = r(L_2) \), since \( L_3 \) and \( L_2 \) necessarily intersect. This contradicts the previous conclusion that \( r(L_1) = r(L_2) \). Thus assuming that the representation in the theorem is possible leads to a contradiction, and so the theorem is
proven. The same argument clearly holds for the FSD relation between arbitrary c.d.f.'s so long as there are at least three c.d.f.'s such that only one FSD relation holds among them.

In many risk analysis applications, including financial risk, normal consequence distributions are especially important. The construction used in the preceding proof yields the following useful results for these distributions.

THEOREM 5.5: Let $N(m',v')$ and $N(m'',v'')$ be two (approximately) normal distributions with means $m'$ and $m''$ and variances $v'$ and $v''$, respectively, over a measurable (difference-preserving) value scale $[0,M]$. (As usual, we assume that truncation constraints can be ignored.) Suppose that preferences are strictly decreasing over $[0,M]$. Then d.m.'s of all types will agree that $N(m',v')$ is riskier than $N(m'',v'')$, i.e.,

$$N(m',v') \leq_{D} N(m'',v'')$$

for ALL nonnegative types $k$, if and only if $m'$ is at least as great as $m''$ and $v'$ is at least as great as $v''$. In this case, $N(m',v')$ is said to "mean-variance dominate" $N(m'',v'')$.

Proof: This is the necessary and sufficient condition for two lines with positive intercepts $m'$ and $m''$ and positive slopes $v'$ and $v''$ not to intersect for any $k$ greater than 0. (By symmetry, the same conclusion holds for strictly increasing preferences as well.)

Theorem 5.5 is similar to the result in financial risk theory that all risk-averse d.m.'s prefer mean-variance dominant (i.e., "efficient") normal distributions to mean-variance dominated normal distributions. However, Theorem 5.5 requires only relative risk aversion, i.e., risk aversion in the attribute "value", rather than risk-aversion for consequences; moreover it shows that mean-variance dominance is necessary as well as sufficient for d.m.'s of all types to prefer one normal distribution to another.
THEOREM 5.6: Suppose that all acts in \( A \) have (approximately) normally distributed consequences in the measurable value scale interval \([0,M]\). Then the nonnegative real numbers can be partitioned into successive intervals, say \([0,x_1), [x_1,x_2), \ldots\) such that

(i) Some act in \( A \) is associated with each interval;

(ii) No act in \( A \) is associated with more than one interval; and

(iii) Any act \( a' \) in \( A \) is a most-preferred act for a player of type \( k \) if and only if \( k \) is in the interval with which \( a' \) is associated.

Only mean-variance undominated acts are associated with intervals. Multiple acts are associated with a single interval only if they induce the same distribution over consequences, and are therefore indifferent.

Proof: Trace out the lower envelope of the lines representing the acts in \( A \). This envelope may be called the undominated, or "efficient", frontier. The intervals described correspond to the faces of this efficient frontier. Q.E.D

This is analogous to the usual construction of efficient frontiers in financial risk theory, but goes beyond efficiency to explicitly relate relative risk attitude to preferred choice of act. We will call the set of intervals and associated acts identified in the above theorem the action partition corresponding to choice set \( A \). This partition will contain a finite number of blocks or segments if \( A \) contains a finite number of acts.

If the r.a. can send the action partition to the d.m., and if the d.m. can identify which segment of this partition contains his own type, then he will know which act to choose. Moreover, no less informative message can inform the d.m. of what act to take regardless of his type. Thus, the action partition represents the minimally sufficient information that the r.a. must convey to a d.m. of unknown type in order to allow the d.m.
to calculate his expected-utility maximizing act. Since the set of action partitions is more complicated than the set of real numbers (for example, it requires the real numbers just to specify the position of the end of the first segment in the partition), there is no way of indexing the set of action partitions by the set of reals, i.e., there is no possible isomorphism between them. Thus, if the message set is \( M(i) = [0,1] \), the r.a. will be unable to send the d.m. sufficient information to allow him to choose among the acts in \( A \). Since this holds when \( D^* \) is the set of all normal distributions over the measurable value scale, it certainly holds in more general situations, e.g., in situations where the assumption of a common deterministic value function is relaxed.

5.4 Comparative Riskiness and Second-Order Stochastic Dominance

Theorem 5.6 implies that undominated acts with consequences normally distributed on the measurable value scale can be unambiguously ranked in terms of comparative riskiness if we define the "comparative riskiness" of act \( a \) with respect to the other undominated acts in \( A \) as the ordinal rank of the block associated with \( a \) in the action partition. In other words, the comparative riskiness of an act is judged not by looking at the c.d.f. over consequences for that act alone, but by looking at the preferences of players of different types (relative risk attitudes) for that act as opposed to other acts. This general idea can be extended to arbitrary c.d.f.'s by defining \( F \) to be (objectively) comparatively more risky than \( G \) if and only if every risk-averse individual prefers \( F \) to \( G \) for any two c.d.f.'s \( F \) and \( G \) in \( D \). For the conventional case of single-attribute utility theory with preferences increasing rather than decreasing in consequence level, it can then be shown that the following are equivalent:

1. Every risk-averse d.m. prefers \( F \) at least as much as \( G \). More precisely, \( E[u;F] \) is greater than or equal to \( E[u;G] \) for every bounded concave increasing \( N-M \) utility function \( u \) on \([0,M]\) (where greater values of \( u \) now correspond to preferred consequences.)
(ii) The integral from 0 to x of F is less than or equal to the integral from 0 to x of G, for each x in [0,M].

(iii) There are random variables X, Y, and Z such that Y has the same c.d.f. as X + Z and E(Z;X) = 0 for all X, and such that F is the c.d.f. of X and G is the c.d.f. of Y. (Thus, "G = F + Noise".)

In any of these cases, F is said to "dominate G in the sense of second-order stochastic dominance", written F SSD G; equivalently, G is said to be "comparatively riskier than" F.

This characterization of comparative risk is due to Rothschild and Stiglitz, 1970; see also Hadar and Russell, 1969, who defined SSD, and Schmeidler 1979 for a generalized mathematical treatment. Machina (1982, Theorem 2) establishes that the equivalence of (i) and (iii) generalizes to the case of local utility functions, i.e., that a d.m. will necessarily prefer F to G whenever G = "F + Noise" in the above sense if and only if the local utility functions u(.;F) are concave in their first argument for all F in D. Precisely analogous characterizations of increasing risk, although with a reversed orientation, hold when preferences are strictly decreasing rather than strictly increasing in c.

Allowing risk comparisons among c.d.f.'s on the basis of SSD changes none of the basic arguments developed in this section for the impossibility of objective numerical measurement of risk. Like FSD, SSD is only a partial ordering on D, even though it can be used to compare some pairs of c.d.f.'s not comparable by FSD. By essentially the same proof as for FSD, therefore, no numerical representation of SSD on D is possible.

5.5 Special Cases and Limited Precision Objective Risk Measurement

The preceding results show that objective numerical measurement of the comparative risks of probability distributions over single-attribute consequences is in general impossible. There is no way to assign numbers to c.d.f.'s over consequences so that all expected utility decision
makers will agree that riskier (i.e., less preferred) c.d.f.'s have been assigned higher numbers. Perhaps more importantly, there is no way to assign numbers to c.d.f.'s so that any d.m. can tell which of two c.d.f.'s he prefers by the numbers that have been assigned to them. Because these limitations hold when all individuals share common deterministic values, consequences are single-attributed, and individuals are expected-utility maximizers, they also hold in the multiattribute case and when individuals have local utility functions or different deterministic preferences. They also apply to risks represented by pairs (p,F), where p is the probability of an accident and F is the conditional c.d.f. over consequences if a c.d.f. occurs.

Given the impossibility of objective numerical risk measurement for c.d.f.'s in general, two coping strategies naturally suggest themselves. One is to seek special, useful classes of c.d.f.'s for which objective numerical risk measurement is possible. In other words, \( M(i) \in [0,1] \) may be an adequately rich message set if the class of c.d.f.'s \( D^* \) to which c.d.f.'s of different acts belong is sufficiently restricted.

The other strategy is to seek nonnumerical representations of risk. Two such representations have already been suggested: the risk function \( r(k;a) = 1 - E[u^*(k);F(a)] \) considered as a function of \( k \), for each act \( a \) (with \([0,M]\) positively oriented) and the action partition labeling consecutive segments of the positive \( k \) axis with the corresponding preferred acts in \( A \). An alternative, limited-precision approach to risk measurement via intervals on a "canonical risk scale" will be developed below.

Each of these two strategies for overcoming the general impossibility of objective numerical risk measurement is now investigated.

5.5.1 Objective Numerical Risk Measurement for Special Classes of C.d.f.'s.

Many of the most important families of probability distributions in applied risk analysis do allow objective numerical risk measurement.
Recall that the requirements for such measurement are as follows. Let $D^*$ be a set of c.d.f.'s, defined over a consequence set $C \equiv [0,M]$, that contains at least all of the c.d.f.'s corresponding to acts in $A$. Objective ordinal numerical measurement of risk for the c.d.f.'s in $D^*$ consists of an assignment of numbers $r(F)$ to the c.d.f.'s in $D^*$ in such a way that for any two c.d.f.'s $F$ and $G$ in $D^*$, $r(F)$ is greater than $r(G)$ if and only if every d.m. prefers consequence distribution $G$ to consequence distribution $F$. Theorem 5.4 showed that such objective measurement is in general impossible. We now identify several restricted classes of c.d.f.'s of practical importance in risk analysis for which it is possible.

**THEOREM 5.7:** Let $[0,M]$ be a measurable value scale and let $M$ in $[0,M]$ go to infinity (i.e., make consequence set $C \equiv [0,M]$ unbounded from above), and assume that preferences are strictly decreasing in $c$. Suppose that $D^*$ is the set of all c.d.f.'s over $[0,M]$ for one the following classes of distributions, or for a subset of one of the following classes of distributions:

- Poisson
- Binomial with fixed parameter $n$ and variable parameter $p$
- Geometric or exponential
- Negative binomial or gamma with either parameter fixed
- Normal with proportional mean and variance, truncated at zero.

Then if all d.m.'s are expected utility maximizers with constant positive relative risk aversions and the same deterministic preferences (e.g., as represented by the canonical measurable value function), objective numerical risk (or preference) measurement for the c.d.f.'s in $D^*$ is possible.
Proof: Examination of the moment generating functions (m.g.f.'s) of these distributions with \(-k\) as the frequency domain variable shows that expected utility (which in effect is this moment generating function, by the constant relative risk aversion hypothesis) is

- monotonic in \(p\) for the family of binomial distributions with fixed parameter \(n\) and variable parameter \(p\);
- monotonic in \(m\) for the family of Poisson distributions with parameter \(m\);
- monotonic in \(p\) for the family of geometric or exponential distributions with parameter \(p\);
- monotonic in the variable parameter for the family of negative binomial or gamma distributions with one fixed and one variable parameter.

For example, the m.g.f. of the binomial distribution with parameters \(n\) and \(p\) is the \(n\)th power of the expression \(p(\exp(-k)) + (1-p)(1)\). Now this is just a weighted average, with weight \(p\), of a quantity \(\exp(-k)\) that is never greater than one for any nonnegative \(k\), and 1. It is therefore monotonic in \(p\). The remaining assertions follow similarly.

Given that expected utility is monotonic in a parameter that indexes the members of \(D^*\), for each of the above specifications of \(D^*\) and for all d.m.'s in \(G\), it follows that the indexing parameter is an objective (for \(G\)) numerical index, or ordinal measure, of risk. Q.E.D.

The family of Poisson distributions and the class \(N(m,m)\) of normal distributions with equal means and variances, which are limiting cases of Poisson distributions, are particularly important in risk analysis. They allow a very strong sort of objective risk measurement.
THEOREM 5.8: Suppose that \([0,M]\) is at least an ordinal consequence scale and that all d.m.'s have preferences over \(D\) that respect FSD. (A sufficient, but not necessary, condition for this is that they be expected utility or generalized expected utility maximizers.) Then any family \(D^*\) of c.d.f.'s that are equivalent under some ordinal transformation of the consequence scale to either

(i) The c.d.f.'s of the family \(\text{Pois}(m)\) of Poisson distributions with parameter \(m\); or

(ii) The c.d.f.'s of the family \(\text{N}(m,m)\) of normal distributions with equal means and variances \(m\)

allows objective numerical risk measurement on a risk scale that is unique up to choice of scaling factor. Any family of normal distributions with proportional means and variances also enjoys this property.

Proof: Each of these two families is a one-parameter class of distributions totally ordered by FSD. For example, let \(\text{Pois}(m)\) denote the class of Poisson distributions with variable parameter \(m\) (e.g., the number of car accidents in a year in a certain county under a given risk management policy.) Then for any two particular Poisson distributions, say \(\text{Pois}(m')\) and \(\text{Pois}(m'')\), \(\text{Pois}(m')\) is preferred to (i.e., is less risky than) \(\text{Pois}(m'')\) for all d.m.'s if and only if \(m'\) is less than \(m''\). That is, \(\text{Pois}(m'')\) FSD \(\text{Pois}(m')\) if and only if \(m''\) exceeds \(m'\), where FSD is the first-order risk dominance (rather than preference dominance) relation.

(To show this, we must show that \(\Pr(X \text{ exceeds } j; m)\) is increasing in \(m\), for all \(j = 0, 1, 2, \ldots\), where \(X\) is a random variable with distribution \(\text{Pois}(m)\). Think of \(X\) as the number of arrivals in the time interval \((0,t]\) in a Poisson process with arrival rate \(L = m/t\); then for any fixed \(t\), this probability is increasing in \(m\) for any \(j\), as claimed.)
Thus, the parameter \( m \) is an objective ordinal numerical measure of risk on the class \( \text{Pois}(m) \) of c.d.f.'s over an undesirable discrete consequence scale.

Now let \( D^* \) be any class of c.d.f.'s that can be transformed into the set of c.d.f.'s for \( \text{Pois}(m) \) by an ordinal transformation of the consequence axis, and assume that this transformation has been made. Let the operation \( * \) denote convolution of c.d.f.'s, and let \( r(F) \) denote the risk index associated with the c.d.f. \( F \) in \( D^* \) (i.e., \( r(F) \) is the parameter of the corresponding Poisson distribution.) Then

\[
r(F * G) = r(F) + r(G)
\]

for any two c.d.f.'s \( F \) and \( G \) in \( D^* \). In fact,

\[
(\text{Pois}(m),FSD,*) \leftrightarrow (\text{Positive reals}, \text{inequality}, +), \text{where } \leftrightarrow \text{ denotes isomorphism. (The family of Poisson distributions is therefore called "infinitely divisible" in probability theory.) Thus, the Pollatsek-Tversky axiom of additive convolutions holds in the family Pois}(m).

It follows that if any particular value of the parameter \( m \) is selected to define a unit risk, then the risk scale for Poisson c.d.f.'s is completely determined. Analogous results hold for the \( N(m,m) \), family since any member of this family may be thought of as a limit of sums of Poisson random variables; this can be shown formally by exploiting the fact that each \( N(m,m) \) is an infinitely divisible. Q.E.D.

Corollary: If the family \( N(m,m) \) of normal distributions in Theorem 5.8 is replaced by the family of gamma distributions with parameters \( m \) and \( n \), with \( n \) fixed, then the conclusions of the theorem still apply. If \( N(m,m) \) is replaced by the exponential family with parameter \( m \), then the ordinal part of the theorem holds.
Proof: The gamma family is also an infinitely divisible one-parameter family totally ordered by FSD and closed under addition of random variables (or convolution of their c.d.f.'s.) The exponential family is an infinitely divisible family totally ordered by FSD, but is not closed under convolution. Q.E.D.

5.5.2 Limited-Precision Objective Risk Measurement and Canonical Risk Scales

The above results show that for several special classes $D^*$, satisfactory numerical risk measurement is possible. Moreover, the one-parameter normal, gamma, and exponential scales are related, since for example, the c.d.f. of a normal distribution on a given consequence scale (briefly, a "normal c.d.f." will have a unique smallest upper exponential c.d.f. and greatest lower exponential c.d.f. that stand in the FSD relation to it. Thus, each normal c.d.f. corresponds to an interval on the exponential risk scale such that the objective risk of the normal c.d.f., as measured on the exponential risk scale, is known to fall within the associated interval. Similarly, every exponential c.d.f. falls within some corresponding objectively determinable interval on the normal risk scale, where the endpoints of the interval are determined by the two parameter values for $m$ in $N(m,m)$ at which the FSD relation between $N(m,m)$ and the given exponential distribution ceases to hold.

In the case of discrete consequences and an absolute (counting) consequence scale, a similar construction holds. Let the consequence set $C$ be isomorphic to some subset of the set consisting of zero and the first $M$ integers, ordered by increasing severity. For any distribution function $F$ over these outcomes, let $PL(F)$ and $PU(F)$ denote the greatest and least Poisson distributions (as ordered by their parameters) truncated at $M$ such that $PU(F) \leq F \leq PL(F)$. Let $r'(F) = r[PU(F)]$ and $r''(F) = r[PL(F)]$ be the corresponding parameters of these two Poisson distributions bounding $F$, i.e., the values of $PU(F)$ and $PL(F)$ on the Poisson risk scale. Then the interval $[r',r'']$ on the Poisson risk scale bounds the possible value of the risk of $F$ on this scale, i.e., under the
usual continuity assumption for preferences, there will exist for each
d.m. some M-truncated Poisson distribution with a parameter value between
r' and r" that is judged exactly as risky as P. Hence, the interval
(r',r") may be called a limited-precision objective numerical measure of
the risk of P. If the intervals constructed in this way for two
distributions do not overlap, then the FSD relation holds between the
distributions (since preference continuity implies that each interval
contains a Poisson "equivalent" to its corresponding distribution.) The
converse, however, does not hold.

Example 5.3: Limited-Precision Objective Risk Measurement on the Poisson
Risk Scale

Suppose that the consequence set is \( C = [0, 1, 2] \), representing the
numbers of fatalities that may occur in car accidents involving a certain
two-passenger car whose safety design is under study. What is the risk
of the probability distribution \( P = [p(0) = 0.2, p(1) = 0.5, p(2) = 0.3] \),
measured on the Poisson risk scale? In other words, what value of m
makes the Poisson distribution over fatalities truncated at 2 just as
risky as (i.e., indifferent to) \( P \)?

The exact answer depends on the risk attitude of the person making the
evaluation. However, a limited-precision answer can be found by varying
m until stochastic dominance in one direction or the other is achieved.
The resulting range is \((m',m'') = (1.1, 1.6)\), or \( r(P) = 1.35 \) plus or minus
0.25, corresponding to truncated Poisson c.d.f.'s of approximately \((0.33, 0.7, 1)\) and \((0.2, 0.52, 1)\), respectively. Two other p.d.f.'s and their
corresponding objective risk intervals on this same consequence set are
as follows:

\[
\begin{align*}
  r(0.33, 0.33, 0.33) & = 1.7 \text{ plus or minus } 0.6 \\
  r(0.5, 0.5) & = 1.675 \text{ plus or minus } 0.005.
\end{align*}
\]

The type of construction used in this example will be crucial throughout
the remainder of this study. We define a **Canonical Risk Scale** for
measuring risks to be any continuous one-parameter family of distributions (or more general objects) totally ordered by FSD (or other dominance relations). Risks of arbitrary distributions are measured with limited precision on a canonical risk scale by dominance comparisons to the members of the one-parameter family used to calibrate the scale. Theorem 5.8 provides Pois(m) and N(m,m) as calibrating families for discrete and continuous consequences, respectively.

5.5.3 Extension to Risk Models of the Form (p,F)

The idea of limited-precision risk measurement on special scales calibrated by known probability distributions can be extended to more useful settings. So far in this discussion we have been treating the objects of choice as probability distributions over consequences. However, as previously discussed, risk models of the form (p,F), where p is interpreted as the probability of an accident and F as the conditional c.d.f of its consequence (or consequence severity) if an accident does occur, are often more useful. Limited-precision objective risk measurement is still possible in this class of models.

Let [0,M] be a single-attribute consequence set. We define a canonical risk scale (CRS) for measuring the risks of pairs (p,F), where each F is a c.d.f. over [0,M], as follows:

- The canonical risk scale runs from $m = 0$ to $m = 1$;
- Point $m$ on the CRS denotes the risk of the pair $(\min(1,2m),U(2m))$,

where $U(x)$ is defined as the uniform distribution running from 0 to $xM$ if $x$ is less than 1, and as the uniform distribution running from $(x-1)M$ to $M$ for $x$ between 1 and 2.

For example, $m = 1$ denotes the most severe possible risk (the worst possible outcome, M, occurs with probability 1), while $m = 0.5$ is the
point at which an accident is first certain to occur, and its consequences are uniformly distributed between 0 and M.

The set of pairs \((m, U(2m))\) is continuous in \(m\) and totally ordered by FSD. Thus, the CRS provides an objective risk scale appropriate for limited-precision measurement of the risks of arbitrary \((p,F)\) pairs. For any individual with continuous preferences, there will be exactly one value of \(m\), say \(r(p,F)\), such that \([r(p,F), U(2r(p,F))]\) is exactly indifferent to \((p,F)\). For the set of all individuals whose preferences for c.d.f.'s respect FSD, only limited-precision measurement is possible. The following result is easily established.

**THEOREM 5.9:** \(r[p,u(a,b)]\) lies between \(\min[p, b/M]\) and \(p + a/M\) on the above CRS, where \(u(a,b)\) denotes the uniform distribution with endpoints \(a\) and \(b\), \(a\) less than \(b\). In particular, the risk of the lottery that gives point consequence \(c\) with probability \(p\) and zero otherwise is between \(\min[p, c/M]\) and \(p + c/M\).

For example, a 40% chance at a loss of \(0.8M\), where \(M\) is the maximum possible loss (normalized to have severity 1 on the CRS) has an associated CRS risk of 0.4 plus or minus 0.2.

As with the other limited-precision canonical risk measurement scales, the CRS defined here may permit a considerable range of possible values for a given risk. Moreover, the FSD relation may hold between two c.d.f.'s that have overlapping ranges on the CRS, so that the imprecision may be crucial. However, this form of measurement requires only two numbers to specify the (approximate) risk of an option \(a\), corresponding to a pair \((p(a), F(a))\), whereas a more discerning alternative such as the risk function \(r(k;a)\) requires an entire function. Moreover, the CRS interval assigned to \(a\), unlike the action partition for a set of acts, can be calculated from knowledge of \(p(a)\) and \(F(a)\) alone. Thus, the use of limited-precision measurement has some advantages over exact nonnumerical measures in terms of simplicity and flexibility.
5.5.4 Practical Implementation: The Canonical Risk Curve

Figure 5.1 shows a suggested format for displaying risks, based on the theory developed in the preceding sections. Associated with any act \( a \) and c.d.f. \( F(a) \) over \([0,M]\) is a curve representing the risk associated with that act, measured on a canonical risk scale, as a function of risk attitude. Any continuous one-parameter family of distributions on \([0,M]\) totally ordered by FSD may be used to define this risk scale, which is the vertical axis. Risk attitude is captured on the horizontal axis by showing the certainty equivalent of a 50-50 lottery between the best and worst possible consequences, 0 and \( M \). In Figure 5.1, it is assumed that this certainty equivalent will be between \( 0.5M \) (for risk-neutral d.m.'s) and \( M \) (for minimaxers); the certainty equivalent scale can easily be extended below \( 0.5M \). Any given certainty equivalent between these values uniquely determines a corresponding coefficient of risk aversion in an exponential utility function, and hence determines the utility function itself. This utility function is then used to determine (i) the expected utility of \( F(a) \); and (ii) the unique distribution in the chosen CRS family that has the same expected utility as \( F(a) \). This CRS risk value is plotted against the chosen certainty equivalent value. The suggested representation of risk is the plot of all such certainty equivalent-CRS value points.

This suggested display, which may be called the **canonical risk curve** for c.d.f. \( F(a) \), has several useful features:

- Any d.m. with constant risk aversion (i.e., with an exponential or linear utility function for consequences) who knows his own risk attitude, i.e., who can locate himself on the horizontal axis, can read off exact CRS values for any set of acts with known c.d.f.'s over consequences.

- Any d.m. obeying the axioms of expected utility theory and who knows the family used to calibrate the CRS, e.g., one of the families in Theorem 5.8, can in principle calculate his own expected utility for act \( a \) from the canonical risk curve for act \( a \). (See Theorem 5.3.)
Figure 5.1: Canonical Risk Curve
The canonical risk curve format can be applied to risks modelled by c.d.f.'s over a continuous consequence scale; by c.d.f.'s over the integers; or by pairs \((p,F)\), as discussed above. Thus, a common format can be used to present, interpret, and compare different types of risks.

Any d.m. whose utility function is bounded (in the sense of FSD) by two exponential utility functions will, assuming continuity of preferences, have a CRS risk value for \(F(a)\) (or \((p(a),F(a))\), depending on the risk model used) that falls within the interval on the CRS axis spanned by the canonical risk curve. Thus, even if his own risk attitude is not constant, limited-precision objective measurement of risk on the CRS scale is possible for such d.m.'s. Moreover, such d.m.'s can detect stochastic dominance among acts by looking for stochastic dominance among corresponding canonical risk curves.

Showing the canonical risk curves for multiple acts on one set of axes allows the action partition for this set of acts to be identified for all exponential utility d.m.'s from the lower envelope of these curves.

Treatment of uncertain risks is particularly convenient. Suppose for example that \(F(a)\) is uncertain and depends on a parameter \(s\), i.e., \(F(a) = F(a;s)\) where \(s\) is a random variable. Then \(r[F(a;s)]\) lies between \(E[r[F(a;s)]]\) at the CRS intercept and \(\max[r[F(a;s)]]\), where maximization is over \(s\).

Finally, unlike a plot of expected utility, for example, both axes in Figure 5.1 have natural intuitive interpretations in terms of risk attitude and standard reference risks.

In summary, the canonical risk curve appears to be a potentially useful theoretical measure of risk for a variety of applications. Whether it will prove useful and acceptable in practice remains to be determined.
CHAPTER 6  COMMUNICATION BETWEEN TWO INFORMED PLAYERS

6.0 Introduction

The preceding chapter explored the problem of one-way risk communication between an informed r.a. and an uninformed d.m. The r.a. was assumed to know the true cumulative probability distribution function (c.d.f.) over consequences, say \( F(a) \), associated with each act \( a \) in the d.m.'s choice set. In reality, however, information is seldom so one-sided. The d.m. may have his own c.d.f. assessed over \( C \) before the r.a. says anything. He may only partially revise, or perhaps not revise at all, his prior beliefs in light of the r.a.'s message. How he adjusts his beliefs, if at all, will in general depend both on what he knows and on what he thinks the r.a. knows. The purpose of this chapter is to explore problems of communication when both the r.a. and the d.m. have (possibly different) knowledge about the c.d.f.'s over consequences associated with different acts.

Section 6.1 introduces the problem of combining knowledge or information from different sources by considering how a d.m. can use expert probability judgements to update his own beliefs -- a topic of considerable current interest in applied risk assessment and statistics. The main observation of this section is that probabilities are in general not sufficient to summarize what an expert knows. The same probabilities can be combined in different ways depending on what is known about what the experts providing the judgements know.

Section 6.2 presents sufficient conditions under which iterative communication of probabilities does suffice to allow two observers with access to different deterministic information channels to communicate their empirical knowledge to each other. The essential condition is that both players share a common initial (prior) probability algebra. That iterative communication (and Bayesian revision) of probabilities leads to eventual agreement on probabilities is one of the central results of the developing theory of "common knowledge" (Aumann, 1976, Sebenius, 1983).
A detailed proof (not given here, but spelled out in the references cited) makes extensive use of the algebra of information partitions introduced in Chapter 2.

Unfortunately, no theorems corresponding to the common knowledge result for probabilities have yet been developed for cases in which observers have different prior probability algebras and/or make observations through noisy channels. Section 6.3 summarizes some classical statistical results that allow an informed observer to communicate the essential contents of his observations without using his own subjective prior, by summarizing them in a "sufficient statistic" based on the likelihood ratio, assuming that this is known. An analog of the impossibility result in Chapter 5 for objective numerical risk measurement shows that one-dimensional numerical sufficient statistics do not exist in general (Theorem 6.4.)

Most of the results in this chapter are already known. They are collected here to provide a balance for the one-sided perspective taken in Chapters 5, 7, and 8, in which an informed risk analyst is attempting to communicate with an uninformed decision maker. Substantially more work needs to be done in this area, however, to develop a useful theory of efficient risk measurement and communication between two or more partially informed players when neither has strictly more information (in the sense defined in Chapter 2) than the other. Extending the theory of common knowledge reviewed in Section 6.2 to cases where players may have different priors and noisy observation channels appears to be a particularly worthwhile research direction, but has not been explored here beyond the elementary results in Section 6.3.

