

A NORMATIVE METHODOLOGY FOR PREDICTING CONSUMER RESPONSE
TO DESIGN DECISIONS: ISSUES, MODELS, THEORY, AND USE

BY

JOHN RICHARD HAUSER

S.B., S.M. (Electrical Engineering), S.M. (Civil Engineering)
Massachusetts Institute of Technology (1973)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF SCIENCE IN
OPERATIONS RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1975

Signature of Author.....*John Richard Hauser*.....

Department of Electrical Engineering and Computer Science, May 21, 1975

Certified by.....*John Richard Hauser*..... Thesis Supervisor

Accepted by.....*John Richard Hauser*.....
Chairman, Departmental Graduate Committee



A NORMATIVE METHODOLOGY FOR MODELING CONSUMER RESPONSE
TO DESIGN DECISIONS: ISSUES, MODELS, THEORY, AND USE

by

JOHN R. HAUSER

Submitted to the Department of Electrical Engineering and Computer Science
on May 19, 1975 in Partial Fulfillment of the Requirements for the Degree
of Doctor of Science in Operations Research at the Massachusetts Institute
of Technology.

ABSTRACT

The design and introduction of innovative products and services determines the fate of many organizations in both the private and public sectors. Such innovation is linked to increased effectiveness and productivity, but often represents high risk to an organization since the success or failure of innovation is dependent upon uncertain consumer response. This dissertation couples quantitative analysis with creative efforts in a methodology to aid managers in the design and implementation of innovation. The methodology produces predictions of consumer response and diagnostics on consumer perceptions, preferences, and choices, and identifies relevant market segments.

Part I discusses the managerial issues, reviews the relevant literature, and structures the methodology. Part II then describes the techniques in detail. Each chapter discusses a functional step in the methodology, i.e., observation and measurement of consumers, identifying and structuring perception, abstracting segments, constructing consumer "utility" functions, predicting individual choice probabilities, and producing aggregate estimates. State of the art techniques from psychometrics, von Neumann-Morgenstern utility theory, and stochastic choice theory are presented for application to the problems. In addition, new theoretic techniques are developed and applied. Together they produce a complete, integrating, usable methodology to facilitate successful innovation. Part III discusses testing the methodology and presents a case study.

The new theory developed is (1) an axiomatic derivation of stochastic preference functions which establishes an isomorphism with von Neumann-Morgenstern prescriptive utility theory to allow its theorems to be applied to descriptive choice, (2) a formal development of an empirical Bayesian probability of choice model to tune rank order effects with the observed "utility" values of the choice alternative, and (3) a unique test for probability of choice models based on information theory and honest reward functions.

The new applications are (1) the use of psychometrics to define performance measures for consumer utilities, (2) feasible measurement techniques for mass assessment of individual specific consumer utility functions, and (3) a direct method to abstract population segments which are homogeneous with respect to preference.

Besides presenting the issues, the theory, and the statistical techniques, the methodology is empirically demonstrated by application to the strategic design of a health maintenance organization for the M.I.F. Health Department.

Thesis supervisor (chairman): John D.C. Little
Professor of Operations Research and
Management

Thesis supervisor (reader): Glen L. Urban
Associate Professor of Management Science

Thesis supervisor (reader): Moshe Ben-Akiva
Assistant Professor of Civil Engineering

ACKNOWLEDGEMENTS

Special thanks to Professors John D.C. Little, Glen L. Urban, and Moshe Ben-Akiva for the creative and provocative comments on this dissertation and for providing me with an invaluable educational experience.

Thanks are also due to the many people who discussed various research questions with me. Among these are Ralph Keeney, Sam Bodily, Nigel Wilson, Frank Koppleman, Charles Manski, Alan Sicherman, and Scott Neslin. Pitu Mirchandani, Gabriel Bitran, Mary McFadden and Florence Hauser provided needed moral support. The artful typing was done by Annie Cooper with a little help from Debbie and Dingo.

Parts of this research were funded by the Sloan School of Management, the Merrill Foundation, the M.I.T. Health Department, Management Sciences for Health, the Urban Mass Transportation Authority, the National Science Foundation and the Army Research Office.

Finally dedications to the Celtics and to my cat, Joubi, who ate many a first draft.

TABLE OF CONTENTS

	<u>Page</u>
PART 1: INTRODUCTION AND OVERVIEW	
CHAPTER 1: INTRODUCTION	17
1.1 Scope of the Methodology	18
1.2 Existing Approaches	19
1.3 Criteria for the Methodology	25
1.4 Topics Covered: Prose Table of Contents	27
CHAPTER 2: DESCRIPTION OF THE METHODOLOGY	30
2.1 Macro Structure of the Methodology	30
2.2 Basic Micro Flow of the Methodology	32
2.3 Summary and Interaction with the Design Process	37
PART II: MICRO DESCRIPTION OF THE METHODOLOGY	
CHAPTER 3: OBSERVATION OF CONSUMERS	39
3.1 General Guidelines for Measurement	39
3.1.1 Respondent Motivation	39
3.1.2 Semantics	40
3.1.3 Pretest	42
3.1.4 Sampling	43
3.2 Choice Alternatives	44
3.2.1 Measurements are Relevant Only with Respect to the Evoked Set	44
3.2.2 Revealed Preference vs. Proxy Choice	45
3.3 Measurement of Input Data	48
3.3.1 Measuring Perceptions of the Attributes of Alternatives	48
3.3.1.1 Properties of the Attribute Set	48

	<u>Page</u>
3.3.1.2 Generation of the Attribute Set	53
3.3.2 Measuring Choice Among the Alternatives	56
3.3.3 Measuring Indifference	56
3.3.4 Demographics and Other Consumer Descriptors	61
3.3.5 Similiarity Measures	61
3.4 Conclusion of Observation of Consumers	62
CHAPTER 4: REDUCTION	63
4.1 The Issues of Reduction	67
4.1.1 Notation	67
4.1.2 Ideal Properties of Reduction: Formal Definitions	68
4.1.3 Reduction in an Engineering Technique: Tradeoffs Must Be Made	73
4.2 Analytic Techniques for Reduction	76
4.2.1 Factor Analysis: An Exploratory Technique to Identify Structure	76
4.2.1.1 Analytic Details of Factor Analysis	77
4.2.1.2 Interpretation of Factor Analysis and Its Use in the Methodology	83
4.2.1.3 Reduction By Factor Analysis: HMO Case	85
4.2.2 Multidimensional Scaling: An Exploratory Tech- nique to Identify the Structure of Similiarity	89
4.2.2.1 Basic Analytic Details of Multidimensional Scaling	90
4.2.2.2 Use of Multidimensional Scaling in the Methodology	91

	<u>Page</u>
4.2.3 Information Theory: A Technique to Select the Most Useful Attributes	93
4.2.3.1 Analytic Details of Calculating Mutual Information	94
4.2.3.2 Use of Information Theoretic Attribute Selection in the Methodology	99
4.2.4 Other Potential Reduction Techniques: An Over- view	100
4.2.4.1 In Depth Utility Assessment	100
4.2.4.2 Professional Judgement	101
4.2.4.3 Null Option	101
4.3 Conclusion of Reduction	101
CHAPTER 5: ABSTRACTION	103
5.1 Criteria for Abstraction	105
5.2 Analytic Techniques for Abstraction	106
5.2.1 Cluster Analysis	106
5.2.2 Automatic Interaction Detection (AID)	110
5.2.3 Discriminant Analysis	119
5.2.4 Other Techniques	121
5.3 Conclusion of Abstraction	122
CHAPTER 6: COMPACTION	124
6.1 Formal Development of Compaction Functions: Defini- tions and their Interpretations	126
6.1.1 Example	126
6.1.2 Compaction Definition	128
6.1.3 Desirable Properties of Compaction Functions	130

	Page
6.1.4 Summary of Definitions	134
6.1.5 Bijection and Monotone Theorems	135
6.2 An Axiomization for Compaction Functions to Establish an Isomorphism with the von Neumann-Morgenstern Utility Axioms	138
6.2.1 Stochastic Preferences	139
6.2.2 The Axioms	142
6.2.3 Interpretations of the Axioms	145
6.2.4 Existence and Uniqueness Theorems	148
6.2.5 Empirical Use of the Axioms Requires Representa- tion of Alternatives as Sets of Performance Measures	150
6.2.6 Independence Assumptions	154
6.2.7 Utility Theorems Identify Unique Functional Forms for Compaction Functions	160
6.2.7.1 Quasi-additive and Multiplicative	166
6.2.7.2 Additive Representations	170
6.2.7.3 Between Multiplicative and Quasi-additive: Aggregability	173
6.2.8 Summary of Axiomization	177
6.3 Examples of Statistical Compaction Techniques	178
6.3.1 Expectancy Value Models	179
6.3.2 Discriminant Analysis	180
6.3.3 Random Utility Models	182
6.3.4 Conjoint Analysis	184
6.3.5 Maximum Score	186

	Page
6.3.6 Preference Regression	187
6.3.7 PREFMAP	187
6.3.8 Summary of Statistical Compaction Techniques	189
6.4 Compaction Techniques as Applied to the Design of a New HMO	189
6.4.1 Regression of Preferences vs. Performance Measures	191
6.4.2 Direct Population Assessment via Personal Interview	193
6.4.2.1 Experimental Design	194
6.4.2.2 Measurement Instrument and the Necessary Mathematics	195
6.4.2.3 Empirical Results	202
6.5 Summary of Compaction	208
CHAPTER 7: PROBABILITY OF CHOICE	211
7.1 Formal Development	214
7.1.1 Important Issues for Probability Models	214
7.1.2 Dependence on the Choice Set	219
7.1.3 Perceptions are Usage Dependent	224
7.1.4 Revealed Preference vs. Proxy Choice	225
7.2 Utility Maximizing Models	232
7.2.1 Explicit Modeling of Error Sources	232
7.2.2 Random Utility Models	235
7.2.2.1 Extended Probit Model	236
7.2.2.2 Multinomial Logit Model	236
7.2.2.3 Discussion and Properties	237

	Page
7.2.3 Aggregate Utility Maximizing Models	239
7.2.4 Conclusion of Utility Maximizing Models	239
7.3 Empirical Bayesian Model	240
7.3.1 Derivation of the Empirical Bayesian Model	241
7.3.2 Desirable Properties of the Empirical Bayesian Model	245
7.3.3 Heuristics For Transforming the Ranked Compaction Vector	252
7.3.4 Two Alternative Logit Equation as an Empirical Bayesian Model	255
7.3.5 An Empirical Example: Aerosol Deodorants	257
7.3.5.1 Calibration	258
7.3.5.2 Comments on Empirical Example	265
7.4 Explicit Modeling of Similarities Among Alternatives	267
7.5 Simultaneous, Independent, and Sequential Choice	268
7.5.1 Simultaneous Models	269
7.5.2 Independent Models	270
7.5.3 Sequential Models	271
7.5.4 Quasi-separable Models	272
7.6 Stability Over Time	274
7.7 Conclusion of Probability of Choice	277
CHAPTER 8: AGGREGATION	279
8.1 Combining Individual Choice Probabilities to Predict Group Response	281

	<u>Page</u>
8.2 Prediction by Changing Attributes, Performance Measures, or Preference Parameters	283
8.3 Correction for the Evoked Set	285
8.4 Dynamics: Trial, Repeat, and Frequency	288
8.5 Empirical Example: HMO Study	294
8.6 Conclusion of Aggregation	297
CHAPTER 9: INTERACTION WITH THE DESIGN PROCESS	298
9.1 Evaluation/Simulation Mode of Interaction	300
9.1.1 Process of Design and Strategy Changes	300
9.1.2 Data Available for Evaluation	302
9.1.3 Models for Evaluation	303
9.1.4 Summary of Evaluation	308
9.2 Refinement	308
9.2.1 Refinement is Guided by the Individual Choice Models	308
9.2.2 Diagnostics Available from the Individual Choice Models	309
9.3 Conclusion of Interaction with the Design Process	311
PART 3: TESTING AND CONCLUSION	
CHAPTER 10: FORMAL TESTS	313
10.1 Aggregate Tests	313
10.1.1 Rank Order Recovery Due to Compaction	313
10.1.2 Tests on Predicted Share	316
10.2 Disaggregate Tests	318
10.2.1 Traditional Tests	318

	Page
10.2.2 Criteria for Disaggregate Tests	321
10.2.3 Formal Test: Theory	323
10.2.4 Formal Test: Use	327
10.3 Conclusion of Aggregation	330
CHAPTER 11: CASE STUDY: THE DESIGN OF A HEALTH MAINTENANCE ORGANIZATION	331
11.1 Summary of Empirical Experience	331
11.2 Managerial Issues	338
CHAPTER 12: CONCLUSION	341
12.1 Summary	341
12.2 Contributions of the Research	342
12.3 Suggestions for Future Research	344
References	348
Footnotes	360
Appendices	
A1: First Questionnaire: Mailed Survey	363
A2: Second Questionnaire: Personal Interview Survey	384
Biographical Note	411

LIST OF FIGURES

	<u>Page</u>
1.1 Process for Development of New Products and Services	19
2.1 Macro Description of the Methodology	31
2.2 Micro Flow of the Methodology	33
3.1 Sample Telephone Call	41
3.2 Concept Description	46
3.3 Proxy Choice as List of Performance Measures	47
3.4 Observation of Consumers	49
3.5 Spanning Set of Attributes	51
3.6 Measuring Perceptions	55
3.7 Measuring Preference and Intent	57
3.8 Schematic of Tradeoff Question	58
3.9 Schematic of Risk Aversion Assessment Question	60
4.1 Relationship of Reduction to the Methodology	64
4.2 Hypothetical Perceptual Space for Transportation Modes	66
4.3 Example of a Disjoint Reduction	73
4.4 Common and Unique Factors	78
4.5 Rotation of Common Factors	81
4.6 Structure Matrix	84
4.7 Factor Loading for Health Care Plans	86
4.8 Perceptual Space for Health Care Plans	88
5.1 Relationship of Abstraction to the Methodology	104
5.2 Hypothetical Cluster Analysis	107
5.3 Automatic Interaction Detection (AID)	111
5.4 Possible Interaction	112

	Page
5.5 Representation of AID Splits	114
5.6 AID Analysis for MIT HMO - Dependent Variable Preference	118
6.1 Relationship of Compaction to the Methodology	125
6.2 Lottery Definition	141
6.3 Schematic of Ordering and Combining Axiom	143
6.4 Schematic of Algebra of Combining Axiom	144
6.5 Schematic of Uniqueness Proof	149
6.6 Two Types of Measurement	154
6.7 Utility Independence	156
6.8 Conditional Utility Lottery	163
6.9 Constant Risk Aversion Lottery	164
6.10 Corner Point Question	167
6.11 Marginality Assumption	171
6.12 Indifference Curves	174
6.13 Direct Perceptions vs. Factor Scores	196
6.14 Schematic of Risk Aversion Assessment Question	198
6.15 Schematic of Tradeoff Question	200
6.16 Indifference Curves for Health Care	207
7.1 Relationship of Probability of Choice to the Methodology	212
7.2 Trial and Repeat	226
7.3 Hyperplane Counter Example	228
7.4 Cell Structure	249
7.5 Histograms of Chip Distribution	260
7.6 Smoothed Histograms of Chip Distribution	261
7.7 Smoothed Histograms of Marginal Distribution	262
7.8 Quasi-Separability: Restriction and Simultaneous Models	273

	Page
8.1 Relationship of Aggregation to the Methodology	280
8.2 Change in the Evoked Set	287
8.3 Dynamics - Macro Flow Model	290
8.4 Dynamics - Usage Over Time	294
8.5 Forecast of MIT HMO Enrollment	296
9.1 Interaction with the Design Process	299
9.2 Decision Tree	304
9.3 GO Fractile	306
9.4 Decision Quadrant	307
9.5 Market Share Sensitivity	312
10.1 Schematic of the Information Test	329
11.1 Alternative Empirical Analysis of the HMO Data	332

LIST OF TABLES

	<u>Page</u>
4.1 Discretized Distribution of an Attribute	95
4.2 Ranking by First Order Mutual Information	96
4.3 Ranking by Second Order Mutual Information	98
6.1 Comparison of Statistical Compaction Techniques	190
6.2 Compaction by Preference Regressions for MIT HMO Case	191
6.3 Rank Order Recovery Table for Preference Regression for MIT HMO Case	192
6.4 Summary Statistics for Direct Assessment of Compaction Functions for MIT HMO Case	204
6.5 Assumption Testing	205
6.6 Rank Order Recovery Table for Direct Assessment of Com- paction Functions for MIT HMO Case	208
7.1 Evoked Set Size and Composition	220
7.2 Average Utility Counterexample	230
7.3 Calibration of Bayesian Model	265
8.1 Dynamics - Macro Flow Model	291
10.1 Rank Order Recovery with Statistical Model	315
10.2 Rank Order Recovery with Random Model	316

Chapter 1

INTRODUCTION

An important problem faced by both the private and the public sector is how to design and introduce innovative products and services. To answer this question it is necessary to know how consumers will respond to the innovation. Consider the following examples:

Transportation

Beset by declining transit ridership and increasing costs a community transportation authority wishes to experiment with drastic service changes. One possible innovation is a computer controlled, dynamically routed minibus system called Dial-a-Ride which provides door-to-door service on demand, but results in wait and travel delays and in shared rides. The authority wants to know how consumers will respond to this innovation and how to design and implement the system based on consumers' perceptions and preference.

Health

A university medical department wishes to provide low cost, high quality medical service to its community. One possible option is to form a group practice, called a health maintenance organization (HMO), which provides complete health care (medicine, surgery, X-rays, etc.) for a fixed monthly fee. To design such a service, the medical department needs to know how consumers' perceive health care, what their preferences

are for the various attributes of health care, and whether everyone has the same preferences. Also for a given plan and price they want to know how many consumers will join.

Package Goods and Other Consumer Products

Deoderants, detergents, shampoo, razor blades, antacids, cake mixes, coffee, tissues, frozen foods, and cereals are all sold in a competitive market place. A firm wants a good indication of consumer response before introducing a new product, but an even more important desire is the creative generation of high potential product ideas. For example, approximately 50% of the growth in sales over a five year period in many industries were accounted for by new products (Booz, Allen, and Hamilton [13]), but new products represent high risk. Approximately 33% of new products fail and over 70% of the resources devoted to new product development are allocated to failures (Booz, Allen, and Hamilton [13]).

Financial institutions, counseling agencies, tourist services, computer hardware manufacturers all share a common need. The need to develop new and existing products and services and the need to understand and predict consumer response to these innovations. This dissertation presents a normative methodology to guide creativity in the development of innovation and to predict consumer response to such innovation.

1.1 Scope of the Methodology

Figure 1.1 is a simple representation of the new product and service innovation process (Hatch and Urban [53]). The first step,

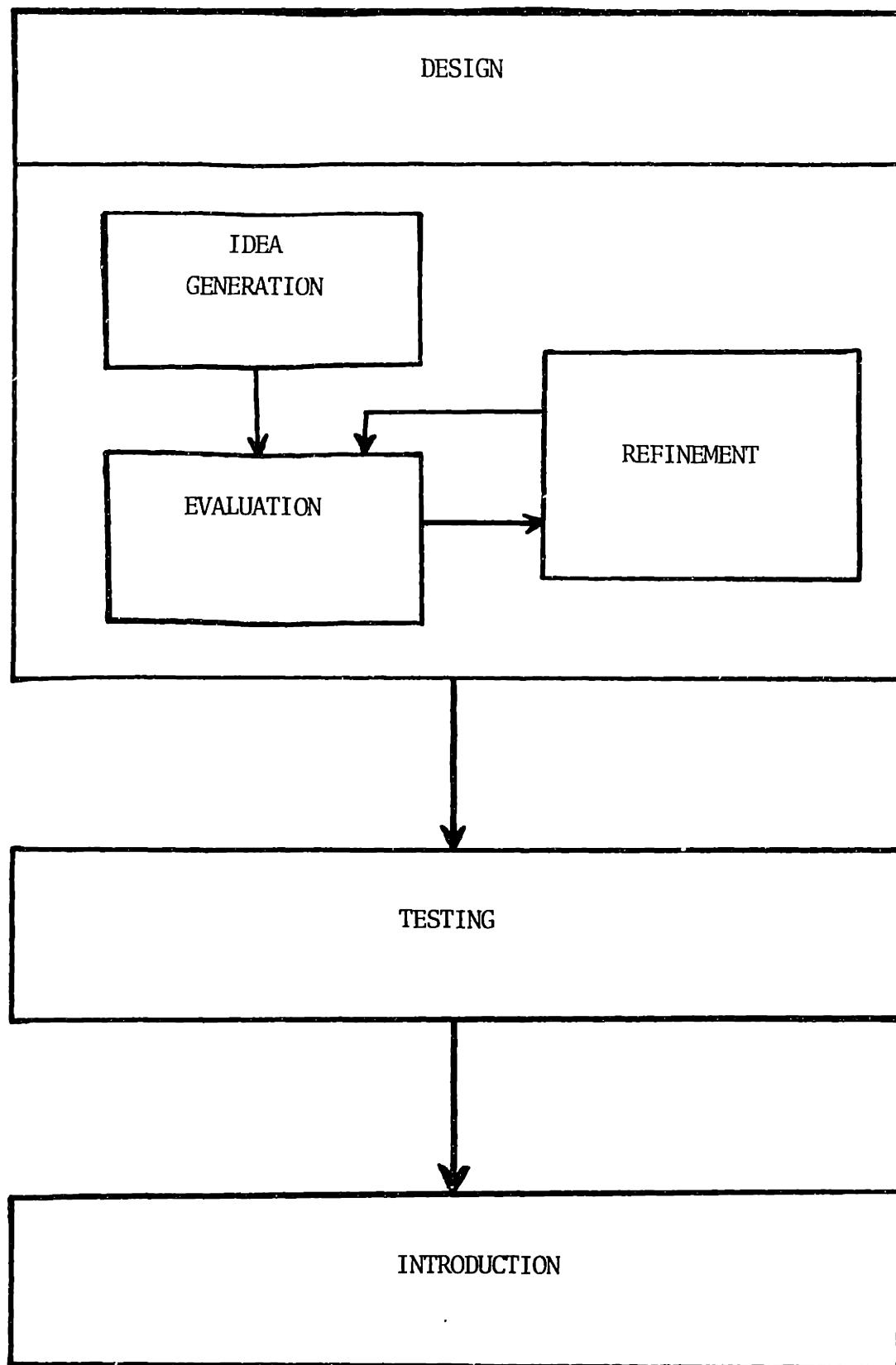


Figure 1.1 Process For Development of New Products and Services

design, integrates consumer studies with technology and creative efforts to generate new ideas. These ideas are then evaluated and the promising ones refined based on predicted and observed consumer reactions, production issues, and financial considerations. The design is iteratively refined until a "best" design and strategy are identified. Then either new consumer measures are taken for evaluation or the innovation is advanced to a pilot test. If this is successful the product is introduced.

The methodology presented in this dissertation concentrates on the design stage, although it could be adapted to the testing stage. Emphasis is on integrating analytic consumer response models with the creative design process to produce better ideas, more accurate evaluation, and clear diagnostics to guide refinement. The methodology identifies the structure of consumers' perception, abstracts strategically relevant segments, measures the importance of various "performance measures," and provides numerical estimates of consumer response.

1.2 Existing Approaches*

The field of consumer response modeling is not new, in fact much work has been done in transportation demand prediction, in marketing research, in mathematical psychology, and in a variety of other consumer choice applications. A brief cross-sectional summary of this work appears in Hauser [55], but this methodology is primarily based on state-of-the-art knowledge in three methodological fields: psychometrics, utility theory, and stochastic choice theory.

*A similar literature review and resulting criteria appears in Hauser and Urban [56].

Psychometrics

Psychometricians concentrate on measuring, structuring and explaining how consumers perceive new products and services. Based on similarity judgements or ratings of attributes multi-dimensional scaling develops perceptual maps which identify the important dimensions of perceptions and indicate the position of each product or service relative to these dimensions (Kruskal [85], Young and Torgenson [155]). Opportunities for innovation appear as gaps in the perceptual space (Stefflre [135], Green and Carmone [47]).

PREFMAP combines preference judgements on existing products with the perceptual data to isolate high opportunity areas by statistically deriving "ideal" points and measures of the relative importances of the perceptual dimensions (Carroll and Chang [24], Carroll [22]). Conjoint analysis uses statistical techniques to identify importances which are revealed by rank order preferences for factorially generated combinations of product attributes (Tversky [144], Green and Wind [49]).

"Expectancy value" models require consumers to directly state importances as well as perceptions of the attributes (Fishbein [33], Rosenberg [124]). Additive combinations of the importances times perceptions are correlated with a measure of preference to assess the adequacy of the formulation.

Utility Theory

Unlike psychometricians who describe consumer choice behavior, utility theorists have developed a normative theory to aid managers make

rational decisions when faced with complex, uncertain outcomes. Utility theory derives its strength from a rigorous set of axioms (von Neumann-Morgenstern [15]) and theorems (Raiffa and Schlaifer [12], Keeney [68,70,71,72,73], Ting [140], Fishburn [35,38]) which specify unique functional forms for utility functions such as additive, multiplicative, and quasi-additive. A utility function combines the performance measures for an alternative to produce a single cardinal measure of goodness for that alternative. The theorems also indicate techniques to directly assess parameters of these functions and to test the behavioral assumptions implied by the functional form. These parameters can be interpreted as directly measuring the relative importances of the relevant performance measures, their interdependencies, and the risk characteristics of the decision process. Since the theory is used to guide the decision rather than describe it, the decision maker chooses the alternative with the highest expected utility value.

To date empirical applications have directly assessed the utility functions of one or a small number of decision makers (Keeney [67], Bodily [11], Horgan [64]). A lengthy in-depth interview is required and the performance measures are chosen by the analyst and the decision maker to accurately describe alternative courses of action with quantifiable measures. This in contrast to the psychometrician who, because he is describing consumer choice behavior, uses techniques with less extensive data requirements, explicitly considers measurement error, and measures and uses perceptions.

Stochastic Choice Theory

Economists, transportation demand theorists, and mathematical psychologists recognize that there will always be uncertainty in any prediction of choice behavior and therefore concentrate on explicit modeling of choice probabilities. Axioms (Luce [92]) and measurement models (Thurstone [139], McFadden [93]) have been developed which determine these choice probabilities from observable "scale" values. Empirically economists and demand theorists parameterize the scale functions and statistically estimate the parameters from observations of actual choice among existing alternatives. The most popular models are the "random utility" models such as the multinomial logit model (McFadden [93,94]). In practice most applications (Ben-Akiva [10], Koppleman [80], Quarmby [118], Lehrman [87], Manski [99]) assume the scale function is linear in its parameters, is based on quantifiable "engineering" variables and is the same for all individuals in a segment. This is in contrast to utility theory with its axiomatic functional forms and idiosyncratic assessment and psychometrics with its perceived dimensions.

Mathematical sociologists model the stochastic choice process directly through diffusion, learning, Bernoulli, and semi-Markov models (Massy, Montgomery, and Morrison [102]). These models describe the dynamics of choice probabilities over time but do not link attributes of products or consumer preferences to choice.

Discussion

It is clear that much diverse work has been done. Each discipline reflects different emphasis, analytic techniques, and measurement

approaches. Psychometricians are concerned with perception and statistical recovery of importances from stated preference. Their axioms and behavioral assumptions are concerned primarily with measurement given some functional form. Utility theorists are concerned with prescriptive choice under uncertainty. Their axioms and behavioral assumptions specify unique functional forms and direct assessment procedures which explicitly incorporate relative importances, interdependencies, and risk characteristics, but the axioms do not necessarily apply to descriptive choice nor are the applications based on perceptions. Choice theorists are concerned with predicting numerical choice probabilities. Their axioms and behavioral assumptions model linkages from scale values to choice probabilities but do not consider linkages between consumer perception and managerial prediction of the attributes or axiomatic specification of functional forms for the utility functions.

The approaches also differ in the level of aggregation. Psychometricians develop average representation of perception and preference, but can explicitly check for homogeneity with respect to perception or preference (Carroll and Chang [23,24], Tucker and Messick [142]). Utility theorists work completely idiosyncratically and although theoretic work has been done on the existence and specification of a cardinal utility function for group decision making (Keeney [69], Keeney and Kirkwood [74], Kirkwood [79], Arrow [5]) not much effort has been directed at the problems of mass assessment of utility functions for describing individual choice. Choice theorists directly model individual response and then aggregate, but their statistical techniques require parameters to be the same for all individuals in the segment¹ and force judgemental specifi-

cation of segments before parameter estimation.

The three approaches are incomplete and diverse, but they are also complementary with each directed at a different phase in the consumer choice process. The methodology presented in this dissertation² integrates these approaches to perception, preference, utility, and choice to form a complete consumer response model. Some initial work has been done to integrate these disciplines but only at the aggregate level (Urban [146], Pessimier [113]).

This integration will require new linkages of perceptual dimensions to utility, improved procedures for homogeneity definitions, an axiomatization of stochastic choice to establish an isomorphism with utility theory and more powerful probability models especially designed to be compatible with direct utility assessment. Because of the complexity of the modeling process new testing procedures and effective methods of linking the analytic models to the management decision process need to be devised.

Next a set of criteria will be presented to define the specific problem.

1.3 Criteria for the Methodology

A normative methodology must be more than a predictor of who will choose what. It must interact well with the design process, reveal why consumers respond the way they do, indicate how to improve this response, and guide creativity in the design of innovation. The following specific criteria were formulated by Hauser and Urban [56]:

Complete and Integrating

The methodology should model the complete choice process and thus be applicable to the wide variety of choice decisions. To do this it must integrate existing approaches into a cohesive but modular process which offers a variety of techniques of varying complexity and data requirements. As such the methodology must respond and adapt to the diverse needs of decision makers and data availability.

Theoretically Sound

The methodology should reflect the acceptance phenomena at a level consistent with what is known about behavior taking into account the degree of modeling simplification required. All models require assumptions, but the methodology should make its assumptions explicit and force submodels to make their assumptions explicit. In doing so, the methodology should isolate weaknesses in existing techniques, and indicate where improvements need to be made. In addition, it should prevent models from being used in applications which violate their assumptions. Finally the conclusions reached by the methodology must result from the use of consistent mathematical logic.

Useful Predictive Powers

It must be technically and economically feasible to obtain the required measurements for the model. The methodology must be able to predict response to changes which are controllable in the design. For example, if the price of a new health maintenance organization (HMO) can

be varied, then the model should include this variable. Finally, the methodology should be extendable to changes or to new alternatives outside of existing consumer experience. For example, the model should be able to predict consumer response to a new HMO, even if none currently exist in the community.

Facilitate Successful Innovation

The design of new products or services requires creativity. The methodology should elicit and focus creativity by identifying characteristics relevant to the choice process and by explicitly measuring the relative importance of these characteristics. No matter how technically accurate the methodology is, it will only be used if it is acceptable to the organization which must design the innovation. This means that although some steps can be "black boxes," the underlying choice process must be understandable to non-technical as well as technical members of the design team. The outputs of each step must be clear and understandable. The methodology should help the design team to visualize the choice process while being sufficiently robust to prevent absurd answers from discrediting the model. Finally the methodology should be normative. It should be oriented toward the design and refinement of alternatives rather than simply describing the choice process. The ultimate objective of the methodology is successful innovation in products and services.

1.4 Topics Covered: Prose Table of Contents

The methodology presented in this dissertation is a joint development with Glen Urban and is published in "A Normative Methodology for Modeling Consumer Response to Innovation" (Hauser and Urban [56]).

The content presented there is repeated here and expanded to a complete detailed description which includes the managerial design and analytic modeling issues, the existing and new models applicable to each stage in the methodology, the formal presentation of the theory on which the methodology is based, and empirical examples of new and some of the existing models. Unique to this document is the formal development of the axiomization necessary to apply von Neumann-Morgenstern utility theory to compaction, the formal development of the Bayesian probability model, and the formal development of the honest reward/information test. Expanded in this document is a formal development of the other modules in the methodology and a presentation and critique of existing techniques including a summary of the mathematics necessary for each technique.

The intention of this dissertation is to present the methodology in a form that can be used to solve real problems. The descriptions of the new techniques are meant to be complete enough to be applied, while the descriptions of the existing techniques are meant to be complete enough to allow a user not familiar with them to judge their applicability to his problem. He can get more complete instructions from the references. The formal definitions in each module are meant to precisely identify the concepts and notation to effectively integrate the various disciplines and to avoid unnecessary ambiguity.

This dissertation is divided into three parts: (1) introduction and overview (chapters 1 and 2), (2) micro description of the methodology (chapters 3 through 9), and (3) testing and conclusions (chapter 10,11, and 12).

Part 1, introduction and overview gives an introduction to the problem, identifies the scope of solution, summarizes existing work, and presents a brief description of the methodology. This description identifies the basic role each module plays in relation to the rest of the methodology, introduces the basic notation, and indicates how to use the methodology.

Part 2, micro description of the methodology is a series of chapters each of which describes a single module in detail. Each chapter discusses the managerial design and analytic issues of the module, formalizes intuitive concepts, and relates the module to the rest of the methodology and to the managerial design process. The issues and mathematics of applicable existing models are presented in the common notation and are critiqued as to their role in the methodology. When necessary new formal theory is developed. Empirical examples are given to illustrate the new models and some of the existing models.

Part 3, testing and conclusion presents formal tests for the models in the methodology and for the predicted choice behavior. It also presents the managerial implications of the empirical case studies, the major conclusions of this research effort, and an indication of related future research topics.

Chapter 2

DESCRIPTION OF THE METHODOLOGY

The criteria specified earlier will now be reflected in a model and measurement based normative methodology to elicit and guide creativity in the design of innovative products and services. First the macro structure will be defined and then a brief description of the micro structure will be given. Part II of this dissertation (chapters 3 through 9) will describe in detail the issues, the models, the mathematical and psychological theory, and the empirical applications of each module in the micro methodology.

2.1 Macro Structure of the Methodology

To facilitate successful innovative it is necessary that the analytic models be closely tied to managerial decisions, thus the overall methodology consists of a managerial design process that is linked on many levels to a parallel consumer response process. (See figure 2.1.) The focus of this dissertation is in the consumer response process. The first step is observation of consumers, which measures consumers' perceptions, preferences, and choice with respect to alternatives and attributes identified as relevant. These measures are used to estimate the parameters of individual choice models which identify the structure of perception, provide indications of relative importances of the measures of perception, identify strategically relevant segments which are homogeneous with respect to preference, and provide numerical indications of the

Design Process

Consumer-Response
Process

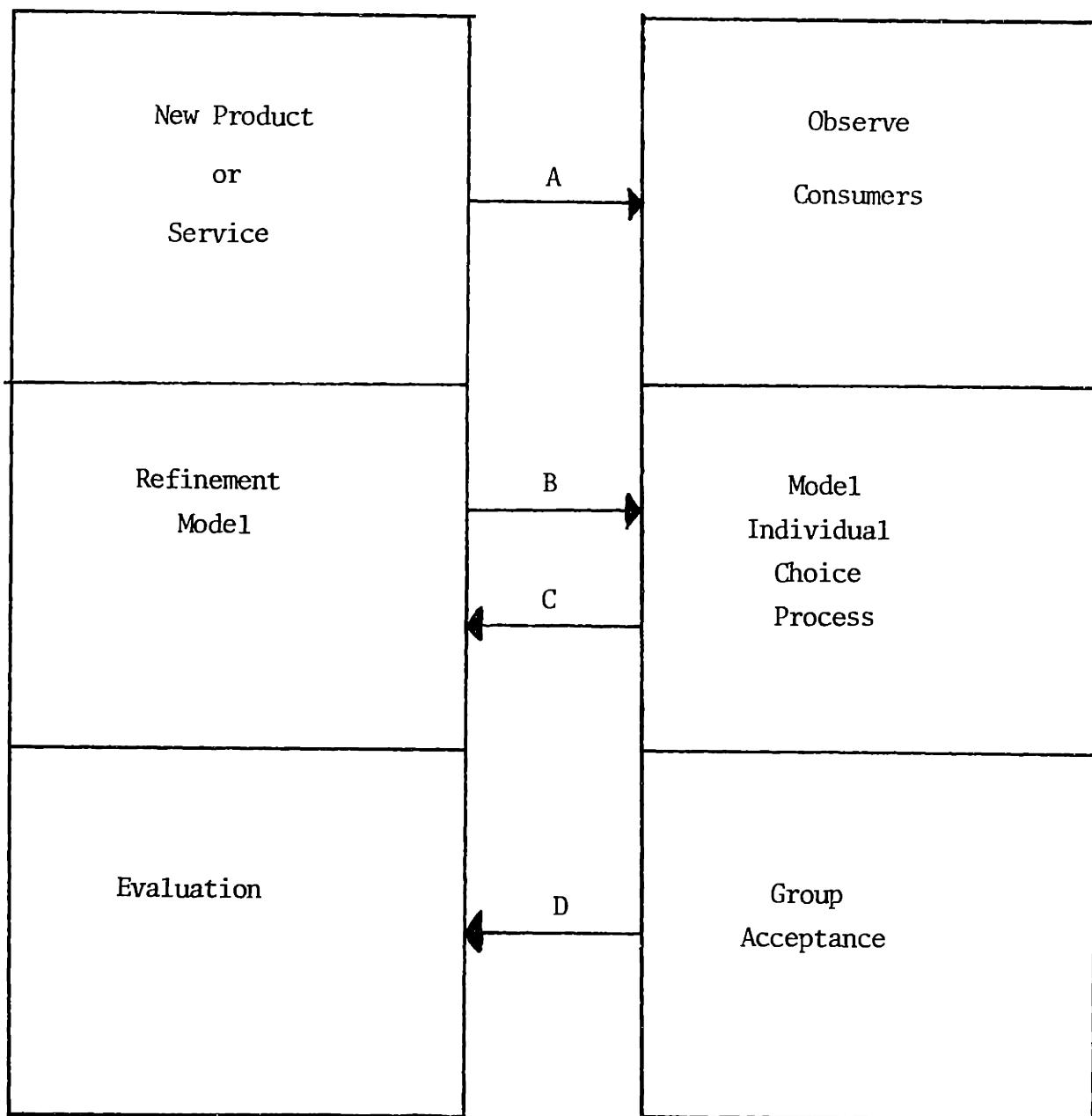


Figure 2.1: Macro Description of the Methodology

individual choice response. Finally an estimate of group response is obtained by aggregating individual choice measures. These measures are then input to a model in the managerial design process which evaluates them relative to investment, operating costs, risk, and externalities.

Although the methodology provides numerical estimates for evaluation of strategy and design, its primary use is to enhance the creative design of new products and services by providing accurate and easily understood diagnostic information about the consumer choice process.

- The first cycle through the methodology acts as a screening process which identifies the most promising alternatives for further consideration. These alternatives are then refined based on the detailed diagnostic information provided by the individual choice models (arrow B in figure 2.1). The refined design is analytically tested by using the choice models to simulate consumer response. Iteration leads to a "best" design (arrows B + C in figure 2.1) which is then tested by taking new consumer measures (arrows A + D in figure 2.1) or advancing the design to pilot test. (Figure 1.1).

2.2 Basic Micro Flow of the Methodology

Observation of Consumers

A good model is dependent upon high quality input. A model is accurate only if the measurements it requires are valid and reliable. It is useful only if the measurements are feasible.

Identify choice alternatives: The first step (See Figure 2.2.) in observing consumers is to identify the relevant choice alternatives

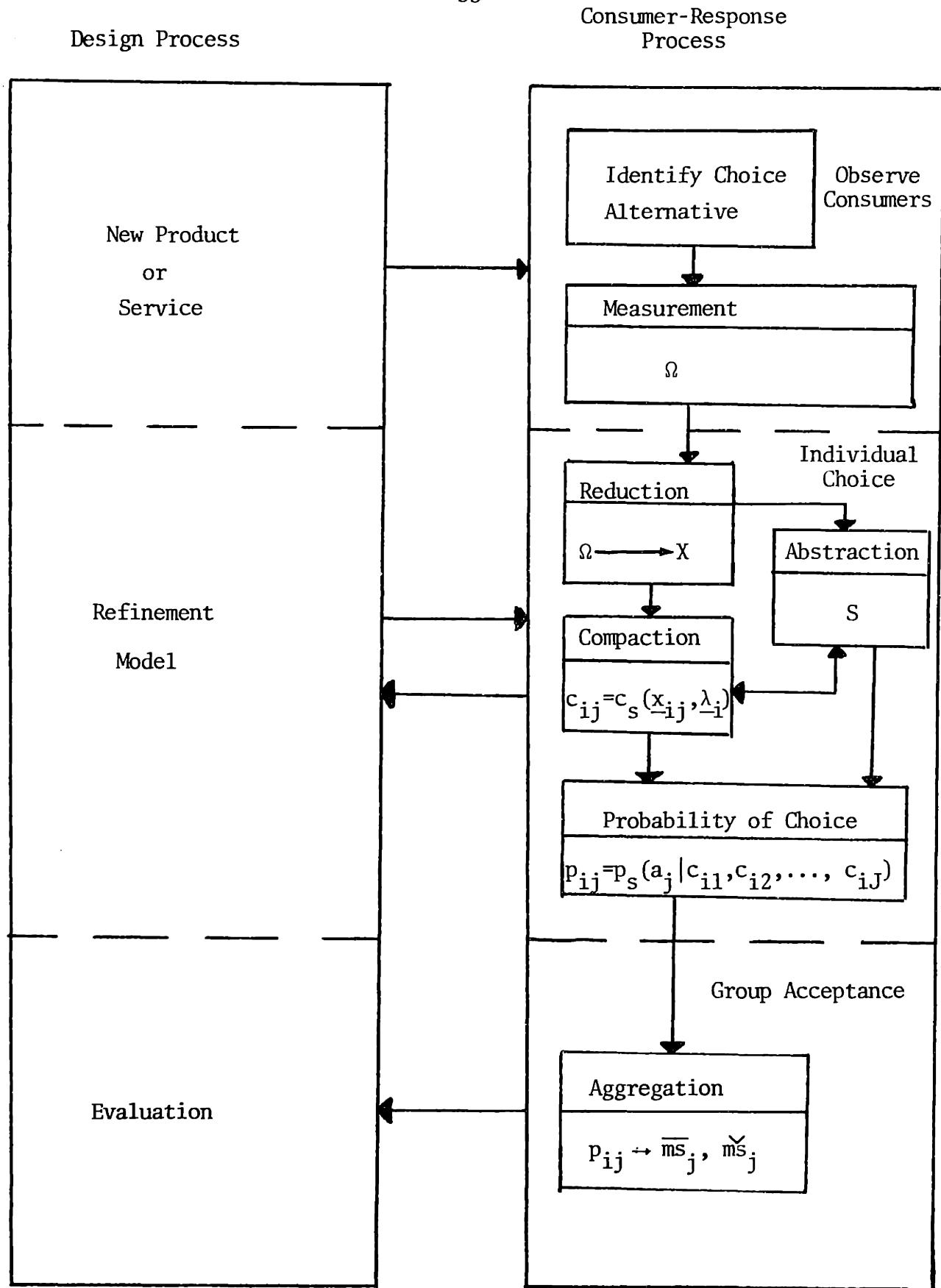


Figure 2.2: Micro Flow of the Methodology

(a_j) whether they are real, such as an existing mode of transportation, or proxy such as a concept statement (See Chapter 3). In either case, the consumer must be aware of the alternative at a level sufficient enough to make a choice. Also in this step informal and formal techniques are used to identify a complete list of attributes and semantics which the consumer uses to describe the alternatives.

Measurement: The second step in observing consumers is to take formal measurements (Ω) of (1) perceptions of the attributes (e.g., rating scales on hospital quality or measuring perceived travel time), (2) preference for the alternatives (e.g., rank order or constant sum) and/or actual choice, and (3) demographics and other consumer descriptors (e.g., sex or existing pattern of health care). If in addition utility theoretic models (See Chapter 6) are to be supported, direct measurements are made of preferences, risk characteristics, and interdependence among a set of "performance measures" which describe the alternatives.

Individual Choice Models

The consumer response process is described by a series of modular individual choice models. This modular process separately models the various parts of the choice process, i.e., perception, preference, segmentation, and choice, so that the managerial design team can better understand and control the choice process. Furthermore, a modular process is more flexible and robust because different models can be used in each module depending on the data available, the diagnostics needed, the choice process, and the money and time available for analysis.

Reduction: The measured attributes, (Ω), necessary to completely describe alternatives, are usually too numerous to enable analysts or managers to truly understand the structure of consumer perception. The first consumer model, reduction, identifies a parsimonious set of performance measures, (X), which describe and structure consumers' perceptions. The model also provides numerical indications of how each consumer perceives each alternative relative to these performance measures. When aggregated these numerical measures provide a clear geometric interpretation called perceptual space. (See Chapter 4.)

Abstraction: Not everyone perceives the performance measures the same, nor does everyone have the same preferences relative to the performance measures. Therefore, rather than designing an average product which may not satisfy the needs of anybody, the methodology abstracts segments, (S), based on homogeneity of preference (and sometimes homogeneity of perception). These segmentations facilitate the design and implementation strategy for innovation by allowing differential targeting of alternatives.

Compaction: Even with only a few dimensions to describe perception, the design team needs to know how important each performance measure is, what the risk characteristics are relative to each measure, and how the measures interact to effect preference and ultimately choice. Specifically a real valued compaction function, $c_s(x_{ij}, \lambda_i)$, is determined which maps individual i 's perceptions of the vector of performance measures, (x_{ij}), for alternative a_j and a vector of individual specific preference parameters, (λ_i), into a scalar measure of goodness, (c_{ij}), for each

individual and for each choice alternative. This measure, c_{ij} , has the property that with all other alternatives held fixed, any set of performance measures yielding the same value, c_{ij} , must also yield the same probability of choice for alternative a_j . In other words, compaction compresses the performance measures into a one-dimensional measure for each alternative, and knowing this value for each and every alternative in an individual's choice set is then sufficient to predict his choice probabilities. This intermediary step before probability prediction provides the diagnostics necessary to understand how consumers determine their preferences for alternatives.

Probability of choice: For both evaluation and refinement it is necessary to know how many consumers will select the new product or service and how many potential consumers will select each of the various competing alternatives. This last step in the choice behavior section mathematically transforms the vector of compaction values, $(c_{i1}, c_{i2}, \dots, c_{iJ})$, into individual choice probabilities, $(p_{i1}, p_{i2}, \dots, p_{iJ})$. The outputs can be Bernoulli probabilities, $p_s(a_j | c_{i1}, c_{i2}, \dots, c_{iJ})$, which estimate an individual's selection probabilities conditioned on the vector of compaction values or they can be Poisson rates. (If γ is the choice rate for an alternative then the probability that the individual will choose that alternative in small time period Δt is $\gamma\Delta t$.) In cases where repetitive choice decisions are made by a consumer, separate trial and repeat choice parameters would be estimated based on the goodness measures before and after use of the new product or service.

Group Response

The individual choice models provide diagnostics for refinement, but numerical measures such as "market share" or the total number of people choosing each alternative are needed for evaluation.

Aggregation: This final step in the consumer response process model combines the individual choice probabilities to produce numerical estimates of the mean \bar{N}_j and variance \hat{N}_j of the total number of consumers choosing each alternative, a_j . In doing so it corrects the probabilities for strength of awareness, and availability. Also if data is available, external models are used to include the dynamic effects of trial, repeat, and frequency. If strategically relevant segments were identified in abstraction, then aggregation is done separately within each segment.

2.3 Summary and Interaction with the Design Process

This methodology facilitates successful strategic design and implementation of innovative products and services by modeling consumer response. First, the consumer response process provides the design process with estimates of the number of consumers choosing each alternative. When integrated with factors such as investment, cost, and risk these estimates result in a screening process which terminates some alternative innovation strategies (NO), identifies some for continued development (ON), and ultimately selects a "best" alternative (if any) to introduce to the market (GO). If initial evaluation results in an ON decision, the design is cycled to a refinement effort. This effort is based on the diagnostic information provided by the methodology, i.e., (1) reduction identifies

the performance measures which structure perception and calculates their average values to give a geometric picture of the "market structure,"

(2) abstraction identifies strategically relevant segments homogeneous with respect to preference or perception, (3) compaction explicitly identifies the relative importances of the performance measures, how strongly they interact and how important risk characteristics are in consumers' preferences and provides geometric interpretations via indifference curves, and (4) probability of choice gives numerical implications of the consumers' preferences. These diagnostics provide the creative managerial design team with a clear understanding and intuition about the consumer choice process. This understanding results in refinements in design and strategy which are iteratively simulated to test and improve understanding and to lead to a "best" design and strategy. Finally either new consumer measures are taken to further evaluate the "best" alternative or the alternative is advanced to pilot test.

The description of the methodology in this chapter is brief and incomplete. Part II of this dissertation will explain in detail the issues, the models, the theory, and some empirical applications.

Chapter 3

OBSERVATION OF CONSUMERS

This methodology is model oriented and good models are dependent upon high quality input. A model is accurate only if the measurements it requires are accurate and it is useful only if the measurements are feasible. This chapter will indicate some of the important issues in measurement as they relate to the methodology. It does not address detailed questionnaire design such as the wording of the questions or the exact type of rating scale used. Instead, based on the issues raised here, the analyst can consult a measurement expert for the detailed survey design. (Some useful references are Payne [112] for semantics and general guidelines, Cannell and Kahn [20] for principals of interviewing, Green [45] for uni-dimensional scaling techniques, and Campbell and Stanley [19] for experimental design and a checklist of factors jeopardizing external and internal validity.)

We begin with some general issues important in any measurement, and particularly important for this methodology. Then the types of choices to give the respondent and the required input data are discussed.

3.1 General Guidelines for Measurement

3.1.1 Respondent Motivation

Some of the potential submodels in this methodology, particularly direct assessment of compaction functions, have extensive and complex

measurement requirements. It is important for accuracy that the respondent has the desire to carefully study and understand the questions and give well thought out responses. Thus it is important that he be highly motivated, feel that his answers are important, and that they will not be misused. A sample telephone call which illustrates respondent motivation is given in figure 3.1.

3.1.2 Semantics

A number of issues such as common trouble words and the importance of not misleading the respondent are discussed in Payne [112] and Cannell and Kahn [20].