6.1 Combining Expert Judgements of Probabilities

Before he learns anything from the r.a., the d.m. will have some beliefs about the probable consequences of the different acts in his choice set, A. For mathematical simplicity, we shall assume for the moment that these beliefs can be represented by a set of probability measures $p(c;a)$
over atomic consequences \( c \) in a finite or countable consequence set \( C \). Uncertainty about \( c \) given \( a \) is due to uncertainty about state \( s \) in the deterministic consequence function \( c(a,s) \), where \( s \) belongs to a finite or countable state set \( S \), and the d.m.'s beliefs about \( s \) are represented by a probability measure \( p(s) \).

Now suppose that the r.a. sends the d.m. a message \( q(c;a) \), e.g., a table giving the probability of each consequence for each act based on the r.a.'s information, where \( q \) denotes the r.a.'s probabilities. What use can the d.m. make of this information?

**Example 6.1: Competing Information about Poisson Risks**

Suppose that \( A \) is a subset of the nonnegative reals, \( C \) the set of natural numbers (zero and the positive integers), and that \( p(c;a) \) follows a Poisson distribution with parameter \( 2a \). Now let the r.a. send a message to the d.m. saying that \( q(c;a) \) follows a Poisson distribution with parameter \( a \). In other words, this is the conditional distribution over consequences given act \( a \), as judged by the r.a. based on his own information. How should the d.m. treat this information? Should he ignore it? Should he abandon his own initial opinion and accept the r.a.'s assessment? Or should he average the two, perhaps adopting a posterior probability measure \( p'(c;a) \) that is Poisson with parameter 1.5?

Rather surprisingly, a considerable literature has developed in the management science literature that attempts to answer such questions, in the simpler context of estimating event probabilities, by proposing axioms and formulas for combining the probability judgements of experts (Schervish, 1986, French, 1986.) A typical such axiom, translated to this context, is that if \( q(c;a) = p(c;a) \), then \( p'(c;a) = p(c;a) \), where \( p' \) denotes the d.m.'s posterior p.d.f. after adjusting his prior, \( p \), in light of the expert's (r.a.'s) message, \( q \). This is sometimes called the "unanimity" axiom for resolution of probability judgements (Dalkey, 1976.) Other axioms lead to aggregation formulas in which the d.m. weights the r.a.'s beliefs according to how well-calibrated and accurate he believes the r.a. to be. If there are multiple experts, their
probability judgements are weighted according to their relative degrees of expertise and calibration, as judged by the d.m., and according to how well they agree with each other. (See e.g., Morris, 1983, Sampson and Smith, 1982, Bordley and Wolf, 1981, Winkler, 1986.)

This axiomatic approach to the combination of probabilistic beliefs leaves out some essential aspects of the problem of combining knowledge.

Example 6.2: Treating Insomnia

A medical doctor (the d.m.) must choose whether to prescribe sleeping pills for a patient who complains of insomnia. The doctor knows that insomnia can be caused by depression or by greed (or both). Moreover, he knows a priori that there is a 50% chance that a randomly selected patient with insomnia will be depressed; independently, there is a 50% chance that he will be greedy. Finally, he knows from background statistics that the success rate of the sleeping pills is

60% among greedy, depressed insomniacs.

40% among greedy insomniacs who are not depressed;

40% among insomniacs who are depressed but not greedy; and

0% among insomniacs who are neither greedy nor depressed.

Also, 30% of greedy insomniacs are depressed and 30% of depressed insomniacs are greedy.

The sleeping pill creates headaches in people for whom it is not effective, so the doctor wishes to estimate the probability that it will be effective before deciding whether to prescribe it. Moreover, it follows from the above that the prior success probability is only 35% -- too large a risk to accept. Accordingly, he calls the patient's banker (who knows whether or not the patient is greedy) and, after explaining the above facts, asks the banker for his expert opinion that the pills
will prove effective. After a short calculation, the banker asserts that this probability is

\[(0.5)(0.6) + (0.5)(0.4) = 50\%\].

From this figure the d.m. can deduce that his patient must be greedy, since otherwise the banker would have announced a probability of

\[(0.5)(0.4) + (0.5)(0) = 40\%\],

and so his own updated probability judgement will agree with the banker's.

Next, the d.m. asks an expert psychologist (who can tell whether the patient is depressed) to interview the patient and then to estimate the probability that the pills will be effective. He first explains the medical and demographic facts of the case, but without mentioning the banker's estimate. After interviewing the patient, the psychologist announces that the probability that the pills will be effective, based on the information available to him, is 50\% -- from which the d.m. concludes that the patient is depressed as well as greedy.

After receiving the banker's estimate, the d.m.'s probability estimate rose from 35\% to 50\%. The psychologist's announced probability agrees with this: it is also 50\%. According to the unanimity axiom, therefore, the d.m.'s final probability should remain unchanged at 50\%. But in fact, the success probability for sleeping pills among greedy depressed insomniacs is 60\%, and this will be the doctor's final estimate.

In summary, the unanimity principle for aggregating probabilities ignores the fact that it is possible for a d.m. to learn from the probability judgements of others not just their opinions about probabilities, but substantive facts that are signalled or encoded in the announced probabilities.
THEOREM 6.1: Suppose that each of N experts independently observes the state s once through his own noisy channel, and that he computes his posterior probability for an event E (a subset of S) on the basis of his observation and Bayes' Rule, starting from a known prior that is common knowledge. Each expert announces his posterior to the d.m. (who may be one of the experts.) If S contains at least three states, then it is in general impossible to assign weights to the experts, say $w_i$ for expert i, and to construct an aggregation function $f$ such that for any set of announced probabilities $p_1',... , p_K', f(p_1', p_2', ... p_N'; w_1', w_2', ... , w_N)$ will give the d.m. the same posterior as he would obtain if each expert reported his observation (and channel matrix) directly.

Proof: This amounts to the observation that the set of channels, or possible "types" of experts, can not be indexed by the reals; see Cox (1984) for details.

Thus, any aggregation scheme based on weighting the announced probability judgements of experts must in general omit information. An exception to this rule is the following:

THEOREM 6.2: Suppose that there are only two possible states, say $s = 1$ and $s = 2$, and let expert i announce $p_i$ for the probability of $s = 1$, based on a single independent observation through a noisy binary channel. Define $L_i = p_i/(1 - p_i)$ as expert i's posterior odds in favor of $s = 1$, and let $L_0$ be the prior odds in favor of $s = 1$ (assumed to be common knowledge.) Then the d.m.'s posterior probability for $s = 1$ is given by

$$\frac{L}{L + L_0^{-N}}$$

where

$L = (L_1)(L_2)...(L_N)$ is the product of the N posterior odds.

Proof: This follows from algebraic manipulation using Bayes' Rule; see Cox (1984) for details.
In this special case, therefore, his probability adequately summarizes what each expert knows.

In summary, the d.m.'s posterior probability, based on his own initial information and on the probability judgements of the experts advising him, should in general depend on their channels, which can not be indexed by any weights (or finite set of weights), as well as on what they announce.

Even when the d.m. knows the experts' channels, some problems of communication may remain. For example, if multiple signals in an expert's channel output set map onto the same posterior probability for E, given the prior q(s), then the expert's announced probability will be an ambiguous indication of what he observed. Hence, the d.m. who uses this probability to update his own prior p(s) will end up with a p'(s) that will necessarily be different from what it would have been if the expert (the r.a.) had simply announced what he observed, instead of encoding it as a probability.

Example 6.3: A Two-Out-of-Three Reliability Problem

A given system will operate as long as at least two of its three components, A, B, and C, operate. Each component, and hence the system as a whole, has a 50% prior probability of operating, and an expert can tell by examining a component whether it will operate. Expert 1 examines components A and B and reports that the system has a 50% posterior probability of working; hence, the d.m. knows that exactly one of A and B has failed. Expert 2 examines components B and C and reports that his posterior probability for system success (i.e., successful operation) is 50%. From these two reports, the d.m. can conclude only that the probability that the system will operate is 50%; he has in this sense learned nothing. If Expert 2 had reported his observation (i.e., whether B or C had failed), then the d.m.'s posterior probability for system failure would have been 0 or 1, respectively. Because Expert 2's announced probability would be the same whether he observed B failed and C working or C failed and B working, however, his announcement is
ambiguous as an indication of what he observed, and the d.m.'s posterior is a weighted average of the two possible posteriors that he would have had under each observation.

(Note that if, after these announcements, a third expert, starting with the same prior as the others, observes component B alone, and announces a posterior probability of 0.25, then the d.m. will conclude that \( p' = 1 \); while if the third expert announces 0.75 (implying that B is operating) the d.m. will conclude that \( p' = 0 \).)

This type of ambiguity in the significance of announced probabilities is one potential problem in the communication of risk information even when the d.m. knows the r.a.'s channel matrix (e.g., the subset of components examined by each expert.) If the state set \( S \) is continuous and nonatomic, however, and the event \( E \) of interest (e.g., two or more components working, in the above example) is chosen at random, then the probability of such ambiguity in the r.a.'s observation-to-announced probability mapping will approach zero as the number of significant digits reported increases.

6.2 Iterative Communication and Common Knowledge

Even if ambiguity of announced probabilities is a potential problem, it can be resolved through iterative communication.

Example 6.3 (Cont.): Common Knowledge as a Communication Equilibrium

Suppose that in the 2-out-of-3 reliability problem described above, the experts are allowed to communicate. Specifically, let there be two experts, 1 and 2 (e.g., the r.a. and the d.m.), and suppose that 1 examines components A and B and 2 examines components B and C. Suppose that both see that only B has failed. Suppose, moreover, that they communicate their probabilities to each other iteratively. Then the following sequence will take place.
STEP 0: Neither player has any empirical evidence, and both have a common prior denoted by (0.5, 0.5, 0.5; 0.5), denoting the judged success probabilities of components (A, B, C; system), respectively.

STEP 1: Player 1 observes components A and B and sees that B has failed. He has posterior probabilities (1, 0, 0.5; 0.5), and announces 0.5 as his updated probability for system success.

STEP 2: Player 2 knows that player 1 knows either that A has failed and B succeeded, or that B has failed and A succeeded. Player 2 observes components B and C and sees that B has failed. Hence, he deduces that player 1 knows that B has failed and that A has succeeded. He announces 1 as the probability of system success.

STEP 3: From this announced probability of 1, player 1 deduces that player 2 knows that player 1 knows that A has succeeded (and that B has failed); and that player 2 knows that C has succeeded. So he, also announces 1 as his new posterior probability of system success.

At this point, both players know what each knows, and the adjustment process terminates. More specifically, (i) each player knows (1, 0, 1; 1); (ii) each player knows that each player knows this; (iii) each player knows that each player knows that each player knows this, and so forth. We may say that the players have achieved "common knowledge" about the state of the system.

This simple example generalizes in a number of interesting ways.

THEOREM 6.3: Let S be a finite or infinite set of states and F an algebra (or sigma field) of events on S. Let E be any event in F, and suppose that

- Two players, 1 and 2, have identical priors $p(s')$ on the states $s'$ in S;
Each player then acquires some private empirical information (or knowledge) about the true state \( s \) by an observation through a deterministic channel (e.g., each player gets to examine certain components of a complex system);

The channel matrices are finite and common knowledge and the join of the corresponding information partitions consists of non-null events (e.g., both players know which components each examines and what it is a priori possible for each to observe); and

After receiving their private information, the two players take turns announcing their updated probabilities for event \( E \) (e.g., the event that the system will operate).

Then

(i) the players' updated probabilities for \( E \) will converge to a common posterior, say \( p^*(E) \), within \( Z \) rounds, where \( Z \) is the sum of the numbers of signals in each player's signal set; and

(ii) The probability \( p^*(E) \) will be common knowledge, in the sense of belonging to a "common knowledge base" \( K^* \) such that

\( (p^*(E)) \) is in \( K^* \);

If (a) is any proposition in \( K^* \), then the propositions \( (K_1(a)) \) and \( (K_2(a)) \) are both in \( K^* \);

No other propositions are in \( K^* \).

Here, of course, \( K_i(a) \) is interpreted as "i knows (a)" in the sense of individual knowledge defined in Chapter 2.

Proof: See Geanakoplos and Polemarchakis, 1982. The proof works by showing that each player's information partition is refined by the other's announcement until the block of their join (coarsest common
refinement) that contains $s$ has been obtained. At this point they have the same information and can learn nothing more from each other. Milgrom (1981) provides an axiomatic characterization of common knowledge $K^*$ in terms of events rather than propositions, and shows that an event is common knowledge for two players when the true state is $s$ if and only if it is a superset of the block of the join of their information partitions that contains $s$.

This result shows that, under the conditions of the theorem, iterative communication of probabilities leads to common knowledge about the smallest set of states that is known to contain $s$. Interestingly, for any integer $n$ it is possible to construct an example in which each player announces the same initial probability in each of $n - 1$ rounds, and only in the last (n-th) round do these probabilities adjust to equal each other (Geanakoplos and Polemarchkis, 1982.) Finally, Bacharach (1985) suggests that the same type of result as in the above theorem holds when the players exchange proposed actions instead of probabilities, assuming that both have a common utility function.

In summary, the common knowledge literature shows that communication of probabilities may be satisfactory, in that it leads to a merging (or join) of the relevant parts of players' information partitions, when the conditions of the theorem are met. How far this result can be extended -- for example, what happens when the players' channels are noisy instead of deterministic -- has yet to be investigated.

6.3 Communication Without Common Priors

6.3.1 Use of Likelihood Ratios as Sufficient Statistics

A second type of ambiguity occurs when the d.m. may have a different prior from the r.a. If the d.m. is uncertain about the r.a.'s prior, then he will not know how to interpret the r.a.'s announced probability even if he knows the r.a.'s channel matrix. Thus, the problem arises for the r.a. of how to communicate his knowledge without using any particular prior.
Suppose again that the r.a. gains probabilistic knowledge about \( s \) by means of one or more observations through a noisy channel, or through a sequence of channels, each with known channel matrix. Then he can summarize the probabilistic evidence that these observations have given him by reporting the likelihood ratios for different states, i.e., the ratios \( \frac{\Pr(y; s')}{\Pr(y; s'')} \), where \( y \) denotes the set of observations that he has made and \( s' \) and \( s'' \) are states in \( S \). More generally, if he has been asked about the probability of an event \( E \), e.g., the probability of system failure in the above example, then he can report the likelihood ratio

\[
L(E;y) = \frac{\Pr(y; E)}{\Pr(y; E')}
\]

where \( E' \) denotes the complement of \( E \) and \( \Pr(y; E) \), for example, is found by summing over all states \( s' \) in \( E \) the channel matrix probabilities \( \Pr(y; s') \). Now \( L(E;y) \) is precisely the information that the d.m. needs to compute his own posterior from Bayes' Rule:

\[
p(E;y) = \frac{\Pr(y; E)p(E)}{[\Pr(y; E)p(E) + \Pr(y; E')p(E')]} = \frac{p(E)}{[p(E) + p(E')/L(E;y)]}
\]

using his own prior \( p(E) \). Thus, knowing the likelihood ratio \( L(E;y) \) alone is as useful to him for this purpose as knowing the r.a.'s observations, \( y \), themselves. (Technically, \( L(E;y) \) is a "minimal sufficient statistic" for \( s \) based on the r.a.'s set of observations \( y \).)

Since the number \( L(E;y) \) is independent of the r.a.'s prior beliefs, say \( q(E) \), the potential problem of discrepant priors between himself and the d.m. has been overcome.

Now suppose that the r.a. does not know which event(s) the d.m. is interested in. Specifically, assume that \( S = [0,1] \times [0,1] \), i.e., that both the act and state sets are continuous intervals, and that the d.m. can partition the consequence set \( C = A \times S \) into blocks such that he is indifferent between all pairs \((a, s)\) in the same block. Then all the d.m.
cares about is the probability distribution over these classes induced by each act a in A. Hence, any states that imply the same probability distribution over these indifference classes are themselves indifferent to the d.m. Without knowing the d.m.'s indifference partition of the consequence set, however, the r.a. has no way of knowing which events, i.e., which induced indifference sets in S, the d.m. is interested in.

Example 6.4: The Perils of Joe

Joe is driving to rescue his girl, Pauline, who is tied to a railroad track 55 miles away. He can drive at any speed, a, between 0 and 110 m.p.h.; however, driving at over 55 m.p.h. makes Joe nervous and depressed. He figures that he should drive at either 55 or 110, though the latter will make him miserable. On the other hand, he knows that a train is heading toward Pauline's position, and that arriving after the train has passed will almost certainly give him insomnia, which will also make him miserable. What he does not know is the remaining time, s, until the train reaches the spot where Pauline is now lying.

Joe partitions the possible consequences into four indifference classes: E1 = speeding and arriving late (s is less than half an hour); E2 = not speeding and arriving late (s is less than one hour); E3 = speeding and arriving early (s is greater than half an hour); and E4 = not speeding and arriving early (s is greater than one hour). These indifference classes (which are numbered in order of increasing preference) induce the following indifference partition of the state set S of possible (as perceived by Joe) arrival times for the train: less than half an hour; between half an hour and an hour; and over an hour. Had Joe been at a different distance from the track, however, he would have had a different partition.

6.3.2 Communicating Arrival Time Distributions

Given a continuous state set S *=[0,1], and assuming that he does not know which events in S the d.m. will be interested in, the r.a.'s problem is to find a message that will convey to the d.m. everything that the
r.a. knows about the probability distribution over states. We shall henceforth concentrate on the case, important for risk analysis, in which the uncertain quantity of interest is the time until the next arrival of an undesirable event. More specifically, $s$ will index the distribution of that time, which is still a random variable at the time the d.m. must make a decision.

We assume that the r.a. has some prior beliefs, which he wants to leave out of his message; in addition, he has a set of known observations, $y$, that give him statistical information about the distribution of the random variable, denoted $T$, representing the time until the next arrival. Specifically, we will assume that the r.a.'s empirical information $y$ consists of $N$ mutually independent sample realizations (i.e., observations) of the random variable $T$. In Example 6.4, $y = (y_1, y_2, \ldots, y_N)$ would be the set of observed interarrival times between trains, where the observed times between trains are assumed to be i.i.d. realizations of $T$, and the r.a. has observed a sample of $N$ of these times.

We assume that the distribution of $T$ is uncertain; the possible distributions are assumed to belong to a one-parameter family indexed by the continuous state variable $s$. (For example, the time between train arrivals may be exponentially distributed with unknown parameter $s$.) Thus, the state $s$ determines (or is) the distribution of $T$, and hence of the d.m.'s observations $y$, which may be thought of as consisting of $N$ observations of the distribution of $T$ (and hence $N$ observations of $s$) through a noisy channel. Note that this is a very different interpretation of "state" than that used above; in particular, knowing the act $a$ and the "state" $s$ (i.e., the distribution of $T$) in the train example only determines a probability distribution over consequences, and not the consequences themselves. But probability distributions over consequences are precisely what are needed for expected utility theories of choice to apply.
Given his observations, y, what message can the r.a. send to the d.m. that will allow the d.m. to calculate his own posterior probability for any (Borel) subset, E, of the time axis that he might be interested in? The answer is that the likelihood function \( L(y;s) \), considered as a function of s, is still a minimal sufficient statistic for s given the r.a.'s observations. (See e.g., Cox and Hinkley, 1974, p. 24.) For any two hypothesized distributions of T, indexed by \( s' \) and \( s'' \), \( L(y;s')/L(y;s'') \) is the likelihood ratio in favor of \( s' \).

The r.a. has empirical knowledge y about the distribution of T, i.e., he knows what he has observed. Let \( M(1) \approx [0,1] \) be a set of numerical messages that he can send to the d.m. Can he encode his relevant knowledge, \( L(y;s) \), from y into a single number, say \( t(y) \), so that from this number alone the d.m. can construct the same posterior for T as if he had observed y himself? If so, then \( t(y) \) will be called a scalar sufficient statistic for the r.a.'s knowledge of s based on his empirical observations, y. The challenge of risk communication in this setting is to find a scalar sufficient statistic for the r.a.'s empirical knowledge about (the distribution of) T. This problem has been treated in classical statistics.

THEOREM 6.4: Assume that the d.m. knows only how many observations, \( N \), the r.a. has made, but not what he has observed. Then a scalar sufficient statistic \( t(y) \) for the r.a.'s empirical information about T based on his observations y exists if and only if the set \( S \) of possible distributions of T is one of the exponential (or Koopman-Darmois) families of distributions, i.e., if and only if the density function for T can be expressed in the form

\[
f(t;s) = a(s)b(t)\exp(c(s)d(t))
\]

where \( a(.) \) and \( c(.) \) are arbitrary functions of s and \( b(.) \) and \( d(.) \) are arbitrary functions of the argument t. In this case, the average value of \( d(t) \) over all \( N \) realized values t of T is such a numerical sufficient statistic.

Proof: See D.R. Cox and Hinkley, 1974, p. 61.
Note that the following practically important families of distributions are exponential families:

- Poiss(m), the family of Poisson distributions with unknown parameter m;
- The family of normal distributions $N(m,v)$ with either parameter fixed;
- The gamma, beta, and negative binomial families with one parameter fixed;
- The family of Bernoulli distributions with unknown parameter p.

Theorem 6.4 implies that in general, a r.a.'s knowledge about the time until the next occurrence of a risk event cannot be summarized by a single number. Even for exponential families, it is in general necessary to specify $N$ as well as the number $t(y)$.

The approach to summarizing knowledge about interarrival time distributions through sufficient statistics, presented in this section, is standard in mathematical statistics. It is applicable when an r.a. is making a judgement at one point in time, based on past observations, about the probable time to the next occurrence of a marker event. To understand risk over time, however, it is necessary to go beyond such one-time predictions and to consider instead the experience of a d.m. continually threatened by the possibility of the next arrival of the marker event. The assessed probability distribution for the remaining time until its arrival will in general be continuously evolving for such a d.m. The problem of representing and measuring such dynamic risks is considered in the next chapter.
CHAPTER 7 MEASUREMENT OF INDIVIDUAL FATALITY RISK

7.0 Introduction and Overview

The preceding chapters have developed abstract mathematical principles for measuring risks for the purposes of communicating decision-relevant information to decision makers. In this chapter, we consider extensions and applications of this theory to the measurement of fatality risks for an individual over time. In the next chapter, this treatment will be extended to cases where many individuals are at risk.

Section 7.1 gives a brief summary of the conceptual framework developed in the preceding five chapters. Section 7.2 introduces three simple models for human fatality risks -- one for individual risk and two for population risks. The remainder of the chapter explores problems of measuring individual fatality risks over time, while Chapter 8 considers population fatality risks. The mathematical techniques developed in this chapter apply to any situation in which the risk event of interest is the first arrival of an undesirable event, which we will refer to as the marker event.

The principal theoretical developments in this chapter are as follows:

- An extension of the canonical risk scale (CRS) idea to known "dynamic" risks using a relation of temporal first order stochastic dominance (FSD), denoted by FSD* (Section 7.3.2). Dynamic risks create a flow of disutility over time that can be measured on a CRS constructed from the exponential family of distributions ordered by FSD*. Moreover, the resulting "hazard rate" (e.g., measured in micromorts per hour) is an exact, rather than a limited-precision, index of instantaneous risk on this scale even when the distribution of remaining time until the arrival of the marker event is not itself exponential.
A distinction is drawn among three types of dynamic risks: known, or deterministically evolving; uncertain; and stochastically evolving. A **known** risk is created by a known fixed probability distribution over the remaining time until arrival of the marker event. An **uncertain** risk is created when this distribution is not known exactly, but is only known to have been drawn from some family of distributions according to some known probability distribution over the family. For both known and uncertain risks, the passage of time conveys information that continually changes the conditional distribution of the remaining time until the marker event arrives. This creates a flow of disutility or risk that can be measured in mors or micromorts per second (assuming continuous observation) and that follows a deterministic time path, or "trajectory", up until the moment when it is truncated by arrival of the marker event. By contrast, a **stochastically evolving** risk creates a nondeterministic flow of risk. Stochastically evolving risks are created when an observer obtains information over time other than that contained in the passage of time itself.

The close connection between **risk** and **observation** is explored through simple examples. In a general theory of risk measurement, it is necessary to define assessed risk with respect to different streams of observations, and no hazard rate may exist if the observation structure allows certain kinds of prediction. (For example, no hazard rate exists if the time of arrival for the marker event is known exactly in advance, even though the experience of living with this knowledge may create a flow of psychological disutility.) However, the observation structure most often used in catastrophic risk assessment is the simple one in which the arrival time of the marker event is not known until it occurs and hazard functions exist. We concentrate on this situation.
A framework for analyzing stochastic risks in discrete time is introduced in Section 7.4. The basic idea of this framework can be understood most easily by supposing that an observer's knowledge and beliefs at any moment can be represented by (i) an initial assignment of prior probabilities to the atomic states (or blocks of states) of the world; and (ii) an information partition and a block within the partition that he knows contains the true state. Given this knowledge, his posterior distribution over states is found by restricting his prior to the block in which he knows the state lies and renormalizing the probabilities of the states in that block by dividing them by the prior probability for the block. His knowledge and beliefs are updated over time by observations taken through a sequence of (possibly randomly selected) deterministic channels. The observer learns which channel he is using before each observation, however.

Central to these developments is a view that the observer's assessed probability for any event, including the arrival of the marker event at a specified future time, will evolve according to a stochastic process determined by the sequence of his observations. Each observation refines his information partition, and probabilities calculated with respect to the new partition may differ from probabilities calculated with respect to previous ones. However, the sequence of probabilities assessed with respect to the sequence of increasingly refined information partitions does constitute a martingale, i.e., the expected probability of any event at any future time, given any sequence of channels through which observations may be taken in the interim, computed with respect to the current information partition is exactly the currently assessed probability for the event. This gives a new significance, unimportant in the usual context of choice among static probability distributions, to the use of expectation in "expected utility" representations of preferences.
The framework introduced in Section 7.4 has the intuitive interpretation and motivation just described, but is more general in that it applies to arbitrary probability algebras instead of to information partitions. It provides a natural generalization of the simple stochastic process for probabilities described at the end of Chapter 2. The stochastic risk framework is applied to define stochastic hazard functions, also called stochastic hazard rate trajectories, and an expected utility theory representation is developed for preferences over them (Theorem 7.10) by applying the pioneering theoretical work of two mathematical economists (Kreps and Porteus, 1979).

The chapter concludes with a discussion of remaining challenges and possible extensions of the theories of individual risk measurement presented here. In particular, the sensitivity of the conclusions to assumptions about the types of control and observation opportunities available is illustrated, and connections to related topics in statistics are discussed.

7.1 Precis

The theory of risk measurement developed in the preceding chapters can be summarized as follows.

- A "risk" is identified as a quadruple \((A, B, C, D)\), labelling the source, target population, effect, and mechanism of the risk, respectively. In this work we are concerned only with \((A, B, C)\), and not with the empirical mechanisms of risk. Each member of \(A\) is a human activity, action, or decision, which we will call an "act." The "effect" of interest is arrival of the marker event.

- The problem of decision analysis is to construct a preference ordering over the objects of choice (acts) in \(A\) from "primitive" preferences over consequences, including the effects in \(C\). The exact nature of consequences has not been specified; in most
applications, they have been taken to be simply the effects in C. Act-state pairs, state-physical outcome pairs, and act-state-physical outcome triples were also mentioned as possible consequences. More generally and abstractly, consequences are the objects of (assumed) primitive preferences.

The problem of risk measurement is to assign a mathematical object, e.g., a probability distribution over C or a number representing such a distribution, to each act in A in such a way that mathematical comparisons between the assigned mathematical objects imply corresponding derived preference comparisons between the members of A. For example, if the "risk" r(a) associated with act a is a number, then the relation of inequality between risk numbers should represent a relation of (derived) preference between the corresponding acts. If the "risk" r(a) is represented by a probability measure over C and if the members of C are ordered by primitive preferences, then the relation of first-order stochastic dominance, FSD, between probability measures over C should reflect derived preferences for the corresponding acts in A.

The derived preference relation over A is defined only with respect to the set C, or, more generally, the set of risk consequences of acts in A: preference relations over A defined with respect to other sets (e.g., deterministic cost or benefit flows) are not considered.

The problem of risk measurement has been posed in the context of communication between an r.a., who knows something about the probable consequences of the acts in A, and a d.m. who must choose an act from A. This forces risk measurement to be "objective" in a way that subjective expected utility measurement for a single d.m., for example, is not. The r.a. may not know all aspects of the d.m.'s preferences, e.g., his deterministic value function for consequences, or his personal risk attitude; moreover, the r.a. may be forced to use a message set with a restricted (e.g., numerical) structure to represent risks for the purposes of communication.
By far the best-studied class of risk measurement problems has been for the case where \( C \) is the consequence set and has the mathematical structure of an interval, \( C \equiv [0,M] \). For this case, we showed that objective numerical risk measurement was in general impossible, although limited-precision and even exact numerical measurement were possible in certain special cases. We developed the idea of a "canonical risk scale" (CRS), calibrated by reference to a one-parameter family of standard risks totally ordered by FSD, for representing a wide variety of risks.

Finally, a "canonical risk curve" plotting the risk of a c.d.f. over consequences, measured on the CRS, against an index of subjective risk aversion such as the certainty equivalent for a standardized lottery, was proposed as a useful representation of risk for an informed r.a. (knowing the c.d.f.) sending a one-way message to an uninformed d.m. The projection of this curve onto the CRS axis could also sometimes be used as a simpler, limited-precision, numerical risk measure. Modifications for cases in which the d.m. as well as the r.a. had prior information, using sufficient statistics and iterative exchanges of probabilities, were examined.

In this chapter, we turn to a more specific class of risk measurement problems commonly encountered in applied risk analysis: those involving human fatalities as potential outcomes. We shall concentrate on the problems of measuring human fatality risks. The "consequences", i.e., the objects of (assumed) primitive preferences, will be shown to have a mathematical structure very different in this case from \([0,1]\) or even from the set of natural numbers (zero and the integers.) In structuring the relevant consequence set and developing appropriate mathematical representations for preferences over this set, it will become clear that the traditional distinction between preferences for acts and preferences for consequences is sometimes no longer useful when decisions and consequences are interleaved over time. Thus, conventional normative decision theories developed for static decision problems require extension.
7.2 Three Fatality Risk Models

One lesson from the theoretical analyses of previous chapters is that there can be multiple normative decision theories (NDT's) and multiple corresponding theories of risk measurement, depending on the mathematical structures of the consequence set and of the probabilistic relation between choice of act and occurrence of consequence. Useful and simple (e.g., numerical) risk measures can be developed only by exploiting such special structures. It is therefore useful to identify the special structures most commonly found in analyses of fatality risks. Starting from a taxonomy of such structures, we can seek to develop useful measures of risk in each.

The following three main classes of fatality risks have been dominant in risk analysis applications to date:

- Individual fatality risks
- Routine population fatality risks
- Catastrophic population fatality risks.

Each is defined informally below. In addition, there are a variety of non-fatality risks to human health and safety, e.g., chronic diseases or debilitating injuries, that are much more difficult to represent and analyze mathematically. An example would be exposure to a risk factor that, while not causing any observable damage itself, leaves the victim more susceptible to other risk factors to which he may be exposed. Or, in the case of injury, it becomes necessary to examine the degree as well as the occurrence of injury -- a process with its own bounded subjectivities, where limited-precision numerical measurement seems to be the best that can be hoped for. In these and other nonfatality cases one loses the simplicity of a dichotomous outcome measure -- death or no death. By focusing on fatalities, we hope to prepare a foundation for studying risk measurement in these more difficult cases.
Each of the three classes of fatality risks just listed is defined more carefully below. The idealized mathematical models introduced for these risks will serve as frameworks for discussing risk measurement.

7.2.1 Individual Fatality Risks

A. Introduction

The simplest type of fatality risk conceptually, though not statistically, is that in which a single individual is identified as the target of the risk. Such a risk can be specified as a triple \((a^*, b, c)\), where \(a^*\) is an act in the d.m.'s choice set whose risk (to \(b\)) is to be measured; \(b\) identifies the individual "at risk", i.e., for whom the risk from act \(a^*\) is to be measured; and \(c\), the effect of interest, is the death of \(b\). More generally, \(c\) may be any observable change in \(b\)'s status; we will call it the \textit{marker event} used to define a risk.

Immediately, more precision is needed. Everyone dies eventually, so perhaps \(c\) needs to specify a death date to be interpreted as an adequate indicator of outcome. Only the risk of \(b\)'s death due to \(a^*\) is to be counted, so perhaps the outcome indicator \(c\) should include an indication of the cause of death. For fatalities with complex etiologies, however, no unique cause of death can be assigned except by fiat: the d.m.'s act \(a^*\) may have been a contributing factor or one part of a synergistic mixture, so that attribution of \(b\)'s death to act \(a^*\) alone, in the presence of joint causes, would be misleading.

Important examples of joint causes arising from joint decisions include the following:

- Consumption or use of a hazardous product (e.g., cigarettes, saccharin, diving boards). Should the risk to the consumer be imputed to his decision to consume or to the manufacturer's decision to produce and market the product? Which "act" should be assigned the role of \(a^*\)?
The risk to an employee in a hazardous occupation. Was it his decision to accept employment or his employer's decision to hire him that created the risk to his health and safety?

The risk to a neighbor of a hazardous facility. Is it his decision to live near the facility or the facility manager's decision to operate the facility in an inhabited area that creates the risk?

These questions do not require either-or answers. They do show the importance of specifying the act with respect to which a risk is to be defined, since the same target, effect, and mechanism can be ascribed to different acts, depending on the purpose of the analysis. Whether \( a^* \) should be specified as the decision of a home-buyer to live in an industrial area or as the decision of an industry to operate in that area depends on whether the home-buyer or the industrialist is seeking advice.

B. Individual Fatality Risk Model 1

To avoid some of these conceptual difficulties and ambiguities, it is useful to focus on the remaining time until b's death as the key quantity of interest, without inquiring into the cause of death. This time, say \( T \), may be looked at as a random variable with a distribution conditioned on whatever is known about \( b \) and about the d.m.'s act, \( a^* \), whoever the d.m. is and however \( a^* \) is defined.

We shall assume that the risk assessor's (r.a.'s) information about individual \( b \) can be represented by either an information partition (resulting from one or more observations taken through a deterministic channel) or as a set of observations taken through a noisy channel. In either case, we will use \( I \) to denote his empirical information. Thus, \( b \) is assumed to have some true "type", \( t(b) \), only part of which is known to the r.a. The information \( I \) may often be conveniently thought of as consisting of the known components of \( t(b) \).
Let \( F(t;a) = F(t;a,I,p) \) denote the conditional probability that \( b \) will die within an amount of time \( t \), given that the d.m. takes act \( a \) and calculated with respect to empirical information \( I \) and some prior probability measure \( p \) over \( b \)'s complete type, \( t(b) \). (\( I \) and \( p \) will usually be left tacit.) Let \( 0 \) denote a "null" act or reference act in \( A \), corresponding to the status quo or reference point. Then one way to model the fatality risk to \( b \) from act \( a^* \) is in terms of the differential shift

\[
dF(a^*) = F(.;a^*) - F(.;0),
\]

where all these expressions also depend on \( I \) and \( p \).

More specifically, we postulate the following decision setting:

**Individual Fatality Risk Decision Model 1**

- Individual \( b \) initially (as of the decision date) has a remaining life distribution of \( F(.;0) \).
- The d.m. chooses an act from his choice set, \( A \). If he chooses act \( a \), then \( b \)'s remaining life distribution changes to \( F(.;a) \). The d.m. knows the effect, \( dF(a) = F(.;a) - F(.;0) \), of each act \( a \) in \( A \).
- No additional acts are taken. Individual \( b \) lives out the rest of his life in accord with distribution \( F(.;a) \).

This simple one-shot model of individual risk appears to be widely used (at least implicitly) in discussing the risks from radiation exposures and from use of certain consumer products. Note that the d.m.'s act may extend over time in this formulation, as in the case where \( b \) is the d.m. and his decision is whether to take up cigarette smoking: nothing has been presumed about its duration.
7.2.2 Population Fatality Models

In individual fatality Model 1, the effect of the d.m.'s choice is to change the probability distribution for b's remaining life. There is a single target, b, of the risk created by the d.m.'s action. In many risk management decisions, however, an essential part of the problem is that decisions affect the remaining life distributions of many people simultaneously and unevenly. For example,

- Building taller stacks on coal-fired power plants decreases the health risks to neighbors of the plant, but spreads much smaller risks over a much wider area. More people are affected, although the effect on each is expected to be comparatively small.

- Substituting capital equipment for labor in the production of an industrial product may substantially decrease the expected numbers of occupational fatalities per unit produced. Thus, the input mix decision, aside from affecting the economies of production by trading off capital against labor costs, affects the risks associated with production.

In assessing the risks of an activity, the numbers of individuals put at risk, and the magnitudes of the risks to which they are exposed, must be considered. The individual fatality risk model presented above will not suffice for examining such "population risks." We therefore introduce two additional models for treating population risks. The main difference between them is that in the first, individual fatalities are statistically independent, while in the second they are correlated.

**Population Fatality Model 1: Routine Population Risks**

- Initially, each individual in a target population B has some remaining lifetime distribution, say \( F(\cdot;0,t(i)) \) for an individual of "type" \( t(i) \). These remaining lifetimes are statistically independent.
Each individual is replaced as soon as he dies with an individual of the same type. (This "population renewal" assumption will be specified more exactly in Chapter 8.)

If the d.m. chooses act $a$ in $A$, then the remaining life distribution for individuals of type $t(i)$ changes to $F(\cdot;a,t(i))$.

We are interested in defining and measuring the effect on total population risk of an act $a$.

This model is intended for applications such as evaluating the public health impacts of a change in automobile tailpipe emissions or the occupational safety impacts of shielding all moving machine parts. In the second population risk model, by contrast, we are interested in the arrival of a single accident that can kill many people.

Population Fatality Model 2: Catastrophic Population Risks

Initially, each individual in target population $B$ is immortal (or, equivalently, he is replaced by an identical individual as soon as he dies from any cause other than the accident of concern.)

The d.m. chooses an act from $A$. If he chooses act $a$, then the time until an accident occurs will follow some distribution $F(\cdot;a)$, possibly putting positive weight at $T = \infty$, where $T$ is the arrival date. If an accident occurs at time $t$, then individual $i$ in the population will die with probability $p(i)$.

The process ends as soon as an accident occurs.

Once again, we are interested in comparing the population risks corresponding to different choices of $a$. 
Clearly, other population fatality risk models can be formalized and studied. For example, an intermediate model would allow the catastrophe-generating process in Model 2 to start over after each arrival of an accident, thus giving it a renewal component. Our two idealizations appear to be particularly useful for applications, however. They or ones like them appear to have been used, at least implicitly, in many discussions of public risks.

In the remainder of this chapter, we analyze individual risks alone. The usual interpretation will be that we are examining the fatality risk to a single individual, although other interpretations (e.g., that the risk is the risk of a plane's crashing) can be applied. Chapter 8 will then build upon this treatment to give a preliminary analysis of population fatality risks similar to those in population risk models 1 and 2.

7.3 Ranking Changes in Lifetime Distributions

When individual risk Model 1 is an adequate formulation of the fatality risk to an individual, measurement of risk can proceed along the lines already developed for single-attribute decision making. The time until b's death is a single numerical attribute, and the d.m.'s choice of act amounts to a choice among c.d.f.'s over the time axis. Assuming that preferences are strictly increasing in remaining life time, the construction of a CRS and of limited-precision risk measures and canonical risk curves for different acts can, in many respects, proceed as in Chapter 5.

Some refinements are needed, however. For one thing, even a well-informed r.a. will typically not know $F(.;0)$. This baseline distribution depends on a variety of factors (wrapped into the concept of $t(b)$, b's "type"), and the r.a. will often be as uncertain about most of these as the d.m.

When there is only one d.m., his uncertainty about $F(.;0)$ can be treated formally by indexing possible c.d.f.'s $F(.;0)$ with an index $s$, and
assessing a subjective distribution over s. In the case of an r.a.
advising a d.m., an alternative approach is available. Without knowing
\( F(.;0) \), the r.a. may still know a great deal about

\[ dF(a) = F(.;a) - F(.;0). \]

If so, he can let the differential shift \( dF(a) \), rather than the complete
distribution \( F(.;a) \), be the mathematical object associated with each act
\( a \), and can concentrate on measuring risks based on these differential
shifts. This is often a much easier task than measuring risks for the
uncertain distributions themselves since, as in other areas of empirical
analysis, the (approximate) change produced by an act will often be much
more accurately known than the base line to which the change is added.

Under the assumptions of the normative decision theories reviewed in
Chapter 3, including ordinary and generalized expected utility analysis,
the differential shifts \( dF(a) \) in the distribution of b's remaining life
span corresponding to different acts in \( A \) can be partially ordered by FSD
(first-order stochastic dominance), and this partial ordering will be
"objective" in the sense that all d.m.'s for whom preferences are
strictly increasing in \( T \) will agree with it. In many important
applications, this ordering will be total, allowing numerical risk
measurement for the acts in \( A \) independent of the base line distribution
\( F(.;0) \). There is a conceptual difficulty with this approach however,
that will be discussed in Section 7.3.2 below under the name "temporal
stochastic dominance." First, however, we examine an important class of
problems in which FSD does suffice for risk measurement.

7.3.1 A Special Case: Competing Risks

By far the most often used special case of Model 1 occurs when \( T \), the
time of an individual's death, is interpreted as the minimum of several
statistically independent times to death from different sources. That
is, deaths may arise from any of several sources, each of which generates
arrivals of fatal events according to some statistically independent
arrival-time distribution. If $T(i)$ is a random variable denoting the (latent) arrival time of a fatality from source $i$, for $i = 1, 2, \ldots, n$, starting from the decision date, then the actual arrival time of $b$'s death will be given by the random variable

$$T = \min\{T(1), T(2), \ldots, T(n)\}.$$ 

This is called a "competing risk" model of $b$'s remaining life distribution.

It is convenient to characterize arrival distributions in terms of their hazard rates, or instantaneous arrival intensities. If $T^*$ denotes the random time at which the next arrival occurs, and if $F(.)$ is the c.d.f. of the time until this next arrival, then the hazard rate implied by $F(.)$, evaluated at any time $t$ prior to the occurrence of $T^*$, is defined by

$$h(t) = \lim_{dt \to 0} \frac{\Pr[T^* \text{ is in } (t, t+dt); T^* > t]}{dt}$$

$$= \lim_{dt \to 0} \frac{[F(t + dt) - F(t)]/[1 - F(t)]}{dt}$$

$$= \frac{f(t)}{1 - F(t)},$$

where $f(t) = dF(t)/dt$.

Intuitively, $h(t)dt$ is (approximately) the conditional probability that the next arrival will occur in the next time increment, $dt$, i.e., that $T^*$ will lie in the interval $(t, t + dt)$, given that it has not occurred by time $t$. We will always assume that $F(.)$ is differentiable, so that the above limit exists, although more general definitions of the hazard rate can be used.

The hazard rate function, or simply hazard function, has several useful properties which will be used without proof. Among them are that

(i) There is a one-to-one correspondence between c.d.f.'s and their hazard functions.

(ii) If $T = \min\{T(1), \ldots, T(n)\}$ where the $T(i)$ are statistically independent, then

$$h(t) = h_1(t) + \ldots + h_n(t),$$
where $h(t)$ is the hazard rate corresponding to $T$ and $h_i(t)$ the hazard rate corresponding to $T(i)$.

(iii) The hazard rate $h(t)$ for an arrival time $T^*$ is constant if and only if the time until arrival is exponentially distributed with parameter $1/h$, where $h$ is the constant hazard rate. In this case, $T^*$ may be thought of as the time of the first arrival in a Poisson process with average arrival rate $h$.

These and other properties are collected and discussed in Chapter 2 of Cox, 1985.

The following observation is simple but very useful.

PROPOSITION 7.1: The hazard rate $h(t)$ is an exact numerical measure of individual risk at time $t$ on a CRS calibrated by the family of exponential distributions. Moreover, the risks from statistically independent competing risk sources are additive on this scale.

Think of each competing risk source as generating a flow of "micromorts" to the individual, with the intensity (measured in micromorts per second) from source $i$ being $h(i)$. (We will use the notation $h(i)$ rather than $h_i(t)$ when there is no need to emphasize the time argument.) Then since each $h(i)dt$ is proportional to $dt$ for sufficiently small $dt$, we can treat $h(i)$ as being constant for the purposes of instantaneous risk measurement. (In effect, each fatality-generating process $i$ is "locally Poisson", or has essentially constant hazard rates over a small enough time interval, although this hazard rate will change over time for all but homogeneous Poisson processes.) The family of (remaining) lifetime distributions with constant hazard rates is the exponential family, and this is a one-parameter family totally ordered by FSD; thus, it defines a canonical risk scale. Intuitively, the instantaneous "risk" on this scale associated with source $i$ at time $t$, namely $h_i(t)$, is the rate at which it is generating micromorts. Every d.m. would prefer to be exposed for the next $dt$ to a Poisson fatality-generating process with average
arrival rate \( r \), instead of to source \( i \), if and only if \( h(i) \) is greater than \( r \).

Two useful implications of the competing-risk model are that

(i) If the hazard rates from different sources are independent of the set of sources to which \( b \) is exposed, so that there is no interaction among sources, then it is not necessary to know what the base line distribution \( F(.;0) \) of an individual's remaining life is in order to calculate the incremental risk added by a new source, \( i \). The added risk at any moment is just the hazard rate, or flow of micromorts, from the new source, which can be calculated independently of what other sources the individual is exposed to.

(ii) If \( b \)'s death occurs at time \( t \) and the cause (source) of his death is unobservable, then the probability that it was due to source \( i \) is just \( h(i)/h \), evaluated at time \( t \), i.e., it is the fraction of micromorts contributed between \( t - dt \) and \( t \) by source \( i \). This is crucially dependent, however, on the assumptions that the time of death is observable, the cause of death is unobservable, and that there is no latency period between an "arrival" from source \( i \) and \( b \)'s death.

**Example 7.1: Fatal Accidents**

Suppose that there are many statistically independent causes of accidental death, e.g., car accidents, falls, choking, electrocution, and so forth. It may be reasonable to model these accidents as arriving in Poisson streams whose exposure variables, rather than being calendar time, are the hours spent by \( b \) in various specific activities, e.g., driving, cleaning the gutters, eating, and so forth, and whose intensities (= average arrival rates) reflect the intrinsic hazard of these activities. Then these hazard rates are measures of the riskiness of different activities, and the total risk from \( X \) hours spent in an activity with hazard rate \( h \) is \( hX \). The total accident fatality risk to which an individual is exposed at any time can be defined as the sum of
the intensities of the activities in which he is participating or to which he is exposed at that time. If this is done, then the additional accident fatality risk due to a new power plant next door, for example, will be simply the arrival intensity of accidents at the power plant that are of sufficient magnitude to kill the individual (corrected for the fraction of the time that he spends at home.)

Example 7.2: Excess Lifetime Fatality Risks

Perhaps the most widely used measure of individual risk in current regulatory and statutory language is the "excess lifetime fatality risk" associated with an activity. Often attention is restricted to cancer fatalities, so that the marker event c in the risk triple (a, b, c) is a cancer fatality for b, rather than death from all causes. For example, regulations applying to foods, drugs, occupational exposures, and so forth often require that the "excess lifetime cancer fatality" risks to consumers or workers be less that 10^{-6}.

Various interpretations can be given to phrases like "excess lifetime cancer fatality risk", "excess lifetime risk of lung disease", and so forth, applied to an activity a*. One is that it is the probability that the marker event c will occur by (or at) the end of b's life, due to activity a*. If c is fatal, e.g., a cancer fatality, and if all sources generate constant flows of micromorts for the occurrence of c throughout his life, then this definition of the lifetime excess risk of c due to a* would correspond to

\[ r(a^*) = \frac{h(a^*)}{[h(1) + \ldots + h(n) + h(a^*)]}, \]

where 1, ..., n are background sources that could cause c; h(1), ..., h(n) are their corresponding risk, or fatality, rates for marker event c; and h(a*) is the added source. In other words, r(a*) is defined as the probability that b will die from source or activity a*.
According to this interpretation, the cancer risk from a particular exposure to radiation, for example, might be defined as the increase in the lifetime probability of a radiogenic cancer fatality, $c$. Thus, if $h(R)$ is the original risk rate from radiation sources other than $a^*$, and $h(B)$ is the "background" risk rate of $c$ from all other sources, then

$$r(a^*) = \frac{[h(R) + h(a^*)]}{[h(R) + h(B) + h(a^*)]} - h(R)/[h(R) + h(B)]$$

could be defined as the lifetime excess risk of $c$ from $a^*$. (Note that there is no need to distinguish between risks and risk rates as long as it is assumed that risk rates are constant or proportional over time.) This type of risk measure can be computed from epidemiological prevalence data on the proportion of deaths of type $c$ in populations exposed and unexposed to the risk created by $a^*$.

A difficulty with this interpretation is that it makes the risk $r(a^*)$ assigned to $a^*$ dependent on the background risks to which $b$ is exposed. For example, it would make the lifetime excess fatality risk to $b$ due to activity $a^*$ (e.g., operation of a power plant next door) decrease if $b$ took up rock climbing or some other hazardous activity, since this would increase the denominators in the above expressions.

A more satisfactory, context-independent, definition of the lifetime excess risk of $c$ from $a^*$ is in terms of the hazard rate $h(a^*)$ alone. In the case of Poisson competing risks, where the competing sources have constant, time-invariant hazard rates independent of the presence of other sources, $h(a^*)$ may itself be defined as the excess lifetime risk of $c$ due to $a^*$. For the set of Poisson processes under superposition and FSD comparison is represented by (i.e. isomorphic to) the nonnegative real numbers under addition and inequality when each Poisson process is represented by its parameter. For the stream of Poisson fatalities generated by activity $a^*$, this parameter is $h(a^*)$, which is thus an adequate measure of excess lifetime risk from $a^*$ (on the CRS scale calibrated by Poisson process parameters) independent of the realized length of $b$'s life, the levels of background risks, or whether other risks are Poisson.
7.3.2 Choosing Among Hazard Functions

The Poisson process model for arrivals of accidents, fatal or not, is probably the most important and most widely used model of accident arrivals in applied risk analysis. It is used routinely in applications ranging from airplane or car crashes to occurrence of power plant transients to arrivals of earthquakes. Where the Poisson arrival assumption holds and the consequence of an arrival is part of its definition (as for the arrival of a marker event c in b's lifetime), the mean of the Poisson process makes a very satisfactory numerical measure of risk to an individual. Its reciprocal, which may be called the "frequency" of the arrival process, is of course an equally satisfactory indication of risk.

For general nonconstant hazard functions, the problem of risk measurement is much more difficult. If act (or activity) a generates fatality arrivals for b according to an arbitrary known hazard function \( h(t;a) \), then the corresponding c.d.f. for the first such arrival (which is the only one that counts) is easily shown to be

\[
F(t;a) = 1 - \exp[-H(t;a)],
\]

where \( H(t;a) \) is \( h(s;a) \) integrated from \( s = 0 \), the moment at which the assessment is being made, up to time \( s = t \). We can think of \( H(t;a) \) as a cumulative exposure variable, and say that the time until the first arrival has the unit exponential distribution, i.e., the exponential distribution with parameter 1, in the exposure variable \( H(t;a) \). Thus, \( h(t;a) \) contains the same information as \( F(t;a) \), and a d.m. who can choose among c.d.f.'s \( F(\cdot;a) \), e.g., on the basis of expected utility or FSD, can in principle choose among corresponding hazard functions \( h(t;a) \).

But this is too facile. For ordinary attributes such as money, the FSD concept among c.d.f.'s is straightforward and static. When the attribute in question is time, however -- specifically, the remaining time until the arrival of a marker event, c -- there is an inherent dynamic
evolution of any distribution over this attribute. As time passes, the
distribution necessarily changes by becoming conditioned on the
information that the marker event has not yet arrived. (We are excluding
additional sources of information here.) The sole exception is the class
of Poisson arrivals, discussed above, in which the passage of time leaves
the (exponential) c.d.f. of the time until the arrival of $c$ unchanged.

To save the FSD concept, for lifetime distributions, we must strengthen
it. Specifically, for any two c.d.f.'s, $F$ and $G$, over $b$'s remaining
lifetime, evaluated at some time $0$, we can say that $F$ FSD* $G$ if and only
if all of the conditional distributions derived from $F$ and $G$ by
conditioning on survival (i.e., nonarrival) for at least an amount of
time $t$, for every $t$ greater than zero, satisfy ordinary FSD. The partial
ordering FSD* may be called "temporal first-order stochastic dominance",
where it is assumed that preferences are monotonic in the time $T$ until
the arrival of $c$.

More specifically, let $X$ and $Y$ be nonnegative continuous random variables
(interpreted as arrival times) with c.d.f.'s $F$ and $G$, respectively. Then
$F$ FSD* $G$ if and only if $\Pr[X > t+k; X > t]$ is at least as great as
$\Pr[Y > t+k; Y > t]$ for all $t$ and $k$ greater than or equal to zero.

There is a close connection between FSD* and the hazard rate functions
for lifetime distributions. The following results are proven in Ross
(1983):

**Lemma 7.2:** For any two arrival time distributions $F(t)$ and $G(t)$, having
hazard rate functions $h(t)$ and $g(t)$, respectively, $F$ FSD* $G$ if and only
if $g(t)$ is at least as great as $h(t)$ for all nonnegative $t$.

Thus, $F$ FSD* $G$ if and only if the hazard rate function for $F$ never lies
above the hazard rate function for $G$. Moreover, $F$ FSD* $G$ implies $F$ FSD
$G$, in the sense of preference (rather than risk) dominance, although the
converse is not true.
THEOREM 7.3: Let $X(t)$ be a random variable denoting b's remaining (or "excess" lifetime) at time $t$. Then

(i) $X(t)$ is stochastically decreasing in $t$, meaning that the c.d.f. of $X(t)$ FSD the c.d.f. of $X(t')$ whenever $t'$ exceeds $t$, if and only if $h(t)$ is increasing in $t$. In this case, we can say that risk is unambiguously increasing over time.

(ii) $X(t)$ is stochastically increasing in $t$ if and only if $h(t)$ is decreasing in $t$.

The use of qualitative properties of hazard rates, such as whether they are increasing, decreasing, or have some other characteristic shape over time, is potentially of great importance in risk management decision making. The following example illustrates some of the links between the shape of a hazard function and the form of an optimal decision rule.

Example 7.3: Deciding Whether to Operate a Hazardous Facility

Consider the problem of deciding whether to engage in an activity, such as operation of a hazardous facility, that produces a constant flow of deterministic costs and benefits over the life of the facility and that in addition produces a flow of $h(t)$ micromorts per second (for the arrival of a fatal accident that will also destroy the plant) at time $t$. Then if $h(t)$ has a "bathtub shape", an optimal decision will be either not to participate in the activity at all, or else to participate for at least a certain amount of time, namely, the amount of time needed for $h(t)$ to return to its initial level. Knowing that this time is greater than a certain length, say $t^{**}$, and that $1 - \exp[-H(t^{**})]$, the probability of an accident by $t^{**}$, is sufficiently small may provide sufficient information to make a decision to operate the plant. Conversely, learning that this minimum time is shorter than some amount $t^*$, or that the probability of an accident by $t^*$ is sufficiently great, may provide sufficient information to decide that some other project with a stochastically dominant hazard function should be funded instead, if there is such an alternative.
Similarly, suppose that \( h(t) \) is monotonically increasing in \( t \). Then the decision problem becomes to decide at what time, say \( t^* \), to close down the facility. Individual risk attitudes will be reflected in different choices of the "optimal" shut-down point, \( t^* \), at which the marginal flow of net benefits from an additional \( dt \) of operation is perceived as being balanced by the marginal flow \( h(t^*)dt \) of micromorts. Less risk-sensitive individuals will tolerate higher values of \( t^* \).

A principle of rationality in this context, assuming that the flow rate of benefits (e.g., from electricity produced by the process) is constant, is that \( t^* \) should depend only on \( h(t^*) \), and not on the history of the risk trajectory \( h(t) \) for \( t \) in \([0,t^*)\). (This is called an "instantaneous look-ahead decision rule.) The reason is that for an increasing hazard rate, the tradeoff between the marginal benefits and marginal risks of operating the process for an additional \( dt \) becomes monotonically less favorable as time (and hence the risk rate \( h(t) \)) increases. The optimal stopping point \( t^* \) is thus the first point at which the tradeoff becomes unacceptable: continuing operation past \( t^* \) would only lead to even less favorable risk-benefit tradeoffs moment by moment. Any point before \( t^* \) is acceptable; any point after \( t^* \) is unacceptable; and the location of \( t^* \) depends only on the worst "acceptable" tradeoff. Whether this instantaneous look-ahead principle holds in practice for risk management decisions is a matter suitable for empirical investigation.

The optimal stopping problem for increasing hazard rates has several potential applications in risk analysis. For example, if the hazard rates for the (latent) arrival times \( T(1), T(2), \ldots, T(n) \) in a competing risk model are increasing over time, then so is the total hazard rate for the actual arrival time, \( T = \min [T(1), \ldots, T(n)] \), and so the stopping problem just described naturally arises.