For this methodology it is important that the questions be phrased in the semantics that the respondent uses to described the alternatives. There are at least four methods to explore for semantics: (1) focus groups, where a number of consumers are brought together with a moderator and asked to discuss the alternatives, (2) open ended questionnaires on "likes" and "dislikes" about alternatives, (3) triad procedures (Kelly [77]) where the set of alternatives are divided into subsets of three and the respondent is asked "Which two are most similiar and how are they similiar?" "Which two are most different and how are they different?" and finally (4) expert opinion of people experienced in working with consumers in the relevant product or service catagory.

The combination and filtering of these procedures is by nature an art rather than a science. In the HMO (Health Maintenance Organization) case discussed throughout this dissertation all four were used in generating both questionnaires (appendices 1 and 2).

Sample Telephone Call

Hello, my name is _____, I am involved in a research effort to determine health care requirements and desires of consumers such as you. This study is sponsored by the Sloan School of Management in cooperation with the M.I.T. Health Department.

Your name has been selected at random to be part of this survey of health care attitudes and preferences in the M.I.T. community. The number of people being asked to participate is small, so your answers are very important. The results of the survey will also be used by the M.I.T. Medical Department to remain as responsive as possible to the needs of the M.I.T. community.

We are asking for help in completing a questionnaire. The questionnaire requests information about the health services now available to you and asks for your opinion on several new methods of delivering health care. The last questions relate to some demographic characteristics that are important in projecting the responses from this small survey to the M.I.T. community as a whole. The questionnaire takes about 20-30 minutes to complete. Most people find it easy and interesting to answer the questions and they think it is important to make their feelings known to those who provide health care. After filling out the questionnaire I would like to ask you some additional questions in person.

It is important to M.I.T. in planning for your needs and it is important for our research project to improve understanding of consumer response to health services. Could I deliver this questionnaire?

Interviewer: Get respondents address and set up a time for a personal interview.

Figure 3.1: Sample Telephone Call

3.1.3 Pretest

No matter how carefully the questionnaire is worded there is always potential for poorly understood or misleading questions. This is especially true in health and transportation where many common words take on special technical meanings and it is easy to overestimate the specialized vocabulary of the respondent. It is imperative that the questionnaire be "pretested" on a small sample before a full scale survey is undertaken. Since the purpose of the pretest is to test understanding of the questions, it is useful to have the respondent read each question and try to explain to the interviewer the meaning and/or purpose of that question.

In direct assessment of compaction functions (Section 6.2), we are not only interested in the accuracy of each question but in the relevancy of the compaction function derived from the questionnaire. In the HMO study compaction functions were assessed over four performance measures, quality, personalness, convenience, and value. In pretest we used an interactive utility assessment program developed by Sicherman [130]. First, various parameters were assessed by different types of questions and were checked for consistency. Then the completely assessed compaction (utility) function was used to rank alternatives and the respondent was asked to verify agreement with the rank order.

This combination of questionnaire and computer was useful in developing a measurement device which we felt could accurately assess compaction functions. For example we found that for discretely valued performance measures, it was easier and more relevant for a respondent to specify a probability for a lottery by setting a probability wheel

than it was for him to specify the certainty equivalent of a 50-50 lottery. (See Section 6.2 for an explanation of these terms.)

3.1.4 Sampling

Care should be taken to get a random, uncontaminated sample. The randomness of the sample is important because in the methodology choice is modeled on an individual level and variation is important in both perceptions and preferences. Cross-sectional variation is necessary for abstraction of segments and projection to the community at large.

Since perceptions and preferences are usage dependent, previous usage of an alternative by the respondent should be noted. For example, the perceptions of people in a pilot HMO could be significantly different from those exposed only to the concept. A related effect is post-purchase rationalization. For example suppose we were measuring perceptions of graduate schools by college seniors. These perceptions should be measured before acceptances (and rejections) are mailed out. If possible measurement should always be made both before and after choice, but if this is not possible, the analyst should judgementally correct for effects of post-purchase rationalization.

Another issue in sampling is the small intensive personal interview versus the large coverage of a mailed survey. In the HMO study a mailed questionnaire was used for the inputs to reduction. This was feasible because the questions were relatively simple attribute ratings and demographics. The resulting large sample (447 out of 1000)¹ allowed for variation in the evoked set (See Section 3.2). Later, for direct assessment of compaction functions, a home interview survey was necessary

because of the complex nature of the questions and because of the need to challenge the answers to indifference questions (See Section 6.2) to insure understanding of the indifference concept. The resulting sample was necessarily smaller (76 out of 80) due to the expense and time involved in personal interviews.

3.1.5 Summary of General Guidelines

Prerequisites for valid measurement are that (1) care be taken to motivate the respondents, (2) that the survey be worded in the semantics consumers rather than professionals use, (3) that even the simplest questions be pretested, (4) that a suitable random sample be drawn, and (5) that the proper technique (personal or mailed) be selected for each measurement.

3.2 Choice Alternatives

The heart of the methodology is a model of individual choice behavior, thus consumer perceptions, preferences, and choice behavior must be observed with respect to the product or service alternatives which are relevant.

3.2.1 Measurements are Relevant Only with Respect to the Evoked Set

An individual's evoked set is basically the set of alternatives which he is aware of and has enough information about to realistically make a choice. Although, there may be a large number of products available, each consumer often can evoke only a few. For example when the evoked set (Allaire [2]) was defined as those brands which the consumer

has used, would consider buying, or would not consider buying, Urban [146], in a study of seven consumer products, found that the average size of the evoked set was three products.

Clearly for perceptions, preferences, and choice to be meaningful, they must be measured with respect to evoked alternatives, even if the analyst forces evoking through the use concept statements, or prototype alternatives. The respondent just does not have enough information to answer any questions about unevoked alternatives.

3.2.2 Revealed Preference vs. Proxy Choice

This methodology uses two types of choice: real and proxy. Real choice is when the consumer is placed in a choice situation that will actually occur outside the laboratory and allowed to make an unconstrained choice. For example, observation of a consumers' last used mode of transportation for a work trip is a real choice. Models calibrated on real choice are usually called revealed preference models. In this methodology such models are used in chapter 7, Probability of Choice, to "tune" predictions based on compaction functions.

One type of proxy choice is the concept statement, which is often necessary in services because the number of existing alternatives is so small that one must force evoking of additional alternatives in order to have sufficient perceptual inputs. For example in the HMO study the evoked set was expanded to four options by specifying three new options in concept form. The options included an M.I.T. HMO, the Harvard Community Health Plan, and a hypothetical Massachusetts Health Foundation. Figure 3.2 is an example of the M.I.T. HMO concept.

Figure 3.2: Concept Description

DESCRIPTION OF THE M.I.T. HEALTH PLAN

M.I.T. announces a new health care plan for YOU AND YOUR FAMILY. By joining the M.I.T. HEALTH PLAN you can get comprehensive health care at a low, fixed monthly charge. Virtually all your medical needs will be met. You will not have to face unexpected doctor or hospital bills and you will not have to worry about finding a good doctor for you or your family.

The cost of joining the M.I.T. HEALTH PLAN is only a little more than regular Blue Cross/Blue Shield health insurance, but you get more services and comprehensive care. There are no charges for doctor visits, nursing and laboratory services, or hospital services. Women in the plan pay nothing extra for prenatal, delivery, or maternity care. The services are comprehensive and include mental health care and emergency services.

The costs are kept low by the utilization of preventive care to keep you well. The plan succeeds by keeping you and your family well and out of the hospital. In addition, the use of trained paramedics and technology helps reduce costs while maintaining the quality of care.

You choose your own personal doctor (specialist in internal medicine for yourself and a pediatrician for your children) from our staff of physicians. Your doctor supervises your total health care at the health center and in the hospital. He will be sure you get the highest quality of care. When you are a member of the M.I.T. HEALTH PLAN you can be sure of getting health care around the clock from the staff of physicians, nurses, social workers and allied health personnel.

The M.I.T. HEALTH PLAN delivers its services from the Homberg Memorial Building on the M.I.T. campus. Parking is available during patient visits. Hospital services are provided by the Mount Auburn and Cambridge City Hospitals. Maternity and gynecological care are provided through the resources of the Boston Hospital for Women. For emergencies outside the Boston area, local hospitals can be used.

You can become a member of the plan by paying \$1.50 per month more than your Blue Cross/Blue Shield coverage if you are single and \$4.00 more per month if you are married. If you are a single student and do not have hospital insurance, the cost is \$8.25/month more than the student health fee you are currently paying; if you are a married student, the cost is \$20.00/month more than the student health fee. These fees cover all of your medical costs except: the first \$50 and 20% of the balance of prescription charges and the excess of \$10 per visit for psychotherapy (over \$5 per visit for group therapy). The plan does not include eye glasses, hearing aids, cosmetic surgery, custodial confinement, or dental care done outside a hospital. If you join the plan, you must remain a member for one year.

The M.I.T. HEALTH PLAN is designed to make comprehensive, high quality health care available to you and your family at a low cost.

Another type of proxy choice is through the use of characteristic lists. Once a complete set of performance measures is specified (See Chapter 4, reduction), alternatives can be presented to the consumer as lists of specified performance measures. See figure 3.3.

Plan A

Quality	=	7(excellent)
Personalness	=	4(satisfactory)
Convenience	=	6(very good)
Value	=	3(poor)

Figure 3.3 : Proxy Choice as List of Performance Measures

These choices are particularly useful when one is trying to find the relative importances of the performance measures. In the methodology, such choices are used in conjoint measurement (Section 6.3.4) where individuals rank order factorial designs, and in direct assessment of compaction functions (Section 6.2) where individuals specify the level of one performance measure of one alternative in order to make that alternative indifferent to another alternative.

Both real and proxy choices have uses in observing choice behavior. A discussion of their relative merits and an indication of how they can be used to complement each other is deferred until section 7.1.4 when they can be discussed in relation to the models used in the methodology. At this stage note that concept statements tend to measure trial whereas real choice tends to measure repeat or steady state choice behavior.

3.3 Measurement of Input Data

After the individual is presented with a choice, perceptions, preferences, and choice must be measured (See figure 3.4).

3.3.1 Measuring Perceptions of the Attributes of Alternatives

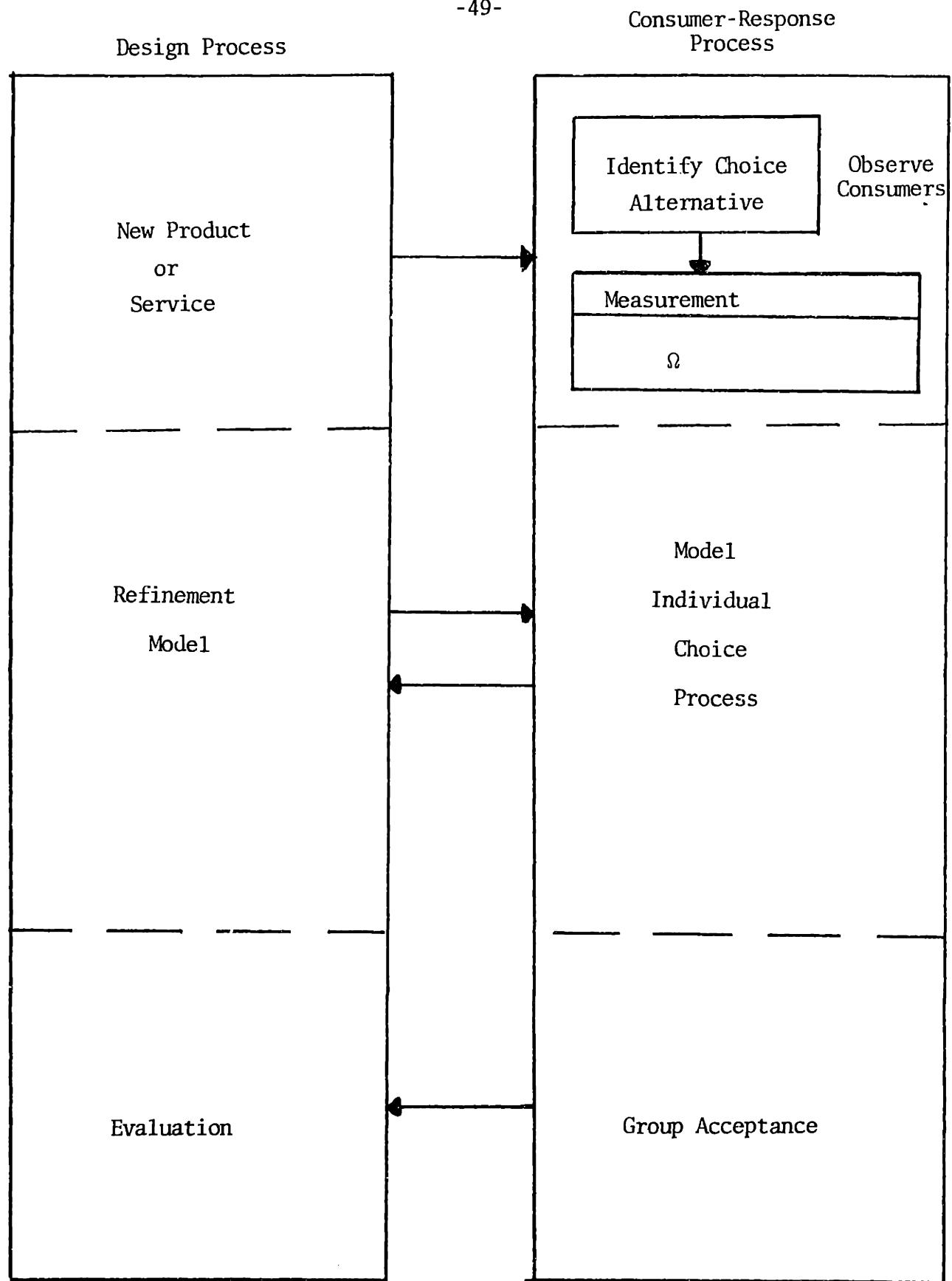
To understand and control individual choice behavior it is first necessary to understand and measure how consumers perceive the attributes of the alternatives. First let us consider certain properties which the set of attributes must satisfy.

3.3.1.1 Properties of the Attribute Set

The set of attributes should be complete and span the set of feasible alternatives. The design team should be able to influence choice by changes in the attributes and it should be possible to measure consumers' perceptions of the attributes.

Complete: A set of attributes is complete if that set alone is sufficient to exactly specify all potential alternatives in the product or service category. In other words, let Y_1 be an attribute and let $Y = \{Y_1, Y_2, \dots, Y_L\}$ be the set of attributes. Let $y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ijL})$ be measures of individual i 's perceptions of these attributes with respect to alternative a_j . A complete set of attributes would then uniquely specify each alternative. Formally:

Definition 3.1: A set of attributes, Y , is said to be complete relative to a category of alternatives if individual choice probabilities relative



3.4: Observation of Consumers

to a choice set A can be uniquely determined by knowing y_{ij} for all $a_j \in A$, and if this is true for all choice sets contained in the category.

A category is a set of existing and potential real and proxy alternatives and is purposely left vague and general at this point.

Notice that the attributes are not required to be disjoint. They can in fact be overlapping measures since one of the purposes of reduction is to detect and correct for overlapping measures.

Spanning: It may be difficult to make predictions for alternatives costing \$100 if all measurements are made on real and proxy alternatives with associated costs of \$10. Thus in addition to the attributes completely describing the alternatives, the range of the measures of the attributes should span the range of alternatives. Formally:

Definition 3.2: The range of measures of an attribute set, Y , is said to span a category of alternatives if the values for the measures of each individual's perceptions of the attributes, y_{ij} , relative to alternatives a_j in the category are contained in the range of Y .

For example, in figure 3.5, if P is the set of all potential alternatives then measurements should cover P or a set like Q rather than a set like M . This may seem obvious but remember that the attributes of new alternatives are often well outside the range of attributes for existing alternatives. Thus measurements on existing alternatives alone may not span the category.

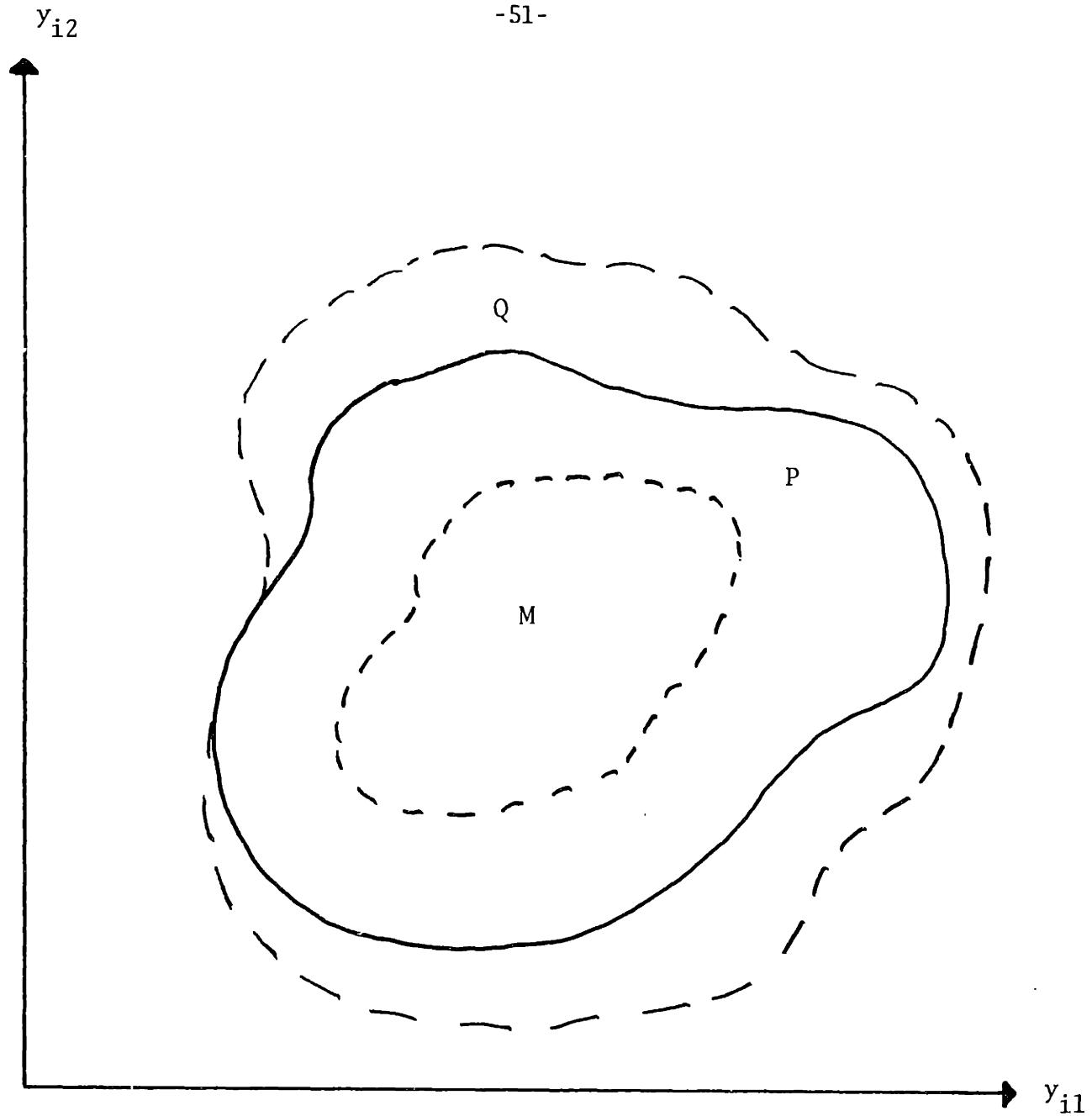


Figure 3.5: Spanning Set of Attributes

Measureable: One of the criteria for the methodology was that it be usable. It might be an interesting mental exercise to describe consumers' perceptions by attributes but will not be very useful for insight or prediction unless some ordinal or cardinal measure can be associated with the attribute. Thus it must be possible to assign a real number to individual i's perception of attribute Y_k for alternative a_j and this number, y_{ijk} , should carry information about individual i's preferences relative to that attribute. (E.g. the bigger the better.)

Specifically there should be a naturally assigned measure for the attribute, or the consumer should be able to directly specify a scale value for the attribute. For example minutes is a natural measure for travel time whereas a seven-point semantic scale might be used for the quality of the hospitals associated with an HMO. Notice that measures such as the number of doctors on the hospital staff or the number of diagnostic errors per million treatments are not used because it is doubtful that consumers use these measures to quantify the quality of a hospital.

In using scales, the analyst should attempt to keep the scales atomic and try to correct for scale bias. Atomic means that only one idea is measured per scale, for example, the five-point agree/disagree scale "I feel that the school is well-known for high quality management training." is not atomic because it blurs the measurement of "well-known" and "high-quality." Scale bias can result from a tendency on the part of the respondent to favor a particular end of the scale. This is partially corrected by combining two techniques. First randomly reverse favorable/unfavorable scales, e.g., for questions 1,3,4,8,9 the right

end of the scale if favorable and for questions 2,5,6,7,10 the left end is favorable. Second normalize the response across alternatives. In other words, if r_{ijk} is individual i's rating of alternative a_j on scale k, then

$$y_{ijk} = r_{ijk} / \sum_{\ell=1}^J r_{i\ell k}.$$

Instrumental: Even if all the attributes are measureable the methodology will not be usable unless design decisions can effect at least some of the attributes and these effects can be planned at least in part. For example travel time is an instrumental attribute for Dial-a-Ride because supply models can be formulated to predict the impact of certain design decisions on travel time. Similiarly the quality of an HMO's associated hospitals is an instrumental variable because the HMO can change its hospital affiliation.

Since we are allowing the methodology to be used in an exploratory mode, the instrumentality of an attribute may occasionally be relaxed such that the only requirement is that the design team can judgementally estimate the impacts of design decisions. The usefulness of this approach is that it may be easier and more robust to judgementally change the value of single attribute's average perceptual measure than to judgementally predict choice.

3.3.1.2 Generation of the Attribute Set

A set of attributes should be generated which are complete, spanning, measureable, and instrumental. To do this preliminary exploratory measurements of consumers are necessary to generate an intial set which

is later refined. Some useful techniques are:

Open-ended survey: One simple technique to get a feel for the attributes is an open-ended survey on likes or dislikes.

Triads: (Kelley [77]) Given three alternatives, which two are most similar and how are they similar? Which two are most different and how are they different?

Brain-storming: If the analyst or someone he can consult is familiar with the product category, an initial set of attributes can be generated by trying to role play and come up with as many as possible by brainstorming.

Focus groups: Bring together consumers from the target group and allow them to discuss the alternatives in the category. Note that triads and open-ended likes and dislikes are useful in the focus groups.

Any combination of the above should be used to generate an initial set of attributes. The analyst should then redefine them if necessary to make them instrumental and if scales are to be used he should word the scales as carefully as possible. These are then resubmitted to focus groups or separately to individual consumers who are asked if (1) there are any missing attributes, (2) if they can relate to the scales, and (3) if they feel any question should be reworded. This process should be iterated until the analyst is satisfied with the set of attributes.

For example, in the HMO study various forms of open-ended surveys, triads, brain-storming, and focus groups were iteratively used to generate the scales in figure 3.6.

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1. I would be able to get medical service and advice easily any time of the day and night.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
2. I would have to wait a long time to get service.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
3. I could trust that I am getting really good medical care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
4. The health services would be inconveniently located and would be difficult to get to.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
5. I would be paying too much for my required medical services.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
6. I would get a friendly, warm, and personal approach to my medical problems.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
7. The plan would help me prevent medical problems before they occurred.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
8. I could easily find a good doctor.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
9. The service would use modern, up-to-date treatment methods.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
10. No one has access to my medical record except medical personnel.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
11. There would not be a high continuing interest in my health care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
12. The services would use the best possible hospitals.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
13. Too much work would be done by nurses and assistants rather than doctors.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
14. It would be an organized and complete medical service for me and my family.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
15. There would be much redtape and bureaucratic hassle.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
16. Highly competent doctors and specialists would be available to serve me.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>

Figure 3.6: Measuring Perceptions

3.3.2 Measuring Choice Among the Alternatives

Real: When there is an actual choice, such as the last used transportation mode, choice is simply measured as a 0-1 variable. I.e., let $\delta_i = (\delta_{i1}, \delta_{i2}, \dots, \delta_{iJ})$ then $\delta_{ij} = 1$ if individual i chose a_j and $\delta_{ij} = 0$ otherwise.

Proxy: When the choice is proxy, is often useful to obtain stronger measures than just 0-1 choice. It is useful to (1) obtain 0-1 choice from a number of subsets of the entire choice set, (2) augment the 0-1 measurement with an intent scale, and (3) obtain rank order preference over the entire choice set. Examples of these measurements are given in figure 3.7.

Constant sum: Whether the choice is real or proxy it is often useful to collect constant sum paired comparison preference. This is done by giving the respondent the alternatives two at a time and having him allocate a fixed number of chips (usually 11) between the two alternatives to indicate his strength of preference. Standard techniques (Torgenson [141]) can then be used to reduce these measures to a single scale.

3.3.3 Measuring Indifference

A limiting case of choice is indifference, i.e., when an individual neither prefers a_j to a_ℓ or a_ℓ to a_j . In the methodology three types of indifference questions are used in direct assessment of individual compaction functions (See Section 6.2).

Figure 3.7: Measuring Preference and Intent

15. You have now read and rated three new health plans in addition to your existing medical service. We are now interested in your preference for these alternatives. Below are listed the alternatives. Place a "1" next to the one which would be your first choice. Place a "2" next to your second choice. Place a "3" next to your third choice. Place a "4" next to your last choice.

- Your existing health services
- HARVARD COMMUNITY HEALTH PLAN
- M.I.T. HEALTH PLAN
- MASSACHUSETTS HEALTH FOUNDATION

16. If only the M.I.T. HEALTH PLAN and your existing health service were available which would you actually choose. Check one.

- Your existing health services.
- M.I.T. HEALTH PLAN

16a. If you selected the M.I.T. HEALTH PLAN, which of the following statements reflects how you feel about your choice: Check one. (Otherwise go to Question 17.)

- I definitely would select the M.I.T. HEALTH PLAN
- I probably would select the M.I.T. HEALTH PLAN
- I might select the M.I.T. HEALTH PLAN

17. If only the M.I.T. HEALTH PLAN and THE MASSACHUSETTS HEALTH FOUNDATION and your existing health system were available, which would you choose? Check one.

- Your existing health services
- M.I.T. HEALTH PLAN
- MASSACHUSETTS HEALTH FOUNDATION

17a. Which of the following statements reflects how you feel about your choice: Check one.

- I definitely would select this alternative
- I probably would select this alternative
- I might select this alternative

18. If you were actually considering joining the M.I.T. HEALTH PLAN would you take actions to get more information?

YES NO

If YES, what would you do?

Tradeoff questions: In a tradeoff question the respondent is given two proxy alternatives. The first is completely and deterministically specified by values for all of its performance measures. The second is specified by values for all but one of its performance measures. The respondent is then asked to specify the value of that one performance measure such that he is indifferent between the two alternatives. In most cases the values of all but two of the performance measures are held fixed and common to both alternatives. (See figure 3.8.)

Risk questions: In a risk question the respondent is again given two alternatives, the first, called the certainty equivalent is completely and deterministically specified. The second is a more complex stochastic alternative which is really a random combination of two alternatives. If the respondent selects the second alternative, then with a probability, p , it takes on one set of characteristics and with a probability,

<u>Plan C</u>	<u>Plan D</u>
Value = 5 (good)	Value = 5 (good)
Convenience = 5 (good)	Convenience = 5 (good)
Personalness = 6 (very good)	Personalness = 2 (very poor)
Quality = 2 (just adequate)	Quality =
	Very poor
	Just adequate
	Satisfactory
	Good
	Very good
	Excellent

Figure 3.8: Schematic of Tradeoff Question

(1-p), it takes on another set of characteristics. He is asked to specify p such that he is indifferent between the two plans. In most cases the values of all but one of the performance measures are held fixed and common to both the certainty equivalent and both parts of the stochastic alternatives. Figure 3.9 gives a schematic description of this question, and section 6.2 gives a much more complete development of this type of question.

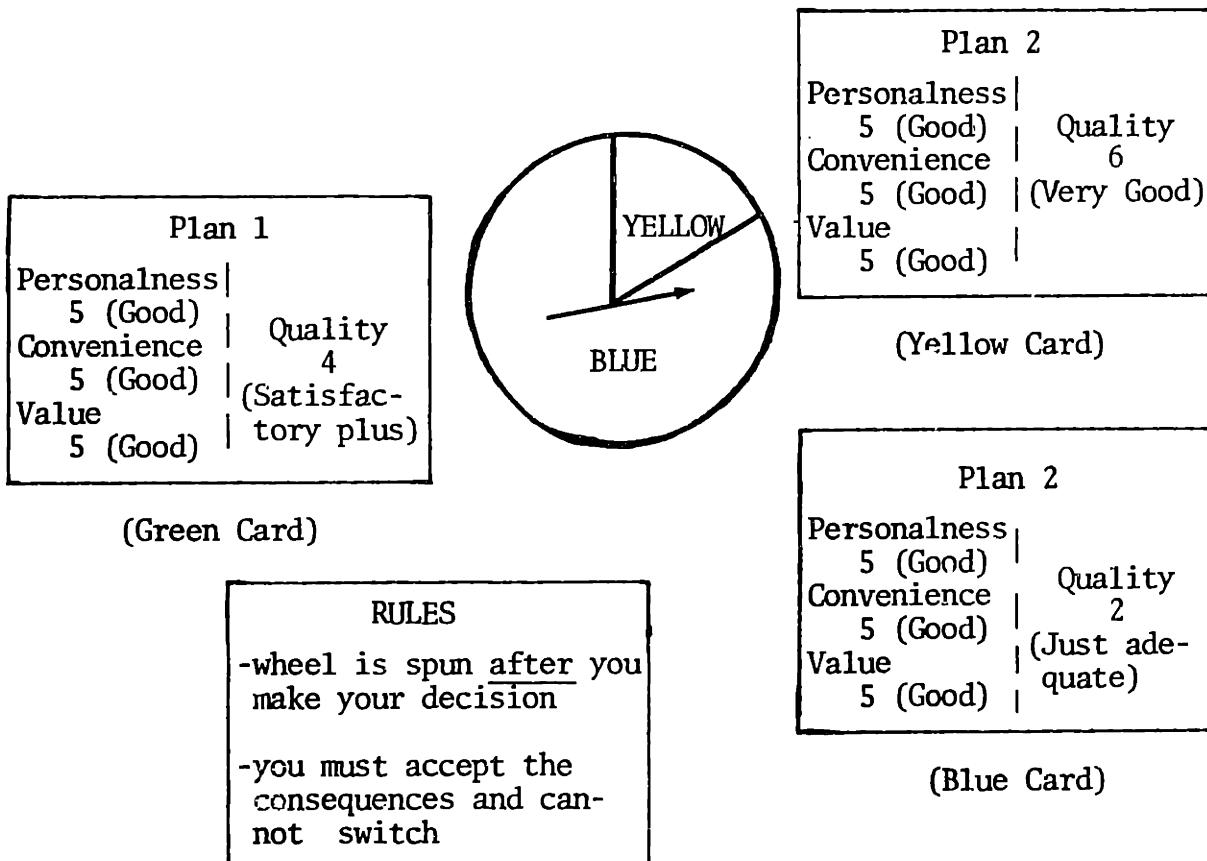
(Note that an alternative technique is to specify p and have the respondent set the value of the performance measures for the certainty equivalent.)

Independence questions: An independence question simply tests to determine if the values specified in either the tradeoff or risk questions are sensitive to the fixed and common values. For example in figure 3.8 one might change the fixed and common levels of value and convenience from 5(good) to 3(poor) and ask the respondent if this would change the indifference setting of quality.

Indifference questions are often very hard for the respondent to understand. Whenever they are used it is important to use simple educational questions to warmup the respondent before asking any of the more complex tradeoff, risk, or independence questions. For example in a warm-up risk question the respondent should be given some probability p_1 such that he definitely prefers the certainty equivalent and some probability p_2 such that he definitely prefers the stochastic alternative. He should then be led iteratively to an indifference setting and once there he should be challenged to make sure he truly understands the questions.

Instruction to Consumer:

Imagine you can only choose between two health plans, plan 1 and plan 2. In both plans personalness, convenience, and value are good (rated 5). You are familiar with plan 1 and know that quality is satisfactory plus (rated 4). You are not sure of the quality of plan 2. If you choose plan 2, then the wheel is spun and the quality you will experience for the entire year depends on the outcome of the wheel. If it comes up yellow, the quality is very good (rated 6) and if it comes up blue the quality is just adequate (rated 2). Graphically this is stated:



Instruction to Consumer:

At what setting of the odds (size of the yellow area) would you be indifferent between plan 1 and plan 2? (Respondent is given wheel and adjusts it until size of yellow area is appropriate. He is challenged by being given the choice with his setting. If he prefers one plan or the other, the interviewer iterates the question until a true indifference setting is determined.)

Figure 3.9: Schematic of Risk Aversion Question

(Note the challenging should continue in the actual data collection.)

In addition visual aids should be used whenever possible to complement the questions. For example it is much easier for a respondent to relate to an area on a two-colored wheel than it is to specify an abstract probability. In the HMO study we used a variable yellow and blue probability wheel and colored the two plan descriptions of the stochastic alternative yellow and blue respectively. The certainty equivalent was green.

The instructions to interviewers used to directly assess compaction functions in the HMO study illustrates many of these questions and techniques. (See appendix 2.)

3.3.4 Demographics and other Consumer Descriptors

To identify segments in abstraction and to adequately project from a sample population it is necessary to measure demographics and other consumer descriptors. For example, in the HMO study, patterns of health care utilization and satisfaction were measured in addition to demographics such as age, sex, family size, and health status.

3.3.5 Similiarity Measures

If multi-dimensional scaling techniques are used similarity judgements must be measured.

The most common way of obtaining similarity judgements is to specify one of the evoked alternatives and have the individual rank order the other evoked alternatives according to their "similarity" to the specified alternative. This is then repeated for all alternatives.

in the evoked set. The criteria for "similarity" is left up to the respondent but he is cautioned to maintain the same general frame of reference throughout the questions.

Similarity judgements can also be obtained directly through the use of an interval scale or indirectly by some "distance" metric in the space of attributes. See Green and Rao [48] for a more complete description of similarity judgements.

3.4 Conclusion of Observation of Consumers

This methodology is model oriented and models are dependent upon high quality data. This chapter began with a few general measurement guidelines: motivate the respondent to carefully consider each question, use the semantics use by consumers, pretest the questionnaire, and use a random unbiased sample for measurements.

The consumer is presented with a real or proxy choice and measurements taken only with respect to his evoked set. First measurements of a complete, spanning set of metric instrumental attributes are taken. Choice is measured and, if utility theoretic compaction models are to be supported, tradeoff, risk, and independence effects are measured. Finally, demographic and other consumer descriptors such as health care utilization are measured.

The next chapter begins modeling of the individual choice process by reducing the set of attributes to a more parsimonious set of performance measures.

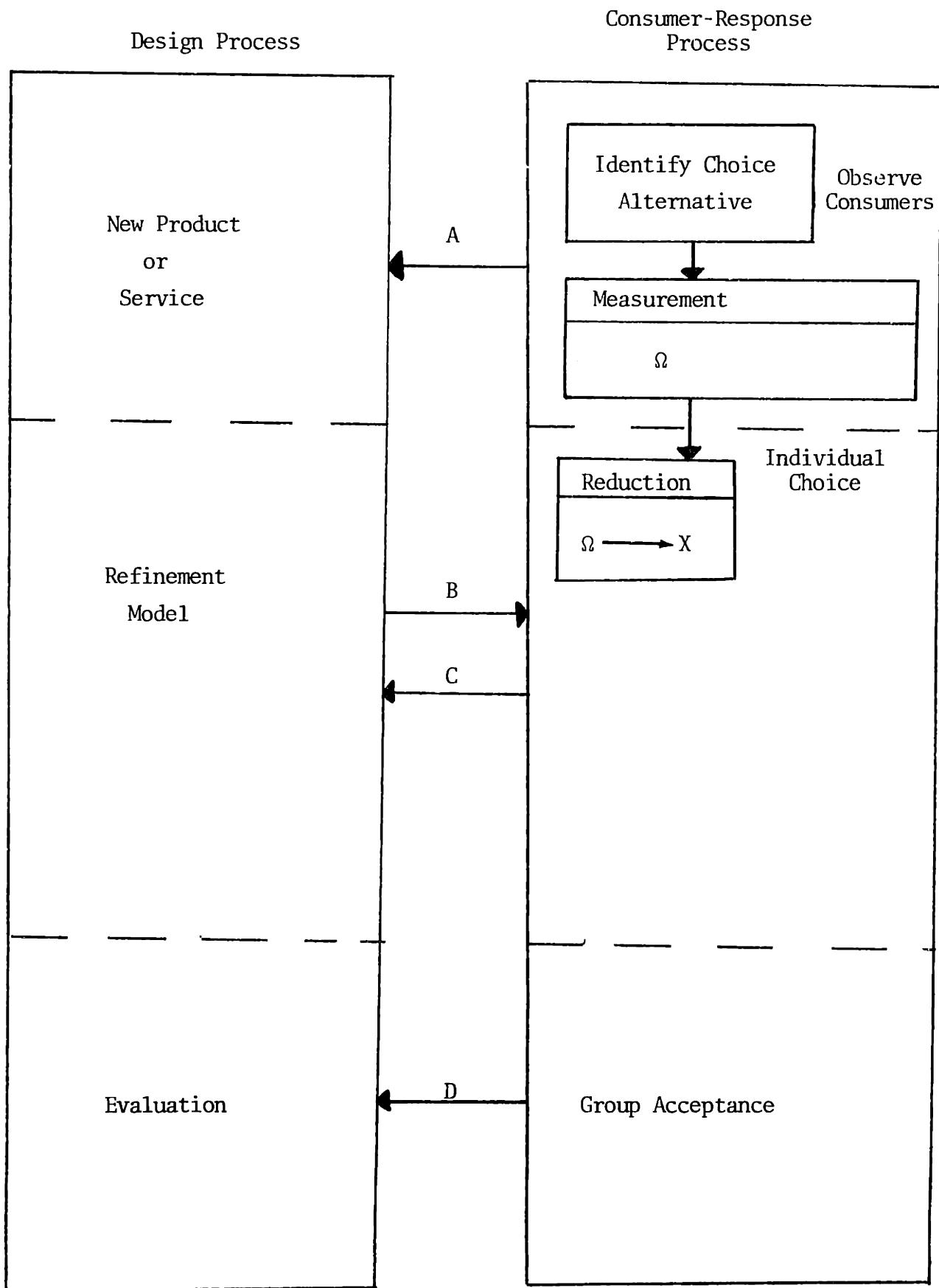
Chapter 4

REDUCTION

The last chapter emphasized the importance of identifying the attributes which individuals use to describe alternatives and consider when making choices among the alternatives. For some types of alternatives there may be only a few attributes but for most types of alternatives, especially major services such as transportation and health care, there will be many attributes. For example, when the choice was among health care delivery systems 16 attribute scales were identified (See figure 3.6).

Even for complex choice decisions it is doubtful that any given consumer will simultaneously and explicitly consider all the attributes in making a choice. Instead he will likely group together various sets of attributes to form a smaller set of performance measures which are easier for him to consider when making a choice. The process of identifying these performance measures is called reduction. (See figure 4.1) Note that the hypothesis here is not that he has an explicit method to do this, but instead that his resulting choice behavior can be described by this process. This is a subtle difference but an important one since most of the techniques described in this chapter uncover the latter but not the former.

Even if all consumers implicitly perform this grouping process, it is theoretically feasible to construct a choice model based on the underlying attributes. But even if the ultimate choice model is based



4.1: Relationship of Reduction to the Methodology

directly on the underlying attributes it is imperative that reduced performance measures be identified (if a reduction exists) because such a reduction can provide managerial insight into the choice process.

In other words, to elicit creativity and to guide the analysis and design process it is important that the model of individual choice provide diagnostics (arrow C in figure 4.1). Since the human mind can only process a few measures simultaneously, reduction identifies performance measures which structure the attributes and allow a manager or analyst to understand perceptions of the alternatives and to mentally manipulate various new concepts. Furthermore, a geometric representation of the data in a few dimensions allows visualization of the choice process and understanding of the market structure. Figure 4.2 shows the placement of automobile, bus, taxi, dial-a-ride, and subway in a hypothetical perceptual space of performance (travel time, wait time, etc.), comfort (seats, privacy, etc.) and cost.

One final reason for reduction is that some of the compaction techniques such as direct assessment have strong data requirements and are feasible in terms of respondent wearout only if the number of performance measures is reasonably small.

After introducing notation this chapter discusses some issues of reduction by defining some ideal properties. In most instances it is not possible to satisfy all these properties and hence some compromises are necessary to find a realizable technique. These tradeoffs and their implications are discussed. The standard marketing techniques of factor analysis and multi-dimensional scaling, as well as an information theoretic technique and an in depth interview exploratory technique are

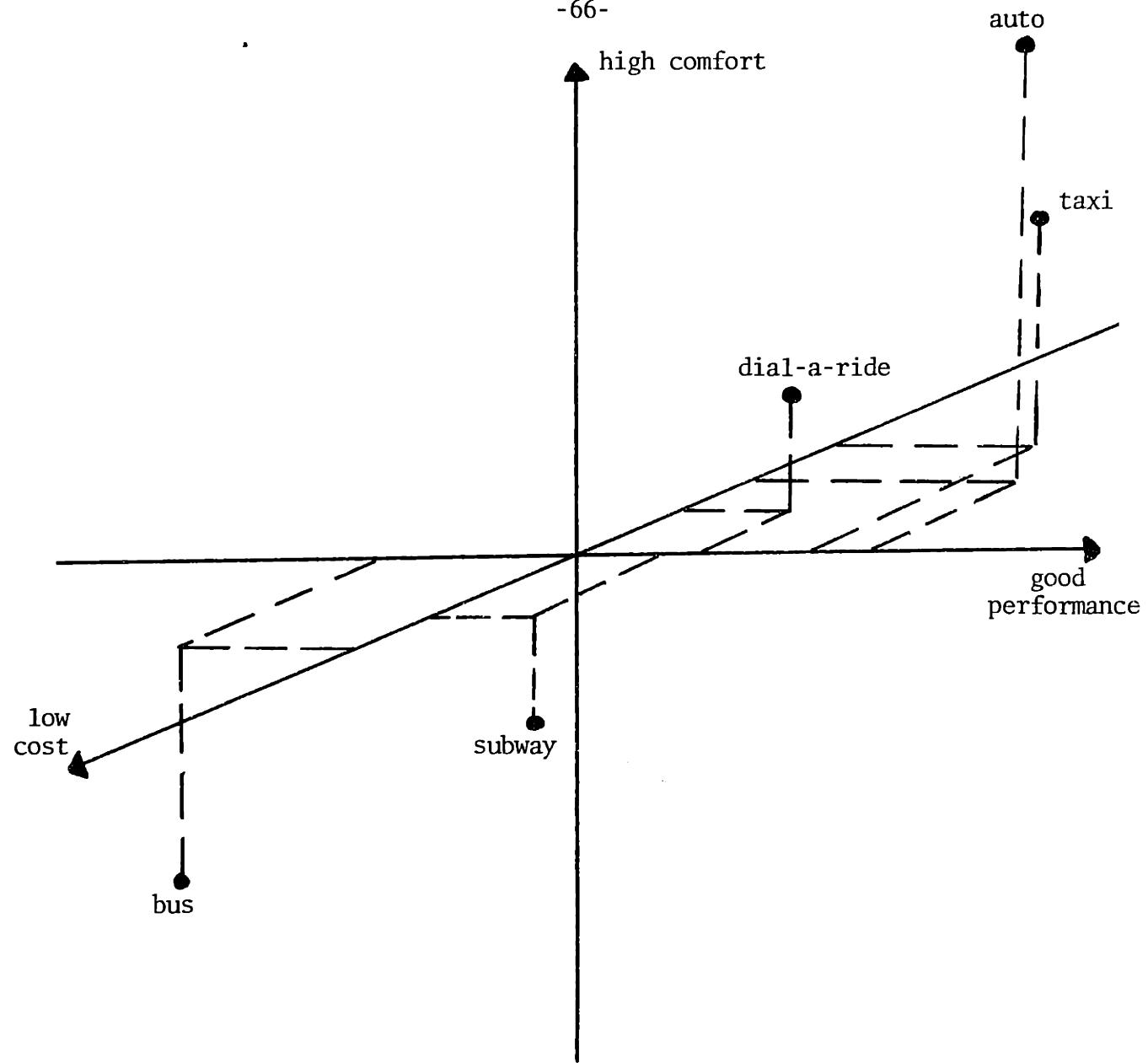


Figure 4.2: Hypothetical Perceptual Space for Transportation Modes

presented and their strengths and weaknesses discussed relative to the methodology.

4.1 The Issues of Reduction

4.1.1 Notation

Let Y_ℓ be an underlying attribute and let $y_{ij\ell}$ be the measure of that attribute as perceived by individual i for alternative a_j . Let X_k be a reduced performance measure and let x_{ijk} be the measure of that performance measure as perceived by individual i for alternative a_j . Let $Y = \{Y_1, Y_2, \dots, Y_L\}$ and $X = \{X_1, X_2, \dots, X_K\}$ be the sets of attributes and performance measures. Define $y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ijL})$ and $x_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijk})$. Let $x_{ijk} = (x_{ij1}, x_{ij2}, \dots, x_{ijk-1}, x_{ijk+1}, \dots, x_{ijk})$, that is the values for all the performance measures except x_{ijk} . Similarly define $y_{ij\bar{\ell}}$. A reduction, \underline{R} , is a mapping from Y into X , $\underline{R}: Y \rightarrow X$, which can be thought of as a set of single valued mappings, R_k , from Y into X_k , $R_k: Y \rightarrow X_k$. In some cases it is desirable that R_k be dependent on only a subset of Y , that is the performance measure X_k if determined by a subset of the attributes. Define $Y(X_k)$ as the subset of Y which determines that value of X_k , in other words R_k really maps $Y(X_k)$ into X_k , $R_k: Y(X_k) \rightarrow X_k$, and the mapping is independent of $(Y - Y(X_k))$. Be careful of this last notation which is set notation and not functional notation.

Perhaps an example will clarify this notation. Imagine you are taking a subway ride and perceive it as defined by (1) waiting time for the train to arrive, (2) travel time in the train, (3) payment when

boarding, and (4) payment when getting off, but imagine you only care about (1) total time and (2) total cost. Then $Y = \{\text{wait time, travel time, boarding charge, egress charge}\}$, $X = \{\text{time, cost}\}$, $Y(X_1) = \{\text{wait time, travel time}\}$, $Y-Y(X_1) = \{\text{boarding charge, egress charge}\}$. In this case R_K is the linear mapping, $x_{ij1} = R_1(y_{ij1}, y_{ij1}) = y_{ij1} + y_{ij2}$. For a trip on the MBTA from Kendall Square to Quincy perhaps $y_{ij} = (10 \text{ min.}, 30 \text{ min.}, 25 \text{ cents}, 25 \text{ cents})$ and then $x_{ij} = (40 \text{ min.}, 50 \text{ cents})$.

4.1.2 The Issues of Reduction: Formal Definitions

The following properties are presented here to raise some important issues to consider when choosing a reduction technique (or choosing whether or not to do a reduction). There is no currently available technique which can guarantee that all will hold simultaneously. Many of these properties are conflicting in that improving one may be detrimental to another, and as one might expect, most of the reduction models discussed in section 4.2 make such tradeoffs. It is the analyst's responsibility (in consultation with the design team) to understand these issues, to weigh their relative importance, and to choose the reduction technique which is appropriate to the choice situation being modeled and the design decisions under consideration.

Computable: A reduction is useful when it is possible to compute a metric which indicates how a consumer perceives a performance measure. Thus one desirable property of a reduction is that it is possible to determine the values of the performance measures, x_{ij} , given observations of the attribute measures, y_{ij} . Formally:

Definition 4.1: A reduction, R , is said to be computable if there exists a mapping from Y into X and if this mapping can be determined for all $y \in Y$.

Essential: If a performance measure has the same value for all currently available and potential alternatives, and if it does not effect the choice, then knowing the value of that performance measure for a given alternative will provide no insight into the choice process. Note that a performance measure might have the same value for all alternatives and still affect choice: consider for example weather affecting the choice between automobile and bus. Thus a performance measure is essential if there exists some conceivable choice situation which it can affect. Formally:

Definition 4.2: A performance measure, x_k , is said to be essential if there exists at least one individual, i , and some alternative set, A , containing at least two alternatives, a_j and a_ℓ , such that the value, x_{ijk} , affects the choice probability $p_i(a_j \text{ from } A)$.

Based on definition 4.2 it seems that almost every conceivable measure should be a performance measure. In reality some tradeoffs must be made and only those performance measures which strongly affect choice will be used.

Complete: Hopefully when the attributes are reduced, no information is thrown away, in other words the performance measures, X , carry as much information about choice as the attributes Y . An attribute set, Y , is complete (definition 3.1) if there is no attribute, Y_{L+1} , not in Y such that a choice probability model conditioned on $Y \cup Y_{L+1}$ yields different probabilities than a model conditioned on Y alone. A similar completeness property for reduction and the performance measures it produces is defined on the attributes, because it is the attributes which carry the information. Reduction is simply a clearer and more concise representation of this information. Formally:

Definition 4.3: Let Y be a complete set of attributes. A reduction, $R: Y \rightarrow X$, and the performance measures, X , that it produces is said to be complete if the values of X produce the same choice probabilities as the values of Y . That is for all individuals, i , and alternatives, a_j , if

$$p_i(a_j | x_{i1}, x_{i2}, \dots, x_{iJ}) = p_i(a_j | y_{i1}, y_{i2}, \dots, y_{iJ})$$

for all $x_{ij} \in X$, $y_{ij} \in Y$, $j = 1, 2, \dots, J$.

Again in most real cases, e.g., $L = 20$ and $K = 5$, completeness will be unrealizable, but it does provide an intuitive yardstick by which we can compare different reduction techniques. One approximation might be to require the ratio (or difference) of the above probabilities to be within ϵ of 1.0 (or 0.0). Another might be to require that the expected

mutual information between choice and X to be the same as the between choice and Y . See chapter 10 for discussion of the information test. Note that even if such tests could be formulated, they would probably not be practical since it is unlikely that both $p(a_j | \underline{x}_i)$ and $p(a_j | \underline{y}_i)$ would be calibrated. None the less this definition is important because compaction (chapter 6) can only be as accurate as the performance measures are complete.

Instrumental: The overall methodology is formulated to be normative, thus it is important not only that the performance measures, X , carry the information of the attributes, Y , but that it is possible for the manager or the design team to affect changes in the values of the performance measures. To do this, the performance measures must be computable from the attributes and the linkages from the attributes to the performance measures must be understandable to the analyst or the design team. Formally we will only require that the attributes be instrumental (section 3.1.1.1) , and that the reduction be computable (definition 4.1).

Definition 4.4: A set of performance measures, X , produced by a reduction, $R: Y \rightarrow X$, is said to be instrumental if R is computable and Y is instrumental.

Parsimonious: Reduction is performed primarily to provide insight to the design team, thus the number of performance measures should be as small as possible. The exact number depends on the choice situation but in most cases the goal is the order of 5 or less performance measures.

Formally we will say that a reduction is parsimonious if it is complete and if the number of performance measures is as small as possible:

Definition 4.5: A reduction, $R: Y \subseteq X$, is said to be parsimonious if R is complete, if all $X_k \subset X$ are essential, and there exists no other complete reduction, $R': Y \rightarrow X'$ such that the cardinality of X' is strictly less than the cardinality of X .

Disjoint : Often a performance measure, X_k , is only determined by a subset of the attributes, $Y(X_k) \subseteq Y$. It is useful both for insight and for prediction if there is no overlap in those subsets, i.e., if a given attribute affects one and only one performance measure. Stated another way, it is often desirable that the attributes be partitioned into K classes such that each class determines uniquely the value of one performance measures. (See figure 4.3.) Formally:

Dinition 4.6: A reduction, $R: Y \rightarrow X$ is said to be disjoint if $Y(X_k) \cap Y(X_\ell) = \emptyset$ for all $k \neq \ell$, where $R_k: Y(X_k) \rightarrow X_k$ is independent of $Y - Y(X_k)$ for all k .

The second part of definition 4.6, which is true by the choice of notation $Y(X_k)$, is stated here for emphasis. Note that definition 4.6 allows the values of X_k to be statistically correlated with the values of X_ℓ across individuals or alternatives, as long as the sets of attributes affecting each are disjoint.

Encompassing: It is possible (and a priori highly likely) that the reduction for individual 1 may be different than the reduction for individual 2. Unfortunately without individual specific measurement it is impossible to determine an individual specific reduction. Thus one

assumption that must be judgementally considered is whether the same reduction applies to all individuals in a given identifiable population

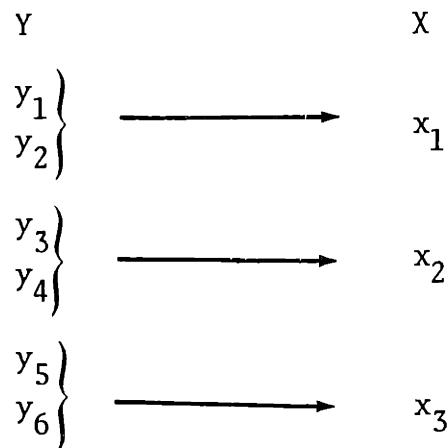


Figure 4.3: Example of a Disjoint Reduction

segment. The tradeoff here is that a general reduction is not as intuitively pleasing as a individual specific one, but it requires a good deal less data. Formally we define encompassing by:

Definition 4.7: A reduction, $R: Y \rightarrow X$, is said to encompass a group of individuals, S , if the mapping is the same for all individuals, i , contained in S . That is, if $y_{ij} \equiv y_{\ell j}$ for $i, \ell \in S$ then $x_{ij} \equiv x_{\ell j}$. Note that encompassing does not imply that the values of the performance measures are the same for all individuals in S , but only that the mapping, i.e., transformation of Y into X , is the same for all individuals.

4.1.3 Reduction is an Engineering Technique: Tradeoffs Must be Made

The last section identified a series of ideal properties that should hold if we are to have unassailable faith in a reduction. Perhaps in the future someone will discover a reduction technique that satisfies

them all, but right now when choosing a reduction technique (or whether to reduce) some engineering judgements need be made. Depending on the application, some properties may be relaxed. For example, a more parsimonious description (say in three dimensions rather than four) may be possible if some completeness is sacrificed (say 96% "explanation" rather than 100% "explanation"). If such a representation yields more insight to the decision maker and if we are in an early design stage which does not yet require perfect prediction then the choice is clear, but in another instance the choice may not be as clear.

In making a choice between reduction techniques the analyst should keep in mind the 7 yardstick properties (computable, essential, complete, instrumental, parsimonious, disjoint and encompassing) but he must also consider (1) the cost of measurement, (2) the feasibility of measurement (e.g., a long questionnaire may not be accurately answered) (3) the complexity of the resulting model, (4) the feasibility of complementary models (e.g., compaction, abstraction, and probability of choice) and (5) the potential for providing insight to and eliciting creativity from the design team. The following are three examples of the many approximations available to an analyst.

Approximate identity: Some techniques allow for overlaps in the sets $Y(X_k)$. Such techniques may make it more difficult to establish a unique causality, e.g., changing hospital affiliation of an HMO may effect a performance measure called quality but it also may effect one called personalness. In this case, the analyst or design team may decide that the performance measures defined on the overlapping sets provides

more insight than those defined on disjoint sets. Still, it may be that the sets are almost disjoint, that is, each attribute effects one performance measure strongly and all others only slightly. Thus, in a fuzzy way, insight is provided as if the sets were disjoint, but the mathematics of prediction is based on the overlapping sets.

Approximate micro-structure: Especially in exploratory stages it is more important to identify and understand the performance measures than to know their exact values. Thus early in the analytic process the analyst may use a reduction technique which identifies the performance measures and identifies which attributes affect which performance measures even if the technique approximates the mathematical mapping from Y to X . For example, a linear approximation to a non-linear mapping may identify the performance measures, X , and the sets $Y(X_k)$ but only approximate the values for x_{ijk} .

Neglect unimportant attributes: An alternative to a transformation-like reduction may be to truncate the set Y . That is to determine some set of attributes, X , which is a subset of Y , such that only the attributes which (by some measure) affect choice the most are in X . More on this in section 4.2.3.

The above are just a few of the engineering judgements facing the analyst and the design team. The next section presents a number of reduction techniques. All make such approximations. Keep this in mind in judging the techniques and in choosing the most appropriate one for a given choice situation.

4.2 Analytic Techniques for Reduction

This section presents the following reduction techniques; factor analysis, multi-dimensional scaling, information theory, and in depth utility theoretic analysis. The assumptions of each technique, the data requirements, and the potential use of each technique are discussed. An empirical example is given of the application of factor analysis to determining the performance measures to describe health care delivery systems.

4.2.1 Factor Analysis: An Exploratory Technique to Identify Structure

Suppose there exists a reduction $R: Y \rightarrow X$, suppose $Y(X_1) = \{Y_1, Y_2, Y_3\}$, that is $x_{ij1} = R_1(y_{ij1}, y_{ij2}, y_{ij3})$, and suppose x_{ij1} is monotonically increasing in y_{ij1}, y_{ij2} , and y_{ij3} . Then one would hope that X_1 would be highly correlated with Y_1, Y_2 , and Y_3 , but less so with Y_ℓ for $\ell > 3$. Of course this might not be the case (e.g., if $\text{cor}(Y_5, X_1) = .99$) but in an exploratory stage it is useful to look at the correlation between some set of performance measures and the attributes. Now if we knew the performance measures and their values for each i, j we could simply look at the correlation between X_k and Y_ℓ for all k, ℓ to determine the structure. But we do not know the values of the performance measures! Instead we can approximate the micro-structure with a linear model using common factor analysis and varimax rotation to yield estimates of the number of performance measures, their values, and the correlations between the performance measures and the attributes. Once the structure is obtained, and the performance measures identified, more elaborate

techniques can be used to complement the linear models. We begin with a presentation of some of the analytic details of common factor analysis. The theory is as presented in Rummel [126], for more details, proofs, examples, and for descriptions of other related techniques see Rummel [126] or Harman [52].

4.2.1.1 Analytic Details of Factor Analysis

Factor analysis starts with a data matrix \underline{Y} of the values of the attributes on a number of stimuli. Ideally, reduction could be individual specific in which case the factor analysis would be across all alternatives (a_j) for each individual (i), (\underline{Y}_i = matrix of $y_{ij\ell}$ with L rows and J columns). In most cases there will not be enough alternatives for this model, in which case factor analysis could be across alternatives and individuals (\underline{Y} = matrix of $y_{ij\ell}$ with L rows and $N \cdot J$ columns).¹ The following discussion holds for both of the above options, but for notational simplicity, i will be used to index stimuli (\underline{Y} = matrix of $y_{i\ell}$ with L rows and N columns).

Common factor analysis assumes there exist some common factors (performance measures), X , in the data matrix, \underline{Y} , with each X_k being some possibly non-linear function of the attributes, Y . It further assumes that each attribute can in turn be represented by a common portion which is a linear combination of the common factors, plus an attribute specific portion, plus a random error (See figure 4.4). That is

$$y_{i\ell} = \sum_k f_{k\ell} x_{ik} + u_{i\ell} + \varepsilon_{i\ell}$$

where u_{il} = specific portion
 ϵ_{il} = random error
 $f_{\ell k}$ = factor loading of factor X_k on attribute Y_ℓ
 x_{ik} = factor score of X_k for stimulus i



Figure 4.4: Common and Unique Factors

Without loss of generality we can assume that each X_k and Y_k are standardized across individuals, that is $\bar{x}_{ik} = 0$ and $\text{var}(x_{ik}) = 1.0$ for all k , and $\bar{y}_{il} = 0$ and $\text{var}(y_{il}) = 1.0$ for all ℓ . Define u_ℓ^2 as the portion of the variance of explained by the unique factor of Y_ℓ and h_ℓ^2 as the portion explained by the common factors. (Note $h_\ell^2 = 1 - u_\ell^2$). Define U^2 as the diagonal matrix of u_ℓ^2 and H^2 as the diagonal matrix of h_ℓ^2 . Let $r_{\ell m}$ be the correlation coefficient of Y_ℓ with Y_m , let R be the matrix of $r_{\ell m}$. Since Y_ℓ is standardized, $R = \underline{Y}^T \underline{Y}$. (\underline{Y}^T is the transpose of \underline{Y} .) Finally, let F be the matrix of factor loadings, $f_{\ell k}$, and I = the identity matrix.

Rummel then shows that:

$$FF^T = R - U^2 = \underline{Y}^T \underline{Y} - I + H^2$$

This equation, known as the fundamental theorem of factor analysis, states that the factor loadings can be found by "factoring" the data correlation

matrix with the communalities, h_ℓ^2 , replacing the 1's in the main diagonal. Now by linear algebra we know that a symmetric matrix ($R - U^2$ is a symmetric) can be factored by its eigenvectors $E = [e_1, e_2, \dots, e_L]$ and a diagonal matrix of its eigenvalues, Λ :

$$R - U^2 = E \Lambda E^T = (E \Lambda^{\frac{1}{2}}) (\Lambda^{\frac{1}{2}} E^T)$$

Now the rank of $R - U^2$ will be the number of factors, K , thus if the eigenvalues, λ_ℓ , are ordered by size, then $\lambda_\ell = 0$ for $\ell > K$. Thus one possible factoring is $F = E \Lambda^{\frac{1}{2}}$ where the zero valued rows are deleted. (This is the principal axes technique.) Note that this is not a unique factoring since any similarity transform T can be applied to $E \Lambda^{\frac{1}{2}}$, i.e.:

$$R - U^2 = (T E \Lambda^{\frac{1}{2}}) (\Lambda^{\frac{1}{2}} E^T T^T)$$

More on this later.

A fundamental paradox of factor analysis is the need to know the communalities, H^2 , before factoring, but the communalities can not be calculated before knowing the factoring. Fortunately this problem has been studied extensively and there exists bounds and estimation techniques for H^2 , (See Rummel [126].)

Clearly in any empirical situation, λ_ℓ will not be identically zero for $\ell > K$. How can we determine the appropriate number of factors, K ? From linear algebra, the eigenvectors are orthonormal and the trace of a matrix (sum along the main diagonal) equals the sum of the eigenvectors thus:

$$\text{trace } (R - I + H^2) = \text{trace } (H^2) = \sum_{\ell} h_{\ell}^2 = \sum_{\ell} \lambda_{\ell}$$

In other words, the sum of the eigenvalues is equal to the common variance. Thus we can tradeoff the number of factors (parsimony in the performance measures) with the percent of common variance explained (one measure of completeness in the performance measures) in choosing the "best" set of performance measures.

In section 4.1 it was indicated that a desirable property of the performance measures was that $Y(X_k)$ be disjoint for all k . Interpreting this in factor analysis we require that each Y_{ℓ} be highly correlated with only one factor. Now for orthogonal (uncorrelated) factors such as obtained by principal axes the factor loading, $f_{\ell k}$, equals the correlation between Y_{ℓ} and X_k . Thus to obtain the desirable property of

disjoint $Y(X_k)$'s we want to choose the transformation of $E\Lambda^2$ which causes each attribute to load heavily on a single factor.

In factor analysis terms this is called obtaining simple structure and the transformation technique is called varimax rotation. (Varimax rotation attempts to maximize the variance in the columns of the squared factor loading matrix since the highest variance is obtained when each Y_{ℓ} is correlated with only one factor, X_k .) The transformation is called a rotation because geometrically it acts to rotate the principal axes in X space. For example, suppose there were two factors, X_1 and X_2 , and ten attributes, Y_1 to Y_{10} . Figure 4.5a represents one geometric interpretation of the factor loadings. Notice that all of the attributes are correlated with both factors, but if the axes are rotated as in figure 4.5b, $Y_1, Y_2, Y_3, Y_9, Y_{10}$ are highly correlated only with X_1 '

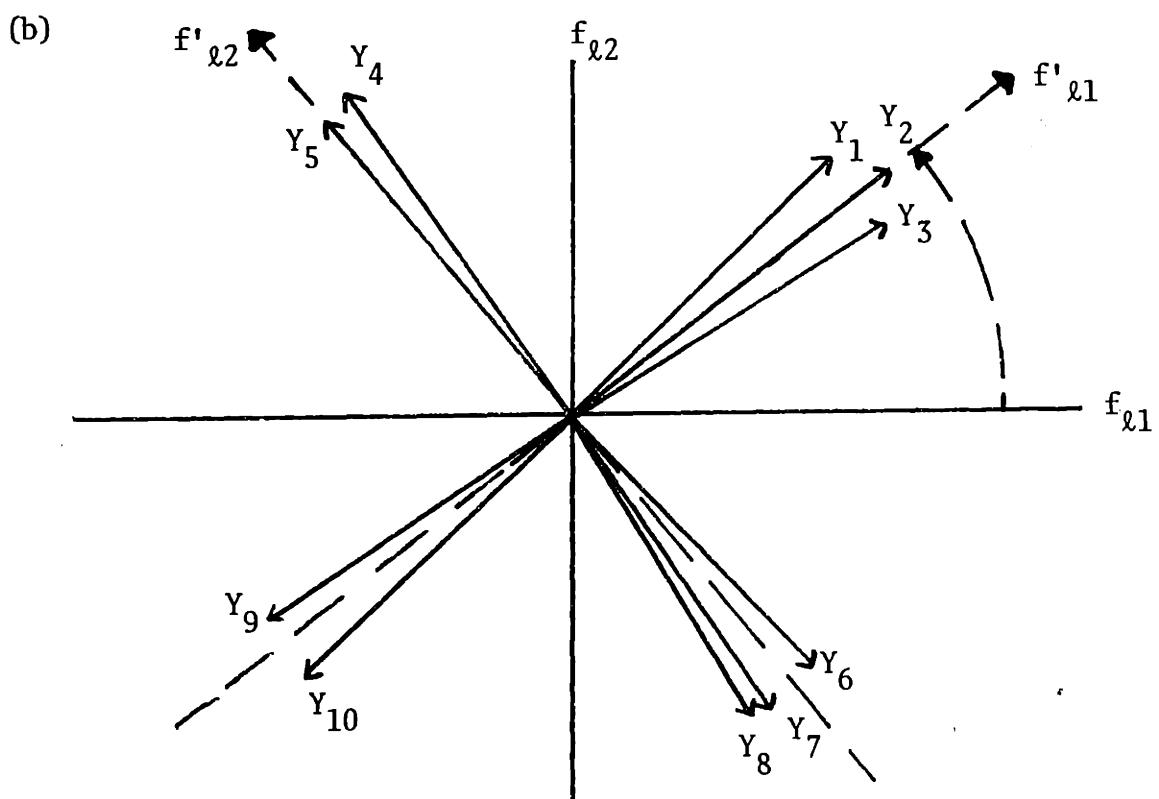
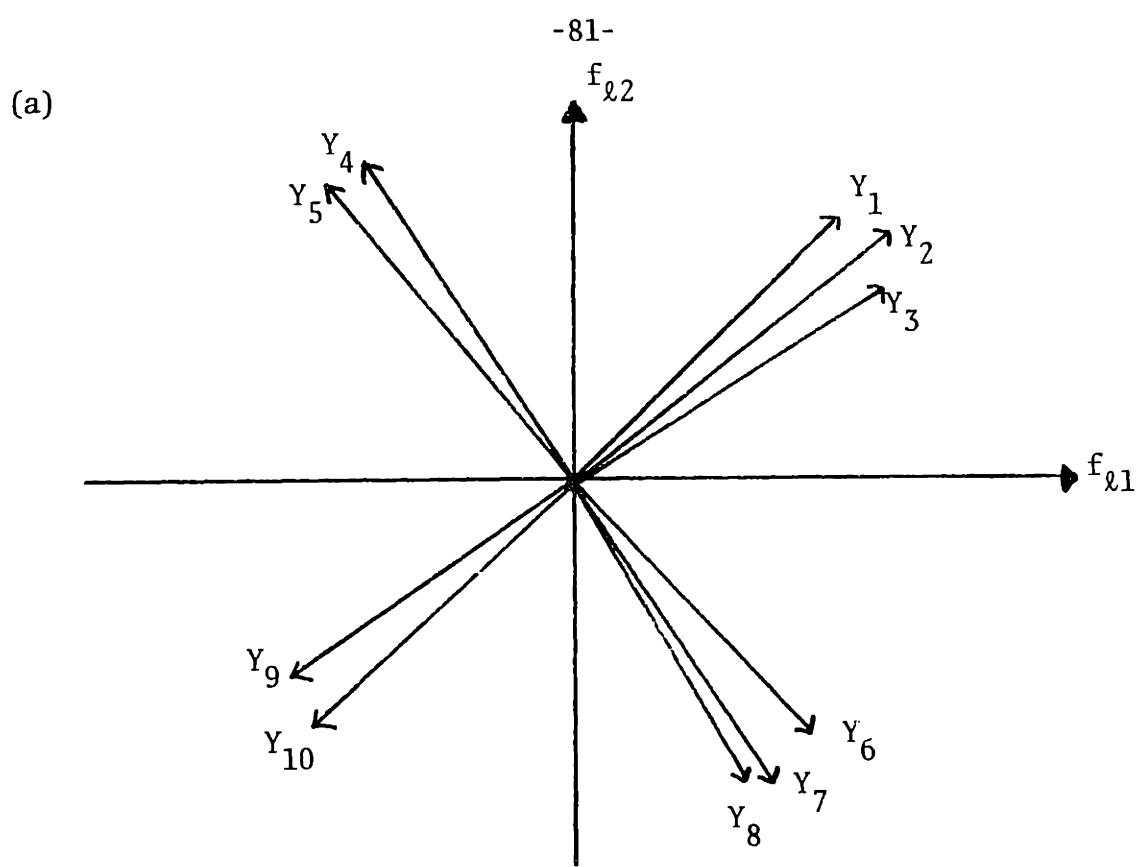


Figure 4.5: Rotation of Common Factors

and Y_4, Y_5, Y_6, Y_7, Y_8 with X_2' .

As presented thus far, factor analysis can determine the number of performance measures and can produce a factor loading matrix which indicates the correlation between each attribute, Y_ℓ , and each performance measure (factor) X_k . To make use of the performance measures we need to know their values, \underline{x}_i for a given set of values of the attributes, \underline{y}_i . If F were of full rank (non-singular) and if there were no unique terms and no error terms then the values of the performance measures could be obtained uniquely, i.e., $\underline{X} = \underline{Y}(F^T)^{-1}$.

Since factor analysis is being used as a reduction technique we expect that $K < L$, hence F will not be of full rank and there will be unique terms and error terms. An alternative technique is to obtain the estimates of the performances measures (factor scores) by multiple linear regression, i.e., estimate x_{ik} by:

$$x_{ik} = \sum_k \beta_{k\ell} y_{i\ell} + \text{error}$$

If B = matrix of $\beta_{k\ell}$ then it can be shown (Rummel [126]) that the least squares estimate of B is $B = (\underline{Y}^T \underline{Y})^{-1} F$, and hence the estimate of \underline{X} is given by $\hat{\underline{X}} = \underline{Y}(\underline{Y}^T \underline{Y})^{-1} F$. A similiar result can be obtained for a rotated factor loading matrix.

To summarize the analytics: (1) factor analysis assumes a linear model $\underline{Y} = \underline{X}F + U + \text{error}$, (2) the factor matrix F is obtained by taking the eigenvectors and eigenvalues of $(R - I + H^2)$, i.e., $F = (E\Lambda^{\frac{1}{2}})$ truncated, (3) F is then rotated to obtain the simple structure of each attribute, Y_ℓ , correlated with as few performance measures as possible,

i.e., $F' = T(E\Lambda^{\frac{1}{2}})$, and finally (4) the individual specific values of the performance measures are estimated by linear regression, i.e., $\hat{X} = \underline{Y}(Y^T Y)^{-1} F'$. The important outputs are the rotated factor loading matrix, F' , whose elements, $f_{\ell k}$, give the correlation of attribute, Y_ℓ with performance measure, X_k , and the estimates of the values of the performance measures, \hat{x}_{ik} .

4.2.1.2 Interpretation of Factor Analysis and Its Use in the Methodology

Factor analysis is primarily an exploratory technique to search for structure in the mapping of the attributes into the performance measures. It approximates the true mapping with a linear mapping and determines the number of performance measures and their correlations with the attributes. Because the true structure is only approximated, and because the attributes are intercorrelated, the resulting correlations of Y_ℓ to X_k will be blurred and thus interpretation of the factor analysis becomes an art.

The analyst should first study the rotated factor loadings, $f_{\ell k}$, note which attributes load heavily on which performance measures, and then try to name and identify the performance measures. His output from this stage is a structure matrix such as shown in figure 4.6. For example in figure 4.6 the analyst has determined that $Y(X_1) = \{Y_1, Y_2\}$. That is, there exists some reduction $R_1: \{Y_1, Y_2\} \rightarrow X_1$, i.e., some function R_1 such that $x_{i1} = R_1(y_{i1}, y_{i2})$.

Depending upon his judgement he can (1) use the linear estimates of the values of the performance measures as obtained from factor analysis,

\hat{x}_{ik} , or (2) assume that the true value, x_{ik} , is some individual specific function of the estimated value, i.e., $x_{ik} = f_i(\hat{x}_{ik})$, or (3) distrust the linear approximation and use some other technique to determine the function $R_k(\cdot)$. The first option is probably the most desirable if no additional data or measurement capability is available. Both the second and the third options require that either the performance measure can be named, identified, and remeasured (the respondent can perceive its value directly) or that there is some external justification for determining the function $f_i(\cdot)$ or $R_k(\cdot)$.

Perhaps this is a bit complex in the abstract. The following empirical example should illustrate some of the options in interpreting a factor analysis.

	name ₁	name ₂	name ₃
	x ₁	x ₂	x ₃
y ₁	x		
y ₂	x		
y ₃		x	
y ₄			x
y ₅			x

Figure 4.6: Structure Matrix

4.2.1.3 Reduction by Factor Analysis: HMO Case

As indicated in section 3.3.1.2, there are 16 attributes useful in describing peoples' perceptions of health care delivery systems. Individuals (447) rated each of four plans on 5 point agree/disagree scales describing each of these 16 attributes. Some data was saved and common factor analysis with varimax rotation² was used to analyze the remaining data. (1200 stimuli = (300 people) x (4 plans)). The result was four factors which explained 97% of the common variance and 55% of the total variance. By examining the high loadings on each performance measure they were named (1) quality, (2) personalness, (3) value, and (4) convenience. Quality correlated to trust, preventive care, availability of good doctors, and hospitals. Personalness reflected a friendly atmosphere with privacy and no bureaucratic hassle. Value was not just price, but rather paying the right amount for the services. Convenience reflected location, waiting time, and hours of operation.

Two techniques were then used to estimate the values of the performance measures. The first was to use directly the estimates obtained through factor analysis. These individual specific estimates were later used in a statistical compaction technique (section 6.4.1) but before that their average values were plotted. (These are shown in figure 4.8.)

First, it should be noted that existing care is perceived as quite good, especially on personalness and value, this as might be expected. Harvard Community Health Plan (HCHP) received a significantly higher rating on quality than did either the M.I.T. HMO concept or the

<u>ATTRIBUTE SCALE</u>		QUALITY	PERSONAL	VALUE	CONVENIENCE
1	DAY + NIGHT CARE	0.37244	0.07363	-0.31379	<u>0.63939</u>
2	WAITING TIME	-0.22082	0.26204	0.15514	<u>-0.64870</u>
3	TRUST-GOOD CARE	<u>0.72125</u>	-0.21828	-0.09556	0.24708
4	LOCATION	0.01144	0.24706	-0.12544	<u>-0.72964</u>
5	PRICE/VALUE	0.03066	0.12810	<u>0.72884</u>	-0.09961
6	FRIENDLY/PERSONAL	0.40986	<u>-0.51317</u>	-0.12285	0.18768
7	PREVENTIVE CARE	<u>0.55403</u>	-0.14187	-0.44353	-0.03653
8	EASILY FIND GOOD MD	<u>0.64412</u>	-0.15036	-0.21491	0.27113
9	MODERN TREATMENT	<u>0.72288</u>	-0.13441	-0.15906	0.08018
10	ACCESS TO RECORDS	0.43412	<u>-0.49053</u>	0.18749	-0.05992
11	CONTINUITY OF CARE	0.20491	0.47900	0.47727	0.04725
12	ASSOCIATED HOSPITALS	<u>0.68006</u>	-0.08256	0.10854	0.00555
13	USE OF PARAMEDICALS	-0.05303	<u>0.67083</u>	0.12888	0.16722
14	ORGANIZED/COMPLETE	<u>0.47725</u>	0.01627	<u>-0.52893</u>	0.14316
15	HASSLE/REDTAPE	-0.13081	<u>0.69824</u>	0.11180	-0.27903
16	COMPETENT MD'S	<u>0.73953</u>	-0.19335	-0.13971	0.18691

* see Figure 3.6 for field rating scale descriptions

4.7: Factor Loadings for Health Care Plans

M.I.T. HMO pilot. This based almost entirely on a low score for M.I.T. on Y_{12} , hospital quality. When you consider that HCHP is associated with the high prestige Boston hospitals like Peter Bent Brigham and the M.I.T. HMO is associated with the smaller, not well known Cambridge hospitals like Cambridge City Hospital and Mount Auburn Hospital this result is not surprising. When comparing the M.I.T. HMO concept with the M.I.T. HMO pilot, a definite perceptual gap is seen. This can be post purchase rationalization but it can also be a good product that is poorly communicated. (The stated intent to re-enroll of over 90% supports the latter hypothesis.) Thus, based on observing perceptions, management is alerted to the potential of shifting hospital affiliation and of more active promotion of the M.I.T. HMO concept.

In chapter 6 a technique is presented to directly assess the importance of each of the four performance measures. This technique requires measurement of direct perceptions of the semantics quality, personalness, value, and convenience aided by dimension descriptions based on the heavy loading attributes. We felt that the factor scores would not adequately reflect direct perceptions on these semantics, but factor analysis was our only managerial link back to the underlying 16 attributes. Instead perceptions of the four performance measures were remeasured and these will be correlated to the factor scores.

4.2.1.4 Relative Advantages of Factor Analysis

Advantages: Although factor analysis requires metric input, its data requirements are feasible. In fact it requires no more than ratings for the underlying attributes. It is an excellent exploratory

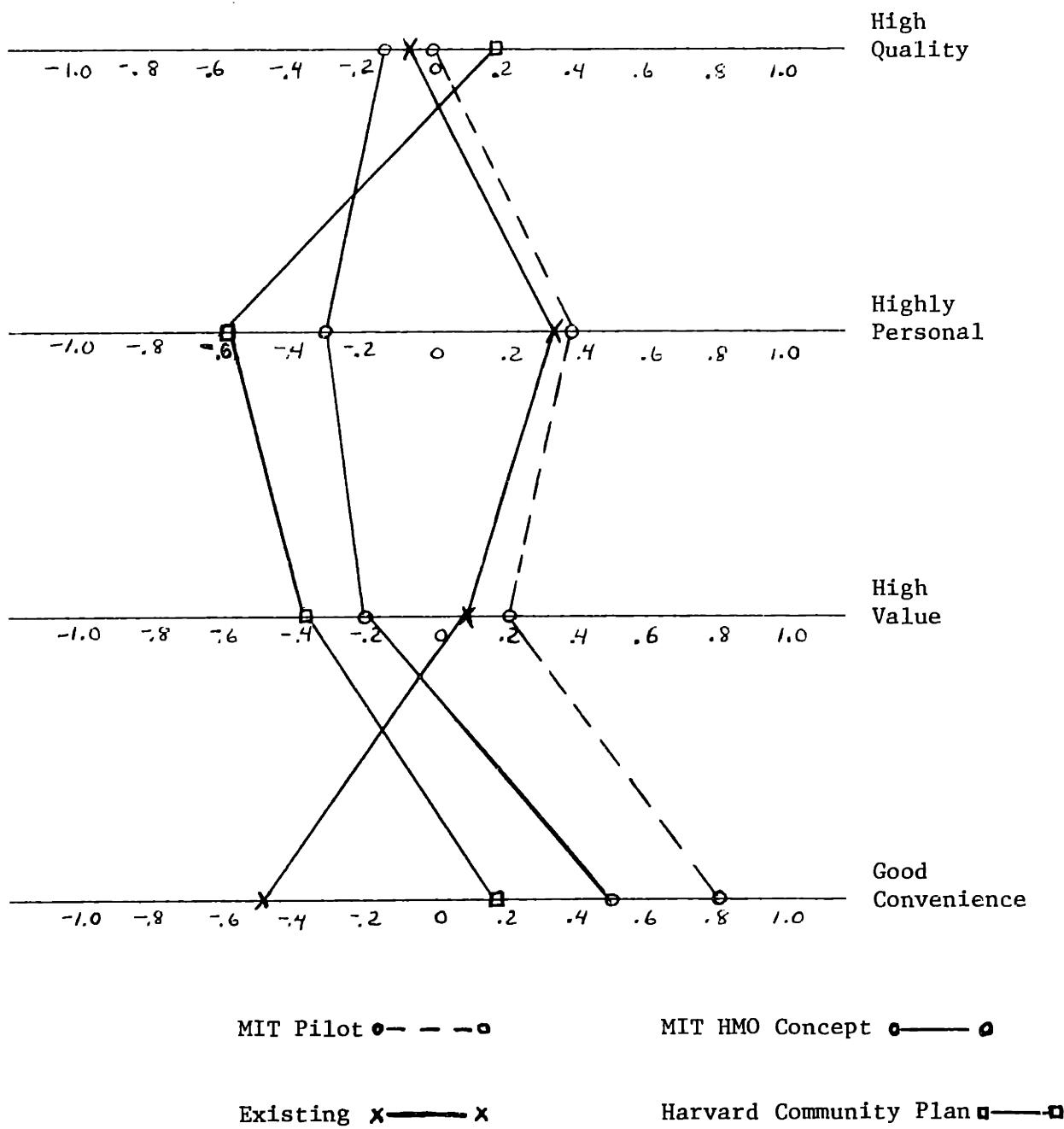


Figure 4.8: Perceptual Space for Health Care Plans

technique providing insight not only on the number of performance measures but also their identity. Furthermore it provides an estimate of the values of the performance measures. It is easy to implement with standard statistical computer packages.

Disadvantages: It approximates the structure with a linear model and it obtains estimates of the performance measures with linear regression. Because of degrees of freedom it is run across individuals rather than being individual specific. (To check sensitivity in the HMO case, factor analyses were run within population segments. The resulting factors had similar structure, i.e., four factors explaining roughly 55% of total variance with mostly the same loadings, but the detailed loadings were numerically different.)

Factor analysis is done on existing concepts and could leave out an entirely new dimension, thus care must be taken to ensure that the existing concepts are diverse enough to span the space of attributes. Also by design, factor analysis investigates the correlation in perceptions of the attributes, arguments can be made that this is not causal and might well miss the true grouping of attributes into perception when choice is to be made.

Overall, despite its faults, factor analysis is a useful, currently available investigative tool for reduction.

4.2.2 Multidimensional Scaling: An Exploratory Technique to Identify the Structure of Similarity

Factor analysis began with the attributes and searched for performance measures with which the attributes were correlated. An

alternative technique is to obtain measures of the similarities between alternatives in the evoked set and use multi-dimensional scaling (MDS) to look for a K-dimensional vector space representation (alternatives are points in this space) where the interpoint distance best recovers the similarities between alternatives. Then by observing the derived locations of alternatives in this space it might be possible to infer the meaning of and name the dimensions. As before once the dimensions are identified, more complex techniques to determine the reduction mappings, R , can be used to complement MDS.

This section describes the MDS techniques based on similarities. There are other techniques based on directly obtained preference judgements. The interested reader is referred to Green and Rao [48] for complete discussion of these techniques. We begin with an indication of the basic elements of MDS techniques. The complete theory with example is presented in Green and Rao.

4.2.2.1 Basic Analytic Details of Multidimensional Scaling

Basically MDS searches for the values of the performance measures, x_{ijk} , for individual i such that the "distances" between alternative, a_j , represented by points, x_{ij} , in the K-dimensional space of performance measures "best" recovers dissimilarities, $\delta_{ij\ell}$, which are measures of how differently individual i views a_j and a_ℓ . For convenience we will drop the i subscript, but remember that MDS can be used for each individual as well as for average dissimilarities.

Similarities: Similarities (or dissimilarities) can be metric (numerical value) or non-metric (rank order) and they can be

directly obtained (judgements by individuals) or derived (squared distance in the attribute space). Various programs can accommodate various combinations of the above.

Distances: Interpoint distances are almost always computed via a p-metric in the space of performance measure. I.e., if x_{jk} = the value of performance measure X_k for alternative a_j then the perceptual distance $d_{j\ell}$, from a_j to a_ℓ is given by:

$$d_{j\ell} = \left[\sum_k |x_{jk} - x_{\ell k}|^p \right]^{1/p}$$

Notice if $p = 2$ then the Euclidean metric results and if $p = 1$ then right angle distance results.

Objective: The objective of each technique is to obtain the values of the performance measures, $(x_{j1}, x_{j2}, \dots, x_{jK})$, so as to represent each alternative, a_j , in K-dimensional space, such that the value of some function of the interpoint distances and dissimilarities is minimized.

Stress: The objective function differs from technique to technique. A typical example is Kruskal's [85] stress formula 2 used in M-D-SCAL V:

$$S = \sum_{j<\ell} (d_{j\ell} - \hat{d}_{j\ell})^2 / \sum_{j<\ell} (d_{j\ell} - \bar{d})^2$$

where \bar{d} is the average of all $d_{j\ell}$ and $\hat{d}_{j\ell}$ is a number chosen to be as

close to d_{jl} as possible subject to maintaining the original monotonicity of the dissimilarities, δ_{jl} . Computationally the algorithm starts with an initial positioning of the points and perturbs the configuration to move in a direction to minimize S . (A steepest descent gradient method is used.)

Other techniques such as TORSCA 8 [48], PARAMAP [48], and INDSCAL [48] differ in either the objective function, the distance metric, or the input data, but the basic principles are the same as in M-D-SCAL V.

4.2.2.2 Use of Multidimensional Scaling in the Methodology

Like factor analysis, MDS is primarily an exploratory technique to uncover the number and identity of the performance measures. If we have no a priori knowledge of the attributes it is a good technique for uncovering the performance measures directly. Even if the attribute ratings are available, the analyst could separately obtain similarity ratings and use an MDS technique to obtain perceptual maps of the alternatives. He should attempt to do so in a variety of dimensions and choose the representation which is most interpretable. Then by examining the location of the alternatives in the perceptual space he can infer the identity of the performance measures. Since MDS provides no direct linkage from the attributes to the performance measures, the analyst must obtain the reduction, $R: Y \rightarrow X$, by other means. For example, by regressing X_k against Y .

Advantages: MDS does not assume any direct structure of the relationship between the attributes and the performance measures, but instead uncovers the performance measures in such a way that they recover

directly obtained similarities. It is a useful exploratory technique which can be used to complement factor analysis or to search for structure when the analyst has no strong a priori convictions.

Disadvantages: The main disadvantage of MDS is the difficulty in providing an instrumental link from the attributes to the performance measures. It is often more difficult to identify the dimensions based on locations of the stimuli rather than the correlation of the performance measures with the attributes. Finally in order for MDS to be statistically significant the evoked set must be of a size ≥ 8 , much larger than is usually the case.

Overall MDS is useful to explore structure, but care must be taken to provide external linkages from the attributes to the performance measures so that predictions can be made for new alternatives.

4.2.3 Information Theory: A Technique to Select the Most Useful Attributes

The previous techniques search for performance measures which were combinations of some of the attributes. An alternative strategy would be to discard some attributes and keep only those which provided the most information about choice. One technique to do this is to use information theory to compute the mutual information $I(Y_b, A)$ between a subset, Y_b , of Y and the alternative set. For a given number of performance measures (cardinality of Y_b) the analyst could choose the subset with the maximum mutual information. He would then examine all cardinalities and use his judgement to select the "best" subset based on the conflicting characteristics of parsimony (fewer attributes) and completeness (larger $I(Y_b, A)$). Due to limitations in the number of

observations this is not possible for the larger cardinalities of Y_b , but reasonable approximations can be made. We begin by discussing the analytic details of selecting the single attribute with the largest mutual information.

4.2.3.1 Analytic Details of Calculating Mutual Information

Suppose we were limited to selecting one attribute, Y_ℓ , from Y . Gallagher [42] shows that the average mutual information provided about choice (selection from the alternative set A) by a discretely valued attribute, Y_ℓ , is:

$$I(Y_\ell; A) = \sum_{y_\ell \in Y_\ell} \sum_{j=1}^J p(y_\ell, a_j) \log \frac{p(a_j | y_\ell)}{p(a_j)}$$

Which could be computed if $p(a_j | y_\ell)$ were known. But by the laws of conditional probability $p(y_\ell, a_j) = p(y_\ell | a_j)p(a_j)$ and $p(a_j | y_\ell)/p(a_j) = p(y_\ell | a_j)/p(y_\ell)$ thus:

$$I(Y_\ell; A) = \sum_{j=1}^J p(a_j) \sum_{y_\ell \in Y_\ell} p(y_\ell | a_j) \log \frac{p(y_\ell | a_j)}{p(y_\ell)}$$

Now if the population of individuals is homogeneous with respect to $p(a_j | y_\ell)$ then $I(Y_\ell; A)$ can be computed for an empirical sample. This homogeneity assumption is perhaps too strong but is useful for a first cut exploratory analysis. Note that if Y_ℓ is continuous then it can be discretized by dividing the range into disjoint intervals. To empirically calculate the needed quantities use a table such as given in table

4.1. Then $p(a_j) = M^j/M$, $p(y_k = k) = M_k/M$, and $p(y_k = k | a_j) = m_{kj}/M^j$.

y_k	choice		
	a_1	a_2	a_3
1	m_{11}	m_{12}	m_{13}
2	m_{21}	m_{22}	m_{23}
3	m_{31}	m_{32}	m_{33}
4	m_{41}	m_{42}	m_{43}

$$M = \sum_{j=1}^3 \sum_{k=1}^4 m_{kj}$$

$$M_k = \sum_{j=1}^3 m_{kj} \quad M^j = \sum_{k=1}^4 m_{kj}$$

Table 4.1: Discretized Distribution of an Attribute

Now to choose the subset Y_b with K elements, we would want to calculate $I(Y_b; A)$. Unfortunately as K becomes larger the number of cells required grows exponentially with K and hence the number of observations in each cell becomes too small for us to have faith in the method. To overcome this problem in dimensionality Boyle [14] proposes the following 1st and 2nd order methods.

First method: Choose the first K attributes with the largest mutual information with choice $I(Y_b; A)$. Boyle [14] applied this technique to the study of personal attributes effecting defaults on bank loans. Table 4.2 is a reproduction of his results.

Note that the most information is provided by the class index. The class index provides perfect information about choice ($p(a_j | \text{index}) = 1$ if a_j is chosen, 0 otherwise) and acts as an upper

Table 4.2: Ranking by First Order Mutual Information

Attribute	$I(Y_\chi; A)$
Y_0 : Class index	.693
Y_{23} : NSB credit score	.207
Y_{16} : Years with former employer	.167
Y_2 : Years at occupation	.158
Y_{15} : Years at former residence	.121
Y_{10} : Years of residence	.111
Y_2 : Age	.038
Y_{18} : Checking account	.037
Y_9 : Own/rent house	.035
Y_{11} : Income	.032
Y_1 : Occupation	.030
Y_{19} : Savings account	.029
Y_{12} : Mortgage/rent	.027
Y_3 : Loan amount	.025
Y_{22} : ability to pay ratio	.020
Y_{20} : auto	.018
Y_{17} : Other income	.015
Y_6 : Marital	.013
Y_8 : Dependents	.012
Y_5 : Purpose	.011
Y_{14} : Telephone	.011
Y_{13} : Total debt	.010
Y_4 : Term	.009
Y_{21} : Total monthly payments	.004

bound on the mutual information measure. Note that substitution in the equation for $I(Y_\ell; A)$ yields that:

$$I(\text{index}; A) = -\sum_{j=1}^J p(a_j) \log p(a_j)$$

which is simply the entropy of the system, $H(A)$. Notice that Y_{23}, Y_{16} , Y_2, Y_{15} and Y_{10} all provide significant information but that there is a dramatic dropoff starting with Y_7 . Also we might expect that Y_{16} could be grouped with Y_2 and Y_{15} with Y_{10} . The next approximation considers pairwise groupings.

Second order method: (1) Choose the attribute with the largest mutual information, $I(Y_\ell; A)$. (2) Then choose the attribute which when combined with the first gives the largest combined mutual information, $I(Y_{\ell 1}, Y_{\ell 2}; A)$. (3) Now choose the attribute which when combined with the second gives the largest combined mutual information, $I(Y_{\ell 2}, Y_{\ell 3}; A)$. (4) Continue until K attributes are chosen.

The combined mutual information, $I(Y_{\ell 1}, Y_{\ell 2}; A)$ is computed in the same way as $I(Y_{\ell 1}; A)$ except that the summation is over cells in $Y_{\ell 1} \cdot Y_{\ell 2}$ space instead of $Y_{\ell 1}$ space. Note that since $I(Y_{\ell 1}, Y_{\ell 2}; A) = I(Y_{\ell 2}; A|Y_{\ell 1}) + I(Y_{\ell 1}; A)$ the second order method can be viewed as choosing the attribute with the largest mutual information when conditioned on the previously chosen attribute. Boyle [14] uses this technique for the same attributes. Table 4.3 is a representation of his results. He also tested the rank ordering of attributes by changing the first attribute chosen, and found that the resulting rank order was only slightly perturbed.

Table 4.3: Ranking by Second Order Mutual Information

Attribute	$I(Y_{\ell m}; A Y_{\ell m-1})$	$I(Y_{\ell m}; A)$	$I(Y_{\ell m}, Y_{\ell m-1}; A)$
Y_{23}	NSB Credit Score	-.207	-.207
Y_{16}	Years with former employer	.117	.167
Y_{12}	Mortgage/Rent	.055	.027
Y_2	Years at Occupation	.162	.158
Y_{15}	Years at former residence	.043	.121
Y_1	Occupation	.041	.030
Y_{10}	Years at Residence	.113	.111
Y_{11}	Income	.038	.032
Y_7	Age	.054	.038
Y_{17}	Other Income	.057	.015
Y_3	Loan Amount	.060	.025
Y_{22}	Ability to Pay Ratio	.060	.020
Y_8	Dependents	.057	.012
Y_{13}	Total Debt	.050	.010
Y_{21}	Total Monthly Payments	.047	.004
Y_9	Own/Rent	.054	.035
Y_{19}	Savings	.024	.029
Y_{18}	Checking	.030	.037
Y_{20}	Auto	.012	.018
Y_4	Term	.015	.009
Y_5	Purpose	.034	.011
Y_6	Marital Status	.025	.013
Y_{14}	Telephone	.009	.011

4.2.3.2 Use of Information Theoretic Attribute Selection in the Methodology

Like factor analysis and multidimensional scaling, information theory can be used to reduce the attributes. Unlike the others it does not find performance measures but instead eliminates those attributes which supply little information about choice. The analyst should use the technique as a guide, leaving the final selection up to a combination of judgement and the information measure.

Advantages: Since there is no mapping introduced the remaining attributes are guaranteed to be instrumental. Furthermore an indication is given of which attributes are very important and which ones hardly effect choice. The technique is not sensitive to distributional assumptions or the particular structure of interaction such as was the case with the linear model of factor analysis.

Disadvantages: Because it eliminates attributes rather than search for structure, information theory does not provide any insight into how the attributes combine to produce performance measures. There are also the approximations introduced by discretizing continuous attributes and the problems of cell size getting too small.

Overall it acts as a useful aid to intuition by identifying the important attributes in the choice process.

4.2.4 Other Potential Reduction Techniques: An Overview

4.2.4.1 In Depth Utility Assessment

Chapter 6 establishes an isomorphism between compaction functions and von Neumann-Morgenstern multi-attributed utility functions. It is shown there that if a set of attributes, Y_b , is "preferentially independent"³ of its complement, $(Y - Y_b)$, then the compaction function, defined here on the attributes, can be written as $c(y) = c(R(y_b), y_{\bar{b}})$ where $y_{\bar{b}} \in (Y - Y_b)$. If enough sets of independence properties can be verified then $c(y) = c(R_1(y_{b_1}), R_2(y_{b_2}), \dots, R_K(y_{b_K}))$ and the "aggregated" variables $R_k(y_{b_k})$ can be identified as performance measures, x_k . Because of the independence properties the functions $R_k: Y(x_k) \rightarrow x_k$ can be assessed independently of the knowledge of $Y - Y(x_k)$.

Thus if we are allowed enough time to perform an in depth interview with each respondent, it is possible to identify directly the structure through independence questions and then assess the reduction mappings in a way similar to the technique described in chapter 6. One potential alternative is to directly assess the compaction functions for a few selected individuals (perhaps "experts"), identify the structure for them, parameterize the reduction mappings, and obtain a distribution of the parameters from the general population.

The advantage of direct assessment is that it is theoretically sound and makes no compromise on completeness. The disadvantage is its extensive measurement requirements.

4.2.4.2 Professional Judgement

Although not a very elegant technique, professional judgement alone can be used if data is unavailable or if there is not enough time to perform an analytic reduction. It is mentioned here for emphasis since it is the most prevalent technique in practice. We do not advocate using it alone but instead coupling it with the insight gained through a more analytic reduction technique.

4.2.4.3 Null Option

If none of the analytic techniques provides insight into possible structure, or if the gains in parsimony are not worth the losses in completeness, or if the chosen compaction technique works just as well with a large number of attributes as it does with a few performance measures then the analyst may decide to skip reduction and go directly to compaction. Note that insight may be gained by showing the positive or negative results of reduction to the design team even if the full attribute set is used in compaction.

4.3 Conclusion of Reduction

The goal of this methodology is to provide insight to the design team. An important step toward this goal is reduction which identifies the structure of consumer perception by reducing the many attributes to a few performance measures. Reduction provides an understanding and visualization (through perceptual maps) of the choice and/or perception process. This difficult problem requires that a number of tradeoffs be made between the insight gained from a parsimonious

representation of perceptions and the completeness of information contained in the set of attributes.

This chapter began by discussing the issues of reduction. I.e., an ideal set of performance measures should be computable, essential, complete, instrumental, parsimonious, disjoint, and encompassing. Any real situation will require the analyst to choose a technique which emphasizes some of these properties and sacrifices others.

This chapter identified three primary reduction techniques. Factor analysis (which is popular in social and psychological research) approximates the structure of reduction with a linear model and uses correlations between the attributes and performance measures to identify which attributes combine to form performance measures. Multidimensional scaling (which is popular in marketing studies for frequently purchased consumer products) uses measures of similarity among alternatives to infer a perceptual map. MDS does not make any explicit structural assumptions but it does not provide any instrumental linkage from the attributes to the performance measures. Information theory (which is popular in pattern recognition but not yet used in marketing) does not identify structure in perception, but does identify those attributes which have the strongest effect on choice. Each of the above makes different trade-offs, the choice of technique depends on the situation.

Reduction structures perceptions, but which performance measures are most important and how do they combine to effect choice? These questions are addressed in chapter 6, compaction, but before compaction is discussed, we will address the question of abstracting population segments based on homogeneity of perception and preference.