Finally, suppose the hazard function \( h(t) \) is hill-shaped. This is the most difficult case. The d.m. must decide whether an initial increasing-risk regime is outweighed by a deferred decreasing-risk regime, and this judgement requires him to consider the global shape of
the hazard function and his attitudes toward both risk and the timing of risks. Is a lower risk later (if he lives that long) worth an increased risk to b now? Such consideration of risk across time was unnecessary for increasing (or bathtub-shaped) hazard functions, since the decision to terminate a monotonically increasing risk can be made on the basis of the local hazard rate alone. A formal theory of preferences among hazard functions is developed in Section 7.4 below.

We now establish some useful results providing for limited-precision measurement of risks (on the Poisson CRS) for arrival times whose hazard rates have certain qualitative properties.

THEOREM 7.4: Let F(t) be the c.d.f. of the time T until b's death, starting from some time origin, and let E(T) and Var(T) be the mean and variance of T, respectively, computed with respect to the c.d.f. F(.). Suppose that F(t) has a decreasing hazard rate function, h(t). Then the risk associated with F(t), measured on the canonical risk scale calibrated by the parameters of Poisson arrival processes, is never greater than

\[ \frac{1}{E(T)} - \frac{[1 - Var(T)/E(T)]}{2t}. \]

In particular, the arrival distribution F(t) is never more risky than the Poisson arrival process with parameter l/E(T) if T has a variance (with respect to F) no greater than its mean, E(T).

Proof: This is obtained by rearrangement of a bound on survival functions often used in reliability theory; see e.g., Ross (1983), p. 265, for a renewal-theoretic derivation of this bound.

THEOREM 7.5: Suppose that all decision makers are risk-averse for the attribute "b's remaining life", in the sense that each has an increasing concave utility function for this attribute, starting from any time point. Suppose also that b's expected remaining life decreases as he ages. Then the risk to b at any time, measured on a CRS calibrated by
the parameters of Poisson fatality-generating processes, is no less than
the reciprocal of his expected remaining life.

Proof: This is an interpretation of another reliability-theoretic bound; see Ross (1983), p. 273 for a proof of this result in terms of the relation "stochastically more variable than" between "new better than used in expectation" distributions and exponential distributions. It is necessary to extend the proof from convex to concave increasing functions, but this is easily done once it is recognized that the "stochastically more variable than" relation is identical with the Rothschild-Stiglitz (1970) definition of "riskier than" between c.d.f.'s for risk-averse d.m.'s.

Although the above theorem has been stated using the idea of a utility function over the attribute "remaining life", the concept of utility must be treated with caution in a dynamic setting. Remaining life can be treated as a numerical attribute for the purposes of asking questions about hypothetical lotteries over remaining lives of different durations if it is assumed that the outcomes of such lotteries, expressed as deterministic numbers of years remaining, are immediately revealed to the d.m. This leads to the type of utility function (assumed to be increasing and concave) envisioned in Theorem 7.5. But the idea of a utility function for remaining life that could be used to represent derived preferences among lifetime distributions via expected utility would require additional specification. For an arbitrary initial distribution $F$ will generate an entire stream of expected utilities, with the expected utility at time $t$ being conditioned on the information that no fatality has yet arrived, and hence being calculated with respect to the updated c.d.f. $P(t+k|k>t)=[F(t+k)-F(t)]/[1-F(t)]$. No process for choosing among such streams of conditional expected utilities is suggested in classical decision analysis.

It may be helpful to note that in this construction,

- The expected utility of remaining lifetime implied by a lifetime distribution $F$ remains constant after $F$ is chosen (or goes into effect) if and only if $F$ is an exponential distribution.
For any two c.d.f.'s \( F \) and \( G \) over the attribute "remaining life, starting from time 0", the stream of expected utilities generated by \( F \) is greater than the stream generated by \( G \), for all times \( t \) greater than zero and for all increasing utility functions over this attribute, if and only if \( F \) FSD* \( G \).

We can continue to exploit the reliability theory literature to gain additional insights into the FSD* relation among time-evolving life distributions.

**THEOREM 7.6:** Let \( F \) belong to one of the following families:

- Gamma
- Weibull
- Normal truncated at zero (i.e., constrained to a nonnegative range.)

Then the risk of death to \( b \), measured on the Poisson parameter canonical risk scale, is objectively nondecreasing (and in general strictly increasing) over time.

Proof: Each of these families has the property that the conditional distribution of remaining life at time \( t \) temporally stochastically dominates (FSD*) the distribution at time \( t' \), for any \( t' \) greater than \( t \). This can be verified by noting that these families have increasing likelihood ratios, which imply the decreasing FSD* property; see Ross (1983), Chapter 8.

The statement that risk is objectively increasing over time in Theorem 7.6 is to be interpreted as follows. Even though the risks of distributions in the above classes can not be measured unambiguously on a CRS, and different decision makers will rank distributions differently within the limited-precision measurement interval (on an appropriate CRS)
implied by the FSD* relation (see Theorem 7.7), it is nonetheless the case that every d.m. who prefers temporally stochastic dominant distributions will agree that risk (or undesirability of the remaining life distribution) is increasing over time. Different d.m.'s may disagree about the magnitude of the risk on a given CRS, but they will all agree about its direction.

Finally, we present a theorem that allows for the construction of canonical risk scales suitable for measuring the risks of (intrinsically time-varying) lifetime distributions.

THEOREM 7.7: Each of the following one-parameter families of lifetime distributions is totally ordered by FSD*:

- The Poisson family Poiss(m) with parameter m. (This allows for risk measurement in discrete time.)
- The binomial family with fixed n and variable p;
- The family of uniform distributions U[0,b] with variable upper bound b;
- The family of gamma distributions with either parameter held fixed.
- The family of exponential distributions with variable parameter. (This is a special case of the gamma family.)

Proof: These families have monotone likelihood ratios, which implies the FSD* property; see Ross (1983), chapter 8.

Thus, any of these families can be used to construct a CRS suitable for measuring the risks of lifetime distributions, where FSD* rather than FSD is used as the basis for comparison. In continuous time, the family of exponentially distributed lives, corresponding to Poisson arrival
processes, is probably the easiest one to use. As usual, only limited-precision measurement will in general be possible for distributions falling outside the selected one-parameter, totally FSD*-ordered family.

The preceding results evidently apply to choice among differential hazard functions, corresponding to different acts in A, when the assumption of competing risks is appropriate. That is, if \( h(B) \) is the "background" hazard function for b's remaining life and act a in A adds an incremental hazard function \( h(a) \), then the FSD* ranking of acts in A based on their incremental hazard trajectories \( h(a) \) will be the same as the FSD* ranking of the corresponding total lifetime distributions, having hazard functions \( h(B) + h(a) \). This is convenient when \( h(a) \) is known and \( h(B) \) is uncertain. However, it depends on the competing risk assumption that the addition \( h(a) \) leaves \( h(B) \) unchanged, i.e., that the (latent) times to death from different sources are statistically independent.

7.3.3 Uncertain Risks: Examples

The results developed above apply when the d.m. knows the effect of each act in his choice set on the distribution (or, equivalently, on the hazard rate) of individual b's remaining lifetime. But this is a very optimistic case. In practice, both the base line distribution of T and its probabilistic response to acts a in A will depend on components of b's "type" that the d.m. does not know or is uncertain about. In such cases, the fatality risks to b from acts in A will be uncertain. The distinction between risk and uncertainty about risk is important in the formulation and statement of risk management policies, and we shall now try to clarify this distinction in the specific context of individual fatality risks.

The fatality risk for an individual depends on what is known about him, and hence on the assessor's history of observations. The risk from an act, like the expected utility from an act, can be change by additional observations that refine the information partition (or that "unaverage"
the conditional probability measure) with respect to which risk and expectation are computed. Thus, research and residual risk are closely linked.

Example 7.4: Schroedinger's Cat

Schroedinger's cat Sam is watching a mouse hole of uncertain occupancy. He knows that there is a 50-50 chance that the hole is occupied by a live mouse. If there is a mouse, it will eventually run out of the hole (thus becoming a fatality if Sam is still there.) The time between excursions for a live mouse is an exponentially distributed random variable with mean $1/h$; equivalently, the mouse arrives according to a Poisson risk with parameter $h$. Sam's probability assessment for the event that the mouse is dead, conditioned on the amount of time that the hole has been observed, increases from 0.5 at time 0 to

$$\Pr(\text{Mouse is dead; } t) = \frac{1}{1 + \exp(-ht)}$$

after the hole has been observed (with no mouse running out) for $t$ periods. For Schroedinger, the probability that the mouse is dead, given that Sam has been locked in the room with the mouse hole for an amount of time $t$, is

$$\Pr(\text{Mouse dead; } t) = 1 - 0.5 \exp(-ht)$$

Mathematically, there are two distinct ways of creating a flow of risk (micromorts per second) to the mouse. The natural way is Schroedinger's: to lock a cat in the room where the mouse takes his runs. The other way is Sam's: to observe an empty mouse hole. Both activities gradually increase the probability (to the relevant observer) that the mouse is dead, although only the first would usually be thought of as creating any additional risk for the mouse.

Example 7.5: Checking for Cancer

Hazel has been exposed to a chemical in her workplace that she fears may produce cancer. She therefore has herself periodically examined for
tumors. Tumors from this chemical have a limited latency period, so that if no tumor appears within Y years, then Hazel will be safe from the delayed effects of the exposure. The hazard rate for tumor arrival is \( h(t) \). What is the "risk" (not the probability) to Hazel of a tumor being found at the kth examination? The answer depends on when she was last examined, and is given by \( H(k) - H(k-1) \), where \( H(k) \), for example, is the hazard rate \( h(t) \) integrated up to the time of the kth observation. We may think of Hazel as accumulating risk (measured in micromorts of tumor arrival probability) between examinations. At each examination where no tumor is found, the current accumulation of micromorts is emptied, and the accumulation process starts again. Thus, the total risk of tumor at time \( t \) (given by \(-\ln(1 - p(t))\) , where \( p(t) \) is the probability that a tumor will be found if an examination is made at time \( t \)) depends on the time since the last examination.

Example 7.6: Stochastic Hazard Rates

The hazard rate \( h(t) \) for arrivals of an undesirable event makes transitions between two values, \( h^* \) and \( h^{**} \), depending on the (unobservable) state of a two-state Markov process (e.g., on whether b is smoking cigarettes at the moment, if \( h(t) \) is the hazard rate for a disease from a source that interacts with cigarette smoke.) What is the risk of this process, measured on the Poisson CRS? Is the risk increased or decreased if the average rate of state transitions in the Markov process is doubled? If the state of the process is monitored through a noisy channel?

Example 7.7: Probabilistic Hazard Rates

Suppose that the hazard rate for arrivals of a fatal cigarette-induced illness is either \( h^* \) or \( h^{**} \), depending on b's biomedical "type". His type is unobservable (or observable only upon autopsy), but the probability that he is of each type is known. What is the risk corresponding to a probability mixture \( (h^*, p, h^{**}) \) between types, measured on the Poisson CRS? (Obviously, it is between \( h^* \) and \( h^{**} \).)
Example 7.8: Superposition of Periodic Hazard Functions

Consider an industrial facility with a sinusoidal hazard function (perhaps reflecting a repair and maintenance cycle.) Suppose that a d.m. knows the shape of the function, but is uncertain about its phase; specifically, he puts 50-50 probability on each of two possibilities, say \( h(a) \) and \( h(b) \), which are 180 degrees out of phase with each other (so that their sum \( h(a) + h(b) \) is identically equal to the average hazard rate for all \( t \).) Does exposure to this cyclical risk become less tolerable if he is informed which process he is exposed to, so that the two processes are "unaveraged" for him?

These examples are intended to suggest some of the rich interactions between preferences for hazard functions, availability of observations, and assessment of hazard rates. They depart from individual risk Model 1 by allowing information about the true distribution (or corresponding hazard function) of remaining lifetime to be learned over time, rather than being perfectly known at the outset.

To answer fully questions of the sorts raised in these examples, a new normative decision theory is needed. The objects of primitive preference in this theory will be known hazard rate trajectories, and the objects of derived preference will be dynamic probability measures, or stochastic processes, over hazard rate trajectories. Before beginning the development of such a theory, however, we will present two results that shed light on Examples 7.6 and 7.7.

**THEOREM 7.8:** Suppose that the hazard rate at time \( t \) for the occurrence of marker event \( c \) from source or activity \( a \) is given by \( h(t;s) \) where \( s \) is a constant parameter drawn at random from a distribution \( f(s) \) before observations begin. Thus, \( h(.;s) \) is drawn at random from a one-parameter family of hazard functions indexed by \( s \). The parameter \( s \) is not directly observable, although its prior distribution \( f(s) \) is known. Define \( h^*(t) \), the "assessed" hazard rate at time \( t \), to be the mean of the posterior distribution of \( h(t;s) \), calculated with respect to the posterior density
f(s;t) for s derived from the prior density f(s) by conditioning on the information that no arrivals have occurred by time t. Then if h(t;s) is decreasing over time for each value of s, the assessed hazard function h*(t) is also decreasing in t.

Proof: It is known from reliability theory, that the family of decreasing failure rate (or hazard rate) distributions is closed under mixtures; see e.g., Barlow and Proschan, 1975, for a proof.

The intuitive content of the above theorem may be clarified by the following example.

Example 7.9: Individual Car Accident Risks

Suppose that the average arrival rate of car accidents in a driver's career, considered as marker events in the exposure stream of his hours behind the wheel, depends only on his age and "skill level", where skill level, whatever it means, can be indexed by (i.e., represented by) a single numerical parameter s. (That the unobservable variable s has this simple mathematical structure is, of course, a strong empirical assumption.) If b is a randomly selected driver from a population of drivers with no previous accidents in which skill, s, is randomly distributed with a known distribution; and if the hazard rate for car accidents is both (i) decreasing over time for each skill level; and (ii) decreasing in skill level for each time, then the assessed hazard rate for b, conditioned on the time that he has driven without an accident, will also be decreasing over time. Intuitively, as time passes with no arrivals of the marker event, it becomes increasingly likely that the randomly selected driver is a low-risk (high-skill) driver. Hence, the hazard rate assessed for him, which at any moment is the mean of the posterior distribution over hazard rates, will be decreasing over time. This holds even though b's skill level may be forever unobservable.

Corollary to Theorem 7.8: Suppose that arrivals of marker event c are generated by a Poisson arrival process with an uncertain (random)
parameter $h$, where $h$ is drawn from some distribution. Then the magnitude of the assessed "risk" associated with this uncertain hazard rate is decreasing over time, even though the "true" (but unknown) hazard rate remains constant.

Again, the general principle is that the assessed risk (hazard rate) at any time is the mean of the posterior distribution over hazard rates, starting from the known initial prior distribution and conditioned on the amount of time that has elapsed with no arrivals. Although decision makers with different priors at $t = 0$ will have different posterior distributions over $h(t)$ for $t$ greater than zero, and hence will assess different numerical values for $h(t)$, all d.m.'s with nondegenerate priors will agree that $h(t)$ is decreasing over time, since the passage of time alone makes it more probable (in a Bayesian sense) that one of the smaller hazard rates has been chosen.

We close with an elegant result, conjectured by Brown and Ross (1982) and proven by Schechner (1982), that generalizes the randomly evolving hazard function in Example 7.6 and that clarifies the relation between total risk (in the sense of integrated hazard) and observations. So far in this chapter, the primary mechanism by which an observer gathers information to update his beliefs about b's survival distribution has been simply the passage of time, starting from a known initial distribution. Suppose, however, that b's hazard rate (or, more generally, the hazard rate for a marker event c) evolves at random according to a process $h[t; s(t)]$, where $s(t)$ is the realization at time $t$ of a stochastic process, and $h(t) = h[t; s(t)] = h(s(t))$ is the hazard rate at time $t$, given that the realized "state" of the stochastic process at time $t$ is $s(t)$. Instead of just observing the time, $t$, suppose that the risk assessor can also observe $s(t)$; thus, he knows $h(t)$ from moment to moment as he observes it, although at any moment he does not know its future values. This is a simple model for a perfectly observed, randomly evolving, hazard function. It is a substantial generalization of the deterministically evolving hazard functions considered hitherto (including the randomly drawn hazard functions of Example 7.7 and Theorem
We will assume that the sample path of $h[t; s(t)]$ is sufficiently well behaved so that it is integrable over any interval $[0, t^*]$. $H(t^*)$ will denote the integral from $t = 0$ to $t = t^*$ of $h(s(t))$.

**Mathematical Note:** A precise statement of the assumed continuity properties requires concepts of martingale stochastic calculus. Namely, $H(t)$ is assumed to be a right-continuous, increasing process starting at the origin, predictable with respect to a complete, right-continuous, increasingly fine family of information partitions (or sigma algebras), and possessing a differentiable sample path.

**THEOREM 7.9:** Under the above conditions, $H(T)$ has a unit exponential distribution (i.e., an exponential distribution with mean 1), where $T$ is the random time at which marker event $c$ (e.g., $b$'s death) arrives.

**Proof:** See Schechner, 1982 for a proof of this result. Schechner's proof uses the modern theory of martingale jump processes, and hence requires a fairly deep understanding of stochastic processes viewed as functions measurable with respect to increasing information patterns (or "filtrations"). A much easier heuristic proof is given in Brown and Ross, 1982.

The intuitive significance of Theorem 7.9 for our theory of risk measurement is this. To an observer who can always perfectly observe the hazard function of a marker event over time (e.g., who has a noiseless, delayless, observation channel), the number of micromorts that will flow past before the marker event arrives is a random variable with a known (exponential) distribution. This holds whether the micromorts are generated

- deterministically according to a known hazard function;
- deterministically according to the posterior mean of an evolving distribution over hazard functions, where the evolution reflects Bayesian updating starting from a known prior distribution; or
stochastically, according to a well-behaved perfectly observed stochastic process.

No matter which way the flow of micromorts is generated for an observer, the total number that will flow past before the marker event arrives or occurs will be a random variable with the standard exponential distribution.

A final comment may help put this result in perspective. In each of the three cases just cited, the time at which c occurs is learned exactly when c occurs: it can not be predicted. This leads to the exponential distribution for total observed hazard. But it is not hard to think of other cases to which the result does not apply. For example, suppose that instead of observation ending with the arrival of the first (and fatal) marker event, many marker events may arrive, each carrying with it a deterministic latent time until death. In effect, a sequence of randomly distributed death dates is observed. (For a specific example, we can let each date lie between 10% and 90% of the way from the date at which it itself arrives to the date announced in the previous arrival; thus, it lies between \( t + 0.1(t^* - t) \) and \( t + 0.9(t^* - t) \) if it arrives at time \( t \) and \( t^* \) was the earliest date of death to arrive hitherto.)

At any time, the observer, b, knows that he will live no longer than until the earliest of the dates that have arrived so far; moreover, he will live exactly until then unless another marker event with an earlier death date arrives first. (For a morbid interpretation, think of a terminal disease patient learning at each of a sequence of examinations whether his projected death date has moved up, either because the disease has worsened or due to other, competing diseases.) In such a case, b will be uncertain at any moment about the exact date of his death, although he knows it is between the present and the earliest date observed so far. Because observation antedates occurrence, no well-defined hazard rate exists, and the above results about the exponential distribution do not apply.
It should be clear that \( b \) may have preference over processes of the sort just described; that well-defined notions of stochastic dominance between such processes exist; and that a one-parameter family of such processes totally ordered by stochastic dominance can be constructed. (Let \( T(k+1) = t + U[0,1-r][T(k)-t] \) and let the average arrival rate of observations be \( -\ln(1 - r) \), where \( U[a,b] \) is the uniform distribution between \( a \) and \( b \), \( t \) is the present moment, and \( r \) is a risk index between 0 and 1. ) Thus, a CRS supporting risk measurement could in principle be developed for this class of problems. However, the case of simultaneous observation and occurrence, i.e., ordinary hazard rates for arrivals of accidents and other risk events, seems vastly more important in practice, and is sufficiently challenging to justify postponing more general formulations.

7.4 Elements of a Preference Theory for Uncertain and Stochastic Hazard Functions

The preceding sections show that (i) known hazard functions make satisfactory indicators of risk over time; and (ii) uncertain hazard functions, e.g., probability distributions over hazard functions, necessarily evolve over time, and thus have a different mathematical structure than known hazard functions. In this section, we attempt to clarify this second point and to develop a normative decision theory for choosing among acts that have uncertain hazard functions as consequences.

7.4.1 Motivation

The nature of the problem is suggested by the following simple decision problem. Suppose you have a choice between two acts:

- \( oA_l \) produces a fatality hazard rate of 0.2 morts per year with certainty. (A "mort" is henceforth defined for continuous time as the unit in which the parameter of a Poisson arrival process is measured; in discrete time, it is the probability parameter for a Bernoulli trial, e.g., the probability of heads in a coin toss if the marker event is the arrival of the first head.)
A2 gives a 50-50 chance between fatality rates of 0.1 and 0.3 morts per year.

Which act would you choose?

To simplify the matter, assume that the arrival process occurs in discrete time. If you choose act A1, a coin will be tossed once at the beginning of each year: in the first year that it comes up "heads", you will die. The probability of heads on any given toss is 20%. If you choose A2, the probability of heads on any given toss is either 10% or 30%, and this probability (which is determined by the toss of a fair coin before you make your decision, and never revealed to you) is the same on all trials. Would you choose act A1 (the known risk) or act A2 (the uncertain risk?)

A traditional line of argument suggests that you should be indifferent between these choices, provided that the outcome of your choice (if you choose A2) is never revealed. For the probability of dying in any year is 0.2 in the first case, and is (0.5)(0.3) + (0.5)(0.1) = 0.2 in the second. The probability of death in any year (the discrete-time hazard rate) is just the expected probability of death in that year. If your preferences depend only on the probability distribution over consequences (life or death), you should be indifferent between A1 and A2.

As stressed earlier in this chapter, such reasoning is incorrect. The relevant consequences are not just "life" and "death", or even the year of death, but rather the entire hazard rate trajectory leading up to (and including) the year of death. For act A1, this hazard rate trajectory is constant at h(t) = 0.2. For act A2, the assessed (or expected) hazard starts out at h(0) = 0.2, and declines steadily toward 0.1 for as long as observations continue. (In the year of death, of course, it jumps up, but by then it is too late for you to revise your opinions.) From this argument, it appears that choosing A2 is a dominant strategy.
Making the numbers more extreme shows that even conditional expected (i.e., "assessed") hazard rate over the period of observation may not be all you care about. Instead of a $(0.1, 50\%, 0.3)$ mixture of hazard rates, suppose that the mixture were $(0.95, 21.3\%, 0.02)$, i.e., with probability just over $1/5$, the hazard rate would be $95\%$, while with a probability of almost $80\%$, the hazard rate would be only $0.02$. Again, the initial expected hazard is $0.2$. But would you still necessarily prefer the uncertain hazard rate to the certain one of $0.2$? Might you not choose the certain $0.2$ hazard rate instead of risking the $0.95$ one, and might not such aversion to potentially large hazard rates be perfectly acceptable to a "rational" person? If so then "consequence" includes not just the posterior mean of the hazard rate, based on observations to date, even though this "is" the probability of death at any time. It must also include other, nonobservable aspects of the probability distribution over hazard rates.

Finally, how would your answers to the above questions change if the outcome of $A_2$, namely $h(t) = 0.1$ or $h(t) = 0.2$, is revealed immediately after $A_2$ is chosen (if $A_2$ is your choice?) In this case, the observable hazard function for $A_2$ no longer necessarily dominates the observable hazard rate from $A_1$: it is equally likely to dominate or to be dominated. Given that you will be locked into whatever hazard rate occurs, should knowing that you will know it change the basis for your initial choice?

These questions address the psychology of temporal risk preferences. Like the questions raised in the previous section and in Chapter 1, they suggest that there is more to risk than probability distributions over consequences. If hazard functions are themselves counted among the consequences of an act, then at the very least a choice among acts requires a choice among stochastic processes -- namely, the trajectories of assessed (conditional expected) hazard rates generated by initial distributions and the passage of time. The remainder of this chapter is devoted to extending von Neumann-Morgenstern (N-M) utility theory to this more complicated set of objects.
7.4.2 A Dynamic Expected Utility Theory for Uncertain and Stochastic Hazard Rates

Our objective in this section is to develop an expected utility type representation for derived preferences over uncertain hazard functions. To this end, we first develop an explicit model of stochastically evolving hazard functions, which we call "stochastic hazard trajectories." We then apply the ideas of N-M utility theory to these objects.

A. Stochastic Hazard Trajectories

It will be mathematically convenient to work in discrete, rather than continuous, time, and to let \( t \) index time periods. It is realistic to assume a finite-horizon problem, and we shall assume that \( t \) takes values in the set \([0, 1, 2, \ldots, N-1]\), where \( N \) is the number of periods over which future hazard trajectories are to be evaluated. (\( N \) might be the operational life of a plant, for example, after which it is decommissioned so that the hazard rate becomes zero; or it might simply be a very large number so that the probability of survival to \( N \) approaches zero.) The following concepts are needed for the construction:

- \( H = [h(.;s'): s' \in S] \) is a set of possible hazard functions, indexed by \( s' \). The actual hazard function will be denoted by \( h(.;s) \). In practice, \( h(.;s) \) may never be known.

- \( h(t) = h(t; I'(t), p(s)) \) is the assessed discrete hazard rate (i.e., the assessed probability of arrival for the marker event) in period \( t \) for a d.m. with information \( I'(t) \) in period \( t \) applied to a prior density \( p(s) \) over hazard functions. That is, \( h(t) \) is the expected hazard rate ( = death probability) in period \( t \), computed with respect to \( I'(t) \) and \( p(s) \). The information in \( I'(t) \) consists of all observations the d.m. has made up through time \( t \); the exact observation process (and the reason for the prime after \( I \) ) are specified below. \( I'(t) \) may also be thought
of as the observer's posterior probability measure (or information partition, in the special case of noiseless observations) on \( S \), starting from \( p(s) \) and after conditioning on all observations; in this case, \( p(s) \) can be dropped from explicit consideration.

- \( H^*(t) \) is the set of all joint probability measures over \([0,1] \times H^*(t + 1)\), for \( t = 0, 1, 2, \ldots, N - 1 \). (Let \( H^*(N) \) be any arbitrary probability distribution on \([0,1]\), since it lies beyond the horizon considered.) The interpretation of \( H^*(t) \), as the set of all stochastic hazard rate trajectories starting in period \( t \), is developed below.

- \( h^*(t) \) is a generic member of \( H^*(t) \), i.e., it is a stochastic hazard rate trajectory starting in period \( t \).

- \( R(t) \) is a preference ordering over the members of \( H^*(t) \).

The key construct in this list is \( H^*(t) \), which we interpret as the set of all stochastically evolving hazard functions that begin in period \( t \). The members of \( H^*(t) \) are stochastic processes. A realization of a process \( h^*(t) \) in \( H^*(t) \) consists of a sequence of probabilities, \( h(t), h(t + 1), h(t + 2), \ldots, h(N-1) \), namely, the set of assessed hazard rates in each future period of interest. (Each \( h(k) \) in this sequence, or "trajectory", is to be interpreted as being assessed according to the information \( I'(k) \) that will be available for assessing it in period \( k \), i.e., it is the probability of an arrival of the marker event in period \( k \), as perceived by an observer in period \( k \).) At any time \( t \), these future probability numbers are uncertain; for example, they will typically depend on the (currently unknown) observations that will become available as periods pass.

A member \( h^*(t) \) of \( H^*(t) \) gives a joint probability distribution (or rather measure, since the second component is not numerical) over \( h(t) \) as the first component, and over stochastic hazard rate trajectories rooted in
period \( t + 1 \) as the second component, based on information available just before the beginning of period \( t \), i.e., based on \( I(t - 1) \). That is, \( h^*(t) \) is a probability measure on \([h(t), h^*(t + 1)]\), based on information \( I(t - 1) \).

Thus, the explicit sequence of events within period \( t \) is envisioned to be as follows.

**Individual Risk Model 2: Stochastic Hazard Functions**

(i) The observer has some prior distribution over \( h(t) \), based on information available to him through the end of period \((t - 1)\). His prior assessed hazard rate for period \( t \) is \( E[h(t); I(t-1)] \). (For example, this may be his subjectively assessed probability that he will develop a cancer in period \( t \).)

(ii) An observation arrives at the start of period \( t \), changing the observer's information from \( I(t - 1) \) to, say, \( I'(t) \). (For example, he may pay a trip to a doctor to obtain prognostic information.) The prime on \( I \) indicates that this is only intermediate information: he will obtain additional information by the time period \( t \) is over. The observation taken at the beginning of period \( t \) may be the "null" observation, in which case \( I'(t) = I(t - 1) \).