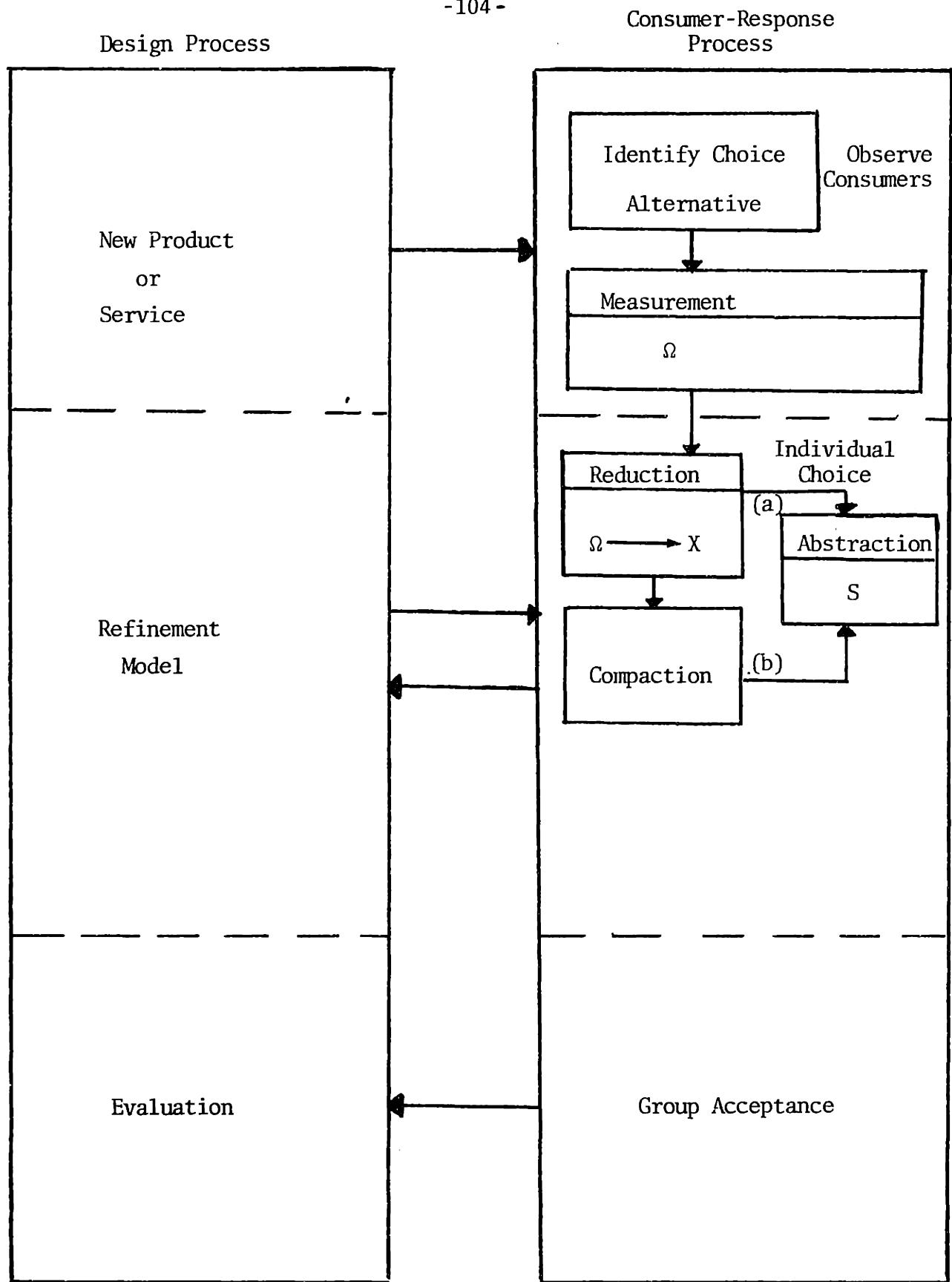
Chapter 5

Abstraction

Reduction identified how people perceive the choice alternatives. In a given situation it might be that these perceptions (and later preferences) are not homogeneous across the population. Abstraction attempts to abstract population segments which are homogeneous with respect to perception (arrow (a) in figure 5.1) and later, after compaction, segments which are homogeneous with respect to preference (arrow (b) in figure 5.1).

Abstraction is managerially useful because segments defined by homogeneity of perception or preference represent groups which either view or value the performance measures differently. These groups represent opportunities for alternative product or service design. Innovation may be more successful by meeting each segment need separately than by designing an average product or service which does not exactly meet the needs of anybody. A second reason for abstraction is that if importances are statistically estimated (See Section 6.3,) homogeneity is required for theoretical soundness.

This chapter begins by identifying some criteria for abstraction. Then the standard techniques of cluster analysis, automatic interaction detection (AID), and discriminant analysis are reviewed in relation to the methodology. Examples are given of the use of cluster analysis and AID in abstracting segments relative to the choice of health care delivery systems.



5.1: Relationship of Abstraction to the Methodology

5.1 Criteria for Abstraction

The following criteria act as qualitative guides for selecting among alternative segmentations. In choosing among segmentations the analyst should consider all four criteria, but the relative importance he attaches to each depends upon the situation and upon his judgement.

1. Strategically relevant: It is important that segments be abstracted if it is feasible to have different strategies for different segments or if identifying segments aids in understanding the consumer choice process. Thus one primary consideration in choosing among segmentations is the potential for product differentiation.

2. Targetable: Segmentation is only useful if the target segments can be reached. Targeting can be accomplished either explicitly, such as with a lower fare for the elderly, or implicitly through self-selection, such as offering different options for an HMO. For prediction it is important that segments be defined relative to independent variables which can be identified and measured.

3. Reduce variation within: One goal of abstraction is to identify homogeneous segments, thus a numerical indicator of the quality of a segmentation strategy is the reduction in variance of the dependent variable or variables within the segment. For example if quality and personalness are the only two performance measures for HMO's, then it might be desirable to identify segments based on the relative importance of quality to personalness. We would want the variance of this ratio to be small within a segment.

4. Significant variation between: Just as it is important that there be little variation within segments, it is important that there be enough variation between segments such that segments are identifiable and strategies are not blurred.

5.2 Analytic Techniques for Abstraction

This section reports on a variety of techniques to abstract segments. For each technique the mathematics are summarized and an indication is given of how to use the technique in the methodology. Empirical examples are given for cluster analysis and AID.

5.2.1 Cluster Analysis

The basic idea behind cluster analysis can be seen in figure 5.2. Suppose that the only two performance measures for health care are personalness and value and that we can measure the importance of each and quantify this importance by an "importance weight," i.e., k_p and k_v . Cluster analysis considers the dependent variables, in this case k_p and k_v , and groups or "clusters" together those with values which are similar relative to some metric, in this case straight line distance in $k_p * k_v$ space. In addition the means and variances of some explanatory variables are calculated and displayed. In figure 5.2 this is shown for the mean \pm standard deviation for age and income. The analyst then interprets the effect of the explanatory variables on the segmentation. If he desires he can then use discriminant analysis to statistically test his interpretation. (See Section 5.3.) In the case of figure 5.2 we would say that the wealthy young value personalness more than value whereas the

elderly poor put a high premium on value. This example if of course contrived, real life is never this simple, but it does indicate the basic technique.

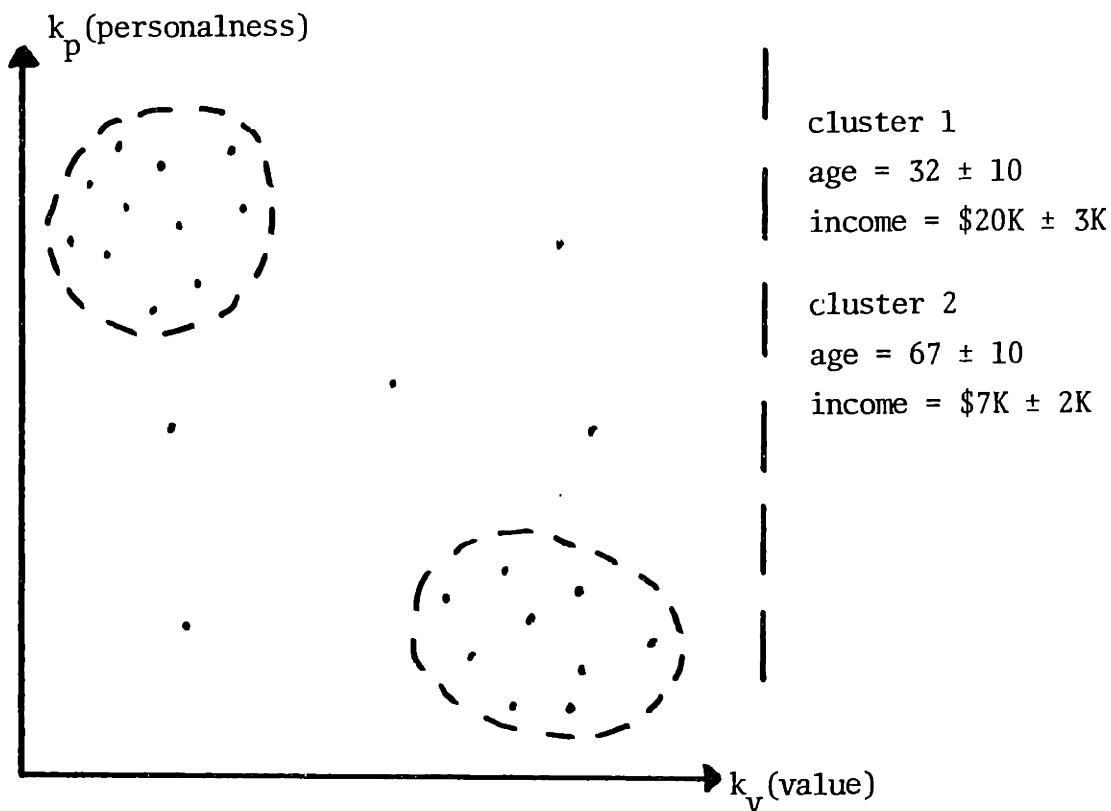


Figure 5.2: Hypothetical Cluster Analysis

Formal technique: (The following is a description of the Howard-Harris clustering program as reported in Green and Rao [48].) Given a set of individuals, S , and a set of clustering variables, $Y = \{Y_1, Y_2, \dots, Y_m\}$ the object is to partition S into p disjoint segments (S_1, S_2, \dots, S_p). Let i index the individuals and let $y_i = (y_{i1}, y_{i2}, \dots, y_{im})$ be the value that Y takes on for individual i . Define the dissimilarity, d_{il}^2 , between individuals i and l , as the square of the Euclidean distance between y_i

and \underline{y}_ℓ in Y. I.e.

$$d_{i\ell}^2 = |\underline{y}_i - \underline{y}_\ell|^2 = \sum_{k=1}^m (y_{ik} - y_{\ell k})^2$$

The total variance, V_T , of Y can be divided into a between group variance, V_B , and an within group variance, V_w . Where Green and Rao show that:

$$V_T = \frac{1}{2n} \sum_{i=1}^m \sum_{\ell=1}^m |\underline{y}_i - \underline{y}_\ell|^2$$

$$V_{S_j} = \frac{1}{2n_j} \sum_{i \in S_j} \sum_{\ell \in S_j} |\underline{y}_i - \underline{y}_\ell|^2$$

$$V_w = \sum_{j=1}^p V_{S_j}$$

n_j = number of individuals in S_j

n = total number of individuals

Ideally for a given p the program would find the partition which minimizes V_w . Instead it is only feasible to find a locally optimal p -fold partition by starting with a $(p-1)$ -fold partition, splitting one segment, S_j , on the basis of maximum within-group variance, and shifting points until a minimum V_w is found. The solution is locally optimal in the sense that shifting any single point would increase V_w .

Use in the methodology: The first use is to cluster individuals homogeneous with respect to perceptions (arrow (a) in figure 5.1). This can be done for each stimuli (alternative a_j), in which case, if there are M performance measures, Y is the M -dimensional space of performance measures. If everyone's evoked set is identical it can be done simultaneously for all stimuli, in which case, if there are J alternatives,

Y is the $M \cdot J$ dimensional joint space of performance measures. For example, such a clustering might find that childless married students perceive "plan 1" (with compulsory pregnancy insurance) as low in value.

A variation in this approach is to cluster individuals with respect to the performance measures of their chosen (or first preference) alternative. The idea behind this approach is that it is likely that the most important performance measures will have high values for the chosen alternative. In a way this is an attempt to identify segments homogeneous with respect to preference.

It is also possible to directly abstract segments based on homogeneity of preference (arrow (b) in figure 5.1). Direct assessment of compaction functions (sections 6.2 and 6.4) identifies preference parameters such as importance weights, k_ℓ , and risk aversion coefficients, r_ℓ , which directly measure the relative importance and risk properties of a performance measure, X_ℓ , such as quality. *A priori* this is probably the more desirable way to define segments because it allows for products to be differentiated by emphasizing different performance measures. For example a quality conscious segment might want dental care in an HMO whereas a value conscious segment might not.

Empirical example: Greer and Suuberg [50] ran a number of cluster analyses on the performance measures identified in section 4.2.1.3. These performance measures are the factor scores for quality, personal-ness, convenience, and value as related to the choice of HMO's. They found that about the best that could be done with perception segmentation is explain 48% of the variance ($V_w/V_T = .48$) with 4 clusters. This can

be compared to 44% explained through an analysis of random data. Thus we reach the conclusion that it is unlikely that any significant segments in this data can be identified based on homogeneity of perception. This does point out the necessity of testing any cluster analysis against random data since no formal statistical tests exist.

Other clustering techniques: Sebestyen and Edie's pattern recognition algorithm as discussed in Boyle [14] can be viewed as a clustering technique. Unlike Howard-Harris which examines the entire sample at once. Sebestyen and Edie's algorithm works with a training sample and "grows" clusters in Y space.

Howard-Harris used squared Euclidean distance, other metrics such as right angle distance are also possible. Johnson's hierarchical cluster program as discussed in Green and Rao [48] uses as input a matrix of "dissimilarities" as measured directly by respondent's similarity judgements on the stimuli.

5.2.2 Automatic Interaction Detection (AID)

(The technique and algorithm described in this section is explained in greater detail in Sonquist, Baker, and Morgan [134].) Basically AID tries to reduce within group variance of a dependent variable, Z , by iteratively making splits on one of the explanatory variables, Y_λ : $\lambda = 1, 2, \dots, m$. To see this in one dimension examine figure 5.3. In this case if the space Y_1 is partitioned into two subsets, $S_1 = \{y_1: y_1 \leq .75\}$ and $S_2 = \{y_1: y_1 > .75\}$ then the sum of the variances of Z in S_1 and S_2 would be significantly less than the variance of Z in the complete set Y_1 . Thus AID, like cluster analysis, partitions the

space of explanatory variables, but the criteria for AID is reduction of within group variance of a dependent variable rather than reduction of within group variance of the dissimilarities of the explanatory variables.

AID is said to detect interactions because successive splits are only on partitions defined by previous splits. For example suppose there are two explanatory variables, y_1 = age and y_2 = income, and the first split is on age = 40. Then $S_1 = \{y_1, y_2\} : y_1 \leq 40\}$ and $S_2 = \{y_1, y_2\} : y_1 > 40\}$. If there were no interaction between age and income then the next split might split both S_1 and S_2 on income = \$10K. If there were interaction then the next split might split S_1 on income = \$7K and the following split might split S_2 on income = \$15K. (See figure 5.4).

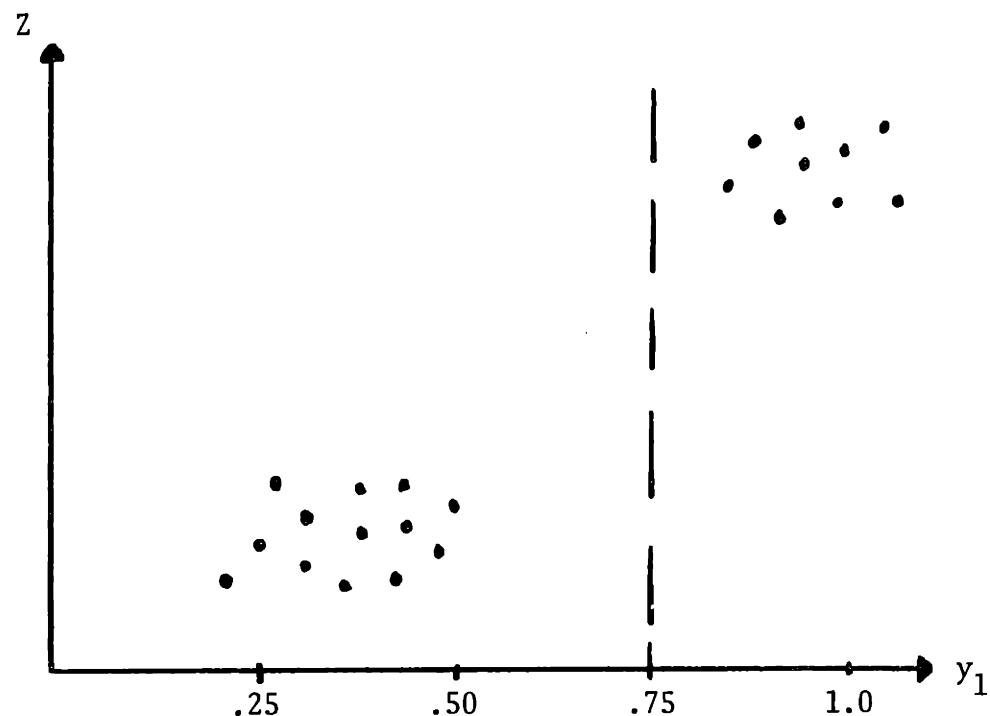


Figure 5.3: Automatic Interaction Detection (AID)

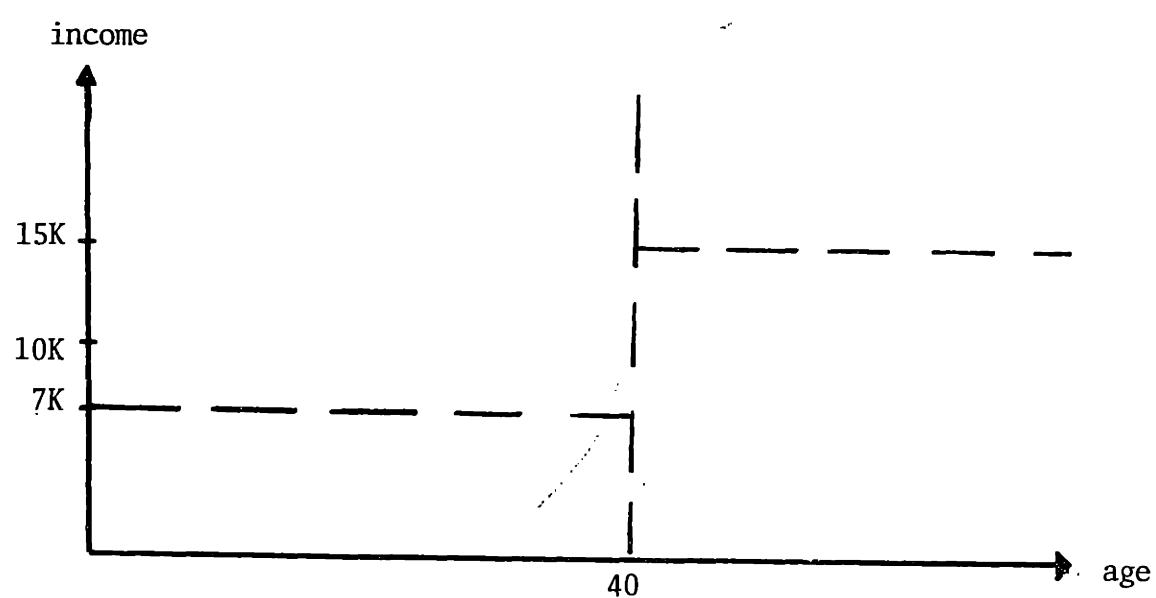
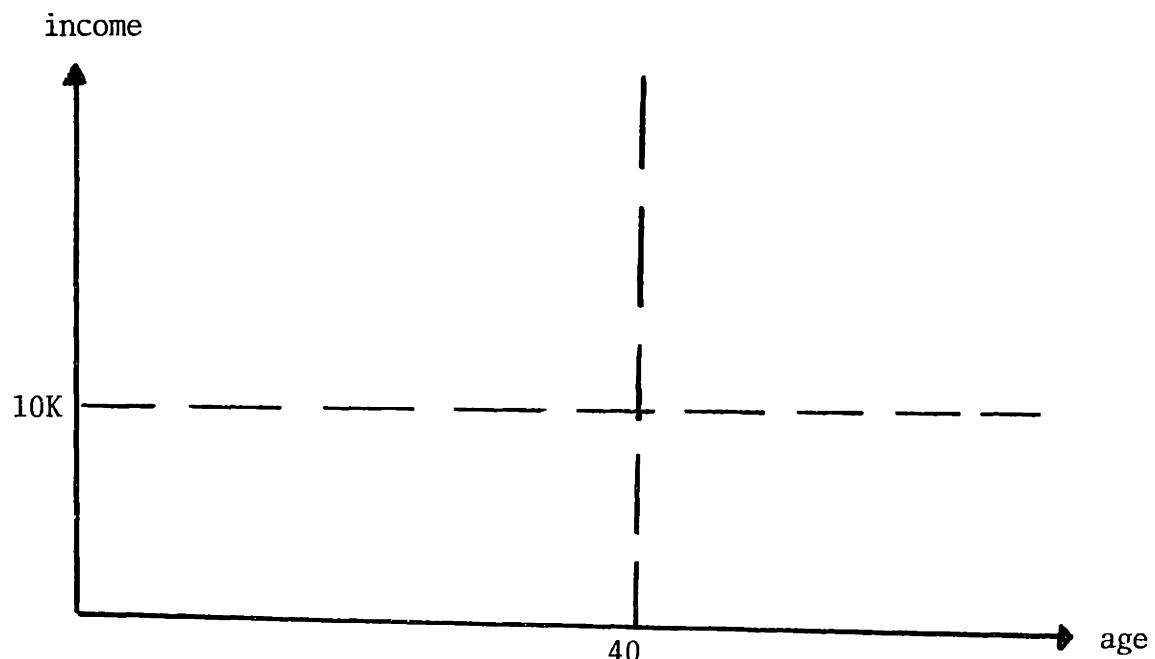


Figure 5.4: Possible Interaction

Notice that in AID splits are always defined by hyperplanes parallel to an axis. A useful representation of successive AID splits is by a tree diagram such as that in figure 5.5, where the dependent variable is distance, d , from home to the central business district. For example the left most partition, $S_1 = \{(y_1, y_2) : y_1 \leq 40, y_2 \leq 7K\}$, has an average distance of 1 mile and contains 35% of the sample.

Formal technique: The following is a description of the simplest version of AID, but does give a good indication of the technique. AID III has many options such as lookahead, pre-set divisions, covariance search, and premiums for symmetry. The interested reader is referred to Sonquist, Baker, and Morgan [134].

Given a set of individuals, S , a dependent variable, Z , and a set of explanatory variables, $Y = \{Y_1, Y_2, \dots, Y_m\}$ the object is to partition S into p disjoint segments, (S_1, S_2, \dots, S_p) by iteratively splitting a "parent" segment into two "children" by a division or split relative to one explanatory variable. The criterion is to minimize within group variance. Consider the first split, then $S = \text{parent}$ and $S_1, S_2 = \text{children}$. The total variance is:

$$V_T = \sum_{i \in S} (z_i - \bar{z})^2$$

where $\bar{z} = \frac{1}{n} \sum_{i \in S} z_i$

After the split, the within group variance is:

$$V_W = \sum_{i \in S_1} (z_i - \bar{z}_1)^2 + \sum_{i \in S_2} (z_i - \bar{z}_2)^2$$

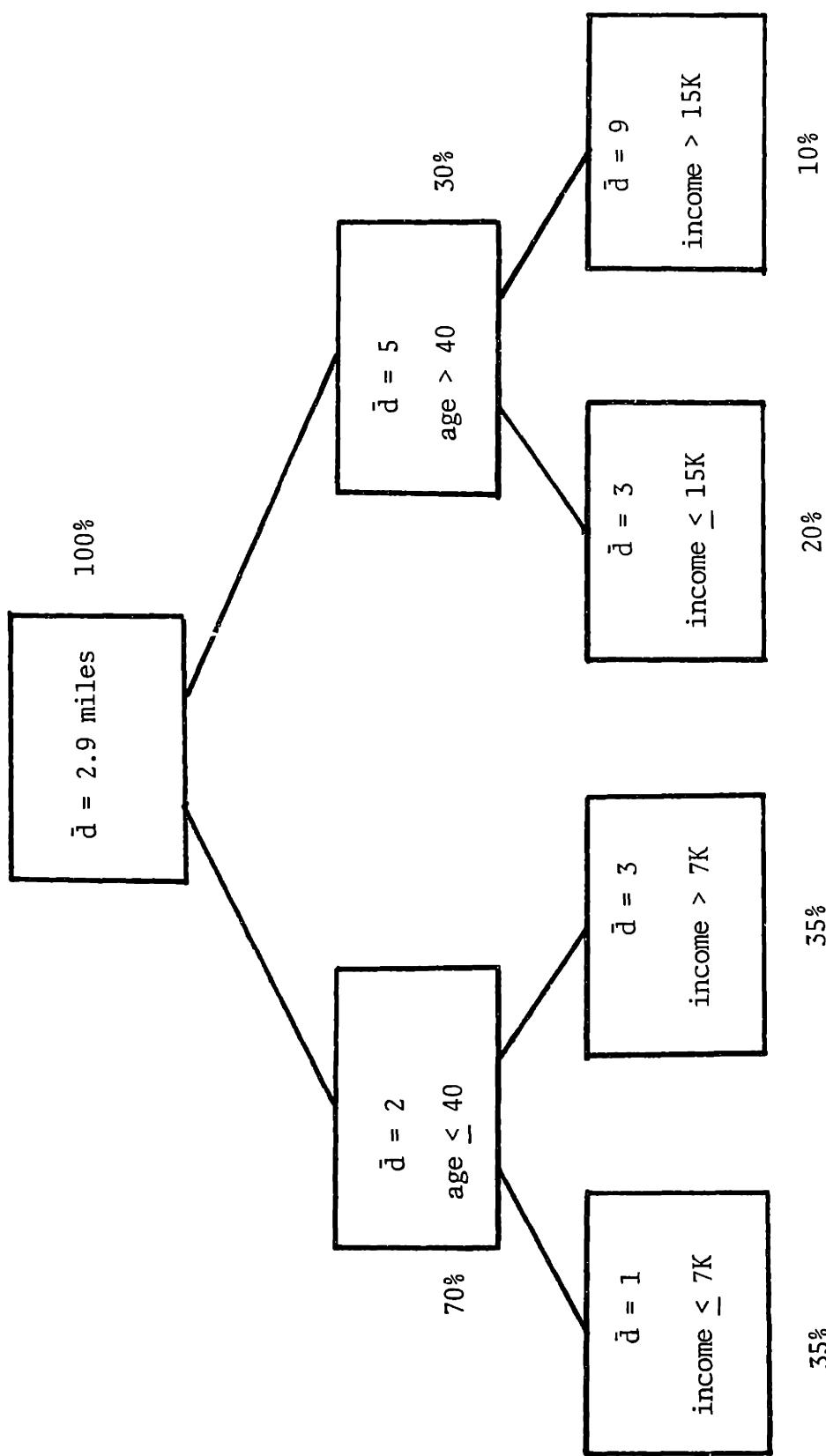


Figure 5.5: Representation of AID Splits

where $\bar{z}_1 = \frac{1}{n_1} \sum_{i \in S_1} z_i$

$$\bar{z}_2 = \frac{1}{n_2} \sum_{i \in S_2} z_i$$

By expanding V_w and V_T and collecting terms it is possible to simplify the expression, but basically the primary version of the algorithm chooses the explanatory variable and the split to minimize V_w .

Use in the methodology: AID analysis is a flexible exploratory technique which can be used to check for homogeneity of perceptions or of preferences. It can also be used in an attempt to look for predisposition to choose one alternative and to implicitly search for indications of the relative importances of the performance measures. The following five choices of dependent and explanatory variables are among the possible uses of AID in the methodology,

(1) The first is to search for homogeneity of perception (arrow (a) in figure 5.1). This can be done by using each performance measure for each alternative as the dependent variable and demographics as explanatory variables. (M performance measures and J alternative means M•J possible AID runs.) Be sure to "normalize" each scale across alternatives to remove scale bias. Clearly for this purpose AID is much less efficient than cluster analysis.

(2) The second use is to search for homogeneity of preference (arrow (b) in figure 5.1). Using an importance weight (or a risk aversion coefficient) obtained in compaction (sections 6.2 and 6.4) as the dependent variable and demographics as the explanatory variables it is

possible to search for segments which value a particular performance measure highly.

(3) One hypothesis often raised is: How useful are the performance measures? Perhaps the strongest effect is a predisposition to select on alternative (attitude) within a population segment? The third use of AID is to search for this predisposition by checking to see if some breakdown by demographics can explain choice of a given alternative (0-1 variable) or intent to choose that alternative (if choice is unobserved).

(4) One indication of the importance of a performance measure is that it be high for the chosen (or first preference) alternative. AID can be used to search for population segments which value a given performance measure highly on their chosen alternatives. The dependent variable is a single performance measure on the chosen alternative and the explanatory variables are demographics.

(5) The fifth use of AID is an attempt to get an indication of the relative importances of the performance measures. This is done by using the rank order (if observed) of a given plan as the dependent variable and the performance measures as the explanatory variables. One indication, although a weak indication because of range and variance effects, of the relative importance of the performance measures is the order in which the splits occur. For example if a split on quality in the choice of HMO's explains 90% of the variance of rank order of a given plan and if splits on all other performance measures yield much

smaller reductions in variance, then this is an indication that quality is an important determinant of choice for that plan.

Empirical example: Greer and Suuberg [50] ran a series of AID analyses on the choice and preference measures described in section 3.3. These are measurements of the M.I.T. community relative to the choice of HMO's. Some of their results are reproduced in figure 5.6. (PAT 2 and PAT 4 were pattern of care variables. See appendix for questionnaire. PAT 2 indicated where you would go for a routine physical and PAT 4 for a dental checkup.)

Examination of figure 5.5 reveals that a split on the demographic variable PAT 2 explained the most variance. This would seem to imply that there is either (1) an inertia effect, that is individuals have a tendency not to switch, or (2) a risk aversion effect, that is individuals have more information about the plan that is currently being used, or (3) self selection and preference, that is these individuals who choose M.I.T. over private for a physical would tend to choose the M.I.T. HMO for much the same reasons. AID can not tell which one of the above is the main reason, but it does serve to elicit and focus creativity with an indication of effects.

Again no statistical significance tests exist for AID, thus a run must be tested against random data. When this was done with variables similar to figure 5.6, AID explained 16.1% of the variance, roughly 50% less. Thus we can conclude that there is evidence for segmentation in the data, but not overwhelming evidence.

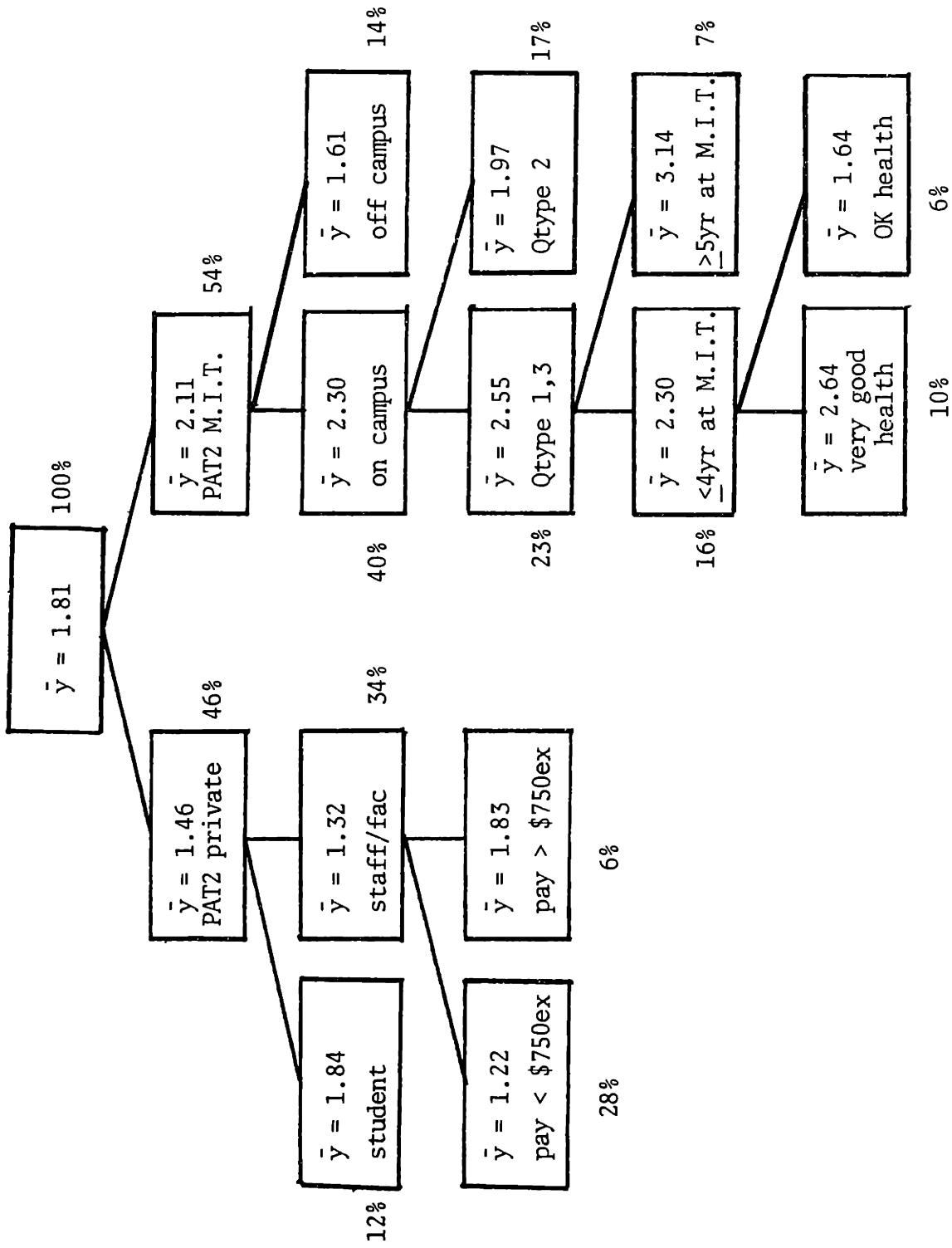


Figure 5.6: AID Analysis for M.I.T. HMO-Dependent Variable Preference

5.2.3 Discriminant Analysis

Unlike cluster analysis and AID which were used to explore segmentations, discriminant analysis is used to test whether segments are significantly different with respect to some explanatory variables.

For example, consider figure 5.1. Here we defined two clusters, S_1 and S_2 . It appears that the demographic variables age and income might "explain" cluster membership. (Obviously k_p and k_v could explain this membership since clusters were defined relative to k_p and k_v , but we are looking for some additional targetable variables to explain the membership.) Discriminant analysis allows us to test this.

Formal technique: Discriminant analysis defines a linear "discriminant function," $d: V \rightarrow R$, which is a linear mapping from the (demographic) explanatory variables, V , into the real line. I.e.:

$$d_i = \sum_{\ell} w_{\ell} v_{i\ell}$$

where

d_i = discriminant value for individual i

$v_{i\ell}$ = value of the ℓ^{TH} explanatory variable for individual i

w_{ℓ} = weight of the ℓ^{TH} explanatory variable

Suppose the w_{ℓ} 's were known, then for each segment, S_j , one could calculate the mean of d_i , and similarly calculate the variance of d_i across the population. I.e.:

$$\bar{d}_j = \sum_{i \in S_j} d_i$$

$$\text{var} = \sum_{i \in S} (d_i - \bar{d}_i)^2$$

For two populations the criteria is then to maximize the ratio $(\bar{d}_1 - \bar{d}_2)^2/\text{var}$, that is the ratio of the squared difference between the means of the two populations to the variance which is assumed common to both populations. Subject to various distributional assumptions on the explanatory variables, V , it is possible to determine the weights, w_ℓ , to maximize this ratio. The extension to more than two populations is notationally and computationally more complex, but the basic idea is similar.

The details of determining the w_ℓ 's can be found in Kendall [78]. Also a description of discriminant analysis with a different viewpoint appears in section 6.3.2.

Use in the methodology: The primary use of discriminant analysis in abstraction is to complement cluster analysis. Cluster analysis identified partitions of the population which were homogeneous with respect to some clustering variables, Y . Sometimes these clustering variables are not necessarily targetable variables, that is, it is difficult to formulate design strategy with respect to segmentation on Y . Thus it is important to identify targetable demographic variables which can "explain" cluster membership.

To do this a discriminant analysis is run on the demographic variables, such as age and income in figure 5.1, given the population segments identified by cluster analysis. The weight, w_ℓ , for a

demographic variable is an indicator of its importance in defining the segment. Statistical tests can then be run to determine if the means of \bar{d}_i , \bar{d}_j , are significantly different between clusters.

5.2.4 Other Techniques

Cluster analysis and AID are the primary segmentation techniques available, but there are others, among these multiple regression and the mutual information measure.

Multiple regression: Multiple regression as suggested by Frank, Massy, and Wind [41] can be used as a segmentation technique. This technique is similar to AID in that it assumes that some dependent variable (choice, perception, or preference) is a function of some explanatory variables, Y . If there are strong a priori convictions on the form of this function, and if the function is linear in some unknown parameters, k , these parameters can be determined by multiple linear regression. The interpretation and usage is then similar to AID and/or discriminant analysis.

Mutual information measure: Just as the mutual information measure was used in reduction to compute a measure of the "information" a performance measure provided about choice, it can be used in abstraction to compute the information that an explanatory variable, Y_k , (e.g., demographics) provides about a dependent variable (e.g., choice, perception, or preference). Again, the interpretation and usage is similar to AID and/or discriminant analysis.

5.3 Conclusion of Abstraction

Abstraction attempts to define managerially useful segments of the population which are homogeneous with respect to either perception or preference. Identifying such segments can facilitate successful innovation because if the needs of the various segments are significantly different it may be better to meet each segment need separately rather than by designing an average product or service which does not meet the needs of anyone.

This chapter summarized the standard techniques of cluster analysis and AID in relation to their use in the methodology. Cluster analysis is used to search for homogeneity in perceptions and/or preferences. If these are not targetable variables, the analyst, by examining means and variance of targetable variables (e.g., demographics) within clusters, tries to infer which ones can "explain" the clusters. He then tests these judgements with discriminant analysis. AID searches for explanatory variables (demographics or perceptions) which can "explain" the variation in some dependent variable (choice, perception, or preference). It does this by successively splitting parent segments into two children with a split (or division) relative to one explanatory variable. Together these techniques provide useful exploratory tools for segmentation.

The ultimate choice of a segmentation strategy must depend upon the analyst and the design team having confidence in the relative homogeneity of the segments and upon the strategic benefits which can be derived from the segmentation. Techniques of reporting segmentation findings to management are discussed in chapter 9.

The next chapter returns to the direct analysis of the individual choice process by examining how the reduced spaced performance measures, defined in reduction, are linked to choice.

Chapter 6

COMPACTON

The previous chapter discussed techniques to reduce the number of attributes effecting choice to a relatively few performance measures in order to allow the design team to better understand and visualize the choice process. But even with only a few dimensions the choice process can be extremely complex. The next step, compaction, explicitly identifies the importance of each performance measure, their interdependency, and the risk characteristics of the choice process (See figure 6.1). To do this a compaction function, $c(x_{ij}, \lambda_i)$, is determined which maps individual i 's perceptions of the performance measures, x_{ij} , and a set of individual specific choice parameters called preference measures, λ_i into a scalar measure of goodness (a real number), c_{ij} . For a given alternative, a_j , this scalar measure of goodness, c_{ij} , has the property that with all other alternatives held fixed, any set of performance measures yielding the same value, c_{ij} , must also yield the same probability of choice for alternative a_j . In other words, compaction compresses the few performance measures for an alternative into a one-dimensional measure. Knowing the value of this measure for each and every alternative is then sufficient to predict choice.

This chapter begins with a rigorous development of the theory of compaction. First a series of general definitions and their intuitive interpretations are given followed by a few simple theorems. Next a choice axiom is postulated, which together with the von Neumann-

Design Process

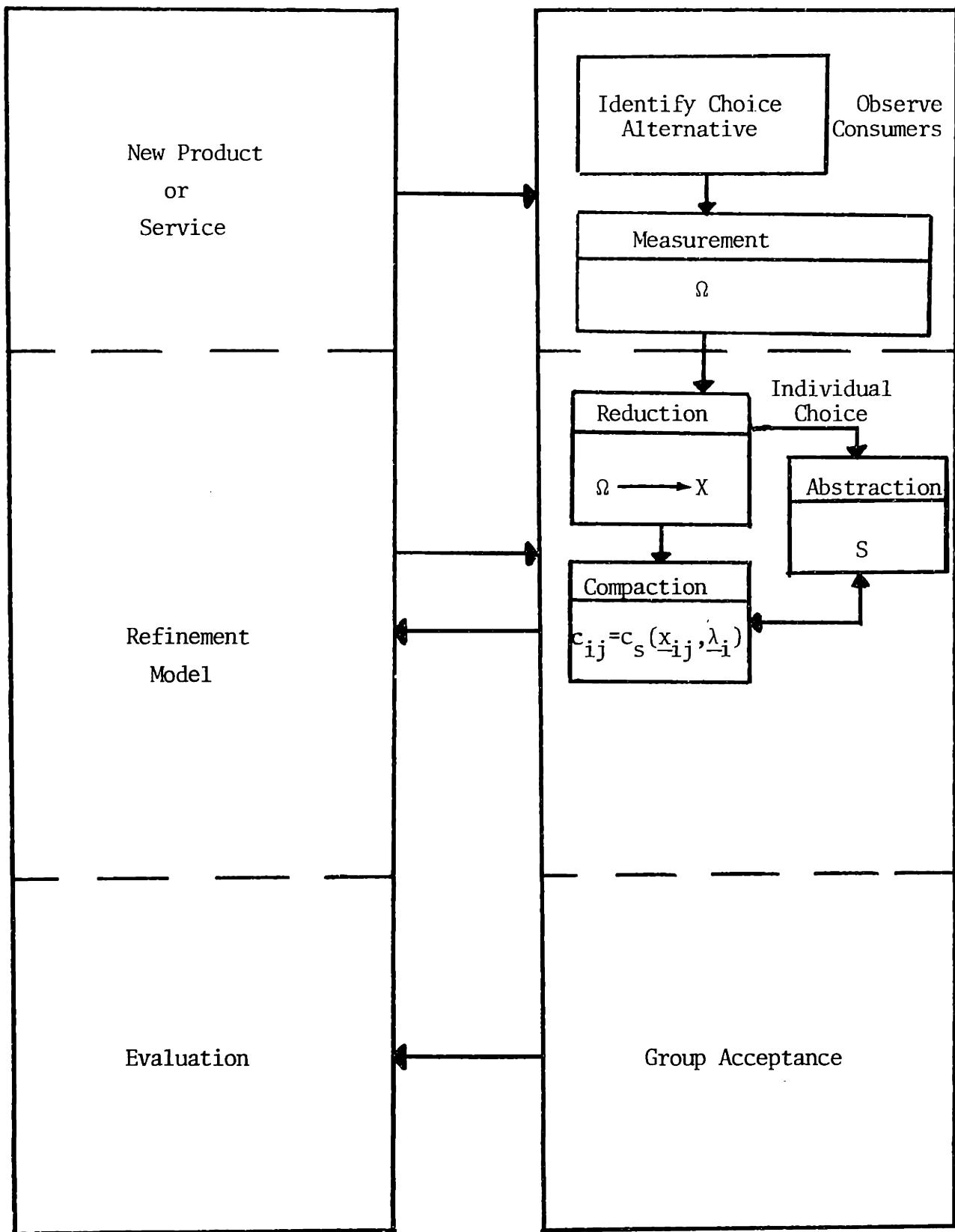


Figure 6.1: Relationship of Compaction to the Methodology

Morgenstern [15] utility axioms establishes an isomorphism between compaction theory and utility theory. This axiomization allows the richness of prescriptive utility theory to be applied to descriptive choice theory. Existence and uniqueness theorems are presented for compaction function and the isomorphism with utility theory establishes functional forms. This leads to a measurement methodology to directly assess individual specific compaction functions. Existing statistical techniques are summarized and finally empirical examples of both statistical compaction and direct assessment are given for the planning of health maintenance organizations.

6.1 Formal Development of Compaction Functions:

Definitions and their Interpretations

A verbal definition of compaction was given in the preceding paragraphs. Before formalizing this definition, let us first consider a simple yet illustrative example.

6.1.1 Example

Suppose we are a team of decision analysts getting together after a hard days work. Earlier today each of us conferred with a (different) zoologist about his purchase of a widget for wild wrinkled wombats who are over-weight and curiously enough each of us assessed our zoologist's utility function and found that the same performance measures, cost, tranquility, and a wrinkle-index apply for each zoologist. Furthermore, each utility function had the same functional form but with different parameters. (Call this function $u(\underline{x}, \lambda_i)$ where $\underline{x} = \{\text{cost},$

tranquility, wrinkle-index} and λ_i are the parameters of zoologist i's utility function). Finally, we were so successful that we convinced each client to use his utility function to make his decision.

A colleague from another division walks in with the known values of the performance measures for all competitive widgets as well as our firm's widget. As an intellectual exercise we wish to predict the choice probabilities for each zoologist.

We know each zoologist's decision rule because they are following our advice. Tomorrow our sales division is going to each zoologist with the same values our colleague has just given us and we know the zoologists will believe our salesmen. We therefore have enough knowledge to compute the choice probabilities. To do this we simply compute the utility value, u_{ij} , that zoologist i will assign widget j. Then with probability 1.0 zoologist i will choose the widget with the largest utility value.

In this example we are lucky enough to know the decision rule and perceptions for each individual. Nonetheless, each zoologist's utility function, $u(x, \lambda_i)$ is a compaction function because it compresses the performance measures into a single number, and because (for each individual) knowing those numbers for all alternatives is sufficient to predict choice probabilities. But what if some zoologists do not select the alternatives with maximum utility, do we then say that utility theory is wrong? Or can we expand the utility concept to get strong but not completely sufficient indicators of choice? In a real situation we may be able to measure consumers' utility functions, but we can not be sure of (1) how they perceive the performance measures, (2) whether

the set of performance measures is complete, (3) whether the utility function completely specifies the choice, or (4) whether we have made no errors in measuring an individual's parameters, λ_i . The concept of a compaction function allows for this uncertainty because it does not require that the choice probabilities are equal to 1.0 or 0.0. The next section presents a formal definition of compaction.

6.1.2 Compaction Definition

Let $A = \{a_1, a_2, \dots, a_J\}$ be a set of alternatives; let X_k be a performance measure, such as "quality," describing at least one alternative, $a_j \in A$.

Let $X = \{X_1, X_2, \dots, X_K\}$ be a complete set of performance measures (definition 4.3) and let $\underline{x}_j = \{x_{1j}, x_{2j}, \dots, x_{Kj}\}$ be the values that the performance measures take on for a deterministic alternative a_j . (The individual subscript, i , is temporarily dropped from the performance measure for ease of exposition). Let $\underline{x}_{-j} = \{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{j-1}, \underline{x}_{j+1}, \dots, \underline{x}_J\}$ where $\underline{x}_1 = \{x_{11}, x_{21}, \dots, x_{K1}\}$, etc. In other words \underline{x}_{-j} is the set of performance measures for all alternatives except a_j . Furthermore, let $p_i(a_j | \underline{x}_1, \underline{x}_2, \dots, \underline{x}_J; \lambda_i)$ be the probability that individual i chooses alternative a_j given the levels of the performance measures, \underline{x} , where $\underline{x} = \{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_J\}$ and given individual i preference parameters, λ_i .

Then we can define a compaction function for individual i as follows:

Definition 6.1: A real-valued function, $c_{ij}(\underline{x}_j, \lambda_i)$, is said to be a compaction function if whenever

$$c_{ij}(\underline{x}_j, \lambda_i) = c_{ij}(\underline{x}'_j, \lambda_i)$$

and \underline{x}_j not necessarily equal \underline{x}_j' , then:

$$p_i(a_j | \underline{x}_j, \underline{x}_j; \lambda_i) = p(a_j | \underline{x}_j', \underline{x}_j; \lambda_i)$$

with \underline{x}_j fixed.

There is a behavioral assumption inherent in definition 6.1, that is, tradeoffs among performance measures for one alternative do not depend upon the levels of the performance measures for the other alternatives. For example, suppose you are indifferent between {wait time bus = 9, travel time bus = 20} and {wait time bus = 15, travel time bus = 16} when subway is {wait time = 20, travel time = 15}. Definition 6.1 implies that you would still be indifferent between {9,20} and {15,16} for bus if subway was changed to {10,15}.

Notice so far that our compaction definition is quite general, in fact, $c_{ij}(\underline{x}_j, \lambda_i)$ can take on a different functional form for each individual and for each alternative, and $p(a_j | \underline{x}; \lambda_i)$ can be different for each individual and asymmetric in its arguments. In some studies some of those generalities may be necessary to appropriately model behavioral phenomena, but if compaction is to be of use to the design team certain of the generalities must be restricted so that useful information and insight is gained about tradeoffs, interdependencies and risk characteristics among the performance measures. To this end a series of desirable properties are identified and definitions developed to facilitate discussion. Later, primitive axioms will be presented which imply one or more of those properties.

6.1.3 Desirable Properties of Compaction Functions

Uniformity: By first going through reduction, an analyst tries to identify a set of performance measures which are complete. He then would hope that tradeoffs and interdependencies among those performance measures would not be alternative specific. In other words, knowing the performance measures, x_j , for alternative a_j and the preference parameters, λ_i , for individual i would be sufficient to compute individual i 's scalar measure of goodness, c_{ij} , for alternative a_j . Thus a uniform compaction function has the same functional form for all alternatives.

Formally:

Definition 6.2: A compaction function is uniform for an alternative set, A , if

$$c_{ij}(x_i, \lambda_i) = c_i(x_j, \lambda_i) \quad \text{for all } a_j \in A.$$

Notice that alternative specific terms can be included as performance measures as long as the functional form is the same for all alternatives in A .

Symmetry: Uniformity deals with the functional form of the compaction function, symmetry deals with the functional form of the conditional probability law. Symmetry implies that a specific value of the scalar measure of goodness has the same implications for each alternative. To better understand this consider the new notation:

$p_i(a_j | c_{i1}, c_{i2}, \dots, c_{iJ})$ = the probability of choosing alternative a_j given that $c_{i1}(x_1, \lambda_i) = c_{i1}, c_{i2}(x_2, \lambda_i) = c_{i2}$, etc. This notation is consistent by the definition of compaction. Furthermore define

$c_{i,j\bar{k}} = \{c_{i1}, c_{i2}, \dots, c_{ij-1}, c_{ij+1}, \dots, c_{ik-1}, c_{ik+1}, \dots, c_{iJ}\}$, that is, $c_{i,j\bar{k}}$ is the set of all scalar measures of goodness (for individual i) except c_{ij} and c_{ik} . Thus formally:

Definition 6.3: A compaction function (and the probability law it evokes) is said to be symmetric for an alternative set A if for all pairs of j and k , $j, k \in A$:

$$p_i(a_j | c_{ij} = x, c_{ik} = y, c_{i,j\bar{k}} \text{ fixed}) = p_i(a_k | c_{ij} = y, c_{ik} = x, c_{i,j\bar{k}} \text{ fixed})$$

Less formally, switching the compaction values for any j - k pair switches the choice probabilities for j and k but leaves all other choice probabilities unchanged. (This can be shown from definition 6.3 since the sum of all choice probabilities must sum to 1.0. Include the null choice, a_0 , as an alternative).

Preference: In chapter 2 compaction was introduced to enable the design team to understand how the performance measures combine and thus guide the decisions they must make in the design of alternatives. To enhance understanding it is convenient if a larger scalar measure of goodness means a larger choice probability (all other alternatives held fixed). This property, preference, is defined:

Definition 6.4: A compaction function is preferential if:

1. $c_{ij}(\underline{x}_j, \underline{\lambda}_i) > c_{ij}(\underline{x}_j', \underline{\lambda}_i) > -\infty$ implies

$$p_i(a_j | \underline{x}_j, \underline{x}_{\bar{j}}; \underline{\lambda}_i) \geq p(a_j | \underline{x}_j', \underline{x}_{\bar{j}}; \underline{\lambda}_i) \quad \text{with } \underline{x}_{\bar{j}} \text{ fixed, and}$$

2. $c_{ij}(\underline{x}_j, \underline{\lambda}_i) = -\infty$ and $c_{ik}(\underline{x}_k, \underline{\lambda}_i) \neq -\infty$ for some k implies

$$p(a_j | \underline{x}_j, \underline{x}_{\bar{j}}; \underline{\lambda}_i) = 0.$$

The boundary condition (2) is added for convenience and to prevent pathological cases when two or more alternatives are "infinitely bad."

Encompassment: Up to this point, compaction functions have been defined for a single individual, but to be useful in a design effort some form of homogeneity within some group of individuals must be identified.

Consider the zoologist example. Each compaction function (utility function) had the same functional form, $u(\underline{x}_j, \underline{\lambda}_i)$, but the parameters of the compaction function varied from individual to individual, i.e., the differences in preference across zoologists were completely specified by the preference measures, $\underline{\lambda}_i$. Furthermore each zoologist had the same probability law (decision rule), he choose the widget with the largest scalar measure of goodness (utility value). In this case the compaction function, $u(\underline{x}_j, \underline{\lambda}_i)$, is said to encompass the group of zoologists because (1) the same functional form applies to each individual and (2) the same probability law is evoked by the set of scalar measures of goodness.

Formally:

Definition 6.5: A compaction function (and the probability law it evokes) is said to encompass a group, G , of individuals if for all $i \in G$:

$$1. \quad c_{ij}(x_{ij}, \lambda_i) = c_j(x_{ij}, \lambda_i) \quad \text{for all } j \text{ and}$$

$$2. \quad p_i(a_j | c_{ij}, \dots, c_{iJ}) = p(a_j | c_{i1}, \dots, c_{iJ}) \quad \text{for all } j$$

Notice that in definition 6.5 the individual subscript, i , has been added to the performance measures, x_{ij} . This allows for individual variation in the perceptions of the levels of the performance measures.

The real power of an encompassing compaction function comes from the fact that it allows individuals to be defined by their preference parameters, and thus paves the way for direction population assessment. (More on this in section 6.4).

Canonicity: A canonical compaction function is simply a compaction function with all four of the previously defined . desirable properties.

Formally:

Definition 6.6: A compaction function (and the probability law it evokes) is said to be canonical for a group of individuals, G , and an alternative set, A , if it:

- 1) is uniform for the set A
- 2) is symmetric for the set A
- 3) is preferential, and
- 4) encompasses the group of individuals, G .

For example, the compaction function, $u(\underline{x}_j, \underline{\lambda}_i)$, in the zoologist example is canonical because (1) it is uniform (each zoologist cares only about {cost, tranquility, wrinkle-index} and $u(\underline{x}_j, \underline{\lambda}_i)$ compacts these performance measures in the same way independent of the specific alternative widget being considered), (2) it is symmetric (the same decision rule, maximize the utility value, and hence the same probability law applies if the indices of the alternatives are changed), (3) it is preferential (if u_{ij} increases and u_{ik} remains the same for all $k \neq j$, then the probability of choosing widget j stays the same or increases, and (4) it encompasses the zoologists (the same functional form of the compaction function and the probability law apply to each zoologist).

6.1.4 Summary of Definitions

Definition 6.1 started with a construct, compaction functions, as an intermediary step before probability of choice. The purpose of compaction functions is to explicitly identify tradeoffs, interdependencies and risk characteristics among the performance measures in such a way as to give insight to the design team. These characteristics would not be as transparent in a conditional probability law of the form: $p_i(a_j | \underline{x}_i, \dots, \underline{x}_J; \underline{\lambda}_i)$. Next a series of intuitive ideas were formalized to emphasize the issues and to avoid ambiguity. These intuitive ideas will later be used to develop "real world" methodologies that are useful to the design process.

The next section presents two simple theorems which will later prove useful in (heuristic) interpersonal normalization of compaction

functions.

6.1.5 Bijection and Monotone Theorems.

Bijection: Suppose a general compaction function has been constructed and some transformation has been identified which gives it one or more desirable properties (definitions 6.2-6.6) Is the transformed function a compaction function, and in general, what class of transformations retain the property of compaction? A sufficient condition is bijection:

Theorem 6.1: Let S_1, S_2 be any two sets of real numbers and let $f: S_1 \rightarrow S_2$ be a bijection from S_1 onto S_2 , then $f(c_{ij}(\cdot))$ is a compaction function, if and only if $c_{ij}(\cdot)$ is a compaction function.

Proof: (If:) Suppose $f(c_{ij}(\underline{x}_j, \underline{\lambda}_i)) = f(c_{ij}(\underline{x}'_j, \underline{\lambda}_i))$. f a bijection implies $c_{ij}(\underline{x}_j, \underline{\lambda}_i) = c_{ij}(\underline{x}'_j, \underline{\lambda}_i)$, and $c_{ij}(\cdot)$ a compaction function implies $p_i(a_j | \underline{x}_j, \underline{x}_j; \underline{\lambda}_i) = p_i(a_j | \underline{x}'_j, \underline{x}_j; \underline{\lambda}_i)$. Thus $f(c_{ij}(\cdot))$ satisfies definition 6.1 and is therefore a compaction function.

(Only if:) Since f is a bijection, f^{-1} exists and is a bijection. Thus by the first half of this proof, $f(c_{ij}(\cdot))$ a compaction function implies $f^{-1}f(c_{ij}(\cdot)) = c_{ij}(\cdot)$ is a compaction function.

Comment 6.1: Since the range of f can be a subset of R , for example $[0,1]$, finite normalization can be performed and the property of compaction maintained.

Comment 6.2: If f is not a bijection, then $f(c_{ij}(\cdot))$ may still be a compaction function for a particular set of alternatives and a feasible

set of values for performance measure, but this property can not be guaranteed.

Comment 6.3: Bijection also maintains other properties such as completeness and symmetry if the same f is used for all alternatives, a_j , and encompassment if the same f is used for all individuals, i. The proofs are similiar. Bijection is not sufficient for preference. Consider for example $f(x) = -x$.

Monotone: Bijection is sufficient for compaction, completeness, symmetry, and encompassment but not for preference. Yet, since preference provides an intuitive link to the design process, it is perhaps one of the most important desirable properties. What class of transformations maintain preference? A sufficient condition is monotonically increasing:

Theorem 6.2: Let S_1, S_2 be any set of real numbers and let $f: S_1 \rightarrow S_2$ be a monotonically increasing function from S_1 onto S_2 , then $f(c_{ij}(\cdot))$ is a preferential compaction function if and only if $c_{ij}(\cdot)$ is a preferential compaction function.

Proof: (If:) f a monotonically increasing function onto S implies f is a bijection onto S , thus $f(c_{ij}(\cdot))$ is a compaction function. Suppose $f(c_{ij}(\underline{x}_j, \underline{\lambda}_i)) > f(c_{ij}(\underline{x}_j', \underline{\lambda}_i)) > -\infty$. f monotonically increasing implies $c_{ij}(\underline{x}_j, \underline{\lambda}_i) > c_{ij}(\underline{x}_j', \underline{\lambda}_i) > -\infty$, and $c_{ij}(\cdot)$ preferential compaction function implies $p(a_j | \underline{x}_j, \underline{x}_j; \underline{\lambda}_i) \geq p(a_j | \underline{x}_j', \underline{x}_j; \underline{\lambda}_i)$. Furthermore if $-\infty \in S$ then $f(c_{ij}(\underline{x}_j, \underline{\lambda}_i)) = -\infty$ and $f(c_{ik}(\underline{x}_k, \underline{\lambda}_i)) \neq -\infty$ implies $c_{ij}(\underline{x}_j, \underline{\lambda}_i) = -\infty$ and $c_{ij}(\underline{x}_k, \underline{\lambda}_i) \neq -\infty$ which implies $p(a_j | \underline{x}_j, \underline{x}_j; \underline{\lambda}_i) = 0$. Thus $f(c_{ij}(\cdot))$ satisfies definition 4.4 and is therefore preferential.

(Only if:) Since f is a monotonically increasing function from R onto S , f^{-1} exists and is a monotonically increasing function from S onto R . Thus by the first half of this proof $f(c_{ij}(\cdot))$ a preferential compaction function implies $f^{-1}f(c_{ij}(\cdot))$ is a preferential compaction function.

Comment 6.4: Since the range of f can be a subset of R , finite normalization can retain preferential compaction.

Comment 6.5: If $g(x) \neq -\infty$ for x finite and if g is monotonically non-decreasing then $g(c_{ij}(\cdot))$ satisfies definition 6.4 (preference) but is not necessarily a compaction function.

Comment 6.6: Since a monotonically increasing function onto S is a bijection, completeness, symmetry, and encompassment can be retained if f is the same for all alternatives, a_j , and individuals, i . This, because f also retains preference, it retains canonicity.

Finally an often used transformation is a positive linear transformation. Since a positive linear transformation is monotonically increasing transformation from R onto R , it retains all properties as noted above. For emphasis this fact is stated as a corollary:

Corollary 6.3: $a + bc_{ij}(\cdot)$ is a (complete/symmetric/encompassing/preferential/canonical) compaction function if and only if $c_{ij}(\cdot)$ is a (complete/symmetric/encompassing/preferential/canonical) compaction function. a, b constants, $b > 0$.

The notation $(//)$ means any or all of the stated properties are retained.

The completes the formal development of the compaction definitions, the next section begins with an axiomization of individual behavior and derives useful relations and functional forms for compaction functions.

6.2 An Axiomization for Compaction Functions to Establish an Isomorphism with the von Neumann-Morgenstern Utility Axioms

In 1947 von Neumann and Morgenstern [151], identified axioms of rational choice which have proven very powerful and useful in the development of prescriptive utility theory. For example, authors such as Fishburn [38], Keeney [68-73], Raiffa [120], Ting [140], and others have traced out the implications of these axioms, identified assumptions inherent in certain functional forms, and have developed to a fine art the structuring of preferences for the purpose of aiding decisions. Uni-attributed utility theory is often used for the explicit incorporation of risk in monetary decisions (Raiffa [120]), while multi-attributed utility theory has been used to structure such large scale decision problems as the siting of Mexico City's airport (Keeney [67]), the New York City air pollution problem (Ellis and Keeney [31]), the use of frozen blood in blood banking (Bodily [11]), and the siting of nuclear power plants (Keeney and Nair [75]). To date, von Neumann-Morgenstern utility theory has not been used to predict choice.

There is however a large body of literature on random utility models for choice theory, for example McFadden [93,94], Tversky [143], Luce [92], Ben-Akiva [10], Manski [99] and others. The dominant concern is with the distribution of the error term, and thus with the form of the

probability law rather than with the functional form of the utility function. Furthermore, the utility function is usually assumed identical for the entire population (within segments) and its parameters are econometrically estimated.

Thus on one hand we have prescriptive utility theory with its axiomatic richness which allows the explicit determination of a function (and its parameters) for each individual but which requires an individual to be rational and always choose that alternative which yields a maximum utility value. On the other hand we have choice theory which recognizes randomness in choice due to (1) measurement error in the performance measures and the parameters of the function, (2) approximations in the functional form and incomplete specification of the set of preference measures, (3) temporal instability (day to day change) in preferences, and (4) "irrationality" in behavior.

Section 6.1 recognized this randomness in its definition of compaction. (Utility functions in random utility models will later be shown to be compaction functions). This section relaxes the utility theoretic restriction of maximal utility choice, restates the von Neumann-Morgenstern in terms of this relaxation, and postulates a choice axiom which is necessary to establish an isomorphism between utility theoretic results and descriptive choice theory.

6.2.1 Stochastic Preference

There are three critical steps in this formalization, the first is in redefining preference:

Definition 6.7 (Stochastic preference among alternatives). Let A be a set of alternatives, let a_j, a_k be elements of A , then $>_A$ is a stochastic preference operator on $A \times A$ if $(a_j, a_k) \in >_A$ implies the probability of choosing a_j from A is greater than the probability of choosing a_k from A . For simplicity write $a_j >_A a_k$ and say a_j is stochastically preferred to a_k (relative to A).

Definition 6.8 (Stochastic indifference among alternatives). Let A be a set of alternatives and let a_j, a_k be elements of A , then \sim_A is a stochastic indifference operator on $A \times A$ if $(a_j, a_k) \in \sim_A$ implies the probability of choosing a_j from A equals the probability of choosing a_k from A . For simplicity write $a_j \sim_A a_k$ and say a_j is stochastically indifferent to a_k (relative to A).

Comment 6.7: Similiarly make the obvious definitions for \geq_A , $<_A$, and \lesssim_A .

Comment 6.8: Definition 6.8 could be extended with the concept of a "just noticeable difference" (jnd) (Fishburn [35]) by stating that $a_j \sim_A a_k$ if the probability of choosing a_j from A is within ϵ (a small number) of the probability of choosing a_k from A . Unfortunately such an indifference operator is not necessarily transitive and is therefore not compatible with our axioms.

Comment 6.9: To simplify exposition $>_A, \sim_A$ will be written as $>, \sim$ when the choice set is clear from context.

Comment 6.10: Note that although A is written as a countable set it is not required to be such and can in general be uncountable.

Consider a set, A , of alternatives and stochastic preference and indifference operators, $>$, \sim , on A . Define a lottery operation, $L(a_j, a_k; p)$ as follows:

Definition 6.9: A lottery, $L(a_j, a_k; p)$: $A \times A \times R \rightarrow A^*$, is an alternative which has the characteristics of a_j with probability p , and the characteristics of a_k with probability $(1-p)$. (A^* is the range of L). (See figure 6.2).

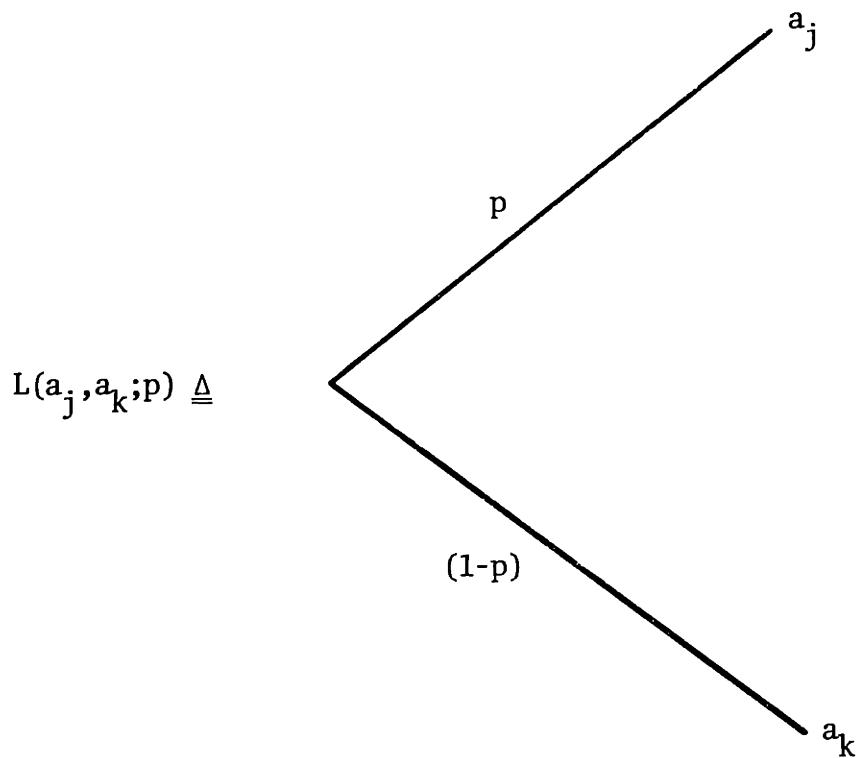


Figure 6.2: Lottery Definition

Comment 6.11: $A \subseteq A^*$ since one possible lottery is the degenerate lottery, i.e., $p = 1.0$.

6.2.2 The Axioms

Suppose A^* , $>$, \sim , and L satisfy the following axioms:

Axiom 6.1: $>$ is a complete ordering on A^* .

(a) For any two a_j, a_k exactly one of the following holds.

$$a_j > a_k, a_j \sim a_k, a_j < a_k$$

(b) $a_j > a_k$ and $a_k > a_\ell$ implies $a_j > a_\ell$

(c) $a_j \sim a_k$ and $a_k \sim a_\ell$ implies $a_j \sim a_\ell$

Axiom 6.2: Ordering and combining:

(a) $a_j \begin{cases} > \\ < \end{cases} a_k$ implies $a_j \begin{cases} > \\ < \end{cases} L(a_j, a_k; p)$ for all $p \in (0, 1)$.

(b) $a_j > a_k > a_\ell$ implies the existence of $p_1, p_2, p_3 \in (0, 1)$ such

that

$$L(a_j, a_\ell; p_1) < a_k$$

$$L(a_j, a_\ell; p_2) \sim a_k.$$

$$L(a_j, a_\ell; p_3) > a_k$$

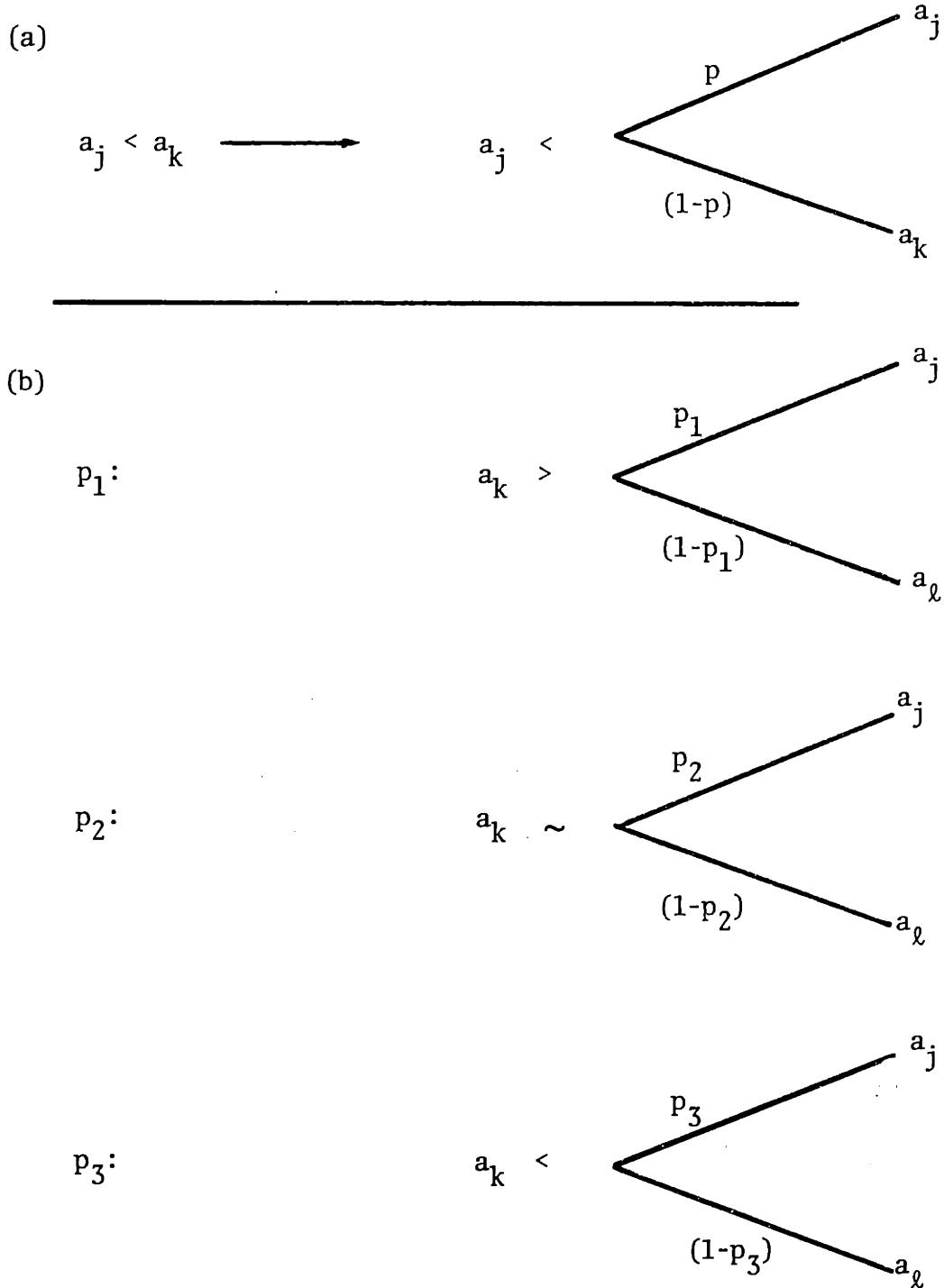


Figure 6.3: Schematic of Ordering and Combining Axiom

Axiom 6.3: Algebra of combining

$$(a) \quad L(a_j, a_k; p) \sim L(a_k, a_j; 1-p)$$

$$(b) \quad L[L(a_j, a_k; p), a_k; q] \sim L(a_j, a_k; pq)$$

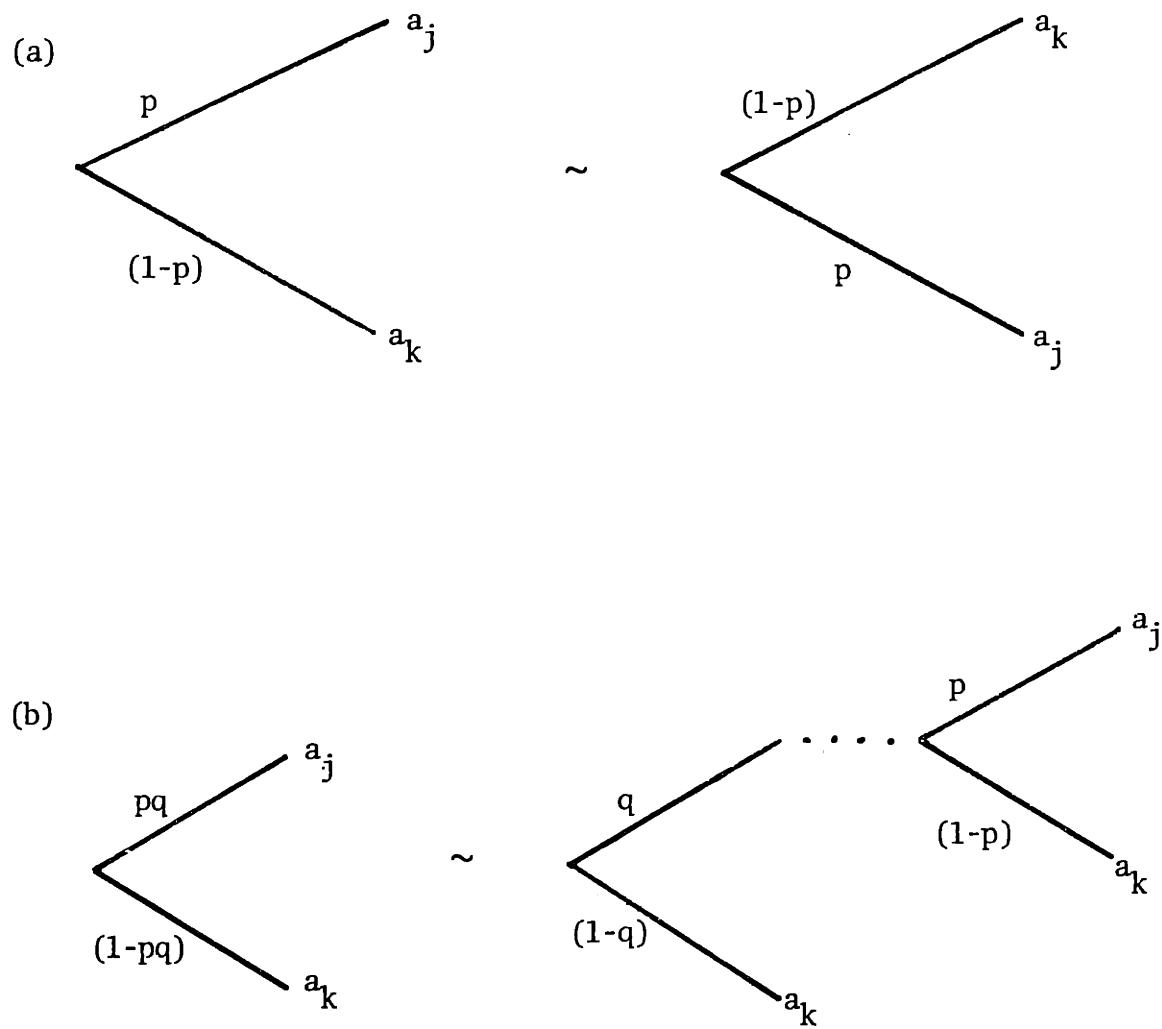


Figure 6.4: Schematic of Algebra of Combining Axiom

Axiom 6.4: Choice axiom. Let α be any subset of A^* , let

$a_j, a_k, a_j', a_k' \in \alpha \subseteq A^*$, then $a_j \sim_A a_j'$ and $a_k \sim_A a_k'$

implies $\text{Prob}\{a_j \text{ from } \alpha - a_k\} = \text{Prob}\{a_j' \text{ from } \alpha - a_k'\}$

where $\alpha - a_k$ is the set α with the element a_k deleted.

Axioms 6.1-6.3 are restatements of the von Neumann-Morgenstern utility axioms, except that the definition of preference is changed. Axiom 6.4 is a choice axiom which allows us to build models based on abstract choices and then apply them to realistic finite choices. The need for such a choice axiom will become clear later in the development. These axioms imply the existence of a compaction function, but before this is shown, let us interpret the axioms.

6.2.3 Interpretation of the Axioms

Axiom 6.1 (Complete ordering): (a) In utility theory this is a reasonably strong assumption, i.e., that an individual can state his preferences and that they are temporally stable. The new preference definition allows stochastic behavior, thus the new interpretation is that an individual's "average" behavior has no uncontrollable long term trend. (b+c) This property is actually induced by the preference definition because $>$ and $=$ are transitive for the real numbers. It is stated explicitly to maintain a parallel with the utility axioms.

Axiom 6.2 (Ordering and combining): (a) This states simply that if a_k is stochastically preferred to a_j then a lottery with even a slight chance

of a_k is preferred to a_j . ("Losing" the lottery gives a_j). (b) If a_k is stochastically preferred to a_ℓ , then given a lottery, $L(a_j, a_\ell; p)$, involving an alternative, a_j , which is stochastically preferred to a_k and a_ℓ , and a_ℓ then the influence of a_j can be made sufficiently small (p_1 close to 0) such that a_k is still preferred to the lottery. (Review figure 6.3). Furthermore, each individual can conceive of some probability, p_2 , which makes him stochastically indifferent between the middle alternative, a_k , and a lottery involving the extreme alternatives, a_j and a_ℓ . Taken together parts (a) and (b) of this axiom imply a reasonable continuity assumption. Since they are stated for stochastic preference these axioms imply properties of the probability model. This will be dismissed in later sections and the next chapter.

Axiom 6.3 (Algebra of combining) (a) This states simply that the lottery operation is commutative, i.e., it does not matter in which order the elements of the lottery are named. (b) This statement of associativity is perhaps the strongest assumption in the utility axioms and hence in our axioms. It states that a series of successive lotteries can be treated as an equivalent one step lottery. In other words it states that every individual can conceive of a complex lottery and that he will rationally react to it as if it were a simple lottery with equivalent probabilities.

Axiom 6.4 (Choice axiom): This axiom states that if the probability of choosing an alternative, a_j , is equal to the probability of choosing another alternative, a_j' , when all alternatives are available then for any subset of the alternatives this equality of probabilities remains

the same if some alternatives (other than a_j, a_j') are deleted from consideration. Furthermore if a_k and a_k' are indifferent on A, then deletion of one is the other is equivalent in terms of stochastic indifference on the respective subsets. In other words if two alternatives are equivalent on the entire choice set, then they are equivalent in their presence or their absence from any subset. This is certainly a reasonable assumption for distinct choices, but be careful, for certain types of choices, particularly hierarchical choices, it can break down.

For example, suppose Michael Doonesburger, a fictitious student, has the following choice probabilities for health care delivery: Boston Group Practice (BGP), .4; private care with Dr. Jones, .3; private care with Dr. Smith, .3, and suppose these choices represent an exhaustive list. Now suppose Dr. Smith runs off to Jamaica. Will BGP still be stochastically preferred to Dr. Jones? Maybe, but perhaps Mike's decision rule is to first choose between group practice and private care and then randomly select a doctor if he decides on private care. This might imply that $Dr. Jones > BGP$ ($.6 > .4$) after Dr. Smith departs.

This example cautions us not to blindly apply models derived from axioms 6.1-6.4. Instead the axioms must be verified before models are built, and if the choice process is hierarchical (sequential) it must be modeled as such.

If this axiom causes trouble with hierarchical choices, why include it? It is needed because alternatives will be represented by sets of performance measures and compaction functions will be inferred from questions about stochastic indifference among abstract alternatives

(represented by values for the performance measures). Thus compaction functions will be determined on uncountable choice sets, $\{X_1, X_2, \dots, X_N\}$, and applied to finite subsets, $\{a_1, a_2, \dots, a_J\}$.

The next subsection discusses some mathematical implications of the axioms.

6.2.4 Existence and Uniqueness Theorems

The first and most significant implication of the axioms is the existence of a real valued function on the expanded alternative set, A^* , which preserves (stochastic) preference and for which mathematical expectation applies. The proof of this result is quite tedious and exactly parallels the proof for utility functions contained in the appendix of von Neumann and Morgenstern [151]. Thus it is stated here without formal proof.

Theorem 6.4 (Existence): There exists a real valued function, c^* , on A^* , $c^*: A^* \rightarrow I \subseteq \mathbb{R}$, with the following properties:

$$(a) \quad a_j \begin{cases} \geq \\ \sim \\ < \end{cases}_{A^*} a_k \iff c^*(a_j) \begin{cases} \geq \\ \equiv \\ < \end{cases}_R c^*(a_k)$$

$$(b) \quad c^*[L(a_j, a_k; p)] = p c^*(a_j) + (1-p) c^*(a_k)$$

where $a_j, a_k \in A^*$, I is an interval, $p \in [0, 1]$.

Idea of proof: The first step is to use axiom 4.1 to choose a^0, a^* such that $a^0 < a^*$ and axiom 4.2 to show the existence of a monotone mapping from the interval $(0, 1)$ onto the interval $a^0 < a < a^*$. Then define $c^*(a^0) = 0$ and $c^*(a^*) = 1.0$ and use axiom 6.3 to show $c^*(L(a^0, a; p)) = pc^*(a)$ and $c^*(L(a^*, a; p)) = p + (1-p)c^*(a)$.

Then extend the mapping to $[0,1]$. Finally extend $c^*(\cdot)$ outside the range $a^0 \leq a \leq a^*$, and show $c^*(L(a,a;p)) = c^*(a)$. Combining all of these together using axiom 6.3b with careful consideration of detail yields the desired result.

In section 6.1 a number of desirable properties of compaction functions were identified and theorems presented to identify which transformations retained these properties, in particular it was shown (corollary 6.3) that a positive linear transformation retains all properties so far defined. The function, c^* , looks similar to a preferential compaction function, but with the additional property of mathematical expectation being appropriate. (This property is known in the literature as cardinality). Later, a related function, will be shown to be a compaction function, but first let us investigate how c^* 's properties behave under transformation.

Theorem 6.5 (Uniqueness): The function $c^*:A^* \rightarrow I \subseteq \mathbb{R}$ is unique up to a positive linear transformation.

Proof: Suppose there exists a function $d^*:A^* \rightarrow \mathbb{R}$. Then there must exist a function $f:I \rightarrow \mathbb{R}$ such that for any $a_j \in A^*$, $f(c^*(a_j)) = d^*(a_j)$. (See figure 6.5) Since c^* and d^* both satisfy the properties of theorem 6.4

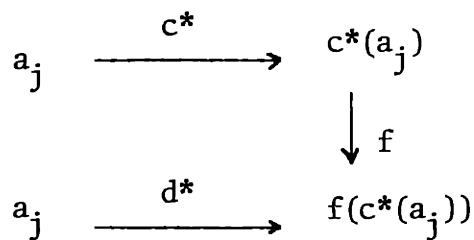


Figure 6.5: Schematic of Uniqueness Proof

then for $x, y \in I$

$$(a) \quad x \begin{cases} \geq \\ \leq \end{cases} R y \text{ implies } f(x) \begin{cases} \geq \\ \leq \end{cases} R f(y)$$

$$(b) \quad f(px + (1-p)y) = pf(x) + (1-p)f(y)$$

But (b) is just the definition of a linear function thus $f(x)$ is linear in x . Furthermore (a) implies that $f(x)$ is monotonically increasing in x , thus $f(x)$ must be a positive linear transformation. Finally a linear transformation of an interval is an interval. Thus if any function, d^* , satisfies theorem 6.4, it must be a linear transformation of another function, c^* , which satisfies theorem 6.4.

(This proof is similar to the uniqueness proof in utility theory).

6.2.5 Empirical Use of the Axioms Requires Representation of Alternatives as Sets of Performance Measures

The methodology described in this dissertation requires that alternatives be represented by sets of reduced performance measures, and to this end compaction was defined in terms of a complete set of performance measures. In other words, definition 6.1 presupposes a correspondence between the alternative set, A , and a finite dimensional real vector space, R^N , i.e., there exists $g: A \rightarrow R^N$, such that for all $a_j \in A$ there corresponds a unique vector, x_j , of performance measures. (Note: Let $X \subseteq R^N$ be the set of performance measures as in definition 6.1).

What we would like to do is define some function, $c: X \rightarrow I \subseteq R$ with the property that c is a compaction function and $c(x_j) = c^*(a_j)$.

This seems trivial but in most cases the mapping g is not onto nor is it one-to-one. In other words some elements of X may not represent actual alternatives or they may represent more than one. Thus we need the following axiom:

Axiom 6.5: Abstract alternatives. An individual can make judgements, and in particular, indicate stochastic preference for abstract alternatives indicated only by values for the performance measures.

If reduction is done correctly, and a complete set of performance measures (chapter 4) is identified, then \underline{x}_j is equivalent to a_j for stochastic preference considerations. Axiom 6.5 simply extends the choice set A to all abstract alternatives represented by elements of X . $c(\underline{x}_j)$ will now be shown to be a compaction function.

Theorem 6.6: Let $A = (a_1, a_2, \dots, a_J)$ be the set of alternatives. Let X be a complete set of performance measures. Suppose axiom 6.5 holds. Let $g: A \rightarrow X \subseteq \mathbb{R}^N$ be a function mapping the alternatives into the performances measures. Let $c: X \rightarrow I \subseteq \mathbb{R}$ be a real valued function on the set of performance measures such that $c(g(a_j)) = c(\underline{x}_j) = c^*(a_j)$. Let $\text{Prob} \{a_j \text{ from } A: c^*(a_j) = -\infty, c^*(a_k) \neq -\infty \text{ for some } k\} = 0$. Then c is a uniform, symmetric, preferential compaction function.

Proof: (Compaction) By axiom 6.5 X can be an alternative set. Call it α . Let $\underline{x}_j = g(a_j), \underline{x}_j' = g(a_j')$, be two elements of X . By hypothesis $c(\underline{x}_j) = c(\underline{x}_j') \Rightarrow c^*(a_j) = c^*(a_j')$ thus $\underline{x}_j \sim_{\alpha^*} \underline{x}_j'$ which means $\text{Prob} \{\underline{x}_j \text{ from } \alpha^*\} = \text{Prob} \{\underline{x}_j' \text{ from } \alpha^*\}$. Let $B = \{\underline{x}_j, \underline{x}_j', \underline{x}_j\}$ which

is also a choice set. Then by axiom 6.4 applied to B

$\text{Prob} \left\{ \underline{x}_j \text{ from } B - \underline{x}_j' \right\} = \text{Prob} \left\{ \underline{x}_j' \text{ from } B - \underline{x}_j \right\}$. Finally since X is a complete set performance measures this implies $p_i(a_j | \underline{x}_j, \underline{x}_j; \underline{\lambda}_i) = p_i(a_j | \underline{x}_j', \underline{x}_j; \underline{\lambda}_i)$ which implies c is a compaction function.

(Preference) By hypothesis $c(\underline{x}_j) > c(\underline{x}_j')$ implies $c^*(a_j) > c^*(a_j')$ which implies $a_j >_{A^*} a_j'$ which implies $\underline{x}_j >_{X^*} \underline{x}_j'$.

Defining B as above and following similar reasoning implies

$p_i(a_j | \underline{x}_j, \underline{x}_j; \underline{\lambda}_i) > p_i(a_j | \underline{x}_j', \underline{x}_j; \underline{\lambda}_i)$. The boundary condition is satisfied by the hypothesis of the theorem.

(Uniformity): Since the alternative set can be X by axiom 6.5 and since c is defined on X and since g is defined on A the same c applies for all alternatives (for all $\underline{x}_j \leftrightarrow a_j$).

(Symmetry): Suppose we have a choice set $a = \{\underline{x}_j, \underline{x}_j', \underline{x}_k, \underline{x}_k', \underline{x}_{jk}, \underline{x}_{jk}'\}$ with corresponding values for $c(\underline{x})$ of $\{y, z, z, y, c_{\overline{1j}}\}$. Since $c(\underline{x}_j) = c(\underline{x}_j') = y$ and $c(\underline{x}_k) = c(\underline{x}_k') = z$ then $\underline{x}_j \sim_{A^*} \underline{x}_k'$ and $\underline{x}_k \sim_{A^*} \underline{x}_j'$. Thus by axiom 6.4, $\text{Prob} \left\{ \underline{x}_j \text{ from } a - \underline{x}_j' \right\} = \text{Prob} \left\{ \underline{x}_k' \text{ from } a - \underline{x}_k \right\}$ and again $\text{Prob} \left\{ \underline{x}_j \text{ from } a - \underline{x}_j' - \underline{x}_k' \right\} = \text{Prob} \left\{ \underline{x}_k' \text{ from } a - \underline{x}_j - \underline{x}_k \right\}$, but the last statement just says $p_i(a_j | c_j = y, c_k = z, c_{jk} \text{ fixed}) = p_i(a_k' | c_j' = z, c_k' = y, c_{jk} \text{ fixed})$ since c is a compaction function. This is the definition of symmetry.

Theorems 6.4, 6.5 taken together state that if a set of alternatives, A, satisfy "reasonable" axioms of choice (axioms 6.1-6.4), if these alternatives can be represented by a set of performance measures, X, and if an individual can conceive of abstract alternatives represented by

values from X (axiom 6.5) then there exists a real valued function, $c(\underline{x})$, on X , (unique up to a positive linear transformation), which is a complete, symmetric, preferential compaction function and for which mathematical expectation applies. This result is very powerful since it allows direction assessment of $c(\underline{x})$ if $c(\underline{x})$ is of some "nice" form.

In other words, suppose $c(\underline{x}_j, \underline{\lambda}_i)$ is known up to a set of preference parameters, $\underline{\lambda}_i$. Then by axiom 6.5 an individual can be asked questions of the form: $\underline{x}_j \sim_A^* \underline{x}_k$ and by theorems 6.4, 6.5, and 6.6 $c(\underline{x}_j, \underline{\lambda}_i) = c(\underline{x}_k, \underline{\lambda}_i)$. This gives one equation in the unknown's, $\underline{\lambda}_i$. Similarly by axioms 6.3 and 6.5 an individual can be asked to find a probability, p , such that $L(\underline{x}_j, \underline{x}_k; p) \sim_A^* \underline{x}_\ell$, and by theorems 6.4, 6.5, and 6.6 $pc(\underline{x}_j, \underline{\lambda}_i) + (1-p)c(\underline{x}_k, \underline{\lambda}_i) = c(\underline{x}_\ell, \underline{\lambda}_i)$. This again gives one equation in the unknowns, $\underline{\lambda}_i$. Summarizing, there are type I and type II questions (see figure 6.6) giving type I and type II equations:

$$c(\underline{x}_j, \underline{\lambda}_i) = c(\underline{x}_k, \underline{\lambda}_i) \quad \text{type I}$$

$$pc(\underline{x}_j, \underline{\lambda}_i) + (1-p)c(\underline{x}_k, \underline{\lambda}_i) = c(\underline{x}_\ell, \underline{\lambda}_i) \quad \text{type II.}$$

Comment 6.11: Type I is actually a degenerate type II equation ($p = 1.0$) but is stated as distinct because of measurement considerations.

Comment 6.12: Measurement issues may cause both type I and type II equations to be estimated rather than exactly determined.

The next section will discuss assumptions necessary to determine what functional forms are correct for compaction functions on X .

Type I

$$\underline{x}_j \sim \underline{x}_k$$

Type II

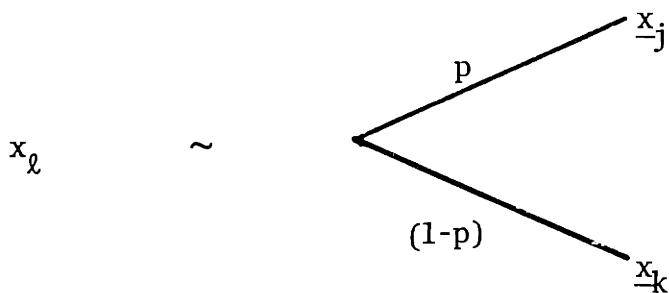


Figure 6.6: Two Types of Measurement

6.2.6 Independence Assumptions

Up to this point we have only begun to tap the power of utility theory. Axioms 6.1-6.5 imply the existence and uniqueness of a complete, symmetric, preferential compaction function which remains valid under mathematical expectation. Not only that, there is a feasible way to measure it if only its functional form were known. One compromise might be to approximate $c(\underline{x}_j, \underline{\lambda}_i)$ by a function transformed to be linear in $\underline{\lambda}_i$ and econometrically estimate $\underline{\lambda}_i$. But what functional form is appropriate?

Another approach is to use the isomorphism established between compaction functions and utility functions and "borrow" the theorems from utility theory which axiomatically establish function forms. To do this, it is first necessary to define two independence properties, preferential

and utility independence.

(These properties were defined by Keeney [70], in 1969. The reader is cautioned that although the word "utility" appears in definition 6.11, this property is defined for compaction functions. The word "utility" is retained to avoid confusion since we are not now inventing new concepts.)

Preferential independence: Suppose that there are N performance measures, $\{x_1, x_2, \dots, x_N\}$ and suppose that tradeoffs among the first two, $\{x_1, x_2\}$ do not depend upon the rest, $\{x_3, x_4, \dots, x_N\} \equiv x_{\overline{12}}$. We would then say that $\{x_1, x_2\}$ is preferentially independent of $x_{\overline{12}}$. For example, suppose that the performance measures for describing dial-a-ride trips are $\{\text{wait time, travel time, cost}\}$ and $\{\text{wait time, travel time}\}$ is preferentially independent of $\{\text{cost}\}$, then if $\{\text{wt} = 10, \text{tt} = 20, \text{c} = .25\}$ is preferred to $\{\text{wt} = 15, \text{tt} = 15, \text{c} = .25\}$ this preference ordering will remain the same if the cost, common to both alternatives, is changed.

The above example is for two performance measures, the formal definition (stated below) is for any partition of the performance measures, i.e., any mutually exclusive and collectively exhaustive subsets of S .

[$\{Y, X\}$ is a partition of X if $Y \cap Z = \emptyset$ and $Y \cup Z = X$]

Definition 6.10: Let X be a complete set of performance measures and let $\{Y, Z\}$ be a partition of X . Then Y is said to be preferentially independent of Z , written $Y \text{ p.i. } Z$, if for all $y_1, y_2 \in Y$ and for some $z_0 \in Z$

$$(y_1, z_0) >_x (y_2, z_0) \text{ implies } (y_1, z) >_x (y_2, z)$$

for all $z \in Z$.

Comment 6.13: The only difference between this definition and the utility theoretic definition is that stochastic preference relative to $X, >_X$, replaces simple preference, $>$.

Utility Independence: Preferential independence deals only with deterministic alternatives, the compaction function of Theorem 6.6 and the type II calibration deal with uncertain alternatives, i.e., lotteries. The next definition, utility independence, simply extends preferential independence to alternatives defined by lotteries. For example, suppose your choice of airlines depends only on {travel time, cost} and {travel time} is utility independent of {cost} then if $L(\{tt = 10, c = 1\}, \{tt = 5, c = 1\}, \frac{1}{2})$ is preferred to $\{tt = 8, c = 1\}$, this preference ordering will remain the same if cost, common to both lotteries, is changed. (See figure 6.7.)

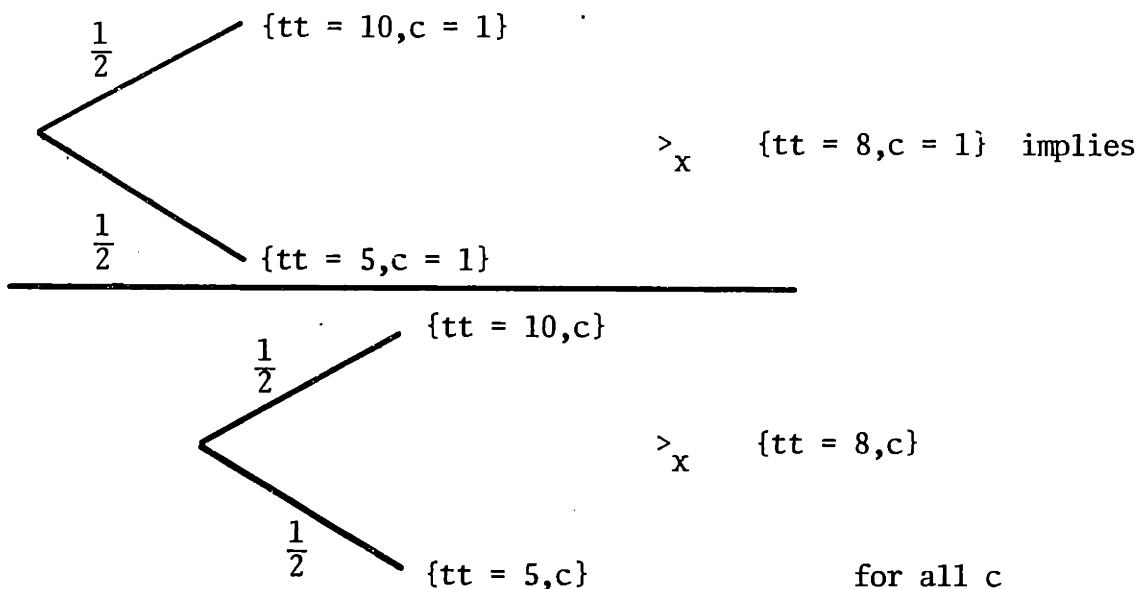


Figure 6.7: Utility Independence

To state utility independence we need a new notation for complex lotteries. Let $\ell(\tilde{x}_j)$ be an alternative which has characteristics \tilde{x}_j where \tilde{x}_j is a random variable with some known probability mass (density) function.

Comment 6.14: The axioms 6.1-6.3 imply continuity as well as mathematical expectation, thus complex lotteries, $\ell(\tilde{x}_j)$ can be built up from simple lotteries, $L(x_j, x_k, p)$ and theorems 6.4-6.6 can be shown to hold for $\ell(\tilde{x}_j)$.

Formally:

Definition 6.11: Let X be a complete set of performance measures and let X^* be the set of all lotteries involving elements of X . Let Y, Z be a partition of X , with corresponding lottery sets Y^*, Z^* . Then Y is said to be utility independent of Z , written Y u.i. Z , if for all $y_1, y_2 \in Y^*$ and for some $z_0 \in Z$

$$\ell(\{y_1, z_0\}) >_{x^*} \ell(\{y_2, z_0\})$$

$$\text{implies } \ell(\{y_1, z\}) >_{x^*} \ell(\{y_2, z\}) \text{ for all } z \in Z.$$

Comment 6.15: Utility independence is defined above for stochastic preference.

Comment 6.16: Neither preferential nor utility independence are reflexive, i.e., Y p.i. Z does not imply that Z p.i. Y . Similarly, Y u.i. Z does not imply that Z u.i. Y .

Comment 6.17: Utility independence is in fact the stronger property, i.e., $Y \text{ u.i. } Z$ implies $Y \text{ p.i. } Z$. This can be shown by considering degenerate lotteries. The converse is not true, i.e., $Y \text{ p.i. } Z$ does not imply $Y \text{ u.i. } Z$.

Suppose $Y \text{ u.i. } Z$, what does this imply? Theorem 6.5 stated that $c(x)$ is unique up to a positive linear transformation thus if $c(x)$ maintains preference ordering among lotteries, $\ell(\tilde{x})$, so does $a + b \cdot c(x)$ where a, b are real numbers and $b > 0$. Now consider lotteries of the form $(\{\tilde{y}, z_0\})$ and $(\{\tilde{y}, z_1\})$, z_0, z_1 fixed. If $Y \text{ u.i. } Z$, preference orderings among the first lotteries directly correspond to preference orderings among the second lotteries, thus¹ $c(y, z_1) = a + bc(y, z_0)$ where a and b depend on z_1 . Extending this to all z implies $c(y, z) = a(z) + b(z)c(y, z_0)$, where $a(z)$ and $b(z)$ are real valued functions of z with $b(z) > 0$ for all z . This is a very simple result, but when other independence assumptions are added, useful functional forms can be identified. Consider the following theorem due to Keeney [70] reformulated by Kaufman [66], stated here for two performance measures. The proof will be stated for compaction functions as an example of how to extend utility theory to compaction theory.