(iii) The observer computes his assessed hazard rate for period \( t \), namely \( h(t) \) (or, more precisely, \( h[t;I'(t)] \), which we abbreviate as \( h(t) \)) based on all information, including his period \( t \) observation, available to him at the start of period \( t \). Thus, he calculates \( h(t) \) after receiving any new information that arrives in period \( t \), and we adopt the timing convention that all such information arrives at the beginning of the period. \( h(t) \) is the hazard rate that he must live with until the end of period \( t \).

The experience of living with this risk for one period may be assumed to create a flow of psychological disutility to the observer. This is our key behavioral hypothesis.
At the end of period $t$, the marker event either occurs or does not occur. The probability that it occurs is given by a true (but in general unknown) hazard rate, say $h(t; s)$, which may differ from the assessed hazard rate, $h(t)$.

The observer updates $I'(t)$, his probability measure over $S$ (or equivalently over the set of possible hazard functions) by conditioning on the outcome of period $t$, namely, the occurrence or nonoccurrence of the marker event. His updated information is denoted by $I(t)$.

Period $t$ ends and period $t + 1$ begins.

Individual risk Model 2 allows for a wide variety of observation mechanisms that can give the observer increased information about the "true" hazard function, $h(.; s)$, as time passes. More general formulations, not addressed here, would allow for $s$ itself to be changing over time, or for observation of a marker event (i.e., knowledge of its arrival date) to precede the actual arrival of the event, as discussed at the end of the previous subsection.

B. Preferences Over Stochastic Hazard Functions

We now consider axioms for preferences over stochastic hazard trajectories. Our treatment applies pioneering work by Kreps and Porteus (1979a, 1979b) in the economic theory of dynamic choice over uncertain consumption streams. It turns out that the mathematical structure of the stochastic hazard trajectories identified above is a special case of the mathematical structure assumed by Kreps and Porteus for their randomly evolving consumption streams, and hence their application of N-M utility theory to such streams applies equally well to stochastic hazard rate trajectories.

Mathematical Note: Kreps and Porteus (K-P) require only that the consumption (or "payoff") set within each period be a compact complete separable metric space. In our theory, this set is the unit interval
(and h(t) is the period t "payoff"), so these properties are trivially satisfied.

In some ways, the K-P analysis applies better to preferences over stochastic hazard trajectories than to dynamic choices. In particular, K-P's interpretation of "actions" as probability measures over next payoff-state pairs, where the state in any period is an index of the set of actions available in that period, seems to raise conceptual difficulties (e.g., why should it be assumed that all such that measures are available for choice?) By eliminating actions over time and focusing on preferences over uncontrolled stochastic hazard trajectories, we avoid this difficulty. Of course, this restricts the scope of application of our theory to choices among stochastic hazard rate trajectories that evolve uncontrolled once selected. However, in many applications, e.g., in deciding whether to submit to potentially damaging medical treatment, to buy a consumer product, to accept a hazardous job, or where to site a facility, the paradigm of an initial choice followed by realization of stochastic consequences seems realistic. The problem of representing preferences over controlled stochastic processes will be addressed in Section 7.5.

We now turn to the problem facing a d.m. who must choose among stochastic hazard rate trajectories in H*(0), where 0 is the decision period. Associated with each set H*(t) of stochastic hazard rate trajectories from period t forward is a preference relation R(t). For any two stochastic trajectories h*(t) and g*(t) starting in t, we will write h*(t) R(t) g*(t) if and only if the d.m. would prefer h*(t) at least as much as g*(t) at the beginning of period t. We make the following assumptions about the preference relation R(t) for stochastic risks:

AXIOM DRL (Weak Ordering and Continuity): R(t) is a complete, transitive, continuous binary relation on H*(t).

These are just the usual assumptions of weak ordering and continuity for preferences, here merged into one axiom for convenience.
AXIOM DR2 (Temporal Consistency and History Independence): Let $h^*(t)$ and $g^*(t)$ be two stochastic trajectories in $H^*(t)$ with known identical first components, i.e., both put probability 1 on some hazard rate $h(t)$ in period $t$. Thus, $h^*(t) = [h(t), h^*(t+1)]$ and $g^*(t) = [h(t), g^*(t+1)]$. Then $h^*(t) R(t) g^*(t)$ if and only if $h^*(t+1) R(t+1) g^*(t+1)$.

Interpretively, this axiom states that preferences exhibit a form of consistency, or stationarity, over time. The decision maker will never disprefer a choice in one period that he preferred in the previous period, other things (e.g., previous-period history) being equal, assuming that he knew what the consequences of his choice would be, and that those consequences indeed occurred. Thus, if he chooses between two trajectories with delayed resolutions of uncertainties, he will not regret or change his choice as the uncertainty resolution date(s) draw near, unless he gains some new information. We propose this as a normative criterion for "rationality" over time. More exactly, axiom DR2 implies that given a choice between any two trajectories that have identical deterministic initial segments and that diverge and become stochastic only after this segment is over, the d.m.'s initial preferences between the two trajectories will coincide with his preferences in the period where they first become uncertain and start to diverge.

A second normative criterion that is implicit in the statement of DR2 should be recognized explicitly. This is that preferences among stochastic hazard rate trajectories starting in period $t$ depend only on those trajectories themselves, and not on the trajectory that led up to period $t$. That is, preferences for future risks, represented by stochastic hazard trajectories, are assumed to be independent of the magnitudes of the hazard rates in preceding periods. Given that he has survived to time $t$, the "rational" d.m. forgets about past risks (except insofar as they conveyed information about the risks that he now faces.) They are treated as "sunk costs." This implicit assumption of "history independence" rules out behavior that makes an individual treat risky prospects differently based solely on the magnitudes of the risks that he has faced (or "overcome") so far.
Because of axiom DR2, we can replace the sequence of preference orderings \( R(t), t = 0, 1, 2, \ldots, N-1 \) with a single preference ordering, say \( R = R(0) \), over \( H^*(0) \). We will use \( R = R(0) \) to denote his preferences among stochastic hazard trajectories starting in period 0. Preferences over sub-trajectories in \( H^*(t) \) are given by restricting \( R \) to trajectories that are identical and deterministic up to and including period \( t - 1 \). By axiom SR2, preferences between any two such trajectories are determined by and agree with preferences \( R(t) \) over the subtrajectories starting in period \( t \).

AXIOM DR3 (Temporal Substitution): Let \( f^*, g^*, h^* \) be any three trajectories in \( H^*(0) \) that are deterministic and identical in periods 0, l, ..., \( t - 1 \), and that become stochastic and diverge in period \( t \). Let \( P \) be the strict preference relation corresponding to \( R \). Then \( h^* \, P \, g^* \) if and only if \((h^*, p, f^*) \, P \, (g^*, p, f^*)\) for all \( p \) in \((0,1)\).

Here, \((h^*, p, f^*)\), for example, is a "temporal probability mixture" of \( h^* \) and \( f^* \) that at the beginning of period \( t \) (and not until then) chooses \( h^* \) with probability \( p \) and \( f^* \) with probability \((1 - p)\), and reveals the outcome to the d.m. Thus, \((h^*, p, f^*)\) is an uncertainty that is resolved at time \( t \) about which of two stochastic hazard trajectories the d.m. will confront from period \( t \) on. Note that \( H^*(0) \) is closed under such temporal mixtures.

THEOREM 7.10: Axioms DR1 to DR3 hold if and only of

(i) There are \( N \) unique one-period canonical \( N \)-\( M \) utility functions, \( U(.,.;t) \), for \( t = 0, 1, 2, \ldots, N - 1 \), each mapping \([0,1] \times [H^*(t+1)]\) into \([0,1]\);

(ii) There are \( N \) unique between-period utility conversion functions, \( u(.,.;t) \), each mapping \([0,1] \times [0, \infty)\) into \([0, \infty)\) and strictly increasing in its second argument; and

(iii) The functions \( U(.,.;t) \), or \( U(t) \) for short, are recursively defined by
\[ U[h(t), h^*(t+l); t] = u[h(t), E[U(t+l)]; t], \]

for \( t = 1, 2, \ldots, N - 1 \) (where \( U(N) \) is some continuous function of \( h(N) \)) such that for any two stochastic hazard rate trajectories \( h^*(t) \) and \( g^*(t) \) in \( H^*(t) \), \( h^*(t) R(t) g^*(t) \) if and only if \( E[U(t); h^*(t)] \) is at least as great as \( E[U(t); g^*(t)] \). In particular, for any two stochastic hazard rate trajectories \( h^* \) and \( g^* \) in \( H^*(0) \),

\[ h^* R g^* \text{ if and only if } E[U; h^*] \text{ is at least as great as } E[U; g^*], \]

where \( U = U(0) \) is the period zero (the decision date) utility function obtained by the above recursive definition.

Proof: The weak ordering, continuity, and substitution axioms guarantee the existence of an \( N-M \) utility function over consequences \([h(t), h^*(t+l)]\) within each period \( t \), by traditional \( N-M \) theory. The temporal consistency axiom, DR2, implies that period \( t + 1 \)'s expected utility enters as an argument into period \( t \)'s utility function via the "linking" function \( u(.,.;t) \). See Kreps and Porteus, 1979 (a), Theorem 2 and Corollary 4 for proof of a more general theorem.

We can summarize Theorem 7.10 by saying that if the d.m.'s derived preferences over stochastic hazard rate trajectories satisfy axioms DR1 to DR3, then they have an expected utility representation. Kreps and Porteus prove some additional qualitative properties, e.g., that a d.m. will prefer earlier to later resolution of uncertainties if and only if the functions \( u(.,.;;) \) linking each period's expected utility to the preceding period's are convex in the second argument. In effect, this requires the d.m. to be risk-seeking in the attribute "next period's utility". For example, if he will face a probability mixture over stochastic hazard rate trajectories starting in period \( t \), after an initial deterministic segment of hazard rates in periods 0 through \( t - 1 \), then he will prefer to learn the outcome of the mixture at time 0 rather than at time \( t \) if the \( u(.,.;k) \) functions for \( k = 0 \) to \( t \) are convex in the
second argument. He will prefer to postpone learning the outcome until period \( t \), when it becomes relevant, if the intervening \( u(\cdot, \cdot; k) \) functions are concave in the second argument. And he will be indifferent about when uncertainty is resolved, so long as it does not affect the (currently unknown) assessed hazard rates that he will face, if and only if \( u(\cdot, \cdot; k) \) is linear.

One possible criticism of this theory is that it treats future hazard functions (and stochastic processes over them) as the sole objects of preferences relevant at any time. The actual date at which the marker event, e.g., b's death, occurs is not explicitly distinguished. If b's death is judged equally undesirable, or "severe", no matter when it occurs, then this focus on future hazard functions seems appropriate. If the severity of b's death (or of the marker event) is judged to change deterministically over time, e.g., to reflect a feeling that deaths at certain ages are especially tragic, then the time-dependent functions \( u(\cdot, \cdot; t) \) must incorporate this fact. The more general problem arising when the marker event is no longer a binary variable, but carries an uncertain severity value with it, is dealt with in the following simple but important corollary of Theorem 7.10.

**THEOREM 7.11:** Let \( r(t) = r[h(t), F(\cdot; t)] \) be the d.m.'s subjectively assessed "risk", evaluated at the beginning of period \( t \) and measured on the unit interval CRS developed in Chapter 5, for an accident that occurs with probability \( h(t) \) in period \( t \), and that produces a random severity in an interval \([0, M]\), if it does occur, according to the c.d.f. \( F(\cdot; t) \) over \([0, M]\). Then the conclusion of Theorem 7.10 still holds when each \( h(t) \) is replaced by \( r(t) \) and each \( h^*(t) \) is replaced by an analogous \( r^*(t) \).

Proof: The domain of \( r(t) \), namely, \([0, 1]\), is the same as for \( h(t) \). This follows from the construction of the CRS. Nothing in the proof of Theorem 7.10 depends on any feature of \( h(t) \) except that it lies in the unit interval. Hence, the problem of representing preferences over "stochastic risk trajectories", \( r^*(t) \), is isomorphic to the problem of representing preferences over stochastic hazard rate trajectories \( h^*(t) \), and Theorem 7.11 follows by reinterpreting Theorem 7.10. Q.E.D.
Theorem 7.10 can now be viewed as a special case of Theorem 7.11 in which severities in each period are deterministic, so that only the hazard rate and the period are important in assessing the risk in any period $t$. Theorem 7.11 only applies to cases in which the CRS is isomorphic to $[0,1]$, however. If accident severity, rather than being indexed by numbers in $[0,M]$, depends on which members of a population are killed, for example, then the preceding theory requires extension. This case will be examined in Chapter 8.

7.5 Remaining Challenges

Theorems 7.10 and 7.11 present a theoretical solution to the problem of individual choice among stochastic hazard rate or risk trajectories. By defining the risk ordering over $H^*(0)$ as the reverse of the preference ordering $R$ (so that an index of subjectively assessed risk would be $r(h^*) = 1 - U(h^*)$ for any $h^*$ in $H^*(0)$, for example), the problem of personal risk measurement for stochastic hazard functions is solved. The d.m. choosing among stochastic hazard rate trajectories can do so on the basis of his risk index $r(h^*)$. A similar construction holds for more general stochastic risks.

Several additional problems remain to be solved, however, before the above preference theory can be developed into a satisfactory theory of risk measurement. These are outlined below.

7.5.1 Nonmonotonic Time Preferences

The theories of objective risk measurement that we have developed exploit the assumption that all d.m.'s share monotonic preferences in some variables, e.g., consequence "severity". This provides a basis for objective risk measurement, using FSD or a variant of FSD to establish an ordering that all d.m.'s will agree with over a reference set of "canonical" risks.
In the case of time varying risks, such monotonicity may not hold. We assumed in Lemma 7.2 and in Theorems 7.3 to 7.7 that it did — that preferences for the arrival date of the marker event were strictly increasing in the time of arrival. Thus, the FSD* relation could be defined and used. While this assumption of monotonic preferences is certainly appropriate for many applications, it would not hold for someone who prefers death before (s)he starts to raise a family to death after (s)he has started to raise a family. Similarly, it would not hold for someone who prefers to die rather than to continue to live in pain or incapacity.

Where preferences for arrival dates of the marker event are not monotonic, the basis that we have developed for objective risk measurement disappears. This could be a real difficulty in certain types of individual health and safety decisions where "fates worse than death" are involved. Even if monotonicity is assumed, the problem of exploiting it in Theorems 7.10 and 7.11 to get from a personal preference theory to an objective risk measurement theory remains to be solved. A possible approach would be to develop a continuous time analog of the above theory, and then to use the stronger mathematical structure of the real line to define a continuous one-parameter family of stochastic hazard (or more general risk) trajectories totally ordered by FSD*. It is not clear now whether this construction is possible.

7.5.2 Deriving Risk Preferences

Theorems 7.10 and 7.11 give a representation for prescriptive or derived preferences over stochastic risk trajectories. That is, they describe what the preferences of someone obeying axioms DR1 to DR3 should look like. But they do not provide a constructive way of explicitly deriving the functions U(t) and u(.,.,.;t) from primitive preferences over something (e.g., known hazard rate trajectories) for which a real d.m. might be supposed to have clearly defined primitive preferences. In other words, the set of known preferences from which the prescriptive preferences are to be operationally derived has not been specified. To
make the theory directly applicable in practice, an evaluation procedure using simple stimuli (e.g., hazard rate trajectories that are zero except in one or two periods) must be developed.

7.5.3 Controlled Processes

Individual risk Model 2 applies to cases in which a d.m. must choose at time 0 among alternative stochastic risk or hazard rate trajectories. There is a single decision date, although the uncertain consequences of the decision may unfold for some time to come as the d.m. makes new observations giving him new information about the risks to which he has committed himself (or other targets of the marker event.) An "act" in this context is just a choice from the subset of \( H^*(0) \) that the d.m. is able to choose from.

In many practical applications, however, the d.m. can take other sorts of acts, or can take a sequence of actions over time. He can control, as well as observe, the risk process. We now show that the expected utility representation of Theorems 7.10 and 7.11 no longer holds when the d.m. has such additional acts available to him. Our argument is based on the idea of "induced preferences" in the economic theory of consumer choice (Spence and Zeckhauser, 1972.)

**Example 7.10: Buying Insurance**

A d.m. is exposed to a constant fatality hazard rate \( h \) is each period. He is initially uncertain about \( h \), but knows that it is drawn from \([0,1]\) according to a c.d.f. \( P \). Thus, \( h \) follows a stochastic trajectory. In period 0, he must decide how much life insurance to buy. In period 1, assuming that he lives to see it, he (and his insurance company) both learn the true value of \( h \). He can then purchase more life insurance if he wants to, but at an increased rate.

We will treat c.d.f's over \( h \) (or, equivalently, trajectories in \( H(0) \)) as the objects of preference and choice, assuming that the d.m. also chooses
his insurance amounts "optimally" given any c.d.f. over \( h \). Let \( F \) and \( G \) be two c.d.f.'s for \( h \) such that the d.m. is indifferent between them in the above situation. That is, he is indifferent between \([a(F), F]\) and \([a(G), G]\), where \( a(F) \), for example, denotes his optimal period 0 insurance purchase given that \( h \) is drawn from \( F \). If an expected utility representation holds for his "induced" preferences over \( F \) and \( G \), then it must be the case that

\[
E(U; F) = E(U; G) = E[U;(F, p, G)]
\]

for any \( p \) in \((0,1)\),

where \((F, p, G)\) is the probability mixture (resolved after \( a \) has been chosen) putting probability \( p \) on \( F \) and probability \((1 - p)\) on \( G \). In other words, if the d.m. is indifferent between \( F \) and \( G \) then under an expected utility representation, he would also have to be indifferent between all probability mixtures of \( F \) and \( G \). This is essentially the substitution axiom.

But in fact, he will in general strictly prefer both \( F \) and \( G \) to any probabilistic mixture of the two. While \([a(F), F]\) and \([a(G), G]\) may be indifferent, \([a(F,p,G),(F,p,G)]\) will typically be dispreferred to each of them. The reason is that uncertainty about whether he faces \( F \) or \( G \) makes it more difficult for him to plan appropriately. For example, if there are quadratic penalties for period 1 adjustments in insurance, and if he chooses \( a(F) \) as an increasing function of \( E[h; F] \), then the expected penalty for period 1 adjustment will be greater for \([a(F,p,G),(F,p,G)]\) than for either \( a(F) \) or \( a(G) \). In short, mixing \( F \) and \( G \) suppresses decision-relevant information, giving the d.m. lower expected utility. The implication is that the d.m.'s induced preferences over \( F \) and \( G \), and hence over the corresponding stochastic hazard rate trajectories, have no expected utility representation.

This example clarifies the conditions needed for the representation of Theorems 7.10 and 7.11 to hold. Specifically, those theorems apply to cases in which (i) The d.m.'s sole decision is which of several stochastic risk trajectories to choose; and (ii) The only consequences of
interest are the successive hazard rates or risks that he will face (and, of course, the final outcome of these risks, e.g., the period in which the marker event finally arrives.) In richer situations where actions, risk trajectories, and non-accident consequences such as payments of insurance premiums are intermingled, induced preferences over stochastic risk or hazard trajectories will not have the simple form of expected utilities.

This result should not be particularly disappointing. It does not mean that derived preferences for acts are any more difficult to compute: it only means that they lack a convenient numerical representation. Under axioms similar to DR1 to DR3, and assuming an infinite horizon for hazard functions and fixed act and state spaces -- perhaps a reasonable model for controlling a hazardous facility, for example -- it is still possible to compute a most-preferred act in each period using ordinal dynamic programming (Sobel, 1975; Blair, 1984.) Thus, risk management decision making can proceed unimpaired, even though numerical measurement of the risks of acts may be made difficult or impossible by the nonexistence of an expected utility representation for controlled stochastic risk trajectories.

7.5.4 Statistical Estimation Issues and Practical Bounds

Throughout this chapter, we have concentrated on conceptual and definitional problems and have stayed away from questions of practical risk measurement or estimation based on realistically available data. Several qualitative results and special models are available, however, that enhance the practical use of hazard trajectories as representations of risk. These models are especially applicable to the study of diseases (such as cancer) having long latency periods and extended effects. For example, the following results are proved in Chapter 5 of Barlow and Proschan, 1975. (For further results of the same type, see e.g., Klefsjo, 1982 and subsequent contributions by various authors appearing in the Naval Research Logistics Quarterly.)
If the fatality hazard rate \( h(t) \) for someone with a terminal illness exists and is known to be continuous and increasing in \( t \), and if the mean survival time is known to be \( T' \) months (starting from the date of onset), then for any time \( t \) between the date of onset and \( T' \), the instantaneous risk \( h(t) \) is no greater than \( 1/T' \), implying that there is at least a \( 1/e = 0.367 \) probability of surviving until \( T' \). Thus, \( 1/E(T) \) gives a (sharp) upper bound on the risk (or hazard) rate when all that is known is that this rate is increasing with time and that \( E(T) = T' \) is the mean time until arrival of the marker event.

Similarly, if the hazard rate is known to be decreasing over time, then the risk rate \( h(t) \) is no less than \( 1/E(T) \) for \( t \) less than \( E(T) \), implying a probability of at most 0.368 of surviving to \( E(T) \); and the probability of surviving for an amount of time \( x \) beyond \( E(T) \) is no greater than \( (1/e)E(T)/[E(T) + x] \). These bounds are also sharp.

Define the average risk or average hazard rate over a time interval \([0,t]\) as \( H(t)/t \), where \( H(t) \) is \( h(s) \) integrated from \( s = 0 \) to \( s = t \). (A more general definition of the cumulative risk \( H(t) \), appropriate when the hazard function \( h(s) \) is not defined, is \( H(t) = -\ln[1 - p(t)] \), where \( p(t) \) is the probability that the marker event will arrive by time \( t \); see Section 4.1.) If average risk \( H(t)/t \) is known to be increasing in \( t \), then the average hazard rate up to time \( t \) provides an upper bound for the risk rate at all moments before \( t \), and a lower bound on hazard rates at all moments after \( t \). If \( H(t)/t \) is decreasing in \( t \), then these bounds hold with the inequalities reversed. The class of lifetime distributions with increasing average hazard rates is of considerable interest in systems reliability theory and disease models since (i) It is the smallest class that includes the exponential distributions (i.e., constant hazard rates) and that is closed under formation of coherent systems and limits in distribution; and (ii) It arises naturally from "cumulative shock" damage models.
In addition to such qualitative results, there is a well developed branch of applied statistics that deals with hazard rate estimation and data analysis for both parametric and nonparametric models (see e.g., Kalbfleisch and Prentice, 1980 for a treatment emphasizing carcinogenic risk assessment applications, or Lawless, 1982 for a somewhat easier treatment with both reliability and biostatistical applications). There is also a rapidly developing, technically sophisticated literature on the application of martingale stochastic calculus theory to estimation of stochastic hazard rates and analysis of lifetime data (see e.g., Harrington, 1983, for an overview of this theory, and Andersen et al, 1982 for a detailed development with applications to lifetime data analysis.) Thus, the conceptual approaches to risk measurement discussed in this chapter are compatible with currently evolving theoretical methods of statistical estimation and data analysis. However, the task of systematically applying these modern theoretical methods to the particular problems of risk measurement (or rather, estimation) for dynamic risks of the sort discussed in this chapter largely remains to be done.

In summary, we have focused on the conceptual aspects of risk definition and measurement as an appropriate starting point for a sound theory of risk analysis. However, a parallel treatment of the statistical aspects of risk estimation from realistic data sets still needs to be developed to allow the concepts of this chapter to be applied in practice.
8.0 Introduction

This chapter deals with two general topics: measurement of routine population fatality risks and measurement of catastrophic population fatality risks. These are the two types of risks usually dealt with in industry- or government-sponsored risk analyses of air pollutants and contaminants, chemical carcinogens, and industrial production processes. Whereas the theories of individual risk measurement discussed in Chapter 7 are potentially useful for expressing the risks to an individual from participating in some activity or expressing the risks to an industrial facility of a catastrophic accident, most applied risk analyses, especially in product safety and public health areas, are concerned with human target populations having thousands or even millions of members. Defining "risk" in such cases requires some new concepts. The purpose of this chapter is to introduce some of the relevant definitional issues and to extend the analytic approach developed in the preceding chapters to apply to population risks.

The chapter is organized as follows. Section 8.1 looks at the problem of defining and measuring routine population fatality risks. The simplest case, in which a d.m. seeks to choose among alternatives that will impose statistically independent, identically distributed (i.i.d.) random times to death on all members of some target populations is considered first, in Section 8.1.1. It is shown that this problem involves a novel aspect of social choice, in that the preferences of the same people at different times must be considered. If eventual, rather than initial, preferences of members of the target population are used as the basis for majority-rule choice, then a simple conclusion is reached: alternative lifetime distributions should be ranked according to mean life length, assuming that this is known.

The crucial assumptions of i.i.d. lifetime distributions are relaxed in two stages. Section 8.1.2 drops the assumption of identically
distributed lifetimes, allowing for heterogeneous populations. Difficulties for classical epidemiological definitions of population risk when populations are heterogeneous are illustrated by simple examples, and an approach to choice among risk distributions based on expected numbers of lives lost per year is proposed and justified.

Finally, Section 8.2 drops the assumption of statistically independent (or "routine") individual fatality risks and considers how best to represent "catastrophic" risks. Section 8.2.1 reviews the classical frequency-severity curve (or "risk profile") for representing compound Poisson processes and suggests that the attribute "number of fatalities" provides an inadequate basis for describing catastrophic risks. Recent work on "risk equity" is discussed in Section 8.2.2, and a new analytic approach to representation and analysis of catastrophic risks based on multivariate lifetime distributions and fatality risk "hypergraphs" is proposed in Section 8.2.3.

8.1 Routine Population Risk

In the risk processes studied in Chapter 7, the analysis of risk over time always ended with the arrival of a marker event, usually interpreted as the death of a particular target individual. In contrast, the study of population risks requires that the analysis continue beyond the death of any specific individual. This can be done conveniently using the idea of a "population renewal process" introduced in Section 7.2. We shall now define such processes more carefully and use them to illuminate some of the problems of population risk measurement.

8.1.1 Population Renewal In a Population of Size One

The simplest population renewal model involves a population with only one person in it at a time. For example, a firm that hires employees into some position, say Chief Inspector, may be interested in the occupational hazards associated with this job. Although only one individual occupies the position at a time, it makes sense in assessing the occupational
risks of the position to consider the fatality rates in the "population" of people who occupy it.

Suppose, then, that the target population B to be analyzed consists of a single person who is replaced immediately by an identical person whenever a marker event (specifically, the death of the current occupant) arrives. This is clearly not intended to be a realistic model of population fatalities, but the analysis for this simplest case generalizes in several ways to larger populations. If each new person independently starts off with the same distribution of remaining life, represented by a known c.d.f. \( F(.\) ), then the result is an ordinary renewal process. The problem then becomes how to define and measure the risk of a renewal process.

From the standpoint of a single participant in the process, the question has already been answered by the analysis of Chapter 7: the risk to him when he enters the process is the risk of the deterministic hazard function (or trajectory), say \( h(t) \), corresponding to \( F(t) \), where \( t \) is the elapsed time since entry. Of course, individuals with different joint risk-time preferences may in general assign different values on the Poisson CRS to a particular trajectory \( h(t) \), i.e., the "Poisson equivalent" of \( h(t) \) as assessed at the moment of entry, \( t = 0 \), may be different for different people. We will assume that everyone agrees with the partial ordering of hazard trajectories induced by \( \text{PSD}' \), the temporal first-order stochastic dominance relation, however.

Given two interarrival time c.d.f.'s, \( F \) and \( G \), under what conditions can we say that the renewal process defined by \( G \) is unambiguously "riskier than" the renewal process defined by \( F \) ? Two possible definitions are

\[ \begin{align*}
0 & \quad \text{The } G\text{-process is objectively riskier than the } F\text{-process if and only if } F \text{ FSD}^* G. \quad \text{(Otherwise, someone can consistently have joint time-risk preferences that would make him prefer participation in either process to participation in the other.)}
\end{align*} \]
The G-process is objectively riskier than the F-process if and only if it generates more fatalities per unit time on average.

The first definition incorporates a democratic perspective: comparative risk is defined in terms of the individual preferences of participants (restricted, as always, to the risk consequence set alone, which in this case has to do with lengths of lives.) This is also the perspective that a single individual might take on choosing which among alternative processes (e.g., corresponding to different occupations) to participate in.