Theorem 6.7: Suppose that $\{X_1, X_2\}$ are a complete set of performance measures. Suppose $X_1 \text{ u.i. } X_2$ and $X_2 \text{ u.i. } X_1$, then $c(x_1, x_2)$ has the following form, labeled quasi-additive:²

$$c(x_1, x_2) = a_1 u_1(x_1) + a_2 u_2(x_2) + (1-a_1-a_2) u_1(x_1) u_2(x_2)$$

where some (x_1^*, x_2^*) and (x_1^0, x_2^0) are chosen such that

$$a_1 u_1(x_1) = c(x_1, x_2^0), \quad a_2 u_2(x_2) = c(x_1^0, x_2) \quad \text{and} \quad u_1(x_1^*) = u_2(x_2^*) = 1,$$

$$u_1(x_1^0) = u_2(x_2^0) = 0.$$

Comment 6.18: Note that $c(x_1, x_2)$ is scaled such that $c(x_1^*, x_2^*) = 1$ and $c(x_1^0, x_2^0) = 0$.

Proof: x_1 u.i. x_2 implies $c(x_1, x_2) = a(x_2) + b(x_2)c(x_1, x_2^0)$ and x_2 u.i. x_1 implies $c(x_1, x_2) = d(x_1) + e(x_1)c(x_1^0, x_2)$. Since $c(x_1, x_2)$ is unique to a positive linear transformation we can scale $c(x_1^0, x_2^0) = 0$, $c(x_1^*, x_2^*) = 1$. Thus $c(x_1, x_2^0) = d(x_1) + e(x_1) \cancel{c(x_1^0, x_2^0)} = d(x_1)$ and $c(x_1^0, x_2) = a(x_2) + b(x_2) \cancel{c(x_1^0, x_2^0)} = a(x_2)$. Choose constants a_1, a_2 and define $u(x_1)$ such that $c(x_1, x_2^0) = a_1 u_1(x_1)$ and $c(x_1^0, x_2) = a_2 u_2(x_2)$. Substituting yields $c(x_1, x_2^*) = a_1 u_1(x_1) + e(x_1) a_2 u_2(x_2^*)$.

Solving for $e(x_1)$ and substituting in previous equations gives

$$c(x_1, x_2) = a_1 u_1(x_1) + u_2(x_2) [c(x_1, x_2^*) - a_1 u_1(x_1)]. \quad \text{Doing similar substitutions gives } c(x_1, x_2) = a_2 u_2(x_2) + u_1(x_1) [c(x_1^*, x_2) - a_2 u_2(x_2)].$$

Using the latter equation yields $c(x_1, x_2^*) = a_2 + u_1(x_1)(1-a_2)$. Substituting this in the former equation yields the desired result i.e.,

$$c(x_1, x_2) = a_1 u_1(x_1) + u_2(x_2) [a_2 + u_1(x_1)(1-a_2) - a_1 u_1(x_1)].$$

Note that except for algebra and fortuitous choice of normalization (possible by Corollary 6.3) the only theorem used was the uniqueness theorem. (Theorem 6.5). In other words it was not necessary to invoke the choice axioms (axioms 6.4, 6.5) or the fact that stochastic preference instead of ordinary preference was used. This is indicative of the other, more subtle and complex proofs of functional forms in utility theory. The next section presents without proof a sample of the theorems available in utility theory that will now be used in compaction.

6.2.7 Utility Theorems Identify Unique Functional Forms for Compaction Functions³

The utility theorems stated below are extremely useful because they (1) identify unique function forms, (2) indicate how to directly assess the compaction (utility) functions, and (3) require independence assumptions that are verifyable. However, before stating the theorems, a new function, a conditional compaction function, will be defined. This function, and its notation (used informally in theorem 6.7), allows for more clairvoyant presentation of the results.

Suppose that x_j u.i. $x_{\bar{j}}$. This means that lotteries involving $\{x_j, x_{\bar{j}}\}$ do not depend on the fixed value of $x_{\bar{j}}$, thus we would expect that there is some function of x_j which incorporates all the lottery characteristics of x_j and which is independent of $x_{\bar{j}}$. This function does exist, is unique, and is called a conditional compaction function. (Conditional because $x_{\bar{j}}$ is fixed). As motivation remember that x_j u.i. $x_{\bar{j}}$ implies $c(x_j, x_{\bar{j}}) = a(x_{\bar{j}}) + b(x_{\bar{j}})c(x_j, x_{\bar{j}}^0)$ which can be written as

$a(x_j) + [b(x_j) \cdot c(x_j^*, x_j^0)] [c(x_j, x_j^0)/c(x_j^*, x_j^0)]$ where $\{x_j^*, x_j^0\} >_x \{x_j^0, x_j^0\}$ and $c(x_j, x_j)$ is scaled such that $c(x_j^0, x_j^0) = 0.0$ and $c(x_j^*, x_j^*) = 1.0$. Note that $[c(x_j, x_j^0)/c(x_j^*, x_j^0)]$ is only a function of x_j and is scaled from 0.0 to 1.0. It will later be shown to be independent of x_j^0 . Formally:

Definition 6.12: Let $c(x_j, x_j)$ be a compaction function scaled from

$c(x_j^0, x_j^0) = 0.0$ to $c(x_j^*, x_j^*) = 1.0$. Let X_j u.i. X_j . Then

$u_j : X_j \rightarrow [0,1]$ is a conditional compaction function, where

$$u_j(x_j) = c(x_j, x_j^0)/c(x_j^*, x_j^0)$$

Above $u_j(x_j)$ appears to depend on x_j^0 . The next theorem shows that it does not.

Theorem 6.8: Let X be a complete set of performance measures. Then if X_j u.i. X_j , the conditional compaction of $x_j, u_j(x_j)$, is independent of the choice of x_j^0 and x_j^* .

Proof: Choose $x_j^0 = y$, $x_j^* = z$ which yields $c(x_j, x_j)$ and $u_j(x_j)$.

Now choose different values for x_j^0, x_j^* , i.e., $x_j^0 = p$, $x_j^* = q$ which

yields $d(x_j, x_j)$ and $v_j(x_j)$. By theorem 6.5 $d(x_j, x_j) = a + bc(x_j, x_j)$,

$b > 0$. $v_j(x_j) = d(x_j, p)/d(x_j^*, p) = [a + bc(x_j, p)]/[a + bc(x_j^*, p)]$ but

$$c(x_j, p) = e(p) + f(p)c(x_j, y) \text{ thus } v_j(x_j) = [a + be(p) + bf(p)c(x_j, y)]/$$

$[a + be(p) + bf(p)c(x_j^*, y)]$. Now if $a + be(p) = 0$ and $bf(p) \neq 0$, the

$$\text{result follows since the proper terms cancel giving } v_i(x_j) = c(x_j, y)/$$

$$c(x_j^*, y) = u_j(x_j). \text{ But by normalization } d(x_j^0, p) = 0 = a + bc(x_j^0, p)$$

$$= a + be(p) + bf(p)c(x_j^0, y)^0. \text{ Finally } b > 0 \text{ and if } f(p) = 0 \text{ then } d(x_j, x_j^0)$$

is independent of x_j which contradicts X a complete set of performance measures.

The strength of this theorem is that if X_j u.i. $X_{\bar{j}}$, an arbitrary $x_{\bar{j}}'$ can be chosen and $u_j(x_j)$ assessed over consequences of the form $\{x_j, x_{\bar{j}}'\}$. Thus if X_j u.i. $X_{\bar{j}}$, the conditional compaction can be assessed over simple lotteries, see figure 6.8, involving only x_j , as long as the consumer is told $x_{\bar{j}}$, i.e., everything else, is fixed.

We will next consider function forms which are appropriate for $u_j(x_j)$. Since by theorem 6.8 $u_j(x_j)$ can be evaluated with simple lotteries, the uni-attribute results of Raiffa [120] can be directly applied. First consider the concept of risk aversion. A consumer is risk averse if he would prefer a guaranteed outcome to an uncertain outcome with the same expected value of the (single) performance measure. In other words he must be paid a premium in order to take a risk. Formally if

$$(p \cdot x_j' + (1-p)x_j'') >_{x_j} L(x_j', x_j'', p) \text{ then a consumer is risk averse.}$$

This condition implies that $c[p \cdot x_j' + (1-p)x_j'', x_j^0] >_R p \cdot c(x_j', x_j^0) + (1-p)c(x_j'', x_j^0)$, thus by the definition of concavity $c(x_j, x_j^0)$ and hence

$u_j(x_j)$ is concave in x_j .

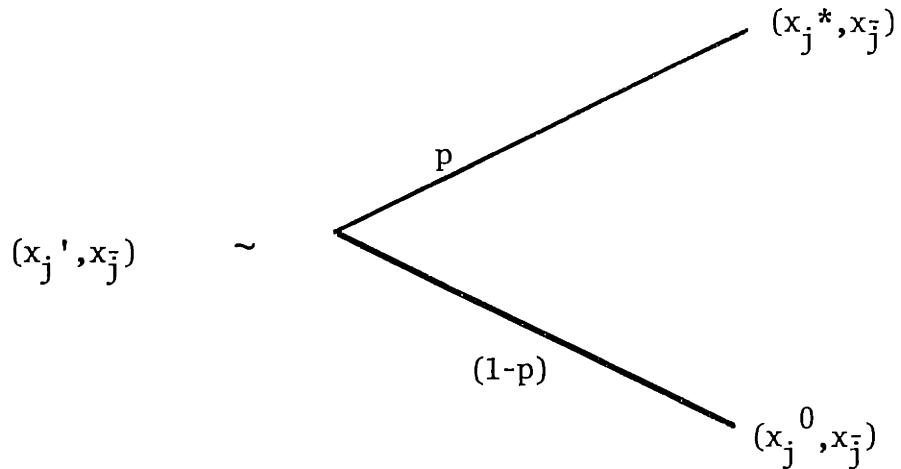


Figure 6.8: Conditional Utility Lottery

Suppose that a consumer's risk aversion does not depend upon his asset level. In other words suppose that if $x_j \sim_{\tilde{X}_j} L(x_j', x_j'', p)$ then giving him Δ units of x_j does not change his feeling toward the lottery, i.e., $x_j + \Delta \sim_{\tilde{X}_j} L(x_j' + \Delta, x_j'' + \Delta, p)$ (See figure 6.9). This condition is known as constant risk aversion and implies a unique function form:

Theorem 6.9: (Raiffa) Let X be a complete set of performance measures, let X_j u.i. $X_{\bar{j}}$, let $u_j(x_j)$ be the conditional compaction of X_j . If

$$(x_j, x_{\bar{j}}) \sim_X L(\{x_j', x_{\bar{j}}\}, \{x_j'', x_{\bar{j}}\}, p)$$

implies that

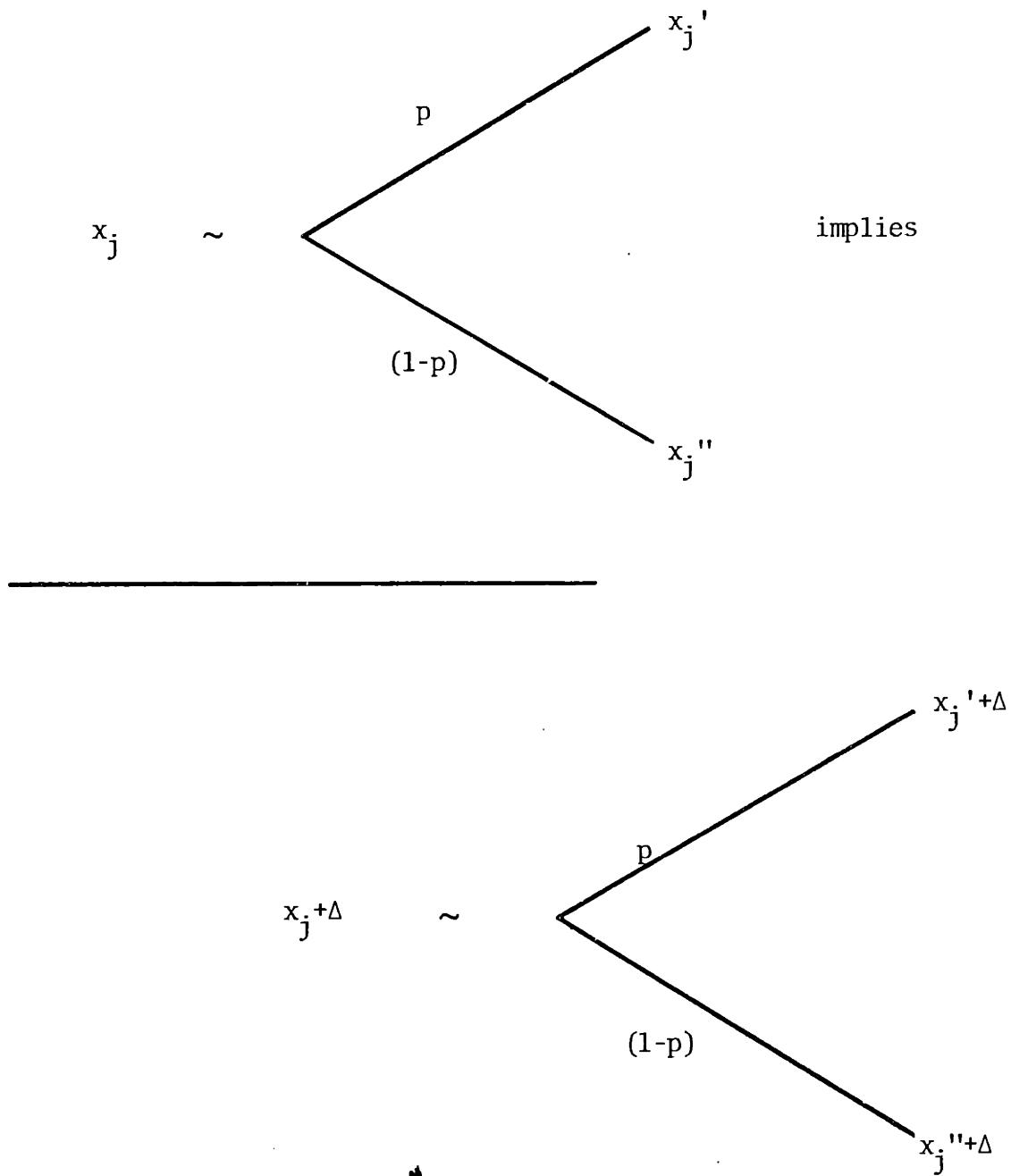


Figure 6.9: Constant Risk Aversion Lottery

$$(x_j + \Delta, x_j^-) \underset{X}{\sim} L(\{x_j' + \Delta, x_j^-\}, \{x_j'' + \Delta, x_j^-\}; p)$$

for all Δ, x_j, x_j', x_j'' , then

$$u_j(x_j) = [1 - e^{-r(x_j - x_j^0)}] / [1 - e^{-r(x_j^* - x_j^0)}]$$

for some real number, r , or

$$u_j(x_j) = [x_j - x_j^0] / [x_j^* - x_j^0]$$

with the latter case only occurring when

$$(p \cdot x_j' + (1-p)x_j'', x_j^-) \underset{X}{\sim} L(\{x_j', x_j^-\}, \{x_j'', x_j^-\}, p).$$

Operationally, the "constantly risk averse" functional form of theorem 6.9 has only one parameter, r . Thus, if a consumer is constantly risk averse, a simple lottery of the form $x_j \underset{X_j}{\sim} L(x_j^0, x_j^*, p)$ where the consumer sets p gives a type II equation which can be solved for r . These simple lotteries were used as part of a procedure used to assess the compaction functions of a random sample of 80 M.I.T. students. See section 6.4 for the results of this experiment.

There are other functional forms besides constantly risk averse, for example $u(x_j) = \log[a \cdot (x_j - x_j^0) + 1] / \log[a \cdot (x_j^* - x_j^0 + 1)]$ is decreasing risk averse. For a more complete discussion of these results see Raiffa [120].

The next section presents theorems which indicate how the conditional compaction functions, $u_j(x_j)$, combine to form a complete compaction function, $c(x)$.

6.2.7.1 Quasi-additive and Multiplicative

Now if x_j u.i. $x_{\bar{j}}$ for all j , all of the conditionals can be easily assessed with simple lotteries. The next theorem due to Keeney [68] shows how one can then build a complete compaction function from conditional compaction functions.

Theorem 6.10 (Keeney) If X is a complete set of performance measures and if x_j u.i. $x_{\bar{j}}$ for all j .

Then:⁴

$$\begin{aligned}
 c(\underline{x}) = & \sum_j k_j u_j(x_j) + \sum_j \sum_{\ell > j} k_{j\ell} u_j(x_j) u_\ell(x_\ell) \\
 & + \sum_j \sum_{\ell > j} \sum_{m > \ell} k_{j\ell m} u_j(x_j) u_\ell(x_\ell) u_m(x_m) \\
 & + \dots + k_{12\dots J} u_1(x_1) u_2(x_2) \dots u_J(x_J)
 \end{aligned}$$

This theorem makes assessment of compaction functions feasible since all one need do is determine the condition compaction functions, $u_j(x_j)$, by simple lotteries and then determine the scaling constants. This can be done by solving simultaneous equations resulting from type I indifference questions, i.e., $\underline{x}' \sim \underline{x}''$, or by asking "corner point" questions,⁵ i.e., $\{y^*, z^0\} \sim L(\underline{x}^*, \underline{x}^0, p)$ where $\{Y, Z\}$ is a partition of X . See figure 6.10 for $X = \{x_1, x_2, x_3\}$ and $Y = x_1$, figure 6.10a gives the lottery, figure 6.10b gives a geometric interpretation of the words corner point. Asking these corner point questions allows simple direct determination of the scaling constants. For example, the lottery in figure 6.10a gives $c(x_1^*, x_2^0, x_3^0) = p$ which when substituted in the

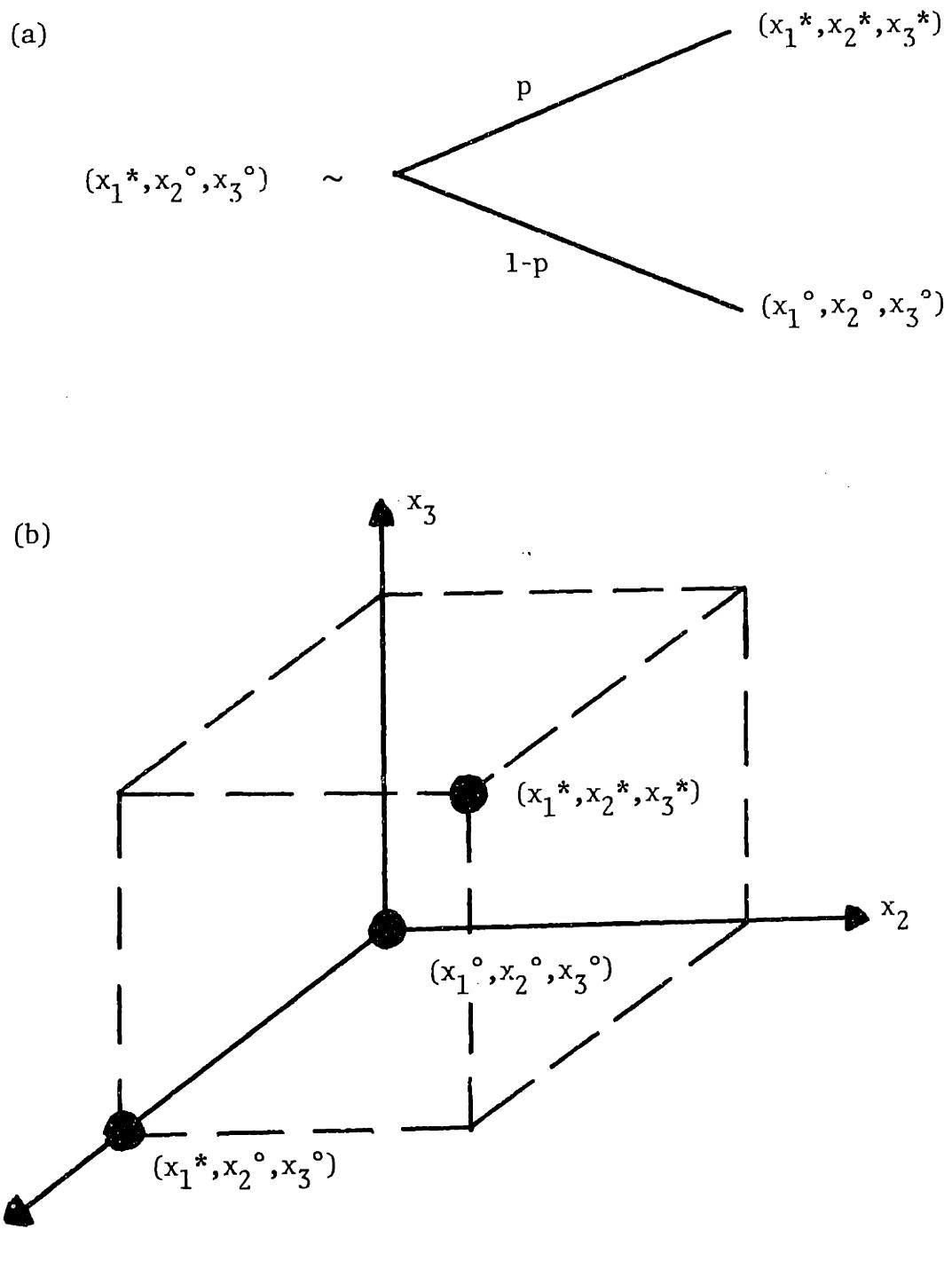


Figure 6.10: Corner Point Question

equation from theorem 6.10 gives directly $k_1 = p$.

Theorem 6.10 makes it theoretically possible directly assess compaction functions, but the task is still quite difficult since 2^{J-1} scaling constants are required. The next theorem due to Keeney [71] reduces this number to $J + 1$.

Theorem 6.1 (Keeney) If X is a complete set of performance measures and if for some j x_j u.i. $x_{\bar{j}}$ and $\{x_j, x_{\ell}\}$ p.i. $x_{\bar{j\ell}}$ for all ℓ . Then:

$$\begin{aligned} c(\underline{x}) &= \sum_j k_j u_j(x_j) + \sum_j \sum_{\ell > j} K k_j k_{\ell} u_j(x_j) u_{\ell}(x_{\ell}) \\ &+ \sum_j \sum_{\ell > j} \sum_{m > \ell} K^2 k_j k_{\ell} k_m u_j(x_j) u_{\ell}(x_{\ell}) u_m(x_m) \\ &+ \dots + K^{J-1} k_1 k_2 \dots k_J u_1(x_1) u_2(x_2) \dots u_J(x_J) \end{aligned}$$

where $k_j \in [0, 1]$ for all j and $K > -1$.

If $\sum_j k_j \neq 1$ this can be factored to:

$$1 + Kc(\underline{x}) = \prod_{j=1}^J [1 + K k_j u_j(x_j)]$$

And if $\sum_j k_j = 1$ then

$$c(\underline{x}) = \sum_{j=1}^J k_j u_j(x_j)$$

Comment 6.19: The above form is termed multiplicative due to the particularly simple representation in product form.

Comment 6.20: It can easily be seen by comparing theorems 6.10 and 6.11 that the multiplicative form is a special case of the quasi-additive form.

Since only $J+1$ scaling constants are required for the multiplicative form, it is quite feasible to obtain them. Again they can be determined by asking J type I indifference questions and solving the simultaneous equations for the relative k_j 's. The $J+1^{st}$ equation is given

by $1 + K = \prod_{j=1}^J (1 + Kk_j)$ because of normalization.

It turns out that there is a particularly simple way to assess the scaling constants for a multiplicative form:

(1) Assess the conditional compaction functions with simple lotteries:

$x_j \sim L(x_j^*, x_j^0, p)$ with $x_{\bar{j}}$ fixed.

(2) Ask type I questions of the following form for all $j > 1$:
The consumer selects x_1^j such that:⁶

$\{x_1^*, x_j^0, x_{\bar{j}}\} \sim \{x_1^j, x_j^*, x_{\bar{j}}\}$

(3) Ask one corner point question: The consumer selects p such that:

$\{x_1^*, x_{\bar{1}}^0\} \sim L(x^*, x^0, p)$

The compaction function is then determined as follows:

(1) Simple lotteries gives $u_j(x_j)$ for all j .

(2) "Tradeoff questions" give $k_j = u_1(x_1^j)k_1$ for all j

(3) Corner point question gives $k_1 = p$.

(4) Using (2) above gives k_j for all j and finally K is

determined by solving⁷

$$(1 + K) = \prod_{j=1}^J (1 + k_j K)$$

Note that the conditions necessary for Theorem 6.11 can be easily verified in the above procedure by asking the respondent whether his answers to the simple lotteries depend on x_j (i.e., verify x_j u.i. $x_{\bar{j}}$) and whether his answers to the tradeoff questions depend on $x_{\bar{1j}}$ (i.e., verify $\{x_1, x_j\}$ p.i. $x_{\bar{1j}}$).

The above assessment and verification procedure was used to assess the compaction functions of 80 M.I.T. students selected at random. The results of this experiment are discussed in section 6.4. The measurement issues are discussed in chapter 3.

6.2.7.2 Additive Representations

Most of the compaction functions in the literature (see section 6.3) have been additive. What conditions are necessary and sufficient for additive representations? From Theorem 6.10 we see that an additive representation is in fact a special case of the multiplicative form, i.e., when the interaction constant, K , equals 0.0. Furthermore this can be

easily verified by the assessment procedure for multiplicative forms because a multiplicative form turns additive when the scaling constants add to 1.0, i.e., $\sum_j k_j = 1.0$.

Fishburn [35] identifies a "marginality assumption" which states clearly the conditions for additivity. It basically says that the consumer does not care how the various performance measures match up, i.e., he would not pay or be paid a premium to ensure that extremely poor values of all performance measures do not simultaneously occur. This is stated formally in the following theorem. (See figure 6.11 for a graphical interpretation of the lotteries).

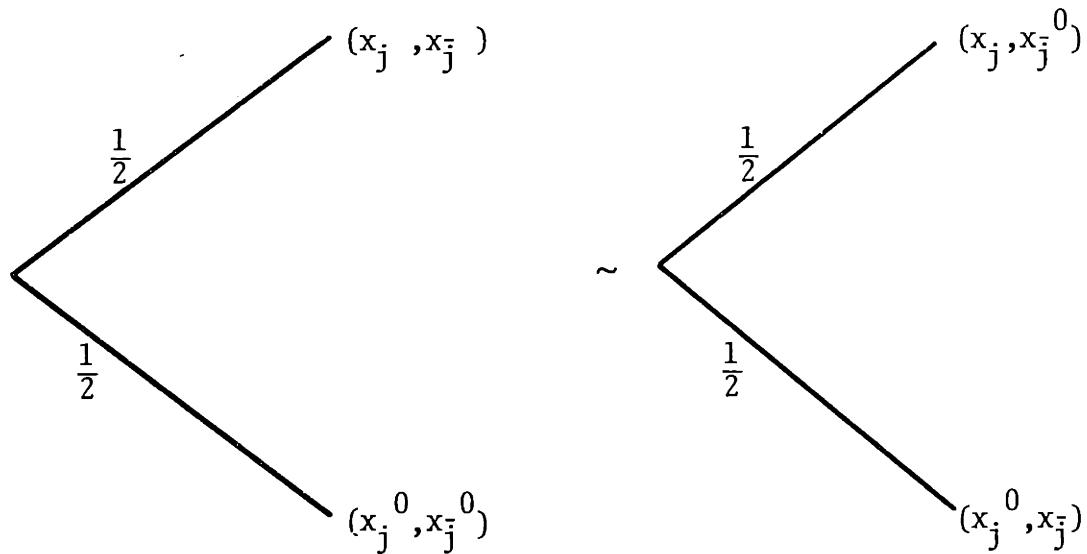


Figure 6.11: Marginality Assumption

Theorem 6.12: (Fishburn) Let X be a complete set of performance measures. If for all j and for all $\{x_j, x_{\bar{j}}\}$ in X it is true that

$$L(\{x_j, x_{\bar{j}}\}, \{x_j^0, x_{\bar{j}}^0\}; \frac{1}{2}) \geq^* L(\{x_j, x_{\bar{j}}^0\}, \{x_j^0, x_{\bar{j}}\}; \frac{1}{2})$$

then

$$c(x) = \sum_j k_j u_j(x_j) \quad \text{where} \quad \sum_j k_j = 1.0, \quad k_j \geq 0.$$

Comment 6.21: Note that the lottery condition implies x_j u.i. $x_{\bar{j}}$ for all j , thus Keeney's quasi-additive theorem applies. The lottery condition then insures that all interaction terms drop out.

Fishburn's theorem is for cardinal compaction functions (since lotteries are involved), but many compaction functions, only claim to be ordinal. Due to theorem 6.2 any monotone transformation retains the properties of compaction and of the special properties defined in section 6.13, thus if some monotone function of an additive representation is a cardinal compaction function, then applying the inverse of that function will result in an additive ordinal representation with all desirable properties except expected value. Ting [140] identifies the conditions under which such a representation exists.

Theorem 6.13 (Ting) Let X be a complete set of performance measures. If $\{x_j, x_{\bar{j}}\}$ p.i. $x_{\bar{j}\ell}$ for all j, ℓ then there exist continuously differentiable functions V, g_1, g_2, \dots, g_J such that

$$c(x) = V[\sum_j g_j(x_j)]$$

Comment 6.22: V is not necessarily monotone in theorem 6.13, because Ting does not require preferences to be monotone in each performance measures. This chapter does and the desired monotone result follows simply.

Comment 6.23: The assumptions for the multiplicative form can be shown (Keeney [71]) to satisfy the assumptions of Theorem 6.12, and as expected, the multiplicative form can be made to be additive by taking logarithms,

$$\text{i.e., } c(x) = K^{-1} \exp[\sum \log(1 + K k_\ell u_\ell(x_\ell))] - 1.$$

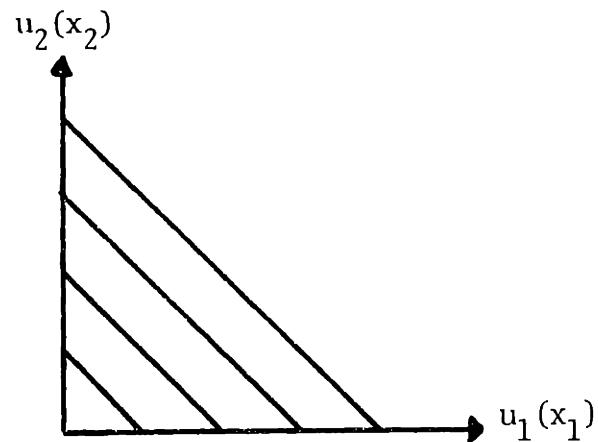
Notice that theorem 6.13 is an existence theorem, but unlike theorems 6.10, 6.11, and 6.12 it does not indicate either the functional forms of the compaction function or how to directly assess them.

6.2.7.3 Between Multiplicative and Quasi-additive: Aggregatability

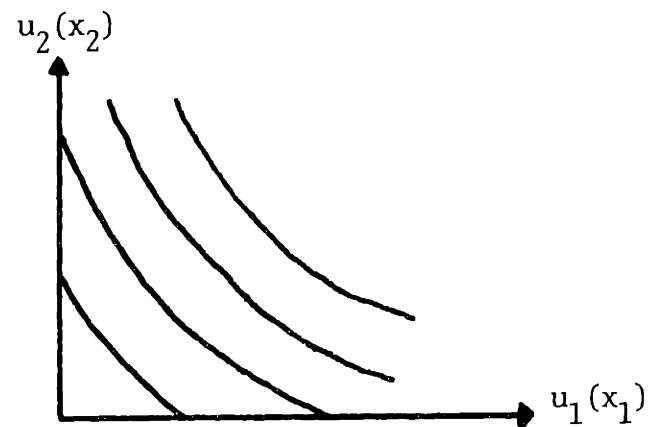
The multiplicative form is very useful because it is easy to assess the scaling constants and since there are only $J+1$ of them, the measurement task is not formidable. Furthermore the multiplicative form is much more general than the additive form since it allows non-linear indifference curves between the conditional utilities of attributes. (See figure 6.12).

Unfortunately the assumptions for the multiplicative form are not always verified. For example suppose the performance measures for dial-a-ride trips are {wait time, travel time, cost}. Now it is reasonable to assume that tradeoffs between wait time and travel time do not depend on cost, i.e., $\{w, t\}$ p.i. $\{c\}$, but tradeoffs between travel time

Additive



Multiplicative



(or)

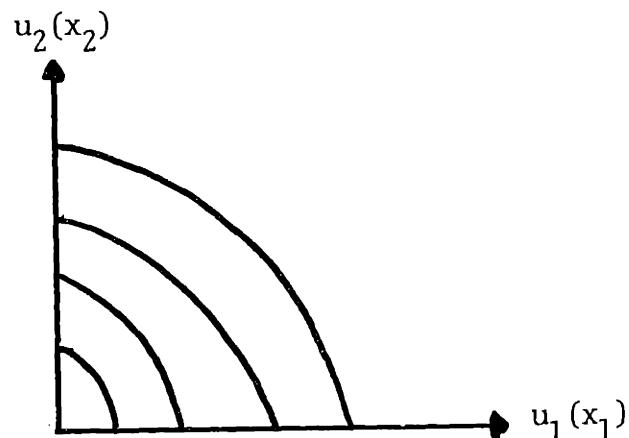


Figure 6.12: Indifference Curves

and cost may depend on wait time, and similarly tradeoffs between wait time and cost may depend on travel time, i.e., $\{t, c\}$ not p.i. $\{w\}$ and $\{w, c\}$ not p.i. $\{t\}$. Suppose now that X_j u.i. $X_{\bar{j}}$ for all j and hence the weaker form, quasi-additive applies. In this case, the quasi-additive form requires $7(2^3 - 1)$ scaling constants, but if J were large then $2^J - 1$ might be for too many scaling constants to assess. (E.g. $2^6 - 1 = 63$ or $2^{12} - 1 = 8,191$). Is there any way that preferential conditions, such as $\{w, t\}$ p.i. $\{c\}$, can be used to simplify the task?

Ting [140] supplies the answer in the following theorem:

Theorem 6.14 (Ting) Let X be a complete set of performance measures. Let $\{Y, Z\}$ be a partition of X . Then if Y p.i. Z , there exists a real valued aggregator function, $g: Y \rightarrow R$, and a real valued function V such that:

$$c(\underline{x}) = V[g(\underline{y}), \underline{z}]$$

What this theorem says is that, the performance measures, \underline{y} , can be aggregated into a single number, $g(\underline{y})$; this number is then traded off with the performance measures, \underline{z} . Unfortunately theorem 6.14 does not indicate how to assess V and g , but if it is known that $c(\underline{x})$ is quasi-additive it can be made operational as follows, shown here for three performance measures, $\{w, t, \$\}$:

(1) $\{w, t\}$ p.i. $\{\$\}$ thus $c(w, t, \$) = V[g(w, t), \$]$

(2) let $g(w, t)$ be quasi-additive, i.e.,

$$g(w, t) = k_w u_w(w) + k_t u_t(t) + k_{wt} u_w(w)u_t(t)$$

(3) let $V[g, \$]$ be quasi-additive, i.e.,

$$\begin{aligned}c(\underline{x}) &= V[g, \$] = k_g g(w, t) + k_s u_s(\$) + k_{gs} g(w, t) u_s(\$) \\&= k_g k_w u_w(w) + k_g k_t u_t(t) + k_s u_s(\$) + k_{wt} u_w(w) u_t(t) \\&\quad + k_w k_{gs} u_w(w) u_s(\$) + k_t k_{gs} u_t u_s(\$) + k_{wt} k_{gs} u_w(w) u_t(t) u_s(\$)\end{aligned}$$

(4) $c(\underline{x})$ is unique under consistent 0 - 1 scaling and since it is known to be quasi-additive, the above form is the only form.

Note that by using the above procedure, only six scaling constants are required whereas the general quasi-additive form requires seven.

Now this savings may not seem worth the effort, but in the case of large J , the resulting savings could be significant. For example suppose $J = 6$ and $\{x_1, x_2, x_3\}$, p.i. $\{x_4, x_5, x_6\}$, then the above procedure results in

17, $(2^3 - 1 + 2^3 - 1 + 3)$, scaling constant rather than 63, $(2^6 - 1)$.

6.2.7.4 Other Theorems

There are other theorems in utility theory which indicate functional forms. For example Pollak [115] and Meyer [105] also derive conditions sufficient for multiplicative form, but their conditions are not as easily verifyable as Keeney's. In addition Fishburn and Keeney [39] have studied the implications of a form of generalized utility independence, and Kirkwood [79] has studied parametric dependence.

6.2.8 Summary of Axiomization

This section began with a set of behavioral axioms which were shown to imply the existence and uniqueness of a compaction function. Taken together these axioms indicated techniques to directly measure compaction functions, or at least to determine equations which could be solved for the parameters of compaction functions. To make these measurement techniques operational it was necessary to determine the appropriate parameterized functional forms. Behavioral assumptions such as constant risk aversion, preferential independence, and utility independence were introduced and the utility theoretic results of Raiffa, Keeney, Fishburn, and Ting were applied to compaction functions to indicate these forms.

The main result of this section is a simple and feasible procedure to represent and directly assess compaction functions for individuals.

The next section examines examples of compaction functions (as per section 6.1) as they exist in the literature, but before going on, let us return to the notation of section 6.1. In section 6.1, compaction functions were across individuals with individual specific parameters, i.e., $c_{ij}(x_{ij}, \lambda_i)$ where λ_i were individual i 's preference measures. In section 6.2, only one individual was dealt with at a time and the compaction functions were shown to be complete and symmetric, hence the notation was $c(x_j)$. We return now to $c_{ij}(x_{ij}, \lambda_i)$. For example if:

$$c_{ij}(x_{ij}, \lambda_i) = k_{i1}u_{i1}(x_{i1}) + k_{i2}u_{i2}(x_{i2}) + k_i k_{i1}k_{i2}u_{i1}(x_{i1})u_{i2}(x_{i2}) \text{ where}$$

$$u_{il} = [1 - e^{-r_{il}(x_{il} - x_l^0)}] / [1 - e^{-r_{il}(x_l^* - x_l^0)}] \quad l = 1, 2$$

then the preference measures, λ_i , are the scaling constants, the risk aversion coefficients, and the interaction constant, i.e.,

$$\lambda_i = \{k_{il}, k_{i2}, r_{il}, r_{i2}, K_i\}.$$

6.3 Examples of Statistical Compaction Techniques

Section 6.1 introduced the concept of a compaction function, $c_{ij}(x_{ij}, \lambda_i)$, mapping the performance measures, x_{ij} , and the preference measures, λ_i , into a single real number, c_{ij} ; with the property that any two sets of performance measures with the same measure of goodness will have the same probability of choice. Section 6.2 then derived one theoretically sound technique, direct assessment, to determine compaction functions and to assess the preference parameters, λ_i .

In some cases, it will not be economically feasible to collect the necessary data for direct assessment and in some cases the analyst may have other reasons for performing statistical techniques. In such cases, the user of the methodology may wish to substitute another compaction technique. This section discusses some of the available statistical compaction techniques which do not require direct assessment. They are

- (1) expectancy value models, e.g., "weights times rates,"
- (2) discriminant analysis,
- (3) random utility models
- (4) conjoint analysis, especially as used with factorial design,
- (5) maximum score models,
- (6) preference regression.

(7) PREFMAP.

6.3.1 Expectancy Value Models

In this technique, consumers are asked to rate each alternative on a set of scales and directly state the importance weights for each scale. The compaction is then a linear additive combination of the importance weights times the perception rates. This technique is based on the mathematical psychological theory of Fishbein [33] which states that if:

A_j = attitude toward alternative a_j

b_{jk} = strength of belief k about a_j

x_{jk} = evaluative aspect of belief k about a_j

then

$$A_j = \sum_k b_{jk} x_{jk}$$

This formulation does not include the risk characteristics of the aspect scale nor the interactions among various scales, but perhaps its biggest weakness is that it assumes that the consumer is acutely aware of and can consistently process the relative scale ranges to produce a mathematically consistent importance weight. In practice, attitudes, A_j , have not correlated well to directly measured preferences such as those obtained from constant sum paired comparison preference data.⁸

6.3.2 Discriminant Analysis

In discriminant analysis, the hypothesis is that there are two (or more) distinct populations and the goal is to identify to which population an individual observation belongs. An extension is to determine the probability that an observation belongs to a given population. In our context, the populations are those choosing a particular alternative.

The first discriminant technique is to use a discriminant value, d_{ij} , which is a linear combination of the observation. I.e.,

$$d_i = \sum_{\ell} k_{\ell} y_{i\ell}$$

where d_i = discriminant value for individual i

$y_{i\ell}$ = value of ℓ^{TH} variable for individual i

k_{ℓ} = weight of the ℓ^{TH} variable

The weights, calibrated on training populations, are determined by maximizing the ratio of the squared distance between the means of d_i relative to the variance of d_i within the population. Based on the assumptions of two populations and multivariate normal distributions of x_i with common covariance matrix, $M_{\ell m}$, Kendall [78] shows that:

$$k_{\ell} = \sum_m \text{cov}(y_m, y_{\ell}) \left(\bar{y}_{\ell}^1 - \bar{y}_{\ell}^2 \right)$$

where \bar{y}_{ℓ}^j is the estimated mean of $y_{i\ell}$ in population j and $\text{cov}(y_m, y_{\ell})$

is the estimated covariance. The observation, d_i , is then classified

to the population with the closest mean value of d_i . Geometrically this is equivalent to establishing a separating hyperplane between the two populations.

The second technique is again to postulate a linear discriminant function, but instead of simple classification, membership probabilities are assigned. Under various assumptions, functional forms for the probability law can be determined and the variable weights estimated with maximum likelihood techniques. For example, Quarmby [118] shows that if d_i is distributed as normal with common covariance in each population, then a logistic model results for binary choice:

$$p(\text{belong to population } j | d_i) = \frac{e^{-a_j + b_j d_i}}{\sum_{\ell} e^{-a_{\ell} + b_{\ell} d_i}}$$

where $a_{\ell}, b_{\ell} = \text{constants}, \ell = 1, 2$

Notice the difference between discriminant analysis and compaction theory. In compaction theory, a scalar measure of goodness, c_{ij} , is determined for each individual and for each alternative. In discriminant analysis, a single number, d_i , is determined for each individual and discrimination is based on that one value. The two approaches are comparable only if there are two alternatives.

Discriminant analysis is presented here because it was used by Quarmby [118] to compute mode choice probabilities in transportation.

In his problem there were two populations, auto users and transit users. He used a linear discriminant function with y_ℓ equal to relative values of performance measures such as travel time and determined the importance weights by the first technique (maximum squared distance/variance). The probabilities were then determined by the second technique (logistic formulation).

For more than two alternatives discriminant analysis is difficult to apply because it is based on a single value rather than a vector of values. The next technique, random utility models is superior for multiple alternatives.

6.3.3 Random Utility Models

Random utility models assume that there exists some utility function, $u(\underline{x}_{ij}, \underline{s}_i)$, where \underline{x}_{ij} are the vector of values of the performance measures, and \underline{s}_i are individual characteristics, such that an individual always maximizes his utility when choosing an alternatives. Furthermore, the utility has an observable component and a random component, i.e.:

$$u(\underline{x}_{ij}, \underline{s}_i) = v(\underline{x}_{ij}, \underline{s}_i) + \epsilon(\underline{x}_{ij}, \underline{s}_i)$$

This model is quite general (McFadden [94]), but in realistic cases certain simplifying assumptions must be made in order to use the model.

Perhaps the most popular assumptions are:

- (1) The observable component, $v(\underline{x}_{ij}, \underline{s}_i)$, is linear in some parameters, a , which are the same for all individuals in a segment, i.e.:

$$v(\underline{x}_{ij}, \underline{s}_i) = \sum_{\ell} a_{\ell} z_{\ell}(\underline{x}_{ij}, \underline{s}_i)$$

(2) The error terms, $\epsilon(\underline{x}_{ij}, \underline{s}_i)$, are independent of alternatives, a_j 's, and are identically distributed for all alternatives.

(3) The error terms have Weibull distributions, i.e.:

$$p_{\epsilon}(\epsilon_0) = e^{-\epsilon_0} \exp(-e^{-\epsilon_0})$$

McFadden [93] shows that these assumptions imply the multinomial logit formulation, i.e.:

$$p(a_j | c_{i1}, c_{i2}, \dots, c_{iJ}) = e^{-c_{ij}} / \sum_{m=1}^J e^{-c_{im}}$$

where

$$c_{ij} = \sum_{\ell} a_{\ell} z_{\ell}(\underline{x}_{ij}, \underline{s}_i)$$

In practice the a_{ℓ} 's are estimated by maximum likelihood techniques, see for example Ben-Akiva [10].

The advantages of this approach are that the calibration procedure uses "revealed preference", i.e., actual choice, and the choice probabilities are determined axiomatically from the compaction values. One disadvantage is that there is no axiomatic specification of the functions, $z_{\ell}(\dots)$. In fact, in practice $z_{\ell}(\underline{x}_{ij}, \underline{s}_i)$ is usually $x_{ij\ell}$. Because only existing alternatives are used for calibration there may not be enough variation in the performance measures to extend the pre-

dictions to radical new alternatives. These disadvantages can be partially overcome by using the compaction function, $c(\underline{x}_{ij}, \lambda_i)$ discussed in section 6.2. It is axiomatic and can be extended to radically different alternatives because the preference values, λ_i , can be determined from proxy choices with ranges of the performance measures comparable with the new alternatives.

If so desired, revealed choice can be coupled with proxy choice by first determining $c(\underline{x}_{ij}, \lambda_i)$ and then using a more general parameterized error term, $\epsilon(\underline{x}_{ij}, \underline{s}_i)$. The relative merits of the two types of choices, and techniques to use both in a calibration process are discussed in chapter 7.

6.3.4 Conjoint Analysis

Conjoint analysis is the general problem of determining parameters of a function given only its independent and dependent values. In other words, given the hypothesis that $y = f(\underline{x})$ and given direct measurements of y and \underline{x} , the problem is to deduce $f(\cdot)$. A case of particular interest is the polynominal measurement model where $f(\underline{x}) = M[f_1(x_1), \dots, f_L(x_L)]$ with M restricted to sums, differences, and products of the $f_\ell(\cdot)$ such that $f(\underline{x}') > f(\underline{x}'')$ means \underline{x}' is preferred to \underline{x}'' . Notice that the quasi-additive, multiplicative, and additive compaction forms are of this form. Fishburn[35] gives conditions identical to those in prescriptive utility for $M[\cdot]$ to be additive. Tversky [44] gives necessary and sufficient conditions for $M[\cdot]$ to be measureable. To state the basic result define a data structure, D , as a partially ordered set of data with each data element represented by specific values, \underline{x} , for the performance measures,

X. Tversky shows that for a data structure, D , to satisfy a polynomial measurement model, M , it is necessary and sufficient that D satisfy a irreflexivity axiom.⁹ Furthermore for any choice of an irreflexive regular extension of $>_x$, the numerical assignment resulting from M is unique. The detailed proof and discussion are tangential to the development in this thesis; the reader is referred to Tversky [144].

The usefulness of this result is that it states axiomatic conditions on the data as to whether a polynomial model can be determined. It is quite general and does not impose many of the "continuity" axioms (6.1 through 6.3) of utility theory, but it does not specify a functional form for the compaction function based on behavioral assumptions such as section 6.2 does nor does it "provide any simple set of empirically testable conditions which can be easily interpreted as a substantive theory. Furthermore, the general theory does not provide any constructive procedure for obtaining the desired numerical representation." (Tversky [144]) An exciting research question is to combine the strengths of the two complementary theories.

One use of conjoint measurement in marketing is to give each consumer a partial factorial design of the performance measures and have him rank order these in terms of preference. $f(x)$ is assumed linear in its parameters and monotonic regression is then used to determine these parameters. (Green and Wind [49])

The main advantages of this technique are that it uses non-metric dependent variables, and that the derived parameters are individual specific. Its disadvantages are that the choices are proxy (sets of values for the performance measures), that the usual choice for $f(x)$,

i.e., $f(\underline{x}_{ij\ell}) = \sum_{\ell} a_{\ell} x_{ij\ell}$ has all the restrictions of the additive risk neutral form, and that the number of factorial combinations can get quite large making it difficult for a consumer to rank order the alternatives. The compaction theory of section 6.2 can aid this technique by specifying a more appropriate function form for $f(\underline{x})$ and by calibrating it with a simpler measurement device. That is, using a few indifference questions rather than many rank order questions. Compaction theory also opens the door for psychometric performance measures identified in reduction.

6.3.5 Maximum Score

Maximum score techniques combine some of the features of random utility and factorial conjoint analysis models. As in random utility models, the consumer is assumed to maximize a utility function which has an observable portion and a random portion. The observable portion is assumed linear in its parameters and the disturbance terms are assumed independent across alternatives. As in conjoint analysis, ordinal rank order data is used, but in contrast to conjoint analysis, data is collected for revealed choice rather than proxy choice.

The basic maximum score technique is quite simple in concept. The objective is to maximize the number of correct classifications according to the utility function, and a mathematical program is set up to determine the parameters accordingly. Extensions are to have more complex scoring techniques than the number of correct classifications.

Relative to the logit formulation, maximum score techniques are more robust because of the use of ordinal data and the weakening of dis-

tributional assumptions on the disturbance term, but they are not as computationally efficient and do not make full use of the observed choice data. [Manski(98)]. Otherwise they share both the advantages and disadvantages of the logit formulation. (See section 6.3.3).

6.3.6 Preference Regression

Conjoint analysis uses a factorial design of the performance measures as alternatives and determines importance weights with monotonic regression. Maximum score uses real choice and determines importance weights with mathematical programming. A related technique is ask questions with respect to the evoked set and to use integer rank order preference (i.e., 1,2,3,...) as a dependent variable and the performance measures as explanatory variables. Enough degrees of freedom are then gained by doing the regression across individuals (and of course across stimuli). Recent simulation research by Green [46] shows that least squares regression closely approximates monotonic regression for rank order dependent variables, thus the analyst can choose the easier, more readily available regression packages.

The advantage of preference regression is that it is easy to implement. The disadvantages are that it assumes the importance weights are the same for all individuals, and has no axiomatic specification of functional form.