The second definition focuses on outcomes, and might be more appropriate for guiding a "social decision maker" (s.d.m.) trying to decide which of two benefit-producing (and fatality-generating) technologies to support. For any renewal process in which the random life length $T$ of each participant has finite mean and variance, $E(T)$ and $\text{Var}(T)$ respectively, the actual number of fatalities (i.e., renewals) in the time interval $(0,t)$, say $N(t)$, approaches a normal distribution with mean $[1/E(T)]t$ and variance $[\text{Var}(T)/E(T)]t$, and the long-run average fatality rate (renewals per unit time) approaches $1/E(T)$ with probability 1. If, in the choice between the F-process and the G-process, it is known that the chosen process will operate for long enough for these asymptotic results to hold to an adequate approximation, then it might be argued that "social rationality" requires selection of the process, with the lower average fatality rate. But this may conflict with the preferences of the individual participants. In other words, although participants may live longer on average in the F-process than in the G-process, it may be that they are deterred from voluntarily selecting participation in the F-process by features of its interarrival time hazard function, e.g., an initial high hazard rate.

A third possible definition that makes especially clear this potential conflict between individual and social preferences for renewal processes will now be developed. Suppose that F and G have respective hazard rate trajectories $h$ and $g$, and that $h(t)$ crosses $g(t)$ exactly once, from
above, at $t^*$. Thus, neither $F$ nor $G$ temporarily dominates the other by the FSD* relation. In choosing between $F$ and $G$, or, equivalently, between $h$ and $g$, a social decision maker (s.d.m.) might take either of two perspectives for combining individual preferences:

(i) He may take an EX ANTE perspective, in which the actual preferences of individuals as of the decision date are used as the basis for choice; or

(ii) He may take an EX POST perspective, in which the presumed preferences that individuals would have if they were perfectly informed are made the basis of choice.

By "perfectly informed" we mean the following. Suppose that the random time at which each individual would die under each alternative could be made known to him at the decision date. Assume that each individual would then prefer the alternative giving him the longest life. The preferences that he would then have are what we are calling his "perfectly informed" preferences.

More realistically, suppose that each individual evaluates over time the alternative that is actually chosen by comparing the distribution of his life length under that alternative, conditioned on however much time has passed, to the life distribution that he would have expected if the other alternative had been chosen. Then his evaluation of which alternative is (or would have been) superior for him may change over time. In the case where $h(t)$ crosses $g(t)$ from above at $t^*$, for example, everyone who survives for long enough if $h$ is chosen will end up preferring $h$ to $g$, i.e., believing that he has done better under $h$ than he probably would have under $g$. On the other hand, individuals who survive for a sufficiently short time may perish while still preferring $g$ to $h$. In the ex post perspective, the s.d.m. weighs these two sets of eventual preferences against each other in deciding which renewal process is preferable.
The problem of deciding whether the F-process or the G-process corresponds to a greater "social risk" (to a socially less preferred risk process) requires a theory of social decision making, e.g., a rule for combining individual preferences into social ones. The study of such rules lies beyond our intended scope. If we simply take "majority rule" as a plausible example, however, then it becomes clear how a conflict between individual and social preferences may arise. Each individual participant ranks alternative renewal processes on the basis of the hazard rate functions of the corresponding interarrival distributions, as judged at the moment of entry into the process, i.e., at a renewal point. But the s.d.m. using an ex post majority rule might choose between processes on the basis of which one will, on a statistical basis, eventually come to be preferred by a majority of participants, as assessed just before the time of death (also a renewal point) rather than at the time of entry into the process. Put otherwise, the preferences of "outgoing", rather than "incoming" participants will be used as the basis for the s.d.m.'s choice among processes. The two may disagree, i.e., a majority of incoming participants may prefer the F-process, while a majority of outgoing participants prefer the G-process. (Of course, both sets consist of the same people, though at an earlier and a later set of times.) Then a s.d.m. who defines the relation "riskier than" between renewal processes to mean "is dispreferred to by a majority of outgoing participants" could find his risk ranking of alternative renewal processes conflicting with the current preference ranking of incoming participants.

In summary, when the FSD* relation fails to hold between two or more renewal processes, it may not be clear how to rank them in terms of comparative risk (or social undesirability.) If the ranking is made by aggregating the individual preferences of potential participants, then the question of when these preferences are to be assessed -- e.g., when individuals are entering the process or when they are leaving it after finding out something about their survival times under the selected alternative -- must be answered. This question arises no matter what preference aggregation rule is used.
Two points may be added.

- Nothing in the argument just outlined depends on the assumption that individuals follow each other sequentially in a renewal process. Exactly the same problem of when preferences are to be assessed arises if \( N \) individuals are trying to decide which of two or more public risks, each imposing identical, statistically independently distributed (i.i.d.) risks on all members of the population, is most preferable.

- If \( N \) is large, then individual lifetime (still assumed to be an i.i.d. random variable) averaged over all \( N \) members of the population will still be approximately normally distributed. The definition of comparative risk that orders sources of i.i.d. risks in terms of average life expectancy will therefore coincide with the definition that orders them by the preferences of perfectly informed outgoing members. (This holds since the mean and median of the normal distribution coincide.)

Thus, analysis of the single-person population renewal process leads to the following conclusion:

**PROPOSITION 8.1:** Consider a choice between several alternative processes, each imposing a known i.i.d. fatality risk (a distribution of remaining life) with finite mean and variance on each of \( N \) individuals. These individuals may live concurrently or at different times, but there are enough of them so that for any alternative that may be selected,

(i) the empirical c.d.f. of their realized lifetimes will approach the theoretical c.d.f.; and

(ii) the average lifetime, averaged over all \( N \) individuals, will be approximately normally distributed.
Then a decision maker who wishes to rank alternatives according to the "perfectly informed" preferences of a majority of the N individuals will rank the alternatives according to the expected values of their corresponding lifetime distributions. In the special case of a one-person renewal process, where the N individuals live consecutively, this is equivalent to ranking alternative renewal processes in reverse order of their average fatality rates.

Proof: The statement of this proposition is deliberately informal, since the conditions needed to prove it depend on the interpretation given to the term "perfectly informed preferences." That a ranking of renewal processes by expected lifetime is equivalent to a reverse ranking by average fatality rate follows directly from the renewal theory result that average fatality rate converges to the reciprocal of expected lifetime with probability one (for lifetimes with finite mean.) If "perfectly informed preferences" are interpreted to mean that each individual eventually learns how long he would have lived under each alternative, and ranks the alternatives accordingly, then the proposition that a majority will come to prefer the alternative giving the greatest expected lifetime follows from the facts that (i) average lifetime will be approximately normally distributed; and (ii) the mean and median of normal distributions coincide. Thus, more than half of the N people will live longer in an (approximately normal) distribution with a higher median than they would have in an (approximately normal) distribution with a lower median. An alternative sufficient condition for establishing this part of the proposition is that each person who ends up living longer under an implemented c.d.f. F than he would have expected to live under an alternative c.d.f. G will end up preferring F to G. Other sets of sufficient conditions for Proposition 8.1 also exist.

Proposition 8.1 provides a possible basis in normative political theory for using expected lifetime values to guide choice among known i.i.d. population risks, even though a majority of the affected individuals may oppose the resulting choices in the short run. According to this reasoning, average frequency of arrivals of fatalities provides an
adequate ordinal measure of risk for both single-person renewal processes and cohorts of N people. More generally, for target populations in which members are born into overlapping generations, the reciprocal of mean life expectancy still provides an adequate ordinal measure of risk under the i.i.d. assumptions.

An interesting aspect of this definition is that the social decision maker (s.d.m.) is assured on a statistical basis (for large enough N) of making what would be the majority choice for perfectly informed participants, even though there may be no way of telling in practice which members of the population will eventually belong to this theoretical majority, e.g., will survive long enough so that the s.d.m.'s choice becomes their own most-preferred choice. The political acceptability of decisions based on such an unobservable statistical majority interest is of course open to question. However, it seems that even in practice, some policy stance must be taken toward the question of whether to use current or eventual (or some form of imputed "perfectly informed") preferences as the basis for public choice among population risks.

8.1.2 Population Renewal in a Heterogenous Population

The renewal process model discussed above contains the essential simplifications that (i) All N members of the population have the same lifetime distribution, which it is the d.m.'s task to choose; and (ii) All individual lifetimes are statistically independent, so that learning that one person has died by (or at) time t gives no information about the likelihoods that others have died as well. In this section, we turn to the more realistic case in which choosing an act may impose different risks on different people. Section 8.2 will allow individual risks to be correlated.

Suppose then that a target population of size N is exposed to risks created by an activity a chosen from the d.m.'s choice set, A. We shall continue to assume that each individual i is replaced at the moment he
dies with a "new" individual of the same type, whom we will for simplicity also call i. \( F(.;i,a) \) is to be understood as the lifetime distribution for individual i at these renewal points. These assumptions define the "population renewal process" model of routine population fatality risks.

Given two N-tuples, say \( F^*(a') = [F(.;1,a'), \ldots, F(.;N,a')] \) and \( F^*(a") \), when can we say that \( F^*(a') \) is unambiguously "riskier than" (or less desirable than) \( F^*(a") \)? Several sufficient conditions suggest themselves. For example, if the order of the component distributions in \( F^*(a") \) can be permuted so that each component distribution in \( F^*(a") \) preference dominates the corresponding component distribution in \( F^*(a') \) in the sense of FSD*, then we could say that \( a' \) is at least as risky as \( a" \). However, this definition would in general be expected to give only a very partial ordering of acts.

Alternative, stronger definitions that are now widely used in epidemiology define population risk in terms of "incidence rates", i.e., average hazard rates. More precisely, let \( h_2 \) and \( h_1 \) be the fatality hazard rates in populations exposed and unexposed to some dichotomous risk factor. For example, \( h_1 \) might be approximated for empirical work as the average fraction of people in population 1 (the unexposed group) who die within any one-year interval; and similarly for \( h_2 \). (Marker events other than death, e.g., occurrences of a specific disease with a specific etiology, can be treated in the same framework.) The following conventional definitions of comparative risk are used (see e.g., Anderson et al, 1980):

1. The **attributable risk** for fatalities due to exposure to the risk factor is defined as the difference between the fatality incidence rates in populations exposed and unexposed to the risk factor, i.e., it is \( h_2 - h_1 \);

2. The **relative risk** of the marker event for individuals in the group exposed to the risk factor is defined as \( h_2/h_1 \);
The odds ratio for the exposed compared to the unexposed group is defined as $\frac{[\text{h2}/(1-\text{h2})] / [\text{h1}/(1-\text{h1})]}{\text{h1}/(1-\text{h1})}$, which will be approximately the relative risk if h2 and h1 are both small.

All three measures reflect the view that population risk is a comparative, rather than an absolute, concept. However, the attributable risk, h2 - h1, is compatible with a definition of (absolute) population risk as the average arrival rate (new cases per person-year) of the marker event in the population. Moreover, for homogeneous populations, it agrees with our use of the hazard rate in Chapter 7 as a cardinal index of individual fatality risk in the special case of Poisson competing risks.

When the Poisson competing risk assumptions are not satisfied, however, either because individual hazard rates are not constant over time or because different people have different hazard rates, both the definition of h2 - h1 as the risk "attributable" to the risk factor and the definition of average hazard rate as an index of population risk become questionable. The problem of extending the definition of "attributable" risk to cases where multiple factors interact nonadditively to create a risk has been addressed elsewhere (Cox, 1985), and will not be considered here. Instead, we shall illustrate with two simple examples some of the difficulties that arise from a definition of population risk in terms of average hazard rate when populations are heterogeneous. Section 8.1.3 then offers a refined definition based on a weighted average of hazard rates for the routine population risk created by an activity.

Example 8.1: Aggregation of Individual Hazard Rates Over Time

Let $F(.;i,a)$ for $i = 1, 2, ..., N$ denote the distributions at birth of the (latent) times to cancer mortality for N members of a population exposed to competing risks. Corresponding to each $F(.;i,a)$ is an age-specific hazard rate for cancer death in person i given a; in addition, there will be a hazard rate for death due to all other causes. Suppressing a, which we assume affects the cancer hazard rate only, we
will denote by $h(t; i)$ and $b(t; i)$ the age-specific cancer hazard (i.e., mortality) rate and the all-other-cause mortality rate, respectively, for individual $i$ at age $t$.

Since epidemiological data are generally available only on a year by year basis, we will use discrete time; thus, $h(t; i)$, for example, is defined operationally as the fraction of people of the same "type" as $i$ who die of cancer between their $(t-1)\text{st}$ and $t$-th birthdays. (This crude estimate may be corrected if necessary for competing-source mortalities in that year, so that cancer mortality per person-year of exposure at age $t$ is adequately approximated. See e.g., Lawless, 1982, Chapter 2 for a discussion of actuarial estimation of age-specific hazard rates.)

Now suppose that it is necessary to estimate the $h(t; i)$ functions for different individuals, and suppose that $i$ now indexes distinguishable "types" of individuals, rather than individuals themselves. Thus, there may be $N(i)$ individuals of type $i$, and $N$ is the the sum of the $N(i)$ over all types $i$. One way to estimate the age-specific hazard rates $h(t; i)$ is to assume that $h(t; i)$ is the same for all types $i$, i.e., $h(t; i) = h(t)$.

Then data can be pooled across types and the hazard rate estimated as a function of the age index $t$ alone. The model specification that $h(t; i) = h(t)$ for all $i$ may be called the assumption or hypothesis of homogeneous population response, since $h(t; i)$ may be thought of as the probabilistic "response" for an individual of type $i$ to the risk factors to which he is exposed over time. If this specification is appropriate, then $h(t)$ can be estimated by tracking one or more birth cohorts over time and noting the fraction of those who survive to year $t$ that perish by year $t + 1$.

However, there is nothing to prevent an analyst from applying the homogeneous response assumption to a population in which individual hazard functions, or probabilistic responses, are actually heterogeneous. Suppose, for example, that each individual in a population has a fixed hazard rate, corresponding to an exponential survival time. Let the hazard rate $h(i)$ for individual $i$ be drawn from some arbitrary continuous distribution with density function $f(h)$, the
same for all individuals. (Since \( f(.) \) is continuous, each individual will with probability one be of a unique type, i.e., have a hazard rate that no one else has.) Thus, the population as a whole is heterogeneous. If the hazard function for this population is now estimated under the assumption of homogeneity, e.g., by tracking the empirical survival fractions from year to year, as explained above, then it will appear that the group hazard function \( h(t) \) is declining with age, \( t \), since in any cohort that is followed, individuals having higher hazard rates will tend to drop out first, i.e., while still comparatively young. Thus, assuming that \( h(t;i) = h(t) \) for all \( i \) when in fact \( h(t;i) = h(i) \) for all \( t \) can lead to the sensible but false conclusion that the age-specific hazard function for the group decreases with age, when in fact there is no (single) hazard function for the group, and each individual's hazard function remains constant with age.

As a concrete continuous-time example (Lawless, 1982, p. 49), suppose that \( h \) follows a gamma distribution in some population, with mean \( m = ab \) and variance \( v = ab^2 \). Then the apparent (unconditional) hazard rate under the homogeneity assumption will be not \( h(t) = m \), but \( h(t) = m/(1 + bt) \).

The point of this example is not just that a misspecified statistical model may lead to incorrect (although perhaps plausible) results. The more important point is that consistent aggregation of individual hazard functions into a single "equivalent" group hazard function is not possible in general. For example, when each individual has a different constant hazard rate, no constant "average" group hazard rate exists that will give the same pattern of fatalities over time when applied to all members of the group.

Finally, it should be noted that the population renewal process produces its own internal dynamics for the evolution of the population. If all individuals do have the same hazard function, for example, so that the homogeneous specification is correct, then the age distribution of the population will adjust over time to a steady state distribution that is
independent of the initial age distribution (see e.g., Feller, Vol. 1, 3rd Ed., 1968, p. 335.) Thus, a definition of the population fatality risk created by an activity that shifts the distribution of hazard functions within a population must in principle take into account the dynamic response of the population over time as it adjusts to the new risk levels.

Example 8.2: An Aggregation Artifact in Comparative Risk Assessment

In epidemiological work, as indicated above, it is usual to compare fatality rates in an exposed (or "treatment") population to fatality rates in an unexposed (or "control") population in order to calculate the "attributable risk", h2 - h1, or the "relative risk", h2/h1, due to exposure to a particular dichotomous risk factor (or to a proxy variable such as employment in a particular occupation.) In practice, each population will consist of a number of different "types" of individuals, with each type having a different hazard rate (or more general risk) function mapping exposure to the risk factor (and to other background and endogenous factors) over time into probabilistic health responses over time. The set of types is usually complicated, e.g., having a mixture of binary, discrete, continuous, and distribution-valued components corresponding to the different types of explanatory factors (such as sex, race, age, and past smoking history, respectively) that at any moment affect the c.d.f. of the remaining time until the occurrence of the marker event. Thus, "type" can not in general be indexed by any index set (e.g., the nonnegative reals or a Euclidean vector space) with a simple mathematical structure.

In practice, the distribution of types in the treatment and control populations is generally ignored, except that matching and adjustment techniques may be used to correct for obvious differences in the distribution of type components, other than the presence or absence of exposure to the risk factor, between the two populations. (See e.g., Anderson et al, 1980.) The effects of residual heterogeneity on the comparison of mean hazard rates between the two groups are then left for randomization to "wash out".
Figure 8.1 presents an extreme hypothetical example of what can happen when mean hazard rates for two heterogeneous groups with different "type" distributions are compared. For simplicity, there are only two types of individuals within each exposure category: "occupational" and "general public". The tables represent the numbers of people in each exposure group and each type who are alive and cancer free at the beginning of year $t$ and who have developed cancer (the marker event) by year $t + 1$. (The numbers are, of course, dramatically inflated for the purposes of exposition.)

Table 8.1a shows that the mean hazard rates for year $t$ and for the exposed and unexposed groups are 46% and 11%, respectively, for an "attributable risk" due to exposure of 35%. Tables 8.1b and 8.1c disaggregate the data by type: they sum, cell by cell, to give the total (aggregate) frequency count data in Table 8.1a. Qualitatively, tables 8.1b and 8.1c tell a very different story from Table 8.1a. The disaggregated "attributable risk" numbers are -41% among people of occupational type, and -5% among people of general public type. Thus, the aggregate data indicate that exposure leads to a positive attributable risk (or to a relative risk greater than 1), while in fact, exposure leads to a negative attributable risk (or to a relative risk less than one) for individuals of both types. The finding of a positive attributable risk is purely an artifact of aggregation. This phenomenon is known in statistics as "Simpson's Paradox" (Blyth, 1972.)

Simulation suggests that while actual sign reversal of this sort is probably uncommon, the bias introduced by omitting "type" as an explanatory variable for observed average hazard rates in heterogeneous populations tends to be substantial. There are several implications for the definition and measurement of population risk. While it is obvious that omitting relevant explanatory variables (i.e., components of "type") is a poor idea when it can be avoided, the important fact is that it can seldom be avoided. Important components of individual type, such as endogenous hormone status, diet and lifestyle histories, exposure histories for various risk factors, genetic makeup, efficiency of
### TABLE 8.1.a

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>505</td>
<td>110</td>
</tr>
<tr>
<td>No Cancer</td>
<td>595</td>
<td>901</td>
</tr>
<tr>
<td>Aggregate Attributable risk = 35%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cancer Incidence Rate: 46% 11%

### TABLE 8.1.b

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>No Cancer</td>
<td>500</td>
<td>1</td>
</tr>
<tr>
<td>&quot;Occupational&quot; Attributable risk = -41%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cancer Incidence Rate: 50% 91%

### TABLE 8.1.c

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>No Cancer</td>
<td>95</td>
<td>900</td>
</tr>
<tr>
<td>&quot;General Public&quot; Attributable risk = -5%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cancer Incidence Rate: 5% 10%
internal repair mechanisms, and so forth that affect the probability of health responses over time are usually unknown or unobservable or both. This means that the hazard function for a particular group of people, even when the group is homogeneous with respect to sex, age, and other obvious variables, is usually best thought of as a distribution over hazard functions, rather than as a single hazard function. Expressions such as $h_2 - h_1$ or $h_2/h_1$ must then be interpreted in terms of derived distributions.

Equivalently, any classification system for assigning individuals to "cells" of a factorial or contingency table for purposes of data analysis will in practice be incomplete. Thus, associated with the population in each cell, where a "cell" is specified by fixing the levels of observable factors, will be residual heterogeneity corresponding to varying levels of other, unobservable factors among the members of the cell. Necessarily, therefore, valid statistical inferences can be drawn only about statistical constructs consisting of cells or aggregates of cells: quantities corresponding to the sub-cell level of detail are not measurable with respect to the partition of the population defined by the cells. In particular, valid statistical inferences about specific individuals can not be drawn from cell data unless it is assumed that the populations within cells are internally homogeneous. Instead, any individual's hazard rate in a particular year must in general be thought of as being drawn from an unknown distribution whose average is the mean hazard rate for the cell to which the individual belongs.

In summary, defining population risk as an average of individual risks (e.g., hazard functions) creates difficulties. On the one hand, the time behavior of the aggregate measure will in general not be consistent with the time behavior of the underlying distribution of hazard functions on which it is based (Example 8.1.) On the other, residual heterogeneity leads to aggregation artifacts which in severe cases can make the qualitative properties of population risk measures disagree with the qualitative properties of every individual's risk measure. It is therefore useful in general to think of population risk as a distribution.
of individual risks, e.g., represented by a distribution over individual hazard functions or "types".

8.1.3 An Alternative Approach to Defining Routine Population Fatality Risks

The above difficulties with conventional definitions of population risk suggest the need for new approaches to the population risk measurement problem. In this section we offer one possibility, based on the theory developed in previous chapters. Since it is largely an application of ideas and principles already developed, our derivation of the proposed measure will be somewhat synoptic.

A. Identifying the Set of Relevant Consequences

A population may be thought of as generating a stochastic stream of fatalities over time. Different choices of acts $a$ in $A$ will induce different probability distributions over time streams of fatalities. We shall take these time streams (or "trajectories") of fatalities as the relevant consequence set $C$ and consider the changes in the probability distribution over $C$ induced by different acts in $A$, relative to some "neutral" (status quo) act, to be the objects for which preferences are to be derived.

B. Developing a Mathematical Representation for Preferences Over Consequences

The next step is to establish a convenient mathematical representation, such as a value or utility function, for preferences over consequences. The argument proceeds in two principal stages: development of a value function for lives lost in a year, and exploitation of the statistical assumptions of the population risk model (specifically, large population size and statistical independence of times of deaths) in the context of this value function to show that the value loss per year will be an approximately normally distributed random variable on an appropriate
value scale. This result is then used to conclude that the effects of a change in routine population risk will be to replace one normal distribution over value with another. Since normal distributions can be characterized with only two parameters, this leaves the way open to tractable measurement of population risk.

C. Developing a Value Function for Routine Mortalities

The following discussion, down through and including Theorem 8.2, is relatively technical. It presents conditions under which the value of having any group of people survive for a year is the sum over all members of that group of the values of having each of them survive for a year. Interpretively, the two key assumptions used to establish this result are that

(i) The value of a mortality, i.e., the loss of value from having the corresponding person die, is assumed to depend only on his "type" at the time of his death. His "type" may include his age, the number and ages of his dependents or other family members, the number of people who know him by name, his health state and expected future earnings, and any other information about him that seems relevant.

(ii) Moreover, the differences (if any) in the values assigned to the deaths of two different people depends only on their types at the times of death.

These assumptions are intended to express the ethical position that the value of a person's life is "intrinsic", rather than extrinsic. In particular, they rule out any assumption that the value of a life lost in some year depends on how many other lives are lost in that year. Decreasing marginal value over "number of mortalities per year", for example, would be excluded. These assumptions are intended only for use in the evaluation of routine fatality and mortality risks (i.e., statistically independent individual risks), and may not apply to "catastrophic" risks. A more formal statement of these assumptions and
of the conclusion that the values of individual lives can be represented by an additive value function will now be given.

Consider a birth cohort consisting of \( N \) individuals. As the members of the cohort die over time, a flow of "social disvalue" occurs. We shall make the following assumptions about this value flow.

**Axiom PRO (Temporal Separability):** Let \( B(t) \) be the set of individuals that are alive (i.e., "at risk") at the beginning of year \( t \). Then the flow of value in year \( t \) depends only on the set difference \( B(t) - B(t+1) \), i.e., on which individuals die in year \( t \).

To state the remaining axioms, we need to introduce a "strength of preference" relation \( PR^* \) between pairs of subsets of \( B \), where \( B \) will be used to denote the population at risk. We will interpret \( PR^* \) to mean that for any four subsets \( Q, R, S, \) and \( T \) of \( B \), \( Q, R \) is preferred to \( S, T \) if and only if (survival of) \( Q \) is preferred to (survival of) \( R \) at least as much as \( S \) is preferred to \( T \). From \( PR^* \) we can define a second relation, \( PR \), between subsets of \( B \), such that for any three subsets \( Q, S, \) and \( T \) of individuals in \( B \), \( Q \) is preferred to \( S \) if and only if \( Q, T \) is preferred to \( S, T \). We interpret \( Q \) to mean that the value of (the survival of) subset \( Q \) is at least as great as the value of subset \( R \), so that the loss of the individuals in \( Q \) (due to routine, statistically independent population fatality risks) is at least weakly dispreferred to the loss of the individuals in \( R \). By definition, \( Q \) is at least as preferred (or "valuable") as \( R \) if and only if the preference difference between \( Q \) and \( T \) is at least as great as the preference difference between \( R \) and \( T \), for any subsets \( Q, R, \) and \( T \) of \( B \). (Recall Chapter 5's discussion of "measurable value functions".)

We will let \( I \) and \( P \) denote the indifference relation and the strict preference relation, respectively, corresponding to \( PR \), and similarly for \( P^* \) and \( I^* \) corresponding to \( PR^* \). The axioms for an additive value function can now be stated as follows.
Axiom PR1 (Weak Ordering): If the d.m. knew the true type of each individual in B, then the preference difference relation PR\* representing his preferences, or valuation of individual lives in B, would be a complete, transitive ordering on the set of all pairs of subsets in B.

This asserts that in principle, given sufficient information about an individual, the relative value or severity of having him become a fatality can be assessed. In practice, of course, adequate information is not available, and probably no one would feel qualified to make the required value judgements. However, as we shall see, considerable mileage can be derived from the assumption that such judgements can be made in principle. In particular, this assumption allows alternatives to the usual conventions that all individuals have the same value, or that each individual's value depends only on his age or expected remaining life. Axiom PR1 leaves room for a variety of other value-relevant distinctions, so that the value of the President's life, for example, may be different from the value of a criminal's.

Axiom PR2 (Positive Individual Values): Let S be any nonempty subset of B containing i as a member, and let S - [i] denote the subset obtained from S by removing i. Then S P (S - [i]).

This says that given a choice between having the members of S die and having all members of S except i die, the latter is strictly preferred.

Axiom PR3 (Intrinsic Values of Lives): Let S' and S" be any two subsets of B that include i as a member. Then S',(S' - [i]) I* S",(S" - [i]).

Interpretively, this condition asserts that person i's life has intrinsic value: the loss when he dies in some year is independent of who else lives and dies, how many others die, who else was alive at the beginning of the year, and so forth. The value of an individual's life depends only on his own type at the time of his death. Note that an individual's type may change over time, e.g., to reflect his age as one component. (Remember that we are dealing with routine fatalities, so that
simultaneous deaths within a family can probably be ignored in deciding whether these axioms are reasonable.)

Axiom PR4 (Intrinsic Differences in Life Values): If $S', (S' - [j]) \succ^* S', (S' - [i])$ for some subset $S'$ of $B$ containing both $i$ and $j$, then $S, (S - [j]) \succ^* S, (S - [i])$ for any subset $S$ of $B$ containing both $i$ and $j$.

Loosely, this says that if $i$ is of greater value than $j$ in one context (i.e., against a particular background of survivors), then $i$ is of greater value than $j$ in all contexts. Again, the motivation is that the difference (if any) between the values of individuals' lives should depend only on their types. Lives have absolute, intrinsic values that are not increased or diminished by the background context of routine fatalities into which they may fall. Thus, lives are different from the goods of conventional economic and single-attribute decision analysis, in which the value of each unit of a good depends on how many other units are consumed or received simultaneously.

THEOREM 8.2: Axioms PR1 to PR4 imply that if the true type of each individual in $B$ were known, then the preference relations $PR$ and $PR^*$ could be represented by an additive value measure $v(\cdot)$ mapping subsets of $B$ into nonnegative numbers in such a way that for any four subsets $Q$, $R$, $S$, and $T$ of $B$,

(i) $v(Q) - v(R)$ is at least as great as $v(S) - v(T)$ if and only if $Q, R \succ^* S, T$.

(ii) $v(Q)$ is at least as great as $v(R)$ if and only if $Q \succ R$.

(iii) The value $v(i)$ assigned to any individual $i$ (where we use $i$ to stand both for individual $i$ and for the set $[i]$ consisting of that individual) is a nonnegative number. The value assigned to the null set is identically zero.
(iv) If $Q$ and $R$ are any two disjoint subsets of $B$, then $v(Q \cup R) = v(Q) + v(R)$, where "$\cup$" denotes set union. For example, $v(Q)$ is the sum over all individuals $q$ in $Q$ of the individual values $v(q)$. The value of having subset $Q$ survive for one period is the sum of the values of having each of its members survive for that period.

Proof: The basis for this result is Theorem 1 of Deutsch and Malmborg (1985), who extend the measurement theory framework of Krantz et al (1971) to the case of "Measurable multiattribute value functions defined on binary attributes." It is straightforward to verify that axioms PR1 to PR4 imply (but are not implied by) Deutsch and Malmborg's sufficient conditions ("weak ordering, essentialism, difference consistency, difference independence, and compensatory independence") for existence of an additive measurable multiattribute value function on binary vectors. We have identified each binary vector with the set of its nonzero elements, so that Deutsch and Malmborg's additive value function (from the set of binary vectors to the nonnegative reals) is replaced by an additive value measure (from the set of subsets of $B$ to the nonnegative real numbers.) We have also strengthened some axioms, e.g., by applying weak ordering to PR* rather than only to PR, so as to streamline the statement of the result and make its interpretation in terms of lives more natural. However, our axioms for sets imply the Deutsch-Malmborg axioms for binary attributes, so that Theorem 8.2 follows as an application of their more general representation theorem. Q.E.D.

We will use the term "value weights" to refer to the atomic values $v(i)$ assigned to the individuals in $B$. Thus, the value of any subset of $B$, according to Theorem 8.2, is just the sum of the value weights assigned to the individuals in $B$, where each individual's value weight depends only on his type.

D. A Statistical Model for the Value of Routine Population Mortalities

Theorem 8.2 establishes conditions under which values of lives can, in principle, be represented on a measurable value scale and combined
additively to obtain the value of the survival for a group. We have argued that in practice, these values will probably never be exactly known, since any individual's complete "type" will typically contain components not observable to the evaluator. However, this need not prevent evaluation of the statistical, or probable, values of large groups. If we assume, following Theorem 8.2, that the value of a group is the sum of the values of its members, and if we treat the value of each member as the realization of random variable, then the value of the group will be a sum of the realizations of random variables. Quite a lot may be known about such sums even though relatively little is known about any individual component. This is the line of reasoning followed below.

In single-attribute value theory for continuous attributes, a measurable value function with a natural zero such as the one identified in Theorem 8.2 is unique up to choice of scale. For binary attributes, this is only approximately true. When the value differences between each pair of individuals in a finite target set B is specified by the PR* preference ordering, a finite number of inequalities constraining \( v(.) \) is produced. (For example, in a three-person society, the relation \([1,2],[1,3]\) PR* \([3,2],[1,3]\) implies that \( v([1,2]) \) is at least as great as \( v([3,2]) \), and hence that \( v(1) \) is greater than or equal to \( v(3) \). In general, such restrictions will not determine a unique set of value weights \( v(i) \) for the members of B, but will instead constrain them to lie in some (polyhedral) solution set. Thus, there is some flexibility in the set of value weights used to represent PR* and PR. As the number of people in B increases, however, the constraints become tighter and the solution set shrinks toward a point. (See Krantz et al, 1971, Chapter 9, esp. Theorem 2.) In what follows, we shall assume that the population in B is sufficiently large so that the individual value weights in Theorem 8.2, and hence the corresponding value scale, are essentially unique (up to choice of scale.) This allows normal distributions on the value scale to be well defined in the sense that the distance between the values of two individuals on the value scale, measured in standard deviations, will be essentially the same no matter what set of value weights representing PR* is used.
In principle, individuals are distinguished by "type". All individuals of the same type have identical value weights and hazard rates. In practice, only observable components of "type" can be used to distinguish among individuals. Given a set of individual characteristics that he can observe for the individuals in B, an observer can partition B into (equivalence) classes such that two individuals are in the same class if and only if the observer can not distinguish between them on the basis of their observable characteristics. We will let I denote the partition of B with these classes as its blocks.

Given M routine deaths in year t, suppose that I(t) is the partition of B into which these deaths are divided for the purpose of assessing the total flow of (dis)value in year t from the loss of these lives. We can think of the M deaths as being distributed over the cells of a multidimensional factorial or contingency table in which each cell represents a block of I(t). (See Chapter 2.) Assuming that each block of I(t) contains several types of individuals, the actual value of the M lost lives is not measurable with respect to I(t). However, if there are sufficiently many mortalities in each cell or block of I(t), then the total loss from any particular block, say block k, will be approximately normally distributed with mean and variance proportional to the number of fatalities in that cell. This follows from the assumptions that

(i) The occurrences of fatalities among members of a cell are statistically independent of each other (our definition of "routine");

(ii) The conditional probability distribution over types within a cell given that an individual from that cell has died remains unchanged (by the population renewal assumption.) Thus, the probability distribution over type for each individual in a cell is identical (for an observer who can tell only which cell an individual belongs to and nothing more.)

(iii) The distribution of values of individuals within any cell has finite mean and variance. (This is a type of "Archimedean" axiom, asserting that no individual is worth infinitely more than any other.)
(iv) The value of the set of individuals dying from cell \( k \) is the sum of the values (i.e., the value weights) of the individuals who have died.

These assumptions imply

**PROPOSITION 8.3:** Under conditions (i) to (iv), the yearly value of the lives lost in cell \( k \) is approximately normally distributed with mean \( N(k)v(k) \) and variance \( N(k)u(k) \), where \( N(k) \) denotes the number of lives lost, \( v(k) \) is defined as the mean of the distribution of values of life in class \( k \), and \( u(k) \) is the variance of this distribution. The total value of life lost in all cells is approximately normally distributed with a mean equal to the sum over all cells, \( k \), of \( N(k)v(k) \) and with a variance equal to the sum over all cells of \( N(k)u(k) \).

Proof: The asymptotic normality of a sum of i.i.d. random variables with finite means and variances is a standard central limit theorem. It is well known that the family of normal distributions is closed under sums in the way indicated. Q.E.D.

If the numbers \( N(k) \), \( v(k) \), and \( u(k) \) can be estimated for each class \( k \), then the total population risk corresponding to any fatality frequency distribution \([N(1), \ldots, N(K)]\) over classes can be calculated. Risk measurement for such normal distributions has been considered in detail in Chapter 5.

### 8.1.4 The Case of Homogeneous Responses Within Cells and Equal Value Weights

Denote by \( N(t;k,a) \) the trajectory of fatalities flowing out of population class or cell \( k \) over time in response to a choice of act \( a \) by the d.m. at time zero. If these trajectories can not be accurately predicted, or if the values \( v(k) \) and \( u(k) \) are not known, then the d.m. is put in the difficult position of having to choose among sequences of uncertain normal distributions over the value scale. At best, he will have to choose among normal distributions (representing the random flow
of value due to routine fatalities per unit time) in which the parameters are uncertain. A normative decision theory for this situation requires considerations of risk attitude and utility functions, as discussed in Chapter 5.

In an important class of cases, however, the $N(t;k,a)$ trajectories are easy to predict. Let $H(t)$ denote the partition of $B$ into classes of individuals whose hazard rate functions (mapping exposure conditions into hazard rates) are the same in year $t$. In particular, the blocks of $H(t)$ are internally homogeneous in terms of their probabilistic fatality responses to acts $a$ in $A$. Denote by $I^*(t) = I(t)^{H(t)}$ the partition such that two individuals are in the same block of $I^*(t)$ if and only if (i) They have the same observable characteristics, and hence are in the same block of $I(t)$; and (ii) They have the same probabilistic response characteristics, and so are in the same block of $H(t)$. Now suppose that $I(t) = I^*(t)$, so that each living person's hazard rate in year $t$ is known on the basis of his observable characteristics, and everyone with the same observable characteristics (e.g., sex, age, race) has the same hazard rate. Then the number of fatalities generated by category $k$ in year $t$, assuming that it is a negligible fraction (e.g., a few percent) of the total number of individuals in cell $k$ in year $t$, will be approximately a Poisson random variable, implying that its mean and variance are approximately the same. If the value weights are the same for all members of the population, then the total value flow in any year $t$ in which $n(t;k)$ people are initially at risk in block $k$ of $I^*(t)$ will be an approximately normal random variable with mean and variance both proportional to the sum over all cells $k$ of $h(k)n(t;k)$, where $h(k)$ is the hazard rate for people in cell $k$ of $I^*(t)$. This sum, which we will denote by $R(t)$, constitutes an adequate numerical index of routine population risk in year $t$, since the normal distribution with proportional mean and variance is a one-parameter family totally ordered by FSD.

To calculate the incremental routine population fatality risks associated with different acts in $A$, let $dh(k;a)$ denote the change in the hazard
rate associated with cell $k$ if act $a$ is implemented instead of the "status quo" or "neutral" act in $A$. Let $dR(a)$ denote the sum over all cells $k$ of $dh(k;a)n(k)$, where $n(k)$ is the equilibrium number of people in cell $k$, given that survivors age deterministically and that age is one of the observable factors used to define cells. (We assume that $n(k)$ is not significantly perturbed by the d.m.'s action.) Then the act $a$ for which the incremental population risk $dR(a)$ is smallest will be most preferred, other things being equal.

We can summarize this discussion as follows: in the special case where the Poisson approximation for fatality arrivals within cells is justified, the parameters of the Poisson arrival processes are known, the value of a life is the same regardless of which cell the life's owner is in, and the system is in steady state demographic equilibrium, then the expected number of lives lost per year is a satisfactory numerical index of routine population fatality risk. If this model holds but the terms $dh(k;a)n(k)$ are themselves uncertain, it will still be known that their sum is approximately normally distributed with equal mean and variance, corresponding to some single point on a CRS calibrated by the normal distributions with equal mean and variance. Since this scale is unique up to choice of unit, probability distributions over it are well-defined, and can be used to represent uncertain $R(t)$ or $dR(a)$ values. Notice that even though the attribute "expected number of fatalities" is an adequate index of population risk when it is known, it would be incorrect to replace probability distributions over this attribute by their expected values. This is an important difference between known and uncertain population risks.

Two generalizations may be made:

- If fatalities occur in the population according to a known Poisson arrival process and if each fatality is of uncertain value, with the values being independently and identically distributed, then the total flow of value per year follows a compound Poisson distribution. By the usual laws for the sum of
a random number of random variables, this distribution will be approximately normal with mean given by the sum over $k$ of $E[N(k)]E(v)$ and with variance equal to the sum over $k$ of $E[N(k)][E(v)]^2$, where $[E(v)]^2$ denotes the mean value of lives. Thus, both mean and variance increase in direct proportion to the expected number of fatalities, $E[N(k)] = h(k)n(k)$, even when the mean and variance of life value are uncertain.

- If the system is not in steady state with respect to the numbers of people in each class, or if the equilibrium distribution is significantly changed by the choice of $a$ (implying a major health impact), then the steady-state value of $dR(a)$ will no longer suffice. Instead, trajectories $dR(t;a)$ must be calculated based on a model of the dynamic adjustment of the population, perhaps using Chapter 7's dynamic expected utility theory for known numerical risk functions over time.

8.1.5 Testing for Homogeneity

The reduction of population risk measurement to evaluation of a simple index, $R(a)$, given by the sum of the products $n(k)h(k;a)$, depends on the simplifying assumption that $h(k;a)$ is the same for all members of class $k$. As a practical matter, there are at least two statistical strategies for testing this key assumption:

- Test whether the variance in the survival fraction for group $k$ in successive years is smaller than would be expected if all members had identical probabilities of death. (Perhaps somewhat counterintuitively, variance around the mean number of deaths is maximized when all individuals have the same probability of death.)

- Test whether the distribution of lives is exponential with parameter $1$ in cumulative hazard. (The cumulative hazard for an
individual who dies in year T is the sum of the age specific hazard rates for him in each year from his time of birth until T.)

Tests for within-cell response homogeneity can be based on either criterion, although the former may be more sensitive in identifying specific cells as ones in which the homogeneity assumption is suspect.

8.2 Catastrophic Population Risks

The routine population risk models discussed so far in this chapter have made heavy use of the assumption of statistical independence in the occurrence of fatalities across individuals. Most of the recent theoretical literature on defining measures of public risk, however (e.g., Keeney, 1980, Fishburn, 1984, Sarin, 1985, Keeney and Winkler, 1985) have concentrated on the very different case in which an accident may kill several people simultaneously, with different subsets of B having different probabilities of being the fatality set. The focus in this literature has been on assessing risk "equity" of different fatality probability measures on the subsets of B, based on the shapes of utility functions over the single attribute "number of fatalities". However, this approach runs into logical difficulties, as pointed out below.

A second approach to "catastrophic" fatality risks has been widely applied in engineering risk analysis since at least the time of the Reactor Safety Study (Rasmussen et al, 1975). This approach, which applies when fatalities are generated by the first arrival in a compound Poisson process, also deals with distributions over the attribute "number of fatalities".

Each approach is reviewed below. We then propose that catastrophic population fatality risks can not be adequately represented in terms of probability distributions over the attribute "number of fatalities". Instead of considering probability distributions and utility functions over this attribute, it is necessary to consider probability (or hazard
rate) measures on subsets of B, and value or utility functions over the set of such measures.

8.2.1 Risk Curves for the Compound Poisson Process

A. The Classical Risk Curve

Suppose that the d.m.'s choice of an act a in A imposes population fatality risks on some target population B, as follows: accidents arrive according to a Poisson arrival process with intensity \( h(a) \), and if an accident occurs, then the cumulative probability distribution function (c.d.f.) for the number of people killed is \( F(.;a) \), which we abbreviate as \( F(a) \). For example, choosing act a (e.g., to operate a certain type of facility in a certain location) might set up a competing-risk process in which different types of accidents can occur, each with a distribution over the resulting number of fatalities. Then the total arrival rate \( h(a) \) is the sum of the accident rates for each type of accident, and the probability \( f(x;a) \) that the first accident to occur will kill \( x \) people is the sum over all accident types of the probability that it will occur first (the ratio of its arrival rate to the sum of the arrival rates) times the probability that it will kill \( x \) people if it does occur. The d.m.'s problem is to choose among alternative pairs \([h(a), F(a)]\). This model is often applied to assess the fatality risks to neighbors of a hazardous facility or transportation route from catastrophic accident events such as fires, explosions, or release of hazardous gases or vapors.

The pair \([h(a), F(a)]\) can be conveniently represented as a "risk curve" in the following way. Let \( x \) denote the event that exactly \( x \) people are killed when an accident arrives. Then each event \( x \) follows a statistically independent Poisson arrival process with intensity \( h(a)f(x;a) \), where \( f(x;a) = F(x;a) - F(x - 1;a) \) is the probability mass put on \( x \) by \( F(a) \). (See Feller, Vol. 1, 1968, p. 292 for a discussion of the mutual statistical independence and Poisson distributions of the numbers of occurrences of each value of \( x \) when the underlying compound Poisson process is allowed to operate over time.) The product \( p(x;a) = \)
h(a)f(x;a) giving the mean arrival rate of marker event x is a satisfactory indicator of the risk of this event, and a plot of p(x;a) against x for x = 0, 1, 2, ... N, where N is the total number of people in B shows visually how relatively likely it is that each x will be the marker event that actually occurs. Finally, let X denote the event that the first arrival kills at least X people. Then plotting p(X;a) against x, where p(X;a) is the sum over all x' greater than or equal to X of p(x';a), gives the classical "risk curve" for catastrophic accident fatalities (Thompson, 1984.)

B. Critique

There are several difficulties with the plot of p(X;a) as a representation of risk. One is that presenting it in cumulative form renders the interpretation of each point on it ambiguous unless the right-hand tail from that point on is shown. For example, knowing that the probability of an accident creating "four or more fatalities" arriving within the next year (assuming that h(a) is measured in discrete time) is 0.01 does not by itself provide any information about whether the likely number of fatalities in this event is five or five hundred. For this information, it is necessary to comprehend the curve to the right of x = 4 as a whole. Plotting the p(x;a) values directly makes it possible to interpret each point on it independently of the rest of the curve; moreover, if p(x;0) is shown, it will put the rest of the curve into perspective by showing the probability of no accident. It seems at least plausible that this format may be psychologically easier to understand and use than the (complementary) cumulative form.

A second difficulty is that the risk curve display can not be generalized to fatality-generating accident processes other than the compound Poisson process. If the joint distribution function of time to arrival and number of people killed is some arbitrary \( F(T,X) \), for example, then the marker events x will not be mutually statistically independent and their marginal arrival (i.e., hazard) rates will necessarily evolve over time to incorporate the information that no accident has yet occurred, as
explained in Chapter 7. Thus, the plot of the marginal probabilities \( p(x|a) \) against \( x \) will necessarily vary nonlinearly with the passage of time unless the compound Poisson fatality-generating process is assumed. (See e.g., Feller, op cit, p. 290.) An alternative representation for this case would be to show separately \( h(t) \), the hazard rate for the arrival of an accident, and \( f(x|a,t) \), the conditional probability distribution over number of fatalities given an accident at time \( t \).

A final, more fundamental, criticism is that the attribute "number of fatalities" is not an adequate description of consequences. The consequence set of interest is not simply the number of people who end up being killed, but the distribution of risk in the population before an accident occurs. It is only this risk, rather than its realization in terms of fatalities, that the d.m. can directly control. The distinction is clarified in the following example.

Example 8.3: Population Risk vs. Number of Fatalities

A d.m. must choose between two acts that impose risks on a three-person target population \( B = \{1, 2, 3\} \), where the numbers index individuals in \( B \). The two acts are as follows:

- Act \( a' \) assigns the following probabilities to subsets of \( B \): 0.5 to \( \{1\} \), and 0.5 to \( \{1, 2\} \); exactly one of these events will occur.

- Act \( a'' \) assigns the following probabilities to subsets of \( B \): 0.5 to \( \{1, 2\} \) and 0.5 to \( \{3\} \); exactly one of these events will occur.

Thus, under \( a' \), the marginal fatality probabilities for individuals 1, 2, and 3 are \((1, 0.5, 0)\), while under \( a'' \) they are \((0.5, 0.5, 0.5)\). The latter is clearly more "equitable" than the former, at least in a literal sense. Yet the c.d.f. over the attribute "number of fatalities" is identical in both cases: in each case, there is a 50-50 chance of 1 or 2 fatalities. If the probabilities are generated by appropriate underlying compound Poisson processes, the risk curves for the two processes will be
identical. In summary, the risk curve treats individuals as "exchangeable", in the sense that sums of fatalities are treated as sufficient representations of consequences without specifying which individuals are included in the sums.

8.2.2 Risk Equity

A. Some Previous Approaches

To deal with this type of difficulty, Keeney (1980), followed by Sarin (1985) and Fishburn (1984), have examined the application of single-attribute and multi-attribute utility theory to the measurement of population fatality risks. Keeney interpreted individual probabilities of death as "attributes", and applied independence axioms from multiattribute utility theory to derive utility functions over the single composite attribute "number of fatalities." He then interpreted preferences for risk equity in terms of the shape (convexity) of utility functions over this attribute. Sarin considered the numerical measurement of equity as one component of a "risk" utility function, and this approach has been continued and refined in Keeney and Winkler (1985).

B. Critique

From our perspective, approaches to risk measurement that use number of fatalities ex post and static individual fatality probabilities (or lotteries over such probabilities) ex ante as summaries of consequences are fundamentally flawed. In the theories of individual risk developed in Chapter 7, the experience of being at risk over time is a crucial component of consequence. Instead of taking binary indicators of individual fatalities and ex ante probability distributions over them as the basic ingredients of risk, therefore, we propose that individual lifetimes and probability distributions or hazard rates over them would be more satisfactory. Using joint distributions of \([T(1),...,T(N)]\), where \(T(i)\) is the random length of individual \(i\)'s life, as the basic objects to which risk or preference numbers are to be assigned is also compatible with the multiattribute approach.
Before pursuing this new approach further, however, we shall present two examples that highlight some of the difficulties with the "risk equity" approaches mentioned above.

Example 8.4: Non-Existence of a Utility Function for Numbers of Lives Lost

The following example is adapted from a discussion by Keeney and Winkler (1985) of a "compatibility theorem" given by Fishburn (1984). Suppose that the target population B consists of two individuals, B = [1,2], and that the d.m. must choose between the following acts:

- a' assigns a probability of 1 to the death of individual 1.
- a" assigns a probability of 1 to the death of individual 2.
- a''' assigns a probability of 0.5 to the death of individual 1 and a probability of 0.5 to the death of individual 2: exactly one of these events will occur.
- a* chooses act a' with probability 50% and chooses act a" with probability 50%.

If an N-M utility function exists over the attribute "number of fatalities", then the utility of acts a' and a" will be the same. Hence, act a* must have the same expected utility as each of them, since it is a probabilistic mixture of them. Finally, a* must have the same expected utility as a''', since both give a 50% chance of the consequence in which 1 dies and a 50% chance of the consequence in which 2 dies, and therefore must have the same expected utility. Thus, a preference for act a''' or act a* over act a' or a" is inconsistent with the hypothesis that an N-M utility function over "number of fatalities" exists. In other words, a preference for "risk equity" is incompatible with the existence of a utility function over this attribute.
Keeney and Winkler conclude from this that "Basically, ex ante equity is a characteristic associated with the decision strategy that should not change as uncertainties are resolved over time." Based on Chapter 7's analysis, however, the opposite conclusion seems more appropriate. The difficulty here is very similar to the one in Example 7.9, and occurs because expected utility theory implies that a d.m. who is indifferent between two consequences must be indifferent among all probabilistic mixtures of those consequences. If the condition of being at risk is itself taken to be part of the consequence, however, then this implication must be refined: uncertainties that resolve at different times are parts of different consequences. From this perspective the resolution of the above counterexample is simple: act $a^*$ will in general not be indifferent to acts $a'$ and $a''$, and will have different values depending on how soon it is resolved. Only if it is resolved at the moment the marker event (death of 1 or 2) arrives will it be indifferent to $a''$. Earlier resolution implies that one of the individuals will live under the shadow of certain (as opposed to probable) death for a longer time, and this is a different consequence than would be implied by later resolution.

Example 8.5: Fatality Risk Hypergraphs

Example 8.3 showed that the risk curve $p(x)$ giving the probability of $x$ or more fatalities for $x = 0, 1, 2, \ldots, N$ failed to distinguish between a case in which one person was certain to die and another sure to live, and a case in which each had a 50% chance of living. (In each case, there was a third person with a 50% chance of living.) Fishburn (1984, p. 907) cites a proposal that would avoid this problem: individual risks would be summarized by an $N$-vector of unconditional individual mortality probabilities, $[p_1, p_2, \ldots, p_N]$, while population risk would be summarized by a risk curve (or actually the underlying p.d.f., $p(x)$, $x = 0, 1, 2, \ldots, N$, where $p(x)$ is the probability of exactly $x$ fatalities.) Thus, in Example 8.3, the two risks would differ in that the individual risk vectors $(0.5, 0.5, 0)$ and $(1, 0.5, 0)$ would be distinct, even though the risk profiles would be the same.
Fishburn also correctly mentions that different probability measures on subsets of $B$ can induce identical risk curves and individual risk vectors. We shall illustrate this possibility with a simple example.

Consider a Poisson competing risk model (see Chapter 7) in which one of several different accident types (or accidents, for short) will be the first to occur. The probability of each type of accident being the one that occurs first can be calculated as usual as the ratio of its hazard rate to the sum of the hazard rates of all types. Associated with each accident is the subset of people that it will kill. This information can be represented in a Boolean matrix having a 1 in row $i$ and column $j$ if and only if individual $j$ is included in the fatality set for accident $i$ (i.e., if and only if accident $i$ will kill person $j$), and having zeros elsewhere. Such a structure is sometimes called a "hypergraph" in the combinatorial literature, since it represents the relations of belonging to a common fatality set among individuals for several different fatality sets. We shall call it the fatality hypergraph matrix, or simply the fatality matrix, representing a catastrophic risk. To complete the specification of problem data, a nonnegative column vector with components summing to 1 may be adjoined to the left-hand side of the fatality matrix. Each component in this left-most column gives the probability that the corresponding accident will be the first to occur; if it does occur, then the fatality set indicated in the rest of that row (i.e., the set of individuals with ones in that row) will die. We will call this data structure a "fatality risk hypergraph matrix", or simply a "fatality risk matrix." Models that take the set of fatality matrices, rather than the set of fatality risk matrices, as the consequence set produce the type of difficulty raised in Example 8.4.

An example of a fatality risk matrix is $D1$:

<table>
<thead>
<tr>
<th>Individuals</th>
<th>( i )</th>
<th>( \Pr(i) )</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ D1 = \begin{pmatrix} 0.3 & 0.7 & 0.7 \end{pmatrix} \]
In the fatality risk hypergraph represented by Dl, there are three individuals at risk (labelled 1 to 3 across the column headings), and three accidents competing to happen. The first accident has a 0.3 probability of being the one to arrive first. If it occurs, it will kill individuals 1 and 2. The second accident has a 0.4 probability of occurring, and will kill individuals 2 and 3. The third accident has probability 0.3 and a fatality set consisting of individual 3 alone. The individual risk vector, \( p = (0.3, 0.7, 0.7) \) is shown beneath Dl. It is found by multiplying the transpose of the accident probability column by the remaining fatality matrix. In other words, the unconditional individual risk, or fatality probability, for any individual \( i \) is the sum of the occurrence probabilities of the fatality sets (rows of the fatality matrix) that include \( i \) as a member. The risk curve \( p(x) \) for Dl evidently puts a 0.7 probability on two fatalities and a 0.3 probability on one fatality. It is found by assigning to each number \( x \) the sum of the probabilities of fatality sets that contain exactly \( x \) fatalities.

Although the fatality risk hypergraph can be generalized, e.g., by replacing the zeros and ones in the fatality matrix with statistically independent conditional probabilities of individual deaths given each accident, the present form is sufficient to demonstrate that essentially different (i.e., non-isomorphic) fatality risk hypergraphs can produce the same individual risk vector and risk curve. For example, the two following fatality risk matrices each give the same individual risk vector and risk curve as Dl:

\[
\begin{array}{ccc}
    & 1 & 2 & 3 \\
\hline
    1 & 0.3: & 1 & 0 & 0 \\
\end{array}
\]

\[
\begin{array}{ccc}
    i & \text{Pr}(i) & 1 & 2 & 3 \\
\hline
    2 & 0.7 & 0 & 1 & 1 \\
\end{array}
\]

\[
\begin{array}{ccc}
    & 0.3 & 0.7 & 0.7 \\
\hline
\end{array}
\]

and
We can think of $D_1$ as being derived from $D_2$ by transferring a probability mass of 0.4 from individual 1 to individual 3. Thus, individual 1 goes from having an individual risk probability of 0.7 in $D_2$ to a probability of 0.3 in $D_1$, while individual 3 goes from having a probability of 0.3 to having a probability of 0.7. Individual 2 has a 0.7 fatality probability in both cases. By symmetry, this transfer might be considered an "equity-preserving" transformation, since in effect individuals 1 and 3 have just traded places as far as individual fatality probabilities are concerned. (Clearly, this argument is unaffected if individual fatality probabilities are expressed in morts rather than probabilities, via the transformation $r(p) = -\ln(1 - p)$, which we have proposed in place of probabilities as a natural unit of individual risk.) It remains an open question (Fishburn, 1984, p. 907) whether all fatality risk hypergraphs that have the same risk curve and individual fatality probabilities are derivable from each other by sequences of equity-preserving transformations.

When individuals are allowed to have different life values, the question of risk "equity" becomes moot. For example, $D_3$ can be thought of as derived from $D_1$ by having each individual redistribute his own fatality risk differently among accident scenarios. If the death of subset $[1,2]$ is preferable to the death of subset $[2,3]$, e.g., because $v(1)$ is greater than $v(3)$ or because of affiliations among the three individuals, then $D_3$ may be dispreferred to $D_1$, even though they are equivalent in terms of risk curves and unconditional individual fatality probabilities. Thus, preferences among fatality risk matrices will depend on more than just

<table>
<thead>
<tr>
<th>Individuals</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Pr(i)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

$D_3 = 0.3\ 0.7\ 0.7$
the matrices, but also on the identities of the individuals in different fatality sets. Thus, the considerations of symmetry inherent in proposed definitions of risk equity (see e.g., the Fishburn, Sarin, and Keeney references already cited) are no longer relevant for choosing among alternative fatality risk hypergraphs.

8.2.3 Limited-Precision Risk Measurement for Catastrophic Fatality Risks

A. Risk Equity Scales

The preceding examples suggest the mathematical complexity of the general problem of evaluating catastrophic fatality risks. The basic objects of preference are (or at least include) fatality risk hypergraphs, and the number of essentially distinct hypergraphs increases explosively with the size of the population, even if all individuals are valued equally. (In this case, two fatality risk matrices are essentially distinct, i.e., nonisomorphic, if and only if it is impossible to transform one into the other by a sequence of permutations of its rows and columns.) In this section, we apply the analytic approach of previous chapters to try to reduce the problem of risk measurement for catastrophic fatality risks to manageable proportions.