6.3.7 PREFMAP

The idea behind PREFMAP is that preferences are related to distances in perceptual space, i.e., the vector space formed by the reduced

set of performance measures. Basically, it is assumed that some measure of preference, s_{ij} , by individual i for stimulus j , is linearly related to the square of the distance between the coordinates of individual i 's perceptions of stimulus j and the coordinates of his "ideal point". I.e.,

$$s_{ij} = a_i d_{ij}^2 + b_i$$

where

$$d_{ij}^2 = \sum_{\ell} (x_{ij\ell} - I_{i\ell})^2$$

$x_{ij\ell}$ = individual i 's perception of performance measure ℓ for stimulus j

$I_{i\ell}$ = individual i 's "ideal" perception of performance measure ℓ

and where the dimensions are differentially weighted and rotated in computing the distances. Quadratic and linear regressions are used to determine the ideal points, the dimension weights and rotations, and the constants a_i and b_i . (Non-metric versions use rank order preference and monotonic regression.) Preference and perceptual data can be from either real or proxy alternatives.

Diagnostically the weights can be interpreted as importance weights for the dimensions and the directional cosines of the rotations can be viewed as measures of interaction among the performance measures. The average ideal point, or clusters of ideal points can serve as a useful communication technique to the design team. The advantage of PREFMAP is that these diagnostics can be determined without the additional data necessary for direct compaction assessment. The disadvantages are in the arbitrary assumptions of the functional form of the compaction

function, and that it requires a large number of stimuli, and hence alternatives for each individual.

6.3.8 Summary of Statistical Compaction Techniques

The seven techniques just described represent a cross section of the available statistical compaction techniques. Their collective advantage is that the data necessary for their calibration can be collected in a single, possibly mailed, questionnaire. This in contrast to the second, usually home interview, measurement necessary for direct compaction. Some have the disadvantage of arbitrary functional forms, others the assumption of the same parameters for everyone, and still others the mixing of effects and, hence difficulties in interpretation (see chapter 7). The user of the methodology is encouraged to choose the technique most compatible with his particular problem and with the computational facilities available to him. Table 6.1 summarizes the features of the various models.

The next section presents examples of two compaction techniques to be used, (1) a statistical technique and (2) direct assessment.

6.4 Compaction Techniques as Applied to the Design of a new HMO

Section 6.2 described direct assessment which is based on axiomatic specification of form, but requires extensive measurement. Section 6.3 described statistical compaction techniques which are less appealing theoretically but have less measurement requirements. This section describes the application of both direct assessment and preference regression to the design of a new health care delivery system at M.I.T.

Technique	Property		
Direct Assessment	functional form	independence questions	Axiomatic
Expectancy Value		✓	✓
Discriminant Analysis		✓	indifference questions
random utility models	probabilistic models	not directly related	Cardinal "utility"
conjoint analysis	existence of measurement	real	Individual specific
Maximum Score		real	Real vs Proxy choice
Preference regression		factorial design	Models measurement error
PREFMAP		robust	Statistical, i.e., no remeasurement
	✓	✓	Separate measurement of different effects
	preference on evoked set	✓	measures risk characteristics
	preference on evoked set	✓	measures interaction of performance measures

Table 6.1: Comparison of Statistical Compaction Techniques

6.4.1 Regression of Preferences vs. Performance Measures

As described in chapter 3, the initial survey to support reduction was mailed to 1000 faculty, students, and staff at M.I.T. Of these 447 returned complete surveys and of those 80 were members of an M.I.T. pilot HMO begun a year earlier. Factor analysis (described in chapter 4) of the survey data yielded factor scores for each performance measure, for each of four alternative health care systems, and for each individual. In addition the survey collected rank order preference for the four plans for each individual. This is the data necessary for statistical compaction by preference regression.

Based on prior segmentation the respondents were grouped into three segments - faculty, students, and staff and in each group the individual rank order preference¹⁰ measures were regressed against the four performance measures.¹¹ Doing this across individuals and stimuli yielded over 800 observations in the regressions. Table 6.2 gives the results.

Table 6.2: Compactions by Preference Regressions for M.I.T. HMO

R ²	Quality	Coefficients (t Statistics)			rank order fit	first preference fit
		Personal-ness	Value	Con-venience		
Overall	.27	6.17(9.16)	3.86(6.4)	5.69(12.2)	3.34(6.00)	.447
Faculty	.38	6.04(5.61)	5.92(5.67)	5.95(9.60)	3.34(3.50)	.48
Students	.31	6.59(6.10)	1.69(1.87)	7.82(8.26)	4.38(4.38)	.54
Staff	.18	4.98(3.60)	5.15(4.32)	3.54(3.98)	2.08(2.06)	.42

All regressions were significant at the 1% level and all regression coefficients were significant at the 5% level.

Since the dependent variable is ranked order, but its predicted value is continuous, the R^2 fit statistic is not the most appropriate. A more appropriate test is to rank order the predicted values and compare these with the actual rank order. Table 6.2 reports the percent of correct first preference predictions, and the percent of matches; i.e., a match occurs when the plan predicted n^{th} preferences was actually n^{th} preference. Table 6.3 is the complete match comparison table for the segment regressions. Tables like 6.3 can be viewed as a Chi-squared contingency tables and tested against random placement (See chapter 10). This test was significant at the 5% level in all regressions.

Table 6.3: Rank Order Recovery Table for Preference Regression for M.I.T. HMO Case

		Predicted Rank Order			
		1 st	2 nd	3 rd	4 th
Actual Rank Order	1 st	90	43	22	18
	2 nd	56	65	34	18
	3 rd	17	44	76	36
	4 th	10	21	41	101

1st preference (fraction) total (fraction)

matches: .479 .52

Notice that overall quality and value were highest but they do not dominate personalness and convenience. Also notice that quality is highest for faculty, but value is highest for students, and personalness is highest for staff. This suggests that there is variation in preferences across the target group and that management may want to consider offering a variety of plans rather than one. Such differentially targeted services may be more successful by meeting each need separately rather than designing an average service which does not exactly meet the needs of anybody.

Finally, as an experiment, interaction terms such as might be expected in a quasi-additive form were added as independent variables and a step-wise regression performed. None entered at any reasonable significance level. This was to be expected since both the statistical compaction model and the factor analysis are linear, but vari-max rotation is used in factor analysis to force the factors to be orthogonal in perception. As will be seen later, the performance measures are not orthogonal in preference and when risk averse conditional compaction functions are used, the interaction terms do become significant.

6.4.2 Direct Population Assessment via Personal Interview

The major result of section 6.2 was that compaction functions could be directly assessed. This technique has the potential to identify theoretically sound and mathematically consistent individual specific compaction functions with easily interpretable parameters. Each parameter characterizes a different effect. Furthermore, since each is determined by a separate measurement, its validity can be directly tested.

(This neatly side-steps problems stemming from multi-collinearity.)

This section describes a study carried out on the M.I.T. population to test the feasibility of direct assessment of a compaction function for the performance measures for health care delivery systems. This study is not meant to test the feasibility of compaction (utility) assessment for this is a proven technique for prescriptive decision making. Instead it is meant to test whether compaction functions for a significant portion of a consumer population can be assessed within the economic and time limitations of a consumer study.

6.4.2.1 Experimental Design

Sampling Process: Care was taken to get a truly random unbiased sample of the M.I.T. students. The population was proportionally stratified by sex (to insure enough females in the survey) and by undergraduate vs. graduate and a sample of 100 was randomly drawn from the list of registered students. The eight interviewers were given lists of the names and instructed to call from the top down until they got 10 positive responses — going to the next name only if a refusal occurred or the person was definitely not available. (Later tests showed no significant variation in refusal rate across interviewers.)

Survey Procedure: The procedure was to telephone the potential respondent, explain the purpose of the study (consumer input to health planning), and request cooperation. (See figure 3.1 for a sample telephone call.) If a positive response was obtained and an interview scheduled, the first portion of the questionnaire was mailed out. The

respondent was to complete this 45 minute questionnaire before the home interview. The home interview, which also takes 45 minutes, then reviewed the mail questionnaire and assessed the respondents compaction function. Finally a 10% validity check was performed to verify that the interviews actually occurred.

Survey Instrument Design and Pretest: Since assessment questions are far from the ordinary survey questions, extreme care was taken to insure that the questions were easy to answer and actually asked what they were supposed to ask. Small convenience samples were used to test various ways of asking each question and in the later stages of design a computer program [Sicherman (130)] was used to immediately give the implications of the questions in terms of preferences so that the respondent could state whether he agreed with these implications. This produced a draft copy. Next "expert" opinion was elicited from the eight students who were to be interviewers, from two graduate students who were familiar with utility assessment, and from an experienced surveyor. This produced a revised copy. The revised copy was pretested on a convenience sample of 16 (2 by each interviewer) for understandability, measurement validity and ease of implementation. The final product was (1) a set of comprehensive instructions to interviewers, (2) a set of response sheets, and (3) a set of props to aid in communication. (See appendix 2.)

6.4.2.2 Measurement Instrument and the Necessary Mathematics

Warmup: Open ended "likes and dislikes" questions were used to review the various health plans and to update our knowledge of consumer

semantics. Then constant sum paired comparison questions were used to get the consumer thinking about comparing plans. (These also provide data for future tests not covered in this dissertation.)

Perception Linkage: Before compaction functions can be used in prediction it is necessary to link the managerially relevant factor scores to consumer perceptions of the semantic scales. Based on ratings in the mailed survey (which was almost identical to the survey for the factor analysis) factor scores were obtained for the performance measures. These are linked to perception by having the respondent rate each plan directly on the performance measures. (This task was not completed for this dissertation.) Monotonic functions are then estimated to quantify the linkage. (See figure 6.13.)

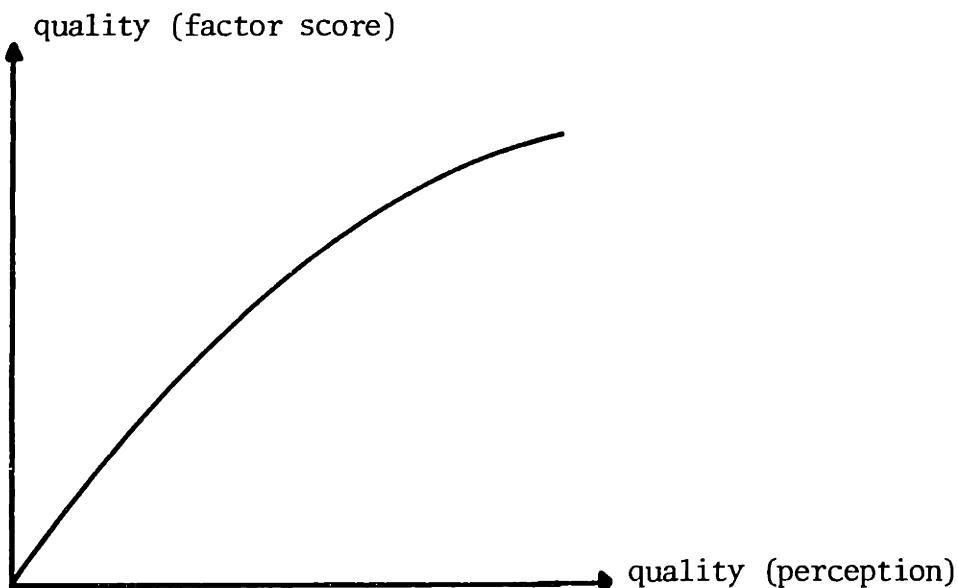


Figure 6.13: Direct Perceptions vs Factor Scores

Risk Aversion Questions: Pretest and previous utility theory indicated that a multiplicative form would not be a bad approximation. Thus the first step in the analysis is to determine the conditional compaction functions for each performance measure. (Due to the difficulty of these questions, they actually appeared later in the survey.) The measurement was to ask a type II (lottery question) of the form $(x_\ell', x_{\bar{\ell}}) \sim_x L[(x_\ell^*, x_{\bar{\ell}}), (x_\ell^0, x_{\bar{\ell}}); p_{i\ell}]$ where the respondent supplies $p_{i\ell}$. (See the mockup in figure 6.14) Constant risk aversion¹² was assumed because it is a very flexible one-parameter curve and $(x_\ell^0, x_\ell', x_\ell^*)$ were set at (2,4,6) to avoid edge effects; $x_{\bar{\ell}}$ was set at (5,5,5) to avoid over reaction to extreme values. It is theoretically possible to verify constant risk aversion but this risks respondent wear out. Mathematically this question involved solving the following equation for individual i's risk aversion coefficient, $r_{i\ell}$.

$$u_{i\ell}(x_{i\ell}') = p_{i\ell} u_{i\ell}(x_\ell^*)^1 + (1 - p_{i\ell}) u_{i\ell}(x_\ell^0)^0 = p_{i\ell}$$

thus by constant risk aversion:

$$u_{i\ell}(x_{i\ell}') = [1 - e^{-r_{i\ell}(x_{i\ell}' - x_\ell^0)}] / [1 - e^{-r_{i\ell}(x_\ell^* - x_\ell^0)}] = p_{i\ell}$$

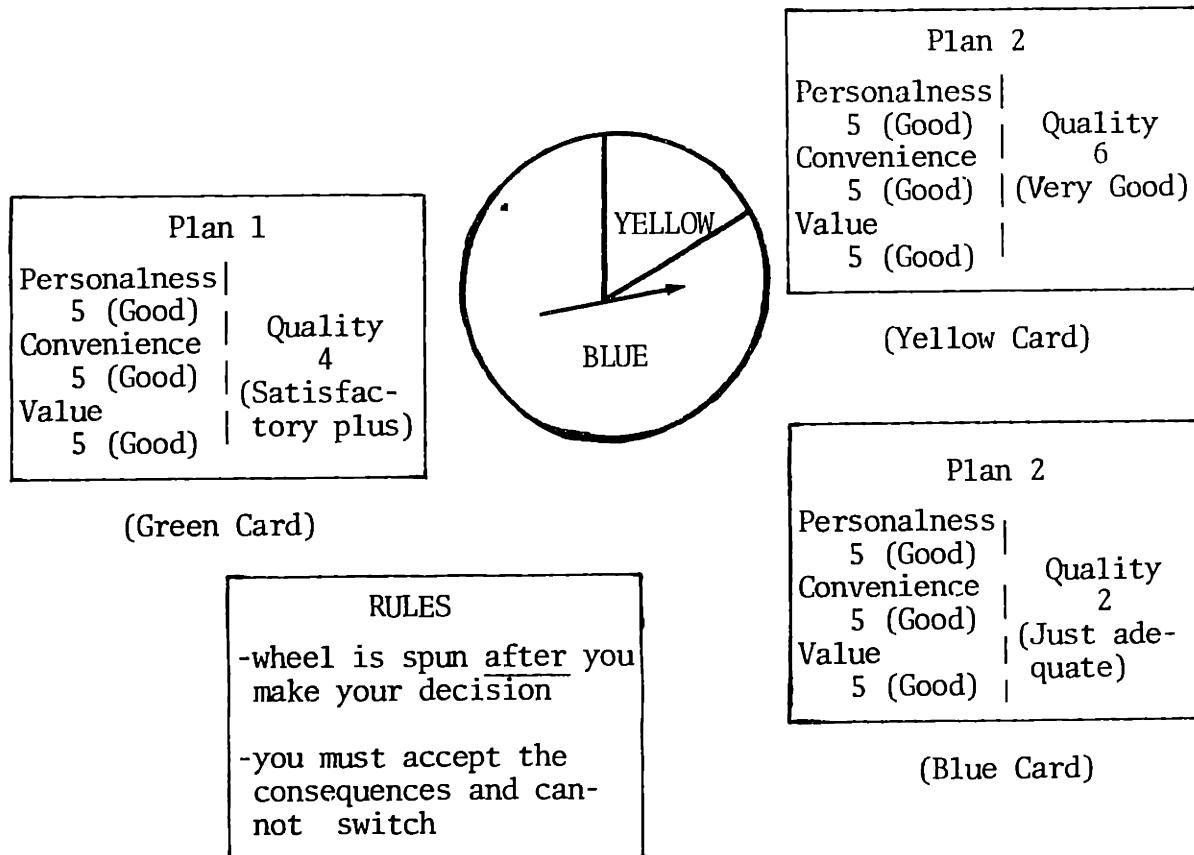
which when solved with (2,4,6) yields

$$r_{i\ell} = \frac{1}{2} \ln\left(\frac{p_{i\ell}}{1 - p_{i\ell}}\right)$$

Utility Independence Verification: One of the assumptions necessary for the multiplicative form is x_ℓ u.i. $x_{\bar{\ell}}$ for some ℓ . The

Instruction to Consumer:

Imagine you can only choose between two health plans, plan 1 and plan 2. In both plans personalness, convenience, and value are good (rated 5). You are familiar with plan 1 and know that quality is satisfactory plus (rated 4). You are not sure of the quality of plan 2. If you choose plan 2, then the wheel is spun and the quality you will experience for the entire year depends on the outcome of the wheel. If it comes up yellow, the quality is very good (rated 6) and if it comes up blue the quality is just adequate (rated 2). Graphically this is stated:



Instruction to Consumer:

At what setting of the odds (size of the yellow area) would you be indifferent between plan 1 and plan 2? (Respondent is given wheel and adjusts it until size of yellow area is appropriate. He is challenged by being given the choice with his setting. If he prefers one plan or the other, the interviewer iterates the question until a true indifference setting is determined.)

Figure 6.14: Schematic of Risk Aversion Assessment Question

survey tested it for each ℓ by changing $x_{\bar{\ell}}$ from $(5,5,5)$ to $(3,3,3)$ and asking another risk aversion question. If $p_{i\ell}$ does not change then this is evidence that x_{ℓ} u.i. $x_{\bar{\ell}}$.

Tradeoff Questions: The next step in the analysis is to determine the relative scaling constants for each performance measure. The measurement was to ask type I (tradeoff) questions of the form

$(x_m^*, x_{\ell}^0, x_{\bar{m}\ell}) \sim_x (x'_{im}, x_{\ell}^*, x_{\bar{m}\ell})$ where individual i supplies x'_{im} . (See figure 6.15.) To be consistent with the risk questions the extreme values were $(2,6)$ and $x_{\bar{m}\ell}$ was $(5,5)$. Mathematically this question resulted in the following equation:

$$c_i(x_m^*, x_{\ell}^0, x_{\bar{m}\ell}) = c_i(x'_{im}, x_{\ell}^*, x_{\bar{m}\ell})$$

Using the multiplicative form and cancelling terms yields:

$$k_{i\ell} = k_{im} \cdot u_{im}(x'_{im})$$

Since $u_{im}(x'_{im})$ is known from the risk aversion questions this gives

Figure 6.15: Schematic of Trade-off Question

Instruction to Consumers:

Now consider the two plans below and choose the level of the value factor in such a way that you are indifferent between the two plans. (Consumer is challenged and the question iterated until a true indifference is determined.)

<u>Plan A</u>	<u>Plan B</u>	
Quality = 5 (good)	Quality = 5 (good)	Excellent
Personalness = 5 (good)	Personalness = 5 (good)	Very Good
Convenience = 6 (very good)	Convenience = 2 (very poor)	Good
Value = 2 (very poor)	Value	Satisfactory
		Poor
		Very Poor
		" Extremely Poor

exactly the relative importance weights. Empirically all performance measures were "traded off" against quality since pretest indicated quality was most likely to have the largest scaling constant and thus most answers for x'_{im} would be in the range (2,6).

Preferential Independence Verification: Another assumption necessary for the multiplicative form is $\{X_m, X_l\}$ p.i. X_{ml} for all l and for some m . (If X_m u.i. X_{ml} .) The survey tested it for each l by changing x_{ml} from (5,5) to (3,3) and asking another tradeoff question. If x'_{im} does not change then this is evidence that $\{X_m, X_l\}$ p.i. X_{ml} .

Interdependency Question: This last question determines the exact value of one of the scaling constants. The other scalings are then known since the tradeoff questions determined their relatives values. Normalization then gives the interdependency coefficient, K_i . The measurement was to ask a type II (lottery) question in which all the performance measures varied simultaneously. (The mockup is similiar to figure 6.14.) It was of the form $(x_l^*, x_{\bar{l}}^0) \sim L(x^*, x^0; p_i)$ where individual i supplied p_i . Mathematically this question resulted in the following equation:

$$c(x_l^*, x_{\bar{l}}^0) = p_i c(\cancel{x^*})^1 + (1 - p_i)c(\cancel{x^0})^0 = p_i$$

Using the multiplicative form and cancelling terms yields:

$$k_{il} = p_i$$

For simplicity quality was chosen as X_l , although any performance measure could have been used. Finally normalization gives $c(x^*) = 1.0$ which

yields the following equation which can be solved for K_i .

$$(1 + K_i) = \prod_{\ell=1}^4 (1 + k_{i\ell} K_i)$$

(Keeney [71] shows that if $\sum_{\ell} k_{i\ell} > 1$ then $K_i \in (-1, 0)$ and if $\sum_{\ell} k_{i\ell} < 1$ then $K_i \in (0, \infty)$ and that in either case there is only one real root in the relevant range.

Finale: After the final question the respondent was thanked for his time and patience. It would have been nice to actually offer a lottery with some nominal monetary value as a reward, but we had no money available for this purpose.

6.4.2.3 Empirical Results

Respondent reaction: Based on a goal of 80 interviews, the interviewers arranged interviews with 76 students. Of these all completed the interview and the general impression was that the respondents understood and could answer the tradeoff and lottery questions. (The educational warmup questions, the props, and challenges of the indifference settings were a major factor in communicating the true content of the question.) Those who had the most trouble with the lottery questions were those trained in probability who were influenced by the expected value of the lottery and found it hard to express their true feelings.

Numerical results: Table 6.4 presents the median, interquartile range, mean, and standard deviation of the importance weights, the risk coefficients, and the interaction coefficient, i.e.

$\lambda_i = \{k_{i1}, k_{i2}, k_{i3}, k_{i4}, r_{i1}, r_{i2}, r_{i3}, r_{i4}, K_i\}$. The median and interquartile range are the better measures of central tendency and dispersion because (1) the distributions of the k_{il} 's are skewed and bounded (2) the risk coefficients are non-linear transformations of the indifference probabilities set by the respondent, thus the median risk coefficient corresponds to the median probability but the same is not true with respect to the mean, and (3) a few individuals had positive interaction coefficients which were large enough to significantly effect the mean.

Interpretation: For the importance weights, the rank order of the mean and the median are the same with quality being the most important attribute but with the other three all significantly important. Managerially this means that the HMO should seriously consider quality but take care not to jeopardize the other three performance measures. The interaction coefficient is negative and close to -1 indicating that the performance measures act as substitutes for each other. An interesting unexpected empirical result is the strong rank order correlation between the risk coefficients and the importance weights, i.e., $\{Q, V, P, C\}$ median risk vs. $\{Q, V, C, P\}$ median importance. Intuitively this suggests that the more important a performance measure is, the less likely an individual is to take a risk with that measure. Finally notice that for every performance measure, the majority of people are risk averse.

Assumption verification: Both tradeoff and lottery questions were asked to check for preferential and utility independence. In each case the respondent was asked if his answers would change if the fixed and common values (x_{lm} for tradeoffs, x_l for lotteries) changed. If

6.4: Summary Statistics for Direct Assessment of Compaction Functions for M.I.T. HMO

	Importances (k_{im})			Risk Aversion Coefficients (r_{im})			Interaction Coeff. (K_i)		
	quality	personal-ness	value	convenience	quality	personal-ness			
Median	.800	.525	.706	.682	.693	.332	.424	.310	-.992
Inter-quartile interval	.650 - .900	.360 - .726	.492 - .800	.525 - .800	.424 - 1.10	.000 - .693	.203 - .549	.000 - .693	-.998 - -.939
Inter-quartile range	.250	.367	.308	.275	.675	.693	.347	.693	.059

he changed his answers he was asked why. Table 6.5 summarizes the number of times the assumptions were verified, (X_λ u.i. $X_{\bar{\lambda}}$ for lotteries, and $\{X_\lambda, X_m\}$ p.i. $X_{\bar{\lambda}m}$ for tradeoffs). "Close" is defined as a change of no more than .05 in the probability setting (utility independence) or a change of no more than 1 unit in the tradeoff setting (preferential independence).

Table 6.5: Assumption Testing

	X_λ	u.i.	$X_{\bar{\lambda}}$	$\{X_\lambda, X_m\}$		p.i.	$X_{\bar{\lambda}m}$
λ	Q	V	C	P	Q	V	C
exact(%)	51	55	39	53	—	64	57
close(%)	66	70	71	68	—	89	84

X_m = Quality

These results intuitively imply that the multiplicative form is not a bad approximation, but no formal mechanism for testing this is available in the utility literature. Note that $K_i \approx -1$ does imply that the additive form is a poor approximation. Qualitatively, when $x_{\bar{\lambda}}$ was changed from {5,5,5} to {3,3,3} the indifference setting probability was slightly increased based on "if I only have one good value for a performance measure, I am less willing to take a risk." This is a reasonable reaction which existing utility forms have no easy way to handle. Some related work has been done by Kirkwood [79] on parametric forms, but this remains an open and provocative research question.

Indifference curves: For consumer modeling the preference parameters are left idiosyncratic, but a useful pictoral representation is to plot¹³ indifference curves based on the mean or median values. Figure 6.16 represents tradeoffs between quality and convenience at various price levels. Because of preferential independence the same curves apply for all levels of personalness. These curves graphically provide management with a view of how much quality and convenience are necessary to support a given price level.

Rank order recovery: Although the compaction values are inputs to the probability of choice models, one early test is to compute the rank order recovery. Table 6.6 is a complete rank order recovery table. Notice that the rank order fit of .474 and first preference fit of .495 are roughly the same as the fits obtained with the statistical procedure. This is encouraging for a first attempt since the parameters of the statistically compaction were chosen to maximize rank order fit whereas the parameters of direct assessment were chosen independently of rank order preference. Because of the differences between the statistical and direct compaction techniques especially in their relation to the rest of the methodology and because of the non-linear relationship between the factor scores, the directly measured performances, and the risk averse scaling function, stronger, more explicit comparison tests need to be devised before importances can be compared. This is the subject of future research.

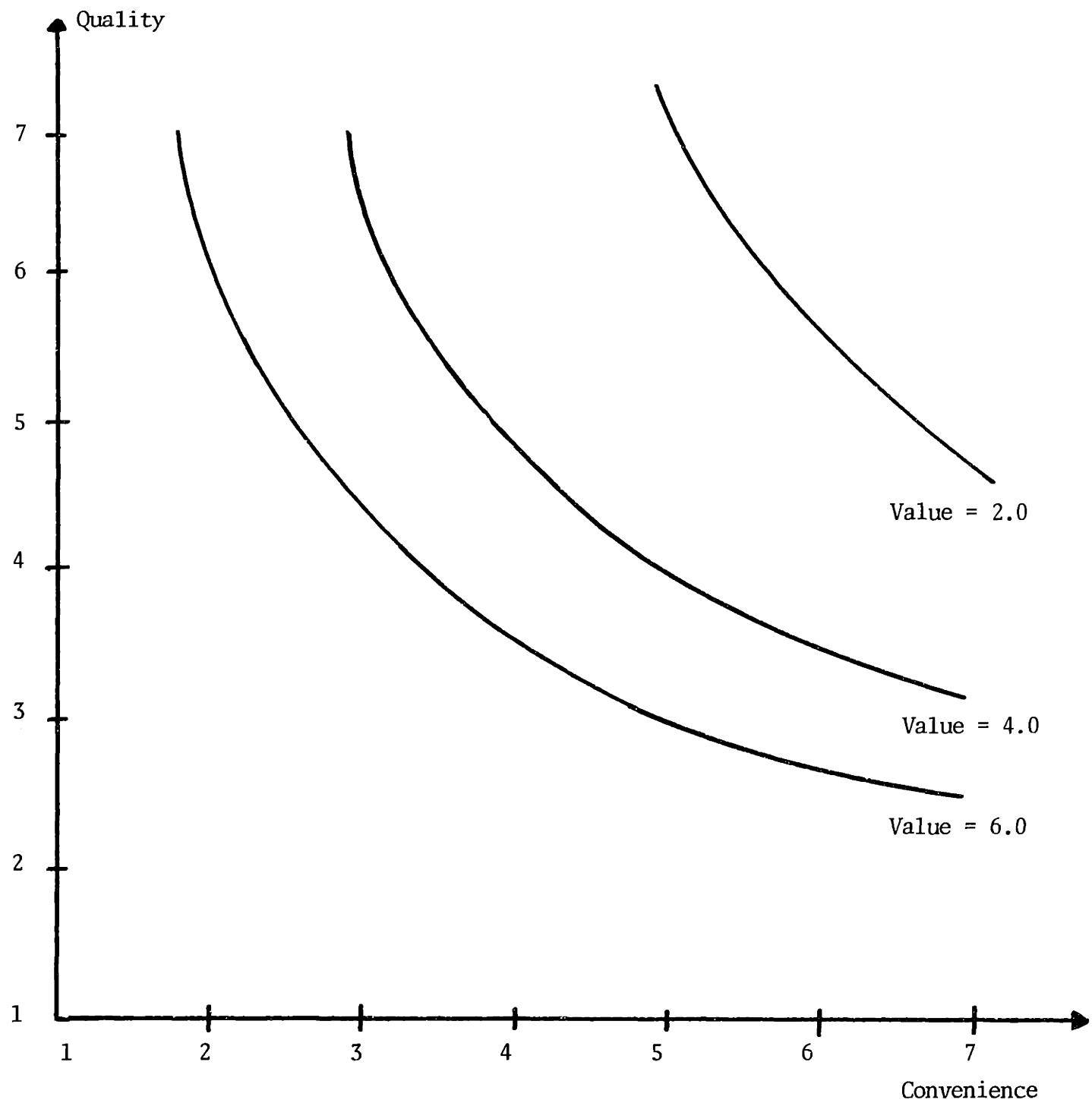


Figure 6.16: Indifference Curves for Health Care

Table 6.6: Rank Order Recovery Table for Direct Assessment of Compaction Functions

		Predicted Rank Order			
		1 st	2 nd	3 rd	4 th
Actual Rank Order	1 st	33.67	14.50	14	5.83
	2 nd	20.17	26.50	12	9.33
	3 rd	7.33	16.17	30.17	14.33
	4 th	6.83	10.83	11.83	38.50
1 st preference (fraction matches:		.495		Total (fraction) .474	

(This table includes ties.)

6.5 Summary of Compaction

An important step in the consumer response methodology is to identify how the reduced performance measures combine to produce a scalar measure of goodness. Furthermore it is important to do this in such a way that valuable diagnostics are provided to aid in the design of innovation.

Section 6.1 formalized the concept of compaction and identified the desirable properties of completeness, symmetry, preference, encompassment, and canonicity. It then showed that certain transformations and in particular positive linear transformations retained these properties.

Section 6.2 began by examining prescriptive utility and found it lacking because consumers did not always choose the alternative with the highest expected utility. Under a different perspective it then re-defined preference and added two choice axioms to the von Neumann-Morgenstern utility axioms. This identified an isomorphism between descriptive compaction functions and prescriptive utility functions which allowed utility theoretic results such as existance, uniqueness, and axiomatic functional forms to be used for compaction. The major output of this section was a feasible, theoretically sound, practical technique to directly assess and verify compaction functions. Furthermore each parameter of the compaction function is separately measured and provides easily interpretable diagnostics about the consumer choice process.

Direct assessment requires a special measurement, i.e., a special survey. Some applications will not have the time and money for such an effort. Section 6.3 surveys the literature and suggests the alternative statistical compaction techniques of expectancy value models, discriminant analysis, random utility models, e.g., logit, conjoint analysis, maximum score, preference regression, and PREFMAP.

Finally section 6.4 relates empirical experience with both statistical compaction and direct assessment with respect to performance measures describing health care delivery systems. The statistical technique was easy to implement with standard statistical packages and provided indications before completion of direct assessment. Direct assessment was found to be feasible and found to provide interpretable results which led to important insights on consumer behavior.

The next chapter examines how to transform the scalar measures of goodness into choice probabilities.

Chapter 7

PROBABILITY OF CHOICE

This last step in the individual choice section of the methodology provides numerical estimates of how many consumers will choose their new product or service and how many of the potential consumers will select each of the various competing alternatives (see Figure 7.1). Mathematically this is done by transforming the vector of scalar measures of goodness obtained in compaction into numerical estimates of individual choice probabilities.

In most cases outputs are one time Bernoulli probabilities, $p_s(a_j | c_{i1}, c_{i2}, \dots, c_{iJ})$, which estimate each individual's selection probabilities for each alternative, a_j , conditional on the scalar measures of goodness for each and every alternative, i.e., $c_i = (c_{i1}, c_{i2}, \dots, c_{iJ})$. In cases where repetitive choice decisions are made by a consumer, separate trial and repeat choice parameters would be estimated based on the goodness measure before and after use of the new product or service. Finally, in some applications the probabilities can be Poisson rates or parameters of more complex models. If γ is the choice rate for an alternative, then the probability that the individual will choose that alternative in small time period Δt is $\gamma \Delta t$.

Empirically, consumers do not always choose the alternative with the largest compaction value (see section 6.4). Reasons for this are effects such as measurement errors, specification errors, and non-

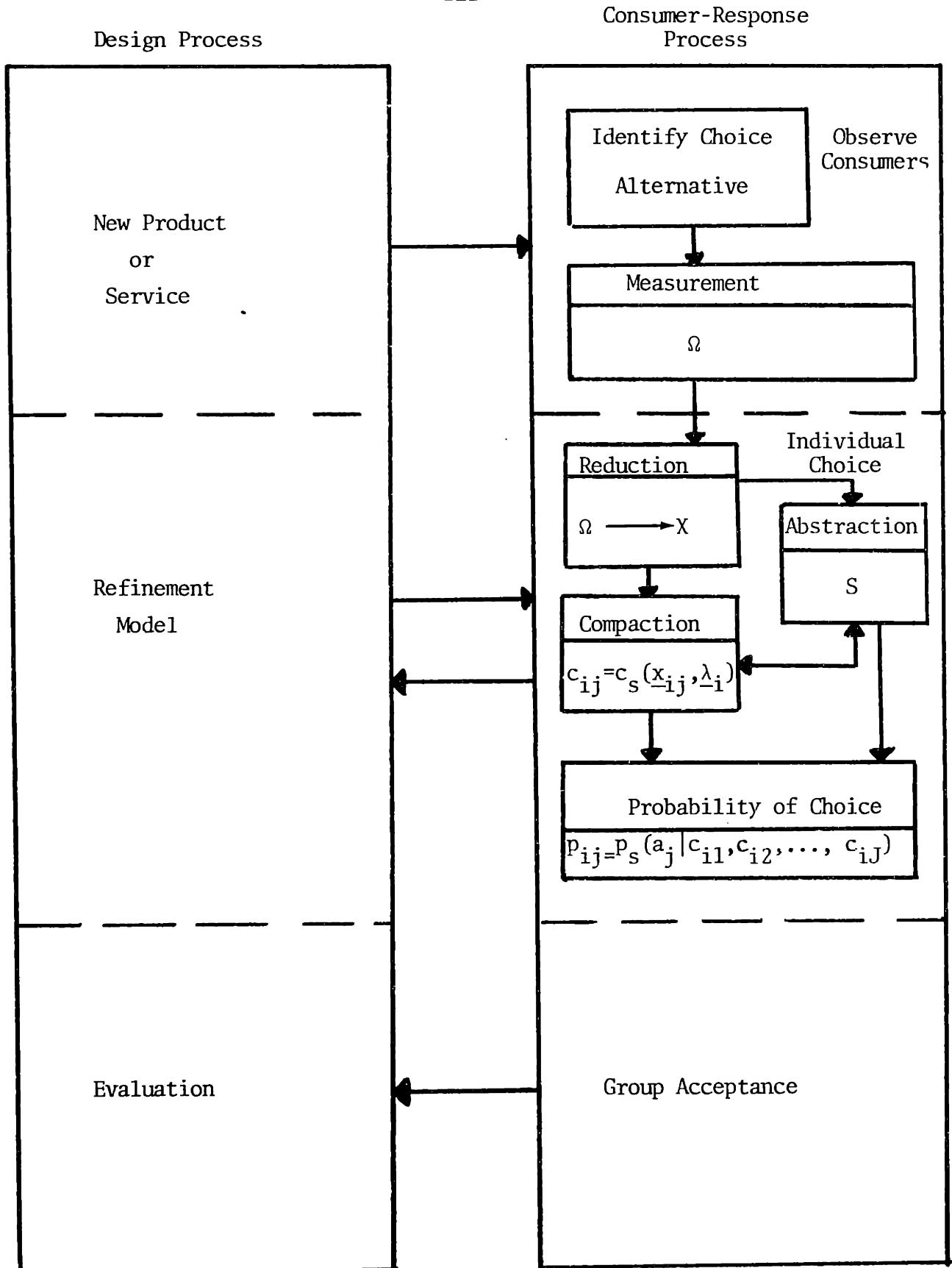


Figure 7.1: Relationship of Probability of Choice to the Methodology

stationarity in consumer response. This chapter discusses these issues in probability prediction and presents techniques which estimate choice probabilities based on both the rank order and the magnitude of the goodness measures. Mostly conditioned Bernoulli models of behavior will be discussed. By externally modeling dynamic effects, e.g., by estimating both trial and repeat probabilities and using a macro-flow model such as Urban's SPRINTER[147], many realistic situations can be modeled with sets of conditional Bernoulli probabilities (see Chapter 8). Extensions to Poisson or to more complex models are not discussed in this dissertation.

This chapter begins with a formal development which identifies the basic issues in probability prediction. Next various utility maximizing models which explicitly incorporate the "errors" are presented. These include general, but not practical, integral equations, random utility models, and aggregate utility models. These techniques are not necessarily compatible with direct assessment of compaction functions. A new Bayesian technique is then presented which is compatible and which implicitly incorporates "errors" via the compaction axioms. An empirical example is given based on data collected prior to national introduction of a now well-known aerosol deodorant. Finally other probability models and the issues of simultaneous choice and stability over time are discussed.

7.1 Formal Development

We begin by identifying and naming some issues of probability models and by relating them to the formal developments in reduction, abstraction, and compaction.

7.1.1 Important Issues for Probability Models

Symmetry: Definition 6.3 introduced this concept in connection with compaction. It is restated here for emphasis. Basically a probability model, conditioned on the goodness measures, c_{ij} , is symmetric if switching two goodness measures while holding all others constant switches the choice probabilities. This means essentially that the goodness measures tell the whole story and that once a goodness measure is known for an alternative the identity of that alternative gives no additional information. As was shown in Theorem 6.5, if the set of performance measures is complete (definition 4.3) and if the compaction axioms hold, then the probability model will be symmetric.

Encompassment: Formally introduced in definition 6.5, encompassment means that the probability model, $p_i(a_j | c_{i1}, c_{i2}, \dots, c_{iJ})$, is the same for all individuals, i , in a segment, s . Note that this allows the compaction function and/or the goodness measures to be different for each individual, i , but it requires that any two individuals with the same vector of goodness measures have the same set of choice probabilities. This property is crucial in forcing enough degrees of freedom in any practical calibration technique that makes interpersonal compar-

isons of choice. It is difficult to relax even if extensive panel data¹ is gathered for each individual.

Equal probabilities for perfect substitutes: If two alternatives have the same compaction value and if the set of performance measures are complete, then one would expect that in terms of that individual's preferences, the two alternatives are indistinguishable and should have the same choice probabilities. But this is just the definition of compaction (definition 6.1) and hence any "correct" probability model must have this property. Note further that this property is implied by symmetry.

Interpolative: Almost by definition, innovation is beyond the range of existing experience, thus we can often expect that the performance measures relative to the new alternatives will have values outside the range of values for existing alternatives. For example, suppose a new mode of transportation is introduced which is faster than any existing mode, but more expensive. If our model is calibrated on only existing modes we might question the accuracy of extrapolating it to the new mode. We would rather calibrate it (somehow) on a range for the performance measures which includes the values for the new mode and then interpolate. Formally:

Let R_1 = the range of calibration, i.e., the range of the performance measures over which the model is calibrated. $R_1 \subseteq X$ where X is the total possible range.

R_2 = the range of prediction, i.e., the range of the performance

measures which includes the values for the new alternatives,

$$R_2 \subseteq X.$$

Definition 7.1: A probability model with calibration and predictive ranges R_1 and R_2 is said to be interpolative if $R_2 \subseteq R_1$. Otherwise it is said to be extrapolative.

Comment 7.1: Most models based on revealed preference are extrapolative.

Comment 7.2: The direct assessment technique based on section 6.2 determines a compaction function calibrated on the entire range X . Thus if a probability model is based on the compaction values and is calibrated on their entire range it is implicitly calibrated on X and hence is interpolative.

Transferability: Suppose we are calibrating a model to be used by small communities for designing health maintenance organizations. We would want to calibrate it once in a representative community and then use it nationwide. We might expect trouble because preferences could be expected to shift from community to community. Perhaps we can capture preferences in the preference parameters of the compaction function and then approximate behavior by using a previously calibrated probability model to transform the goodness measures, c_i , into probabilities. In other words suppose that $p_s(x_j)$ is the probability distribution of the performance measures² for alternative a_j in segment s of community T_1 . Similarly define $p_s(x_1, x_2, \dots, x_J)$. Furthermore let $p_s(\lambda)$ be the distribution of preference parameters in T_1 . Taken together these

result in a derived distribution for the goodness measures, $p_s(c_1, c_2, \dots, c_J)$. We would like to then use a probability model calibrated in representative community, community T_2 , to obtain $p_s(p_1, p_2, \dots, p_J)$, i.e., the choice probabilities. To facilitate later discussion we will call a probability model for which this is valid transferable from T_2 to T_1 . Formally:

Definition 7.2: A probability model, $p(a_j | c_{i1}, c_{i2}, \dots, c_{iJ})$ is transferable from community T_2 to community T_1 if whenever $c_s(x_{ij}, \lambda_i) = c_t(x_{hj}, \lambda_h)$ for all $j = 1, 2, \dots, J$ then the choice probabilities for i and for h are the same. Where i is an individual in T_1 and h is an individual in T_2 and where s and t are equivalent segments.

Comment 7.3: The specification of the concept "equivalent segments" is purposely left vague. The idea here is to relate the segment definitions to observable measures such as demographics or other individual descriptors. See Chapter 5, abstraction.

In applying this definition one must be careful that the alternative sets may differ from community to community. This presents no problem if the probability model is symmetric because then only the goodness measures and not the identity of the alternatives are used in the probability model. Another problem that may arise is that the number of alternatives in T_1 is different than the number in T_2 . This can be avoided if the compaction function is preferential (definition 6.4) by using artificial alternatives with compaction values equal to minus infinity ($-\infty$). Note that the axiomization of section 6.2 implies

both symmetry and preference.

Extendability: If an entirely new alternative, say a new health maintenance organization (HMO), is offered, then everyone who becomes aware of this HMO will have his choice set augmented by the new service. Imagine that the probability model was calibrated on the J existing services. To be useful for the new HMO it must be valid when $J+1$ services are offered. This property of being valid for an entirely new alternative is called extendability. Formally:

Definition 7.3: Suppose a probability model, $p(a_j | c_{i1}, \dots, c_{iJ})$ is valid for J alternatives. It is K th order extendable if it is possible to construct without recalibration a valid probability model for $J+K$ alternatives. If this is true for all $K=1, 2, \dots$, the model is said to be extendable.

Independence Among Individuals: How often have you heard: "Everyone knows that word of mouth is the best form of advertising."? Such comments seem to imply that any probability model, $p(a_j | c_i)$, should be conditioned on other individuals. But the strongest effects of peer group influence is to create awareness, update perceptions of the performance measures, x_{ij} , or alter preference parameters, λ_i . External macro-flow models such as Urban's SPRINTER [147] can model the propagation of peer group effects on x_{ij} , λ_i , and awareness if the behavioral propagation phenomena are known. Since these parameters are then enough to compute the compaction values, c_i , the methodology can predict

choice. In transportation there are congestion effects, such as lowered speed at rush hour, which alter the values of the performance measures and hence effect choice. These too must be externally modeled.

7.1.2 Dependence on the Choice Set

One would expect that the choice probability for a given alternative would depend on the choice set, i.e., on which competing alternatives are known and available to the consumer. This section indicates some of the practical problems of calibration and prediction and discusses some of the theoretical issues of choice set dependence. Certain independence properties will be identified which are necessary for any probability model consistent with the compaction axioms.

Evoked Set: Any rational consumer will not consider choosing an alternative if he does not know enough about it to make a choice. When measuring consumers' perceptions of alternatives, when calibrating a probability model, and when predicting choice, one must be careful to ensure that alternatives are considered only if the consumer can realistically make judgements about them. These relevant alternatives are called the evoked set. For example in a study (Urban [146]) of seven consumer products, the evoked set was small compared to the number of brands available (See Table 7.1). In calibration and prediction this concept is incorporated by conditioning choice probabilities on the evoked set and by estimating awareness of the new alternative by another model. Chapter 8 (aggregation) discusses prediction when there are

Table 7.1 Evoked Set Size and Composition*

Product	Median Evoked Set Size	Total Number of Brands Evoked	Number of Brands Necessary to Account for 80% of Market
Canadian Beer	7	15	7
Aerosol Deodorant	3	20	6
Skin Care Product	5	30	12
Over the Counter Medicinal Product	3	20	5
Pain Relief Product	3	18	6
Antacid	3	35	6
Shampoo	4	30	20

changes in the evoked set.

Independence of Irrelevant Alternatives: When deterministically comparing two alternatives, one would expect that the preference ordering among these alternatives would not change if a third alternative is added to the choice set. This concept, introduced by Arrow [5] is known as independence of irrelevant alternatives.

In stochastic choice it is not always true, especially if the choice decision is hierarchical. See for example the health example in section 6.2.3, when the choice was first between group practice and private practice and then the choice of plan or doctor. But, if we are careful to explicitly model the hierarchical nature of certain choice decisions we can invoke independence of irrelevant alternatives in terms of stochastic choice within each level of the hierarchical

* Reproduced from Urban [146].

choice process. In fact this property is implied by axiom 6.4, which stated that when an alternative is deleted from the choice set the stochastic preference ordering did not change for the remaining alternatives.

For example suppose you are purchasing a bicycle. Suppose further that you are in some mythical society (not Boston) where bicycle theft is not a problem. In this case it is reasonable to assume that the class of ten-speed bicycles is stochastically preferred to the class of three-speed bicycles, e.g., $\text{Prob (10-speed)} = .6$, $\text{Prob (3-speed)} = .4$. But suppose you walk into a store with the intent to purchase a bicycle and the store has three identical 10-speeds but only one three-speed. Under reasonable assumptions $\text{Prob}(\text{purchase } \underline{\text{the 3-speed}}) = .4$ but $\text{Prob}(\text{purchase } \underline{\text{a particular 10-speed}}) = .2$. In this case if the nature of the choice process is not explicitly modeled, axiom 6.4 will be violated because a reversal of stochastic preference will occur.

Thus to be consistent with the compaction axioms any probability model using directly assessed compaction functions must have the property of independence from irrelevant alternatives in the stochastic sense.

A stronger form of this property, sometimes also referred to as independence of irrelevant alternatives is Luce's choice axiom [92]. This states essentially that:

$$\frac{p(a_1 \text{ from } \{a_1, a_2\})}{p(a_2 \text{ from } \{a_1, a_2\})} = \frac{p(a_1 \text{ from } A)}{p(a_2 \text{ from } A)} \quad \text{for any } A \text{ such that } a_1, a_2 \in A.$$

This axiom implies [Theorem 3, page 23 of Luce (92)] that there exists a real-valued function $v(a_j)$ defined on A such that

$$p(a_1 \text{ from } A) = \frac{v(a_1)}{\sum_{j=1}^J v(a_j)} \quad A = \{a_1, a_2, \dots, a_J\}$$

This is clearly much stronger than axiom 6.4 and will not be required for any probability model consistent with the compaction axioms, although some models in section 7.2 will exhibit this property.

Simple Scalability: A generalization of Luce's choice axiom, first investigated by Krantz [83], is simple scalability. Simple scalability states that there exists a real-valued function $u(a_j)$ defined on A and a sequence of real-valued functions $F_J: \alpha \rightarrow \mathbb{R}$ for all $J = 2, 3, \dots$ where $\alpha \subseteq A$ and α contains J alternatives such that

$$p(a_1 \text{ from } \alpha) = F_J[u(a_1), u(a_2), \dots, u(a_J)]$$

where F_J is strictly increasing in $u(a_1)$ and strictly decreasing in $u(a_j)$ for $j > 1$. (Non-decreasing and non-increasing if $p(a_1 \text{ from } \alpha) = 0$ or 1.)

Theorem 6.6 implies that $c(x_{ij}, \lambda_i)$ is a real-valued preferential compaction function. But, the compaction definition implies that the choice probabilities are completely determined by the compaction vector which implies that the compaction function is a scale function. Furthermore the property of preference implies that $p(a_1 | c_{i1}, c_{i2}, \dots, c_{iJ})$ is

non-decreasing in $c(x_{i1}, \lambda_i)$ and non-increasing in $c(x_{ij}, \lambda_i)$ for $j > 1$.

Thus any probability model consistent with the compaction axioms must satisfy this weaker form of simple scalability.

Independence from Zero Probability Alternatives: A concept related to independence of irrelevant alternatives (IIA) is that choice probabilities should be unaffected by alternatives which will never be chosen.

In other words suppose $\text{Prob}(a_j \text{ from } A) = 0.0$, then one would expect that $\text{Prob}(a_k \text{ from } A - a_j) = \text{Prob}(a_k \text{ from } A)$ for all $a_k \in A$, $a_k \neq a_j$.

This is different from (Arrow's) IIA because IIA deals only with stochastic preference ordering while independence from zero probability alternatives deals with the values of the choice probabilities. Interpreting this in terms of compaction functions (see Theorem 6.6), if $c(a_j) = -\infty$ then $\text{Prob}(a_k \text{ from } A - a_j) = \text{Prob}(a_k \text{ from } A)$ for all $a_k \in A$, $a_k \neq a_j$. (Note it is possible that $c(a_j) > -\infty$ and $\text{Prob}(a_j \text{ from } A) = 0.0$.) Thus, Theorem 6.6 requires that any probability model consistent with the compaction axioms must exhibit independence from zero probability alternatives in terms of compaction functions.