Accordingly, suppose that a target population of size \( N \) is put at risk by an activity, such as operation of a hazardous facility. We shall use the population renewal assumption to keep this population stationary, and shall assume a competing-risk accident process, as in the development of fatality risk hypergraphs. Let \( h(t;i) \) denote the marginal hazard rate for the \( i \)th accident at time \( t \), conditioned on survival of the facility to date, and let \( S(t;i) \) denote the fatality set at time \( t \) for accident \( i \). Thus, if accident \( i \) arrives at time \( t \), which happens with probability \( h(t;i)/h(t) \), where \( h(t) \) is the sum over all accidents \( j \) of \( h(t;j) \), then the people in set \( S(t;i) \) will die. The problem is to develop a measure of risk for the hypergraph trajectory defined by \([h(t;i),S(t;i)], i = 1, 2, ..., K\), where \( K \) is the number of different accidents.
We will focus on the problem of measuring the "risk" of this multivariate trajectory at a single point in time, \( t \). In addition, we will initially make the simplifying assumption that all individuals are valued equally. Then following our usual strategy, we seek a continuous one-parameter family of hypergraphs \( [h(i), S(i)], \ i = 1, \ldots, K \), totally ordered by some form of dominance, to serve as a numerical risk scale. In addition, we need a simple way of taking attitudes toward risk equity into account. To this end, we define two canonical risk scales, representing extremes in terms of risk equity.

- The first scale is indexed by a parameter \( r' \), running from 0 to positive infinity. Any given value of \( r' \) represents a risk in which (i) Accidents arrive according to a Poisson process with average arrival intensity \( r' \); (ii) If an accident occurs, the number of fatalities follows a Poisson distribution, possibly truncated at some large number \( N \), with parameter \( r'N' \), where \( N' \) is some conveniently chosen scaling constant; and (iii) Individuals in the target population have been numbered from 1 to \( N \). If there are \( x \) fatalities, then individuals 1 to \( x \) die.

- The second scale is indexed by a parameter \( r'' \), representing a family of risks in which (i) Accidents arrive with intensity \( r'' \); (ii) The number of individuals killed if an accident occurs is a Poisson random variable with parameter \( r''N' \), where \( N' \) is the scaling constant chosen above; and (iii) If there are \( x \) fatalities, then each subset of \( x \) people is equally likely to die. In other words, the fatality set of size \( x \) is randomly selected from the population.

On each scale, preferences are monotonically and continuously decreasing, i.e., risk is increasing. The first scale is strictly less equitable than the second, however, in that risk is concentrated more on some individuals than on others in the first case: individual 1, for example, will necessarily die if anyone at all dies. Each successive individual belongs to one fewer fatality sets than the preceding individual. On the
second scale, by contrast, risk is always equitably distributed, no matter how large its magnitude is. (An interesting deterministic alternative for this scale, if \( N \) is effectively infinite, is to number the individuals from 1 to \( N \) and to have individuals \( 1 + x(x - 1)/2 \) to \( 1 + x(x + 1)/2 \) die if the total number of fatalities is \( x \). Then each individual belongs to exactly one fatality set. However, we will not explore this possibility here.)

Given the two scales indexed by \( r' \) and \( r'' \), we assume that a d.m.'s attitude toward risk equity can be represented by a continuous, monotonically increasing function, \( f(.) \), starting at the origin, mapping each value of \( r'' \) into an "equivalent" value of \( r' \). This "equity indifference curve" presumably lies everywhere above the 45 degree line, and indicates the expected number of inequitably distributed risks that are considered indifferent to any specified expected number of equitably distributed risks.

Either of the two canonical risk scales can be used to measure the approximate risk of a fatality risk hypergraph \( [h(i), S(i)] \), \( i = 1, 2, \ldots, K \), once the risk indifference curve to be used has been specified. The procedure is as follows:

- Let the risk curve corresponding to the hypergraph be denoted by \( p(x) \). Use FSD to obtain bounds for \( p(x) \) on each scale, assuming that the distribution of risk for the hypergraph is the same as the distribution (equitable or inequitable) for the scale on which risk is being measured.

- Use the equity indifference curve \( f(.) \) to map the endpoints bounding \( p(x) \) on each scale onto corresponding endpoints on the other scale. Thus, each scale will now have four points on it: two from the initial FSD bounds, and two from the image under \( f \) or \( f\)-inverse of the two FSD endpoints on the other scale.

- Take the extreme pair of endpoints on each scale as the final bounds for the risk value of the hypergraph on that scale.
If the risk equity indifference curve is unknown, then plotting the upper point from the equitable scale $r^*$ on the inequitable scale $r'$ will give an upper bound for the risk of the hypergraph on that scale.

B. Evaluation

The analytic strategy outlined here for measuring catastrophic fatality risks can be refined in various ways. For example, if the value of a fatality set is the sum of the values of its members and if each individual has an intrinsic value, as in Theorem 8.2, then the individuals may be numbered in order of decreasing value for the inequitable risk scale, so that the most risk is concentrated on the most valuable individuals. However, there are more basic issues to be considered:

- On the theoretical side, the proposed canonical risk scales for hypergraphs is only applicable to cases where there are no certain fatalities in the event of an accident. By using the family of Poisson distributions over fatalities to calibrate the risk scale, we have restricted the method to cases where the c.d.f. over fatalities given an accident is bounded in the sense of FSD by the c.d.f.'s for Poisson distributions.

- On the practical side, it is not clear how useful the bounds given by the Poisson family under FSD will be in distinguishing among the risks of different hypergraphs. If the bounds are too wide, they will overlap for most or all of the alternatives of interest, and thus give inconclusive results.

- No work has yet been done on whether and how coherent, consistent equity indifference curves can be elicited and used. However, the above construction suggests that attitude toward equity can not in general be represented by a numerical index, but requires at least a function or curve, $f$. 

244
The first two difficulties, raised by the choice of the Poisson family can perhaps be addressed in applications by constructing other one-parameter families (e.g., binomial distributions with an added number of certain fatalities, all increasingly monotonically in the risk parameter) that give sharper bounds to the particular risks being measured. However, complicated families quickly lose the intuitive interpretation which is one of the chief values of canonical risk scales. The problem of how to elicit risk equity attitudes, or even of determining whether coherent risk attitudes exist, will require empirical research.

8.3 Summary and Conclusions

The two main problems of population risk measurement considered in this chapter have very different mathematical structures, are developed to different degrees, and differ in their implications for current risk analysis practice. The theory developed for routine (i.e., statistically independent) population risks in large populations is comparatively straightforward. Its principal conclusion is that under certain conditions, expected values, e.g., expected numbers of fatalities, can serve as adequate indices of population risk. For example, if the number of fatalities per year is a Poisson distributed random variable and all lives are valued equally, then the mean of this distribution makes an adequate measure of risk. When values of lives differ for different individuals, however the d.m. may have to choose among (possibly uncertain) normal distributions over the expected value of lives lost, evaluated on the value scale of Theorem 8.2. No normative decision theory for this case, analogous to the one developed in Chapter 7 for individual risks, has yet been worked out.

The theory of population risk measurement for catastrophic (multifatality) accidents is even less well developed. We have proposed certain conceptual tools -- the fatality risk hypergraph and the risk equity indifference curve -- that help to illuminate the essential combinatorial nature of the problem and to suggest at least one way of
avoiding the associated combinatorial complexity by establishing approximate bounds on risk. However, the area is still ripe for further conceptual investigation.
As outlined in Chapter 1, the essence of decision making in classical decision analysis is choice from a set. The decision maker (d.m.) faces a set A of acts having uncertain consequences and his decision problem is to choose a member from A. This problem is solved indirectly by first representing each act as a probability distribution (over relevant consequences) and then deriving a preference ordering over these probability distributions, and hence over the corresponding acts.

Our work has modified this traditional approach in three essential ways:

- Acts are represented by mathematical objects other than probability distributions. Specific models are reviewed below.
- Consequences are reinterpreted to include time spent "at risk", whether or not the risk event ever occurs. The condition of being at risk as well as any resulting physical harm are treated as parts of the consequences of a risky activity.
- Instead of deriving complete subjective preference orderings for acts, we have concentrated on the "objective" part of those orderings, defined as the most complete partial ordering that all individuals (with preferences satisfying certain assumptions) would agree with. The objective partial ordering of acts is induced by dominance relations such as first order stochastic dominance (FSD) and temporal first order stochastic dominance (FSD*).

The purpose of these changes has been to help develop a foundation for normative decision theories that will be more useful for practical risk management decision and communication problems than the classical Von-Neumann Morgenstern Utility Model. Each of these changes is reviewed below.
9.1 Risk Models and Decision Models

In practical applications ranging from systems safety and reliability engineering to epidemiology and dam safety, "risks" arising from human activities are represented by a relatively small set of mathematical models. These models are usually best thought of as stochastic processes in which the arrivals of "marker events" defining the risk of interest are of chief concern. Classical decision theory, emphasizing choice among probability distribution functions over a single attribute (e.g., dollars), has developed largely in the context of business and financial decisions. Health and safety risk management decisions however, require decision theories applicable to choices with consequences represented by stochastic processes. As examples,

- The time until the next occurrence of an earthquake;
- The time (possibly infinite) until a latent cancer in an individual expresses itself as a tumor; and
- The time until the first failure on demand of the auxiliary feedwater system in a nuclear power plant

are all better thought of in terms of (possibly multivariate) stochastic processes characterized by hazard rate trajectories than in terms of static probability distributions.

One goal of our research has been to begin the development of normative decision theories (Chapter 7 and 8) that will apply to such cases. A more modest goal has been to explore the mathematical structure of such risks, to determine how and under what conditions they can be given useful numerical representations. The rest of this section discusses the mathematical structures we have examined. Section 9.2 reviews the interpretations of these models, emphasizing the new questions they raise for risk management that do not appear in "classical" decision models. Section 9.3 concludes with a summary of our results on numerical and nonnumerical representations of risk.
9.1.1 Models of Risk

We have found it useful to enrich the structure of the set $A$ of acts (and the set $C$ of consequences) in the classical decision analysis paradigm so that acts are no longer represented only by probability distribution functions, but by mathematical objects that are better suited to describing health and safety risks. A variety of such objects have been proposed and analyzed in this study. The most important are as follows, in order of increasing generality and putative usefulness.

1. $p = \text{probability of occurrence of a dichotomous consequence (e.g., occurrence of a marker event.)}$ We have proposed (Chapter 5) that risk can be conveniently measured in this situation in units of "morts", related to the probability $p$ (which has no units) by the transformation $r(p) = -\ln(1 - p)$.

2. $F = \text{probability distribution function over a single-attribute consequence interval, } [0,M]$. This is the classical case emphasized in Chapter 3 and treated from a fresh perspective in Chapter 5; it is also the model most commonly studied in the mathematical psychology literature reviewed in Chapter 4.

3. $[p,F] = \text{the "compound" risk in which an accident (or marker event) occurs with probability } p, \text{ and in which the probability distribution function over (single-attribute) consequences if it does occur is given by } F$. Implicit in this formulation is the idea that $F$ is resolved only after $p$. Having to live with the uncertainty represented by $p$ (for some unspecified amount of time) makes the object $[p,F]$ different from the corresponding unconditional distribution over consequences, in which nonoccurrence of an accident is treated as occurrence of a zero consequence. $[p,F]$ is the model axiomatized by Fishburn (1984) in his treatment of the mathematical foundations of risk measurement, reviewed in Chapter 4; it is also treated at the end of Section 5.4.2, where a canonical risk scale appropriate for risks of this form is developed.
4. \( h(t) \) is the hazard rate trajectory for arrival of a dichotomous marker event. This is a dynamic version of model 1, where risk is represented by the probability \( p \) of occurrence of the marker event. \( h(t) \) is already in units of morts per unit time, however. Each trajectory \( h(t) \) corresponds to a family of probability distributions, indexed by \( t \), for the remaining time until arrival of the marker event, conditioned on the amount of time that has already passed. Thus, models of the form \( h(t) \) have a mathematical structure more complicated than any static univariate probability distribution. Moreover, it is necessary to consider preferences evaluated at different points along \( h(t) \) in evaluating the desirability of \( h(t) \) as a whole. Chapter 7 takes some first steps toward developing a theory of preferences over hazard rate trajectories by deriving a dynamic expected utility representation for the discrete-time case.

5. \([h(t;s),f(s)]\) is an "uncertain risk" in which the actual hazard trajectory depends on the (uncertain) value of \( s \), and in which only the probability density function \( f(s) \) for \( s \) is known. Clearly, this includes Model 4 as a special case. Uncertain risks induce corresponding "assessed" hazard trajectories, where the assessed hazard rate at time \( t \) is \( h(t) = E[h(t;s);t] \). Under the assumptions of the theory developed in Chapter 7, assessed hazard trajectories can be evaluated just like deterministic ones using the same dynamic expected utility representation.

6. \( h[t;I(t)] \) is a "stochastic" hazard rate trajectory in which the assessed hazard rate at time \( t \) depends on all information from observations taken up to and including time \( t \). If the observations up through time \( t \) consist solely of the amount of time that has passed with no arrival of the marker event, then this reduces to Model 5. If \( I(t) \) contains additional information, e.g., on the states of monitored components within a complex system whose failure is the marker event of interest, then the (assessed) sample path followed by \( h[t;I(t)] \) up to the arrival of the marker event (if it occurs) may be highly uncertain at the decision date, \( t = 0 \). This distinguishes stochastic hazard rate trajectories from both deterministic and uncertain hazard rate trajectories.
trajectories. However in principle, the dynamic expected utility theory of Chapter 7 applies to this case as well. In practice, the theory can not be applied until some way has been found to assess utility and "linkage" functions from expressed preference data.

7. \( r[t;I(t)] = r[h(t),F(t);I(t)] \) is the stochastic "risk" trajectory followed by a process in which the (assessed) hazard rate for arrival of an accident or marker event at time \( t \) is \( h[t;I(t)] \), and in which the (assessed) probability distribution function over physical consequences if it does arrive at time \( t \) is \( F[t;I(t)] \). This is a dynamic stochastic version of the compound risk represented in Model 3 by objects of the form \([p,F]\). Again, the dynamic expected utility theory of Chapter 7 would be appropriate if the pragmatic preference assessment burden that it imposes could be met.

8. \([h(t;i), i = 1,2,\ldots,N]\) = a set of statistically independent hazard rate trajectories for the remaining times until the deaths of \( N \) individuals indexed by \( i = 1 \) through \( N \). Chapter 8 considers several variations of this model, e.g., to deal with uncertain individual hazard rates, and argues that under certain ethical assumptions, routine population risk should depend only on weighted sums of individual hazard rates.

9. \([h(t;j),S(j)]\) = a set of hazard rates, possibly one for each subset of individuals in a population exposed to a "catastrophic" fatality risk. Here these subsets are indexed by \( j \), and \( h(t;j) \) is the hazard rate at time \( t \) for arrival of an event that will kill subset \( S(j) \) of the population. This model of catastrophic population risk and an approach to reducing its complexity by concentrating on bounds obtained from a "most equitable" and a "least equitable" distribution of risk are explored in the second half of Chapter 8. It differs from the "risk curve" or "risk profile" traditionally used to represent catastrophic fatality risks (modelled by compound Poisson processes) by distinguishing between individuals. For example, two risks that give the same hazard rates and conditional probability distributions over numbers of people
killed need not be indifferent if they impose different individual fatality risks. Model 9 includes models 4 to 8, applied to human fatalities, as special cases.

The items on this list provide an array of models for thinking about and formally describing the risks from human activities. As risk analysts have become more sophisticated about the complexity of risk concepts and more realistic in explicitly recognizing and modeling uncertainties about risk, attention has slowly been shifting down the list. Over the last ten years, models closer to our Model 5 (uncertain risk) have started to supplant ones like Model 3 (known risk) in many policy analyses and risk assessment software systems. Recent theory has touched on Model 9. However, stochastic risk models and models that explicitly recognize population heterogeneity (our Models 7 and 8) are only beginning to be applied in practice.

9.1.2 Decision Theories for Risk Management

The above models provide a relatively small set of mathematical objects that appear to be particularly useful for representing health and safety risks. They are not intended to constitute an exhaustive taxonomy. For example, Chapter 7 presents a counterexample (see discussion following Theorem 7.9) showing that if the time of occurrence of a marker event can be estimated very precisely from observations taken before the event occurs, then the hazard rate for that event need not exist, and none of the models just presented may apply. Similarly, if risk control opportunities are distributed over time, then our basic framework of decision making as a one-time choice from a set is no longer appropriate (except in the trivial case of deterministic optimal control). The preceding models must be generalized to allow for input control trajectories and feedback to the decision maker, who acts as a combined controller and observer implementing some (typically closed loop) strategy. "The risk" from an activity is no longer defined independent of the control strategy in which it is embedded, and may lack any convenient mathematical representation (see e.g., Example 7.9.)
Our claim, however, is that many practical social risk management decision problems, such as deciding whether to allow a hazardous facility to be built and operated at a particular site, whether and how much to tighten emissions standards or passive restraint standards on new automobiles, whether to allow a new drug to be marketed, and so forth, share the following two characteristics:

- The decision must essentially be made at a single point in time;
- Its consequences (e.g., accident or no accident, or changes in population fatality rates, depending on the application) will not be learned until they have occurred.

In such cases, our choice-from-a-set of acts paradigm, as opposed to the optimal control paradigm, is likely to be appropriate. But acts may be represented by any of the models just summarized. It is therefore necessary to extend classical decision theory to apply to choice sets consisting of these objects. This is the essential message of Chapters 7 and 8.

In this work, we have not been primarily concerned with the development of new decision theories, but with identification and mathematical representation of the objects to which such theories will apply. Natural next steps, as suggested in Chapters 7 and 8, will be to develop normative decision theories for these classes of objects, in conjunction with practical statistical tools for identifying and estimating corresponding risk models and preference functions.

9.2 Expression of Risk and Interpretation of Consequences

Part of what goes or should go into a description of risk is a characterization of the potential consequences of the risk event. A second component deals with the probable timing of event arrivals. If the event of interest is an accident, the probable timing of accident
arrivals may be conditioned on the "size" (i.e., the consequences) of the accident, as described in Chapter 8. The two components of consequence and probable timing, traditionally referred to as accident "severity" and "frequency", respectively, are usually assumed to be numerical variables and to jointly determine accident "risk".

Our analysis suggests that neither the consequences nor the probable timing of accidents (or other marker events) can in general be adequately represented by numbers. As a rule, neither set is simple enough to be uniquely indexed by the real numbers, although there are noteworthy exceptions (such as the Poisson arrival process with a dichotomous consequence.) Instead, it is necessary to use objects such as hazard rate trajectories to represent risks over time, and any systematic method for projecting such objects onto a numerical scale is bound to introduce some imprecision in the process.

In the remainder of this section, we summarize our own conclusions about how risks ought to be presented and about how the current frequency-severity paradigm might be modified to more clearly indicate the nature of risks.

9.2.1 "Frequency" and Time-Varying Hazard Rates

The risk models outlined in the preceding section suggest that there is no universal language for communicating risk. Rather, risks can be described and quantified only within the context of an assumed family of models, and different descriptors are appropriate for different models. Perhaps the most important distinction is between models such as \( h(t) \) that explicitly recognize the passage of time, and models such as \( p \) that don't. In describing the cancer risk from a new food additive, for example, analysts and regulators can focus on either the "lifetime" cancer risk generated by the product -- e.g., the increase in the probability that a randomly selected consumer will get cancer before dying -- or on the age-specific cancer risks, defined as the increase in the incidence rate of cancer among different age groups in response to
consumption of the product. A product whose risks are low compared to background according to the first measure may have very high risks in certain age groups according to the second. Such information on the variation of the increase in hazard rate with age can be important in deciding how production (or consumption) of the product should be restricted. Using a single "lifetime" risk figure ignores the distribution of the risk over age groups.

A similar argument applies to the quantitative expression of accident risks for industrial facilities, transportation, dam failures, and so forth. The overwhelming emphasis in risk analysis applications to date has been on looking at "annual" risk or event "frequency", defined as the reciprocal of the mean time between occurrences. As discussed in Chapter 7, this information can be very useful in setting bounds on risk. But more than expected interarrival time should go into the risk estimate. For example, whether $1/E(T)$ is an upper or lower bound will depend on whether hazard rate is increasing or decreasing over time. In the latter case, Theorem 7.4 gives an improved bound on risk that adjusts $1/E(T)$ to take into account the variance of the interarrival time distribution and the current age of the plant. Only in the special case of Poisson arrivals is the unadjusted annual risk $1/E(T)$ an accurate measure of risk.

It is important to realize that incorporation of information about risk over time into a risk estimate or bound need not impose a great information burden. Qualitative properties such as whether risk is increasing or decreasing, whether the age of a unit is greater or less than the average survival time for similar units, and so forth are often sufficient to turn estimated annual risks into useful bounds. Perhaps the main message is that treating annual frequency estimates as point estimates of annual risk (i.e., hazard rate) is apt to be misleading. Treating them as upper or lower bounds (and specifying which they are) will in many cases be more realistic and more useful. Known bounds on risk may allow confident decisions in situations where point estimates with unknown error characteristics would not.
9.2.2 Temporal Resolution of Uncertainty and Ex Ante vs. Ex Post Perspectives

In our view, time enters into the determination of risk not only through time-varying hazard rates, which generalize the usual "frequency" component in traditional definitions of risk, but also through the "consequence" side. Specifically, we have emphasized that the duration of uncertainty about physical outcomes and the condition of being at risk should logically be included as (nonphysical) parts of the "consequences" of risky decisions with delayed resolution of uncertainty about outcomes. Aversion to such uncertainty can affect preferences among decision options with delayed consequences or delayed learning of consequences.

This proposition — that the experience of living with uncertainty is in itself an important potential consequence of risky decisions — is familiar to students of the psychology of risk and risk perception. But it has largely been ignored in expected utility theory, which by construction (through axioms such as "reduction of compound lotteries") applies to instantaneously resolved uncertainties. For example, a conventional multiattribute utility approach to time streams of consequences takes as the basic objects of preference probability measures (or distribution functions) over dated physical outcomes. Any two decisions that give identical probability distribution functions over dated consequences as of the decision date must be indifferent, according to this approach. In risk analysis applications, however, the process by which this uncertainty is resolved over time (if it ever is resolved) is crucial. Identical initial probability measures over physical outcomes that lead to different time streams of information about physical outcomes, e.g., because of different observation opportunities, may be valued quite differently. To capture this element of preferences analytically, we recommended taking the stochastic processes consisting of sequentially resolved probability distributions over future sequences of observations as the basic objects of preference and choice, representing the acts in the d.m.'s choice set. Chapter 7 presents a
discrete-time formalization of this approach, leading to a dynamic generalization of expected utility theory.

Even without this analytic machinery, however, just recognizing extended uncertainty as a consequence suggests some useful perspectives on several open problems in risk analysis. For example,

- It is not sufficient to look at probability distribution functions over numbers of fatalities as a descriptor of the probable consequences of a catastrophic population risk. To capture equity concerns, this "ex post" perspective, which only recognizes physical outcomes, must be supplemented by an "ex ante" perspective that looks at how exposure to risk is distributed in the population.

- In addition, as developed in Chapter 8, the question may arise as to whether an ex ante or an ex post perspective is more appropriate for social decision making when the eventual (ex post) consequences of decisions can be confidently predicted on a statistical basis ex ante. The possibility of divergence between an individual's preference orderings over the same set of acts evaluated at different times does not arise in conventional expected utility theory, where the passage of time is essentially ignored.

- The passage of time in a heterogeneous population in which different members have different hazard rates can induce changes in the age structure of the population over time, making it difficult or impossible to define any meaningful static index of "population risk". (In the special case of statistically independent individual hazard rates with all fatalities equally valued, we have recommended the sum of individual hazard rates over all members of the population as being an appropriate index of instantaneous population risk if certain ethical axioms are accepted for the valuation of statistically independent lives.
However, there are many situations of population risk that cannot be handled by this expedient. For example, if an activity were to expose one or more people to certain death, then the sum-of-hazard-rates index would become infinite, since a certain death is counted as being "infinitely more risky" than any uncertain death on the mort scale."

These issues are important in developing normative risk management decision theories and useful descriptions of individual and population risk. There are various ways of resolving them, e.g., by taking stochastic processes \([h(t);I(t)]\) and population hazard rate trajectories \([h_1(t),...,h_N(t)]\) as the basic objects representing risks, and developing corresponding individual and social normative decision theories for choosing among such objects. Introducing such objects as models for risk naturally raises key issues of health and safety risk analysis that tend to be suppressed in more traditional static formulations.

9.3 Risk Measurement

We have just argued that it is useful to have dynamic mathematical models, such as \([h(t);I(t)]\) or \([h(t),F(t)]\), to represent realistic health and safety risks. However, adopting such objects as the basic objects of preference and choice creates a need for normative decision theories that will apply to choices among them. And practically implementable theories for such complicated objects are not easy to come by. In addition to the theoretical problem of establishing mathematically tractable representation theorems for normatively derived preferences, there is the pragmatic problem that objects of the types we have recommended are intrinsically complex. They therefore make poor stimuli for eliciting primitive preferences. Some way of reducing their complexity is needed.

To this end, we have considered a number of techniques for representing complex mathematical objects by intervals on a numerical scale, where the natural partial ordering of intervals by inequality (rightward intervals
outrank leftward ones) represents the partial ordering of the corresponding objects by "dominance", or "objective" preference. We first established (Chapter 5) that exact numerical representation of objective preference relations is impossible even for the simple family of risk models consisting of univariate distribution functions (c.d.f.'s) over a "consequence severity" axis. In fact, an exact interval order representation in which the relation of dominance among c.d.f.'s is represented by the relation "lies strictly to the right of" among corresponding intervals is also impossible: projecting c.d.f.'s onto intervals will in general result in overlapping intervals being assigned to c.d.f.'s that are separated by dominance.

Despite these limitations, we established a "canonical risk curve" representation for risks of c.d.f.'s that is (arguably) both intuitively appealing for nonspecialists who would be put off by expected utility calculations, and that clearly shows dominance relations and all other decision-relevant information. The canonical risk curve plots a numerical risk measure of risk attitude against a numerical measure of risk, measured on a "canonical risk scale" that has the same interpretation for all observers.

Partial extensions of the canonical risk curve idea to some of the other families of risk models are made in Chapters 5, 7, and 8. Chapter 5 establishes a canonical risk scale for risks of the form \([p,F]\), and Chapter 7 introduces a similar scale for discrete-time dynamic risks of the form \([h(t),F(t)]\). However, no analogue to a numerical measure of risk attitude (e.g., an index of joint time-risk preference) has yet been developed, so that full development of a canonical risk curve for dynamic risks must await additional research. Chapter 8 considers canonical risk scales that incorporate notions of risk equity.

The basic contribution of the canonical risk scale and canonical risk curve ideas is that they illuminate exactly how much of comparative risk judgements can be separated from subjective risk attitudes, and how much is intrinsically subjective. Moreover, they show that the objective
portion of risk in general requires a curve to represent it: simpler mathematical structures can not do the job except in special cases. Extensions of this reasoning show that some lines of current research on risk representation are unlikely to prove successful (unless it turns out that people have unexpectedly homogeneous subjective risk attitudes.)

Finally, the canonical risk scale idea makes use of comparatively weak axioms (continuity, first order stochastic dominance) that are shared by a wide variety of normative decision theories (reviewed in Chapters 3 and 4). These conditions will plausibly be incorporated into whatever normative decision theories are eventually developed for the types of risk models discussed above. Thus, our investigation of the requirements for mathematical representations of risk, the limitations of numerical (or interval-valued) representations, and the minimum complexity needed to represent dominance relations among risk objects is likely to remain applicable as more highly structured representations and more detailed models of risk are developed and explored.

Our results are in this sense very weak, in that they are intended to apply across many risk models and normative decision theories: much stronger results are to be anticipated in special cases. For this reason, our analysis does not in itself contain any dramatic implications for decision making, beyond an awareness of some of the things (duration of uncertainty, ex ante risk equity) that ought to be included in analysis and that are often left out. Rather, we have attempted to clarify the boundaries of the set of risk management decision problems, as distinct from other sorts of decision problems, and to provide a conceptual framework in which useful normative decision theories for these problems can be developed.
REFERENCES


Hintikka, J., and P. Suppes (Eds.), Information and Inference, Reidel, Boston, MA, 1970.


Loewenstein, G., "Expectations and Choices Involving the Timing of Outcomes," Yale University Department of Economics (Undated).