Independence from Identical Substitutes: In the bicycle example, we did not expect the probability of choosing the 3-speed to depend on the number of 10-speeds available. Generalized this property is independence from identical substitutes, i.e., if $x_{iJ} = x_{iJ+1} \neq x_{i1}$ then $p(a_1 | c_{i1}, \dots, c_{iJ}) = p(a_1 | c_{i1}, \dots, c_{iJ}, c_{iJ+1})$ for an extendable probability model. Clearly we would not expect this property to hold if $x_{iJ} = x_{i1}$. But what if $x_{iJ} \neq x_{iJ+1}$ but $c(x_{iJ}) = c(x_{iJ+1})$, does

this property still hold? What if $c(x_{iJ}) \approx c(x_{iJ+1})$? In general it may not, but if it does the property will be called independence from perfect substitutes. Some of the probability models discussed later exhibit this property, others do not. At this point it remains an unanswered question whether probability models should exhibit this property.

Summary of independence properties: This section has raised a number of independence issues. Some of these will be used to evaluate probability models introduced later in this chapter, because any probability model consistent with the compaction axioms must exhibit the properties of (1) independence of irrelevant alternatives (in terms of stochastic preference ordering), (2) weak simple scalability, and (3) independence from zero probability alternatives. Finally the evoked set must be explicitly considered in both calibration and prediction.

7.1.3 Perceptions are Usage Dependent

How many times did you try a new product that did not live up to your expectations? Did you ever try it again? Perhaps your perception of the product and the values of its performance measures were changed after using it.

Trial vs. Repeat: In general, each time a consumer uses a product his perceptions are updated. A simple approximation to this phenomena would be to assume that only the first usage, i.e. trial, changes perceptions. This would result in a trial-repeat model similar to that of

Figure 7.2. Such a model is capable of indicating the time dependent diffusion of acceptance of a new alternative. This will be discussed in detail in Chapter 8, aggregation. (Of course more complex models can also be used.)

None-the-less this indicates the importance of measuring perceptions of the performance measures before and after first usage or higher order usage if more complex models are used. Furthermore, one must be careful in interpreting the scalar measures of goodness and the resulting choice probabilities when they are obtained from concept statements or by other techniques used to force evoking of a new alternative.

7.1.4 Revealed Preference vs. Proxy Choice

Direct assessment of compaction functions requires consumers to specify stochastic preference relative to proxy choices, i.e., alternatives represented by lists of attributes. This is in contrast to the econometric methods which work on revealed preference, i.e., by statistically determining parameters of models based upon observing choices among actual alternatives. Each method has its problems, and hence different choice sets require different techniques.

Arguments pro revealed choice: There are three major arguments for revealed choice: (1) consumers do not always choose what they say they will, (2) the concept statement or list of performance measures may not be accurate in specifying an alternative, and (3) any list of attributes may be incomplete and hence statistical estimation will

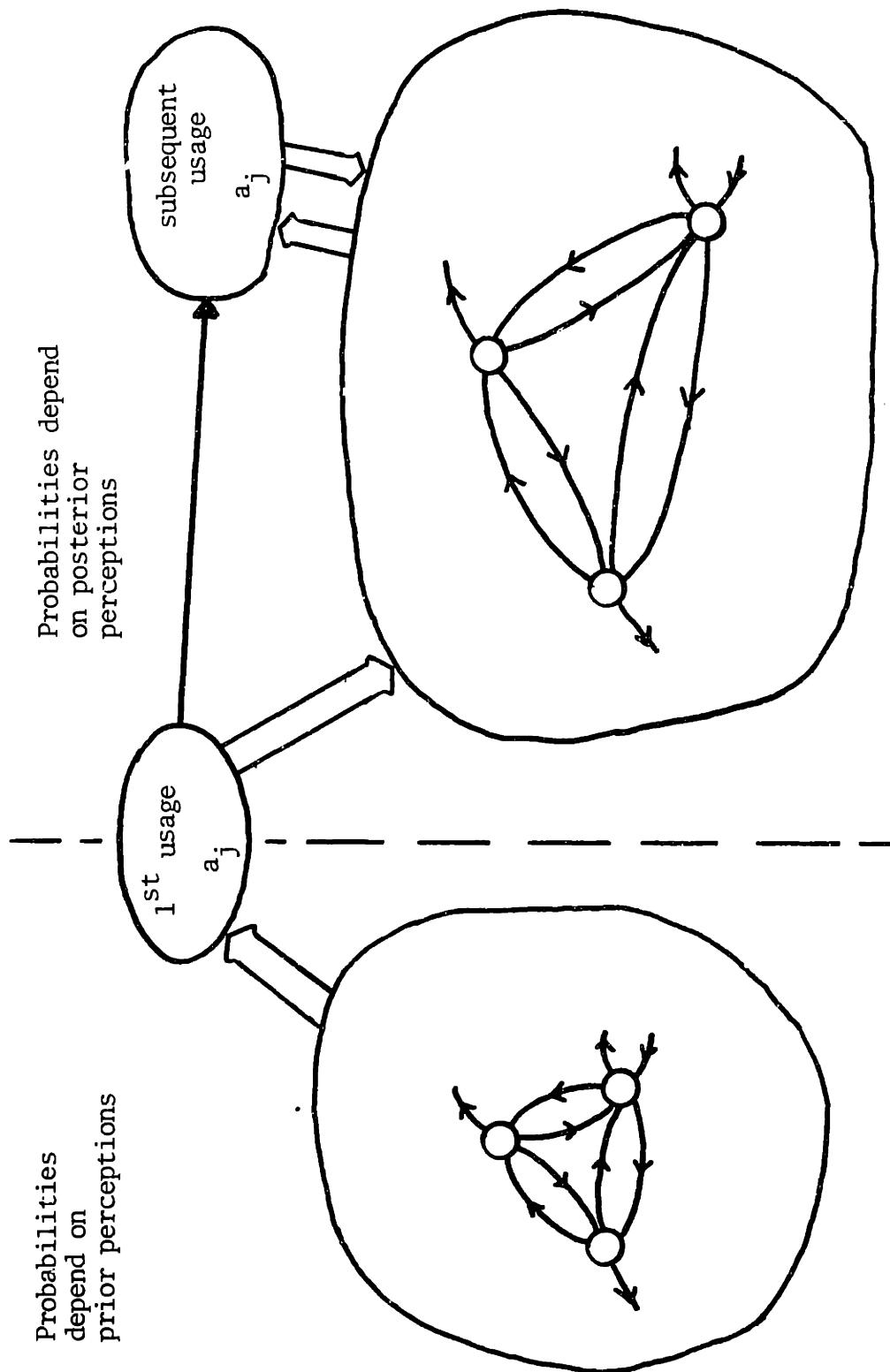


Figure 7.2. Trial and Repeat

capture and correct for this incompleteness by having alternative specific parameters.

Much of argument (1) stems from consumers changing perceptions when faced with a real alternative. This is a real issue which must be faced if proxy choice is used. Argument (2) is similar. Concept statements give trial probabilities. Usage changes perceptions.

Argument (3) is very important if the attributes and/or performance measures are restricted to "engineering" measures such as travel time or wait time. One of the goals of reduction (Chapter 4) is to get a complete set of performance measures. If reduction is successful, then alternative specific parameters should not be necessary.

Arguments pro proxy preference: There are four arguments pro proxy preference: (1) Most statistical techniques using revealed preference require compaction functions to be linear in their parameters. This may not adequately model choice. (2) Statistical techniques based on revealed preference require the parameters to be the same for all individuals in a segment. This may not adequately model choice. (3) There is often multicollinearity among performance measures. This may make interpretation of statistically determined preference parameters difficult. And (4) misspecification noise, i.e., neglected performance measures, can bias statistically determined preference parameters.

The first argument (due to Zelany [156]) is apparent from Figure 7.3.

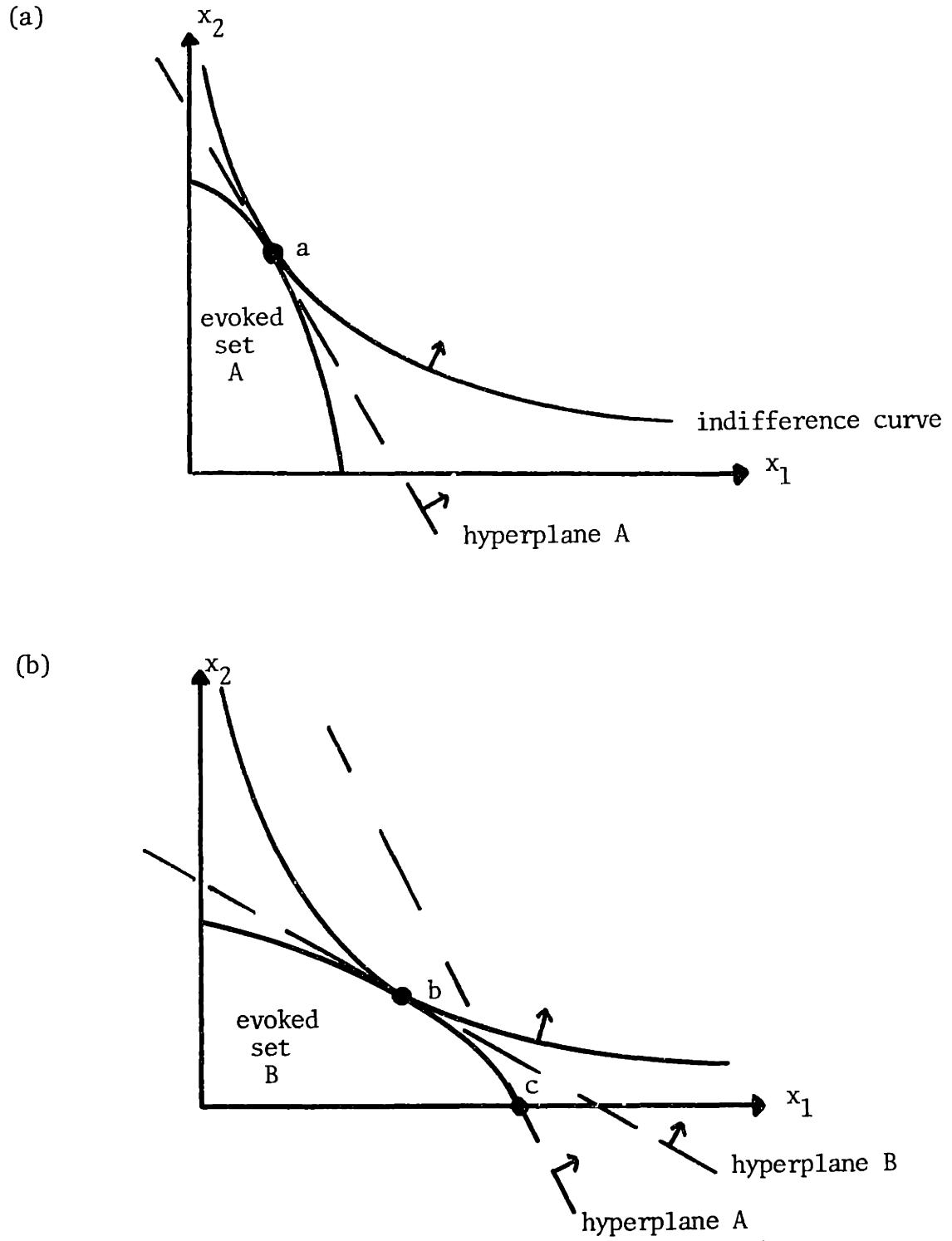


Figure 7.3: Hyperplane Counterexample

Suppose there are only two performance measures, x_1 and x_2 , and suppose the indifference curve for these measures is as shown in Figure 7.3. If the evoked set A is convex, then a linear approximation, hyperplane A, will reveal the same preference as the non-linear compaction function. Suppose now the evoked set changes to evoked set B. Again under convexity a new approximating hyperplane can be found, but note that (1) it is quite different from hyperplane A and (2) that had hyperplane A been used it would have predicted point c as preferred, not point b. This indicates the importance of specifying and testing the function form of the compaction function. With a correct compaction form, statistical techniques are not as sensitive to this problem, but complex forms such as risk averse and quasi-additive make statistical techniques infeasible. Proxy choice enables direct assessment of these more complex forms.

Argument (2) is stronger. In fact it is possible to show (see Table 7.2) that for two individuals an average compaction function can invert everyone's preferences. In other words suppose $c_s(x_{ij}, \lambda_i) = \lambda_{i1} x_{ij1} + \lambda_{i2} x_{ij2} + \lambda_i x_{ij1} x_{ij2}$ is individual i's compaction function. Then an average compaction function is $\bar{c}_s(x_{ij}, \lambda) = \lambda_1 x_{ij1} + \lambda_2 x_{ij2} + \lambda x_{ij1} x_{ij2}$ where $\lambda_1 = \frac{1}{n} \sum_{i=1}^n \lambda_{i1}$, etc., Table 7.2 presents an example where each individual prefers a_1 to a_2 ($a_1 > a_2$), but the average compaction function implies each prefers a_2 to a_1 ($a_2 > a_1$). Note that the numbers are not pathological, in fact the only main difference is that for one individual the performance measures

are complements ($\lambda_i > 0$) and for the other they are substitutes ($\lambda_i < 0$). One way to avoid this is to first abstract segments based on homogeneity of preference, but this can not be done directly if only revealed preference is used. Proxy choice can create enough choices so that individual specific compaction functions can be estimated. A possible solution is to abstract homogeneous segments based on preference parameters determined through proxy choice and then use revealed preference statistical techniques within segments.

Table 7.2. Average Utility Counter Example

Individual 1

$$c_s(x_{ij}, \lambda_1) = \frac{1}{6} x_{1j1} + \frac{1}{3} x_{1j2} + \frac{1}{2} x_{1j1} x_{1j2}$$

Individual 2

$$c_s(x_{ij}, \lambda_2) = \frac{1}{3} x_{2j1} + x_{2j2} - \frac{1}{3} x_{2j1} x_{2j2}$$

Average

$$\bar{c}_s(x_{ij}, \lambda) = \frac{1}{4} x_{ij1} + \frac{2}{3} x_{ij2} + \frac{1}{12} x_{ij1} x_{ij2}$$

$$x_{ij1} \quad x_{ij2} \quad c_s(x_{ij}, \lambda_i) \quad \bar{c}_s(x_{ij}, \lambda)$$

Individual 1

a_1	1/2	1/2	9/24	67/144
a_2	0	1	8/24	96/144
a_1	0	3/4	27/36	72/144
a_2	1/4	2/3	25/36	75/144

Individual 2

Argument (3) is multi-collinearity. For example suppose the alternatives are transportation modes. In any statistical sample we would expect that cost and travel time would be related, thus a statistical technique would suffer from multi-collinearity. Proxy choice allows measurement based on alternatives where all but two of the performance measures are held fixed. If the individual views these two performance measures as preferentially independent of the other measures, then the relative preference parameters can be determined directly. See for example the tradeoff questions (Figure 6.15) used in direct assessment.

The last argument for proxy choice is similar. Each effect can be separately determined. For example with direct assessment tradeoff questions give importances one at a time, lottery questions separately investigate reliability of each performance measure, and corner point questions directly uncover interdependency. In other words, by using proxy choices the rest of the world is held constant and only those effects which are modeled are measured. Noise from misspecification does not enter the preference parameters. It is instead picked up in the probability model.

In summary, there are both advantages and disadvantages to the use of revealed preference and proxy choice. The analyst must select the technique and type of choice which is best for the situation being modeled.

The next section (section 7.2) discusses statistical models which can be used with revealed preference or proxy choice. Section 7.3

presents a technique which uses proxy choice for compaction and then uses revealed preference to "tune" the predictions.

7.2 Utility Maximizing Models

Because of measurement error, specification error, and day-to-day fluctuations in preference consumers will not always choose the alternative with the largest scalar measure of goodness. One approach, due to McFadden [93], is to assume that true utility has a measurable component, the compaction function, and an unobservable component, the error term. A consumer then always maximizes true utility.

I.e. $u_{ij} = u_s(x_{ij}, \lambda_i) = c_s(x_{ij}, \lambda_i) + \epsilon_s(x_{ij}, \lambda_i)$
and

$$p(a_j | c_{i1}, c_{i2}, \dots) = \text{Prob}[u_{ij} = \max_k u_{ik}]$$

Note that $c_s(x_{ij}, \lambda_i)$ is not necessarily a compaction function relative to this model since two individuals with the same compaction vector, c_i , may have different choice probabilities because $\epsilon_s(x_{ij}, \lambda_i)$ is a function of x_{ij} and λ_i .

There are at least two ways to incorporate utility maximization into a probability model; (1) explicitly model the various sources of error and (2) make assumptions about the error term.

7.2.1 Explicit Modeling of Error Sources³

Measurement error enters through noise in the estimates of the perceptions, x_{ij} , and the preferences, λ_i . Suppose that the true

perceptions are distributed according to $p_i(x_{ij})$ and true preference parameters according to $p_i(\lambda_i)$. Unbiased (zero mean) specification error can be modeled without loss of generality by an additive error term as long as the distribution of that error term is conditioned on x_{ij} and λ_i . In other words, $p_i(\epsilon_{ij}|x_{ij}, \lambda_i)$. Non-stationarity in preferences can be incorporated in $p_i(\lambda_i)$ by allowing the parameters of the distribution to be time varying. Putting these explicit errors together, it is possible to write down the integral equations necessary for a utility maximizing model.

First suppose $\underline{x}_i = \{x_{i1}, x_{i2}, \dots\}$ and λ_i are known of certain. Then the only error is $p_i(\epsilon_{ij}|x_{ij}, \lambda_i)$. Thus $\text{Prob}[u_{i1} = \max_j u_{ij}] = \text{Prob}[c_{i1} + \epsilon_{i1} \geq c_{ij} + \epsilon_{ij}, \forall j] = \text{Prob}[\epsilon_{i1} \geq c_{ij} - c_{i1} + \epsilon_{ij}, \text{ for all } j]$.

If $\epsilon_{i1}, \epsilon_{i2}, \dots$ are not independent then this probability must be calculated by the appropriate integration in ϵ_i space. If $\epsilon_{i1}, \epsilon_{i2}, \dots$ are independent then:

$$\text{Prob}[\epsilon_{i1} \geq c_{ij} - c_{i1} + \epsilon_{ij} \text{ for all } j] =$$

$$\int_{\epsilon_{i2}} \dots \int_{\epsilon_{iJ}} \left[\int_{\epsilon_{i1} = \gamma(\epsilon_i)} p_i(\epsilon_{i1}|x_{i1}, \lambda_i) d\epsilon_{i1} \right] p_i(\epsilon_{i2}|x_{i2}, \lambda_i) p_i(\epsilon_{i3}|x_{i3}, \lambda_i) \dots$$

$$p(\epsilon_{iJ}|x_{iJ}, \lambda_i) \cdot d\epsilon_{i2} d\epsilon_{i3} \dots d\epsilon_{iJ}$$

$$\text{where } \gamma(\epsilon_i) = \max_j [c_{ij} - c_{i1} + \epsilon_{ij}].$$

Now to calculate p_{il} it is sufficient to integrate out \underline{x}_i and $\underline{\lambda}_i$. I.e.,

$$p_{il} = \int_{\underline{x}_i} \int_{\underline{\lambda}_i} \text{Prob}[\varepsilon_{il} \geq c_{ij} - c_{il} + \varepsilon_{ij} \text{ for all } j | \underline{x}_i, \underline{\lambda}_i]$$

$$\cdot p_i(\underline{x}_i) p_i(\underline{\lambda}_i) d\underline{x}_i d\underline{\lambda}_i.$$

Of course for any reasonable compaction function, e.g., multiplicative with concave conditional compaction functions, it is much easier to write down the integral equations than to perform the integrations. In fact to the author's knowledge it is an unsolved problem to find any naturally conjugate distribution to anything more complex than a compaction function which is linear in its parameters. The next subsection presents a model which makes simplifying assumptions in order to use the above integral equations.

7.2.2 Random Utility Models

A random utility model as defined by McFadden [94] is a special case of the integral equations of the last section. In McFadden's random utility model the observable portion of the utility, $c_s(x_{ij}, \lambda_i)$, is linear in its parameters, i.e., $c_s(x_{ij}, \lambda_i) = \sum_k \lambda_{ik} z_k(x_{ij}, s_i)$

where the $z_k(x_{ij}, s_i)$ are known functions of the performance measures, x_{ij} , and "socio-economic" variables, s_i , such as income. Such a representation allows for statistical estimation of the λ_{ik} 's, but as with the value functions of section 6.2.7.2, there is no normative theory to specify the form of $z_k(x_{ij}, s_i)$.

Random utility models assume no measurement error for x_{ij} , assume that λ_i is drawn from some probability distribution rather than individually measured, and make assumptions about the distribution of the error terms. The functional form of the probability model is then derived

using the integral equations of section 7.2.1. Two special cases, both due to McFadden, are presented here.

7.2.2.1 Extended Probit Model

Assume that λ_{ik} for all k and $\epsilon(x_{ij}, s_i) = \epsilon_{ij}$ for all i, j are independent normal with $E[\lambda_{ik}] = \lambda_k$, $E[\epsilon_{ij}] = 0$, $E[(\lambda_{ki} - \lambda_k)^2] = \sigma_k^2$, and $E[\epsilon_{ij}^2] = \sigma_0^2$. Then $[u(x_{i1}, \lambda_i), u(x_{i2}, \lambda_i), \dots]$ is multivariate normal with mean $[\sum_k \lambda_k z_k(x_{i1}, s_i), \sum_k \lambda_k z_k(x_{i2}, s_i), \dots]$ and covariance matrix $V_i = \sigma_0^2 I + z_i^T D z_i$, where $I =$ identity, $D =$ diagonal

$(\sigma_1^2, \sigma_2^2, \dots)$, and $z_i =$ the $K \times J$ matrix of $z_k(x_{ij}, s_i)$. Under these conditions the integral equations reduce to an iterated integral of standard normal densities. In the two-choice case, the choice probabilities can be expressed as an extension of the closed form Probit formula, i.e.,

$$p_{i1} = \Phi [\Lambda^T z_i t / \sqrt{t^T V_i t}]$$

where $\Lambda^T = [\lambda_1, \lambda_2, \dots]$, $t = [1, -1]$, $\Phi(\cdot)$ is the cumulative normal distribution, and the superscript T indicates transpose.

7.2.2.2 Multinomial Logit Model

Assume that there is no measurement error in x_{ij} and λ_i , and assume that $\epsilon_s(x_{ij}, \lambda_i) = \epsilon_{ij}$ are independent Weibull distributed,

i.e., $p_{ij}(\epsilon_0) = e^{-\epsilon_0} \exp(-e^{-\epsilon_0})$. Then it can be shown that the choice probabilities are expressed as the closed form multinomial logit model:

$$p_{ij} = \frac{e^{c_{ij}}}{\sum_{l} e^{c_{il}}}$$

where $c_{ij} = c_s(x_{ij}, \lambda_i)$.

7.2.2.3 Discussion and Properties

Probit: The extended probit model recognizes that preference parameters vary across the population, but it does not assess them idiosyncratically. Unfortunately without individual specific assessment of the λ_{ik} 's, there is no easy way to estimate $E[(\lambda_{ik} - \lambda_k)^2]$. This model does point out how "simple" assumptions about the errors can cascade into a very complex and analytically untractable model. It emphasizes the importance of finding analytically tractable assumptions to couple with the integral equations of section 7.2.1.

Logit: This popular probability of choice model has been used in countless applications in transportation, bioassay and marketing. In normal use the compaction function is assumed linear in its parameters (McFadden [94]), i.e., $c_s(x_{ij}, \lambda_i) = \sum_k \lambda_k z_k(x_{ij}, s_i)$, and the parameters are statistically estimated via maximum-likelihood techniques. This formulation requires a homogeneity assumption on the λ_k 's, and lacks a

normative theory to specify the form of the $z_k(x_{ij}, s_i)$'s.

The logit model is not particularly suited for direct assessment because with the multinomial logit model once the c_{ij} 's are known the p_{ij} 's are completely specified leaving no degrees of freedom with which to calibrate the probability model. A more reasonable model would use proxy choice to assess $c_s(x_{ij}, \lambda_i)$, but would leave "tuning" parameters which would make use of observations on choice. Such a model is discussed in section 7.3.

Properties of the logit model: In the language of section 7.1, the logit model is symmetric, assumes independence among individuals, and is assumed to encompass the individuals in the segment. If $c_s(x_{ij}, \lambda_i)$ is directly assessed, the logit model is interpolative, if it is calibrated from a linear-in-the-parameters form based on revealed preference, it is extrapolative. Similarly direct assessment implies extendability by adding another term to the summation in the denominator, but statistical, linear-in-the-parameters cannot guarantee extendability even if $z_k(x_{ij}, s_i)$ are not alternative specific. ($z_k(x_{ij}, s_i)$ is alternative specific if its functional form depends on the alternative under consideration.) See Figure 7.4 for hyperplane counterexample.

Dependence on choice set: Implicit in the functional form of the logit model is the strong assumption of independence of irrelevant alternatives in the sense of Luce [92]. I.e.,

$$\frac{p(a_1 | c_{i1}, c_{i2}, \dots, c_{iJ})}{p(a_2 | c_{i1}, c_{i2}, \dots, c_{iJ})} = \frac{e^{c_{i1}} / \sum_{j=1}^J e^{c_{ij}}}{e^{c_{i2}} / \sum_{j=1}^J e^{c_{ij}}}$$

$$= \frac{e^{c_{i1}} / (e^{c_{i1}} + e^{c_{i2}})}{e^{c_{i2}} / (e^{c_{i1}} + e^{c_{i2}})} = \frac{p(a_1 | c_{i1}, c_{i2})}{p(a_2 | c_{i1}, c_{i2})}$$

As discussed earlier, this implies that hierarchical choices must be explicitly modeled.

7.2.3 Aggregate Utility Maximizing Models

The integral equations used to explicitly model error sources can be viewed as aggregate equations. I.e., one can have the firm judgmentally set $p(x_{ij})$ for their product, alternative a_j , and judgmentally estimate $p(\lambda_i)$ as a distribution across consumers. The integral equations then produce estimates of market share rather than individual choice probabilities. Such a model could prove useful as an initial screening model before any complex analysis is done. This model will not be discussed here, but the interested reader is referred to Keeney and Lilien [74a] for a more complete description.

7.2.4 Conclusion of Utility Maximizing Models

By assuming that there is a true utility which consumers maximize, it is possible to explicitly model the various forms of errors. This leads to a set of general integral equations. With complex compaction forms these equations are untractable, but under major simplifying assumptions analytically tractable approximations can be derived. The most popular is the logit model.

The linear-in-the-parameters logit model is quite useful when used with statistical calibration of revealed preference on large data

sets, but it suffers from the theoretically bothersome problems of extrapolation and unknown extendability and from the implicit assumption of Luce's strong independence of irrelevant alternatives.

When used with direct assessment, the logit model is interpolative over the entire range of performance measures and is extendable to new alternatives, but there are no degrees of freedom left to "tune" the model based on observations of choice.

The next section introduces a new empirical Bayesian model which was especially formulated to be compatible with direct assessment. It assumes only encompassment of a valid compaction function and implicitly models error sources by using empirical observations of past behavior to calibrate a model which "tunes" rank order phenomena with the cardinal values of the compaction vector, c_i .

7.3 Empirical Bayesian Model

If there were no uncertainty, then each individual would always choose the alternative with the largest scalar measure of goodness. Hopefully a successful assessment of an individual's compaction function will "explain" as much as possible and the resulting choice model will approximate the certainty decision rule. Therefore it is desirable that a probability of choice model make use of the rank order phenomena inherent in the certainty rule. There are many reasons to expect uncertainty, but one might expect that the uncertainty is related to the relative cardinal values of the scalar measures of goodness in

the evoked set. For example with only two alternatives it is expected that $p(a_1 | c_{i1} = 10, c_{i2} = 1) > p(a_1 | c_{i1} = 10, c_{i2} = 9)$.

The basic idea behind the empirical Bayesian model is to "tune" predictions based on the ordinal rankings of the compaction values with their cardinal properties. This is done by exploiting the tautology of Bayes Theorem to derive a feasible empirical probability of choice model, $p(a_j | c_{i1}, c_{i2}, \dots, c_{iJ})$, based on observing consumers in real choice situations.

7.3.1 Derivation of the Empirical Bayesian Model

Ideally we could observe each individual making repetitive choices from alternative sets with compaction vectors, \underline{c}_i , over the entire range of \underline{c}_i . The choice model, $p_i(a_j | \underline{c}_i)$ could then be directly fit. More likely we will have one observation for each individual making a choice from an alternative set with a particular corresponding compaction vector. Since he will choose a single⁴ alternative the observations will be $p_i(a_j | \underline{c}_i) = 1.0$ or $p_i(a_j | \underline{c}_i) = 0.0$. Thus it is nearly impossible to directly fit $p_i(a_j | \underline{c}_i)$.

One of the primary motivations behind direct assessment is that individuals have varied perceptions and preferences. Indeed direct assessment can be viewed as an attempt to make a population homogeneous by explicitly incorporating individual differences into a compaction function $c_s(x_{ij}, \lambda_i)$. If this is successful, the probability model will encompass the population segment. (Encompassment means that any two individuals with the same compaction vector, \underline{c}_i , will have the same vector of probabilities, $p_i = [p(a_1 | \underline{c}_i), p(a_2 | \underline{c}_i), \dots, p(a_J | \underline{c}_i)]$.)

This hypothesis of encompassment allows the i subscript to be dropped from the probability model, $p(a_j | \underline{c}_i)$, and, because observations across individuals can now be used for calibration, it opens the door to empirical fitting of $p(a_j | \underline{c}_i)$.

Even with encompassment $p(a_j | \underline{c}_i)$ cannot be fit directly, because such a model would neither be symmetric nor extendable to new alternatives. Furthermore it would neglect the rank order phenomena. Instead consider the following transformation.

Definition 7.4: Let e_1 be the event than an individual chooses the alternative with the maximum scalar measure of goodness. Let e_2 be the event that an individual chooses the alternative with the 2nd largest scalar measure of goodness. Similarly define e_j , for $j=3,4,\dots$. Call these events the rank order events.

Definition 7.5: Let cr_{i1} be individual i 's maximum scalar measure of goodness. Similarly let cr_{ij} be the j^{th} largest scalar measure of goodness for $j=2,3,\dots$. Call the vector $\underline{cr}_i = (cr_{i1}, cr_{i2}, \dots, cr_{iJ})$ the ranked compaction vector and call the reordering of \underline{c}_i to obtain \underline{cr}_i the rank order transformation.

Definition 7.6: Let $p(e_1 | \underline{cr}_i)$ be the probability that individual i chooses the alternative with the largest scalar measure of goodness given the ranked compaction vector \underline{cr}_i . Similarly define $p(e_j | \underline{cr}_i)$ for $j=2,3,\dots$. Call the model which yields these probabilities the ranked probability model, and define $\underline{pr}_i = [p(e_1 | \underline{cr}_i), p(e_2 | \underline{cr}_i), \dots, p(e_J | \underline{cr}_i)]$.

Clearly the simple probabilities, p_i , can be readily obtained from the ranked probabilities, pr_i , thus if $p(e_j | \underline{cr}_i)$ can be calibrated so implicitly can $p(a_j | \underline{c}_i)$ be calibrated. A direct calibration of $p(e_j | \underline{cr}_i)$ has many properties such as symmetry and extendability which $p(a_j | \underline{c}_i)$ may not. (See section 7.3.2 for further discussion of special properties.)

Now, if enough individuals happen to have the same ranked compaction vectors, then $p(e_j | \underline{cr}_i)$ for all j could be (1) observed over the range of \underline{cr}_i , and (2) parameterized, and (3) calibrated. Unfortunately in any reasonable sample this will be rare and for any particular value of \underline{cr}_i either $p(e_j | \underline{cr}_i) = 1.0$ or $p(e_j | \underline{cr}_i) = 0.0$. Consider instead the use of Bayes Theorem:

$$p(e_1 | \underline{cr}_i) = \frac{p(\underline{cr}_i | e_1)p(e_1)}{\sum_{j=1}^J p(\underline{cr}_i | e_j)p(e_j)}$$

Empirically most evoked sets are small, e.g., a mean of about three alternatives (See Table 7.1) and in the only empirical case where this Bayesian model was used $p(e_j) = 0$ for $j > 3$. This suggests that one might expect that the important effects will be carried in the first m values in any ranked compaction vector. Suppose that the smallest $(J-m)$ compaction values are ignored, i.e., $p(e_j) = 0$ for $j = m+1, m+2, \dots, J$, and that a multivariate probability distribution with n parameters per variate⁵ is fit to $p(\underline{cr}_i | e_j)$ for $j=1, 2, \dots, m$. Then with N observations there are the order of (N/m) observations to

fit ($m \cdot n$) parameters. Typical numbers are $N=200$ and $m=3$ and $n=2^6$ yielding the order of 67 observations for 6 parameters. Thus in many cases there will be enough degrees of freedom to fit the necessary $p(\underline{cr}_i | e_j)$'s. Note that rather than having an equal number of observations in each cell, (i.e., N/m), it is likely that there will be more for the rank order events with small j , e.g., more for e_1 than for e_2 . Rather than hurt the calibration process, this effect helps because it means that there are more observations for the important events. To see this examine the Bayesian equation: if $p(e_j)$ is small then $p(e_1 | \underline{cr}_i)$ is not very sensitive to $p(e_j | \underline{cr}_i)$. On the other hand $p(e_j | \underline{cr}_i)$ will be sensitive, but the overall "loss" from this will be small. (See Chapter 10 on testing). Thus we can posit that $p(\underline{cr}_i | e_j)$ can be empirically determined and used with little error.

The next step is to determine $p(e_j)$. If there are N observations and if the unconditioned distribution $p(\underline{cr}_i)$ does not change then $p(e_j)$ can be estimated by n_j/N where n_j = the number of individuals who choose the alternative with the j^{th} largest scalar measure of goodness. If $p(\underline{cr}_i)$ does change then $p(e_j)$ will change correspondingly. At first glance this is a significant problem since any innovation will actually try to change $p(\underline{cr}_i)$. Fortunately $p(e_j)$ and $p(\underline{cr}_i | e_j)$ for all j will all change simultaneously in such a way as to keep the derived model $p(e_j | \underline{cr}_i)$ valid over the new range. In other words a calibration based on Bayes Theorem is a consistent estimate of a "true" probability model. This fact will be proven in section 7.3.2.

To summarize, if all individual differences can be incorporated

into the compaction function, then $p(a_j | c_i)$ can be determined empirically by first determining the ranked probability model, $p(e_j | \underline{c}_i)$ as follows:

- 1) Observe the distribution of ranked compaction vectors for each rank order event.
- 2) Empirically observe m and parameterize $p(\underline{c}_i | e_j)$ and statistically determine the parameters for $j=1, 2, \dots, m$, where \underline{c}_i is truncated after m terms, $m \leq J$.
- 3) Estimate $p(e_j)$ by n_j/N for $j=1, 2, \dots, m$.
- 4) Using Bayes theorem and multiplying through by N :

$$p(e_j | \underline{c}_i) \simeq \frac{p(\underline{c}_i | e_j) n_j}{\sum_{\ell=1}^m n_{\ell} p(\underline{c}_i | e_{\ell})} \quad \text{for } j=1, 2, \dots, m.$$

$$p(e_j | \underline{c}_i) = 0 \quad \text{for } j=m+1, m+2, \dots, J.$$

7.3.2. Desirable Properties of the Empirical Bayesian Model

The Bayesian model has the properties of symmetry, extendability, predictability, weak simple scalability, independence from zero probability alternatives, and independence from (irrelevant) alternatives with ordinal rankings greater than m . It is sensitive to identical substitutes and therefore hierarchical decisions must be explicitly modeled. Finally it assumes independence among individuals and that the probability model encompasses the population segment. Explanations follow:

Symmetry: By the very nature of the rank order transformation, if the compaction values of two alternatives are switched with the others fixed, c_i will change but cr_i will not. Thus pr_i will not change and only the assignment of probabilities will change by switching.

Extendability: If a new alternative is added to an individual's evoked set then cr_i will change, but because $p(e_j | cr_i)$ is not alternative specific and because there is a cutoff for $j > m$, the new probabilities can be calculated. This of course means that the alternative which had been ranked m^{th} may now be ranked $m+1^{\text{st}}$ and have zero probability of being chosen.

Interpolative: If $c_s(x_{ij}, \lambda_i)$ is valid over the entire range of X , then if $p(e_j | cr_i)$ is interpolative over the range of c_i , the probability model will be interpolative over the entire range of X .

Sensitivity to Calibrating Observations: As was stated earlier, the calibration technique depends on $p(e_j)$ which is sensitive to $p(cr_i)$. It turns out, as will be shown in the next theorem, that the probability model itself, $p(e_j | cr_i)$, is insensitive to $p(cr_i)$.

To show that the probability model is insensitive to $p(cr_i)$ it is sufficient to show it is insensitive to the particular observations chosen for calibration as long as the number and variations of those observations are sufficient to calibrate the parameters for $p(cr_i | e_j)$. For ease of presentation there are two simplifications in theorem 7.1. First, it is stated for discrete distributions of $p(cr_i | e_j)$; the extension to the continuous case is with incremental arguments when

$p(\underline{cr}_i | e_j)$ is well behaved. Second, the proof is for $m=2$; the extension to arbitrary m is straightforward but notationally complex.

The idea behind theorem 7.1 is that there is some "true" probability model, $p^T(e_j | \underline{cr}_i)$. What we would like to show is that two separate sample calibrations of $p(e_j | \underline{cr}_i)$ yield $p^T(e_j | \underline{cr}_i)$ even if $p(\underline{cr}_i)$ is quite different in both calibrations. Since we would like to do this independently of the calibration procedure used for $p(\underline{cr}_i | e_j)$ we will need to assume that the estimate $\hat{p}(\underline{cr}_i | e_j)$ converges to the true value $p^T(\underline{cr}_i | e_j)$ for a given $p(\underline{cr}_i)$. In other words that the estimation of $p(\underline{cr}_i | e_j)$ is consistent. Based on this it is possible to show that the Bayesian calibration of $p(e_j | \underline{cr}_i)$ maintains this consistency and is in fact independent of the distribution of observations, $p(\underline{cr}_i)$.

Theorem 7.1: (Consistency and independence of observation distribution.)
Suppose that $p(\underline{cr}_i)$ is discretely valued. Suppose that two separate samples are drawn for calibration. Suppose that the number of observations, N , in the first sample is sufficiently large and varied such that any estimate, $\hat{p}^N(\underline{cr}_i | e_j)$, of $p^N(\underline{cr}_i | e_j)$ is consistent for $j=1,2$. ($p^N(\underline{cr}_i | e_j)$ is the true distribution given $p^N(\underline{cr}_i)$ for the first sample.) I.e., $\hat{p}^N(\underline{cr}_i | e_j) \rightarrow p^N(\underline{cr}_i | e_j)$. Let M be the number of observations in the second sample and suppose that $\hat{p}^M(\underline{cr}_i | e_j) \rightarrow p^M(\underline{cr}_i | e_j)$. Then Bayesian models calibrated on either the N samples or the M samples are consistent and independent of $p^N(\underline{cr}_i)$ or $p^M(\underline{cr}_i)$.

Proof: Relax the integrality of the observations since N is assumed sufficiently large and varied. Suppose $p^T(e_j | \underline{cr}_i)$ is the true model.

Consider the cell structure in Figure 7-4a. The expected number of observations in the qr^{th} cell for the event e_j will be the expected number of occurrences of $c_{i1}=q$ and $c_{i2}=r$ times the true probability of event e_j conditioned on $c_{i1}=q$ and $c_{i2}=r$. (Use state indicator random variables.) Notationally $E[\hat{n}_{qr}^j] = [N p^N(q, r)] \cdot [p^T(e_j | q, r)]$. Since $\hat{p}(\underline{cr}_i | e_j)$ is consistent, $\hat{p}^N(q, r | e_j) \rightarrow p^N(q, r | e_j)$. But $p^N(q, r | e_j) = E(\hat{n}_{qr}^j)$, thus $\hat{p}^N(q, r | e_j) \rightarrow N p^T(e_j | q, r) p^N(q, r)$. For notational simplicity write $\hat{n}_{qr}^j = p^N(q, r | e_j)$. Consider now the cell structure in Figure 7.4b. Similar arguments and notation show that $\hat{m}_{qr}^j \rightarrow M p^T(e_j | q, r) p^M(q, r)$. Now by Bayes Theorem $\hat{p}^N(e_j | q, r)$ is given by:

$$\begin{aligned}
 \hat{p}^N(e_j | q, r) &= \frac{\hat{p}^N(q, r | e_j) \hat{p}^N(e_j)}{\hat{p}^N(q, r | e_1) \hat{p}^N(e_1) + \hat{p}^N(q, r | e_2) \hat{p}^N(e_2)} \\
 &= \frac{\frac{\hat{n}_{qr}^j}{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j}}{\frac{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j}{\sum_{j=1}^2 \sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j}} \\
 &\quad \times \frac{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j}{\sum_{j=1}^2 \sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j} \\
 &= \frac{\frac{\hat{n}_{qr}^1}{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^1}}{\frac{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^1}{\sum_{j=1}^2 \sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j}} + \frac{\frac{\hat{n}_{qr}^2}{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^2}}{\frac{\sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^2}{\sum_{j=1}^2 \sum_{s=1}^2 \sum_{t=1}^2 \hat{n}_{st}^j}}
 \end{aligned}$$

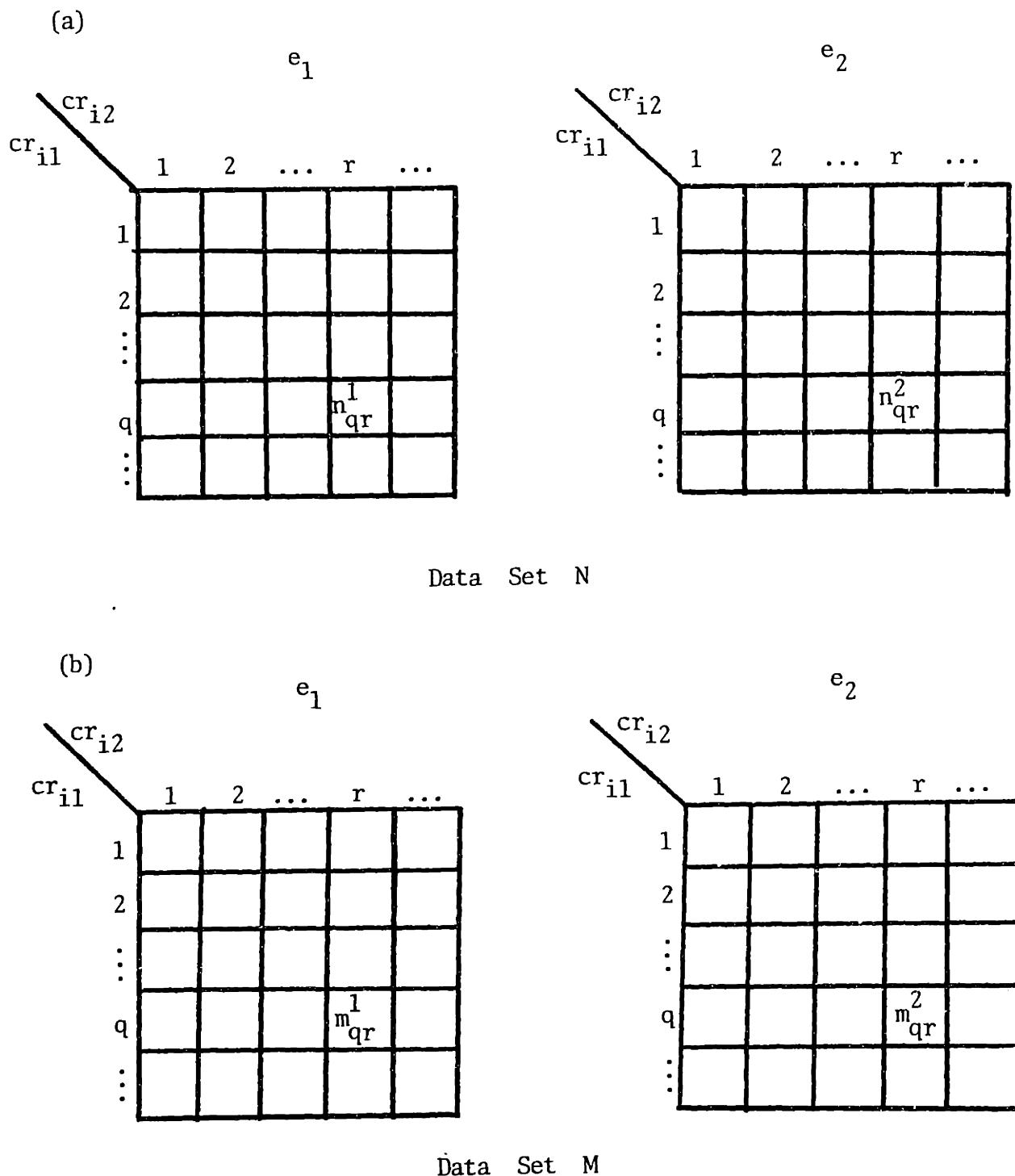


Figure 7.4: Cell Structure

$$= \hat{n}_{qr}^j / (\hat{n}_{qr}^1 + \hat{n}_{qr}^2)$$

$$\rightarrow \frac{N p^T(e_j | q, r) p^N(q, r)}{N p^T(e_1 | q, r) p^N(q, r) + N p^T(e_2 | q, r) p^N(q, r)}$$

$$= p^T(e_j | q, r)$$

Similarly,

$$p^M(e_j | q, r) = \hat{m}_{qr}^j / (\hat{m}_{qr}^1 + \hat{m}_{qr}^2)$$

$$\rightarrow \frac{M p^T(e_j | q, r) p^M(q, r)}{M p^T(e_1 | q, r) p^M(q, r) + M p^T(e_2 | q, r) p^M(q, r)}$$

$$= p^T(e_j | q, r)$$

Therefore $\hat{p}^N(e_j | q, r)$ and $\hat{p}^M(e_j | q, r)$ converge to $p^T(e_j | q, r)$ but q , r , and j were arbitrary, thus $\hat{p}^N(e_j | \underline{cr}_i)$ and $\hat{p}^M(e_j | \underline{cr}_i)$ converge to $p^T(e_j | \underline{cr}_i)$ for $j=1,2$.

Weak Simple Scalability: The first part of simple scalability is easy, $p(a_j | \underline{c}_i)$ depends only on \underline{c}_i because of the rank order transformation. The second part is true because by examining the proof of theorem 7.1 it is seen that $p^N(e_j | \underline{cr}_i) \rightarrow p^T(e_j | \underline{cr}_i)$, thus if the true probabilities satisfy simple scalability then the large sample calibration will also. But if $c_s(x_{ij}, \lambda_i)$ is a valid compaction function, then

$\bar{p}^T(e_j | \underline{cr}_i)$ will be monotonically non-decreasing in cr_{ij} and monotonically non-increasing in cr_{ik} for $k \neq j$.

Independence from Zero Probability Alternatives: If $p(e_j | \underline{cr}_i) = 0$ then either $j > m$ in which case $p(e_j) = 0$ or \underline{cr}_i implies it is absurd for i to choose e_j in which case $p(\underline{cr}_i | e_j) = 0$. Either way the product $p(\underline{cr}_i | e_j)p(e_j) = 0$ and by Bayes Theorem it will not effect $p(e_\ell | \underline{cr}_i)$ $\ell \neq j$.

Independence from Irrelevant Alternatives: If the rank of an alternative is greater than m , it will not effect the probabilities for the other alternatives in the evoked set. If the rank of an alternative is less than or equal to m it will effect the probabilities, which is quite reasonable, but by the same arguments used for weak simple scalability it will not effect the stochastic preference ordering. Even if the size of the evoked set is less than m , the ordering will be unchanged.

Independence from Identical Substitutes: The rank order transformation is extremely sensitive to identical substitutes. For example suppose $c_i = \{c_{i1} = 10, c_{i2} = 3, c_{i3} = 1\}$ and $m = 3$. Now suppose an identical substitute to a_1 is added to the evoked set yielding $c_i = \{c_{i1} = 10, c_{i2} = 3, c_{i3} = 1, c_{i4} = 10\}$. Since $m=3$, the rank order transformation will cause $p(a_3 | 10, 3, 1, 10) = 0$ whereas before $p(a_j | 10, 3, 1) \neq 0$. Thus we must concern ourselves with explicit modeling of the hierarchical nature of the choice situation.

Independence among individuals: The model assumes independence among individuals.

Encompassment: As discussed in the derivation, the Bayesian model assumes that all individual differences in perceptions and preferences are accounted for in the compaction function and that once the compaction vector, c_i , is known, the probability model conditioned on c_i encompasses the population segment.

7.3.3 Heuristics for Transforming the Ranked Compaction Vector

Compaction functions are unique only up to a positive linear transformation, thus the encompassment assumption in the Bayesian derivation implicitly makes interpersonal comparisons of compaction vectors. For the cardinal tuning if this assumption is not made, real choice cannot be used for calibration because there are not enough degrees of freedom to fit a model. Thus empirically if individual specific assessments are made some heuristic attempt must be made to rationalize this assumption. This section presents a series of intuitively pleasing heuristics to make interpersonal comparison. Note that the ordinal nature of the compaction function is insensitive to the positive linear transformation.

Since the Bayesian model deals mainly with the first m ranked compaction values, the following normalizations will also.

Common endpoints: Select the same maximum and minimum performance measures for each individual and set $c_s(x_{\min}, \lambda_i) = 0$ and $c_s(x_{\max}, \lambda_i) = 1.0$.

Share of total with fixed lower endpoint: Select the same minimum performance measures and set $c_s(x_{\min}, \lambda_i) = 0$. Then compute the share:

$$cr_{ij} \leftarrow \frac{cr_{ij}}{\sum_{\ell=1}^m cr_{i\ell}}$$

Note that this normalization is not sensitive to the upper endpoint because once the lower endpoint is chosen all the upper endpoint does is fix the positive multiplicative constant which cancels when the share is computed. It does make the behavioral assumption that share is an important effect. As an indication of how strong this assumption can be, notice that if $cr_{ij} \rightarrow e^{cr_{ij}}$ then this share is the multinomial logit. Remember that these transformations are done before the probability model is calibrated. Calibration and prediction are then done relative to the transformed goodness measures.

Relative value with fixed upper endpoint: Select the same maximum performance measures for each individual and set $c_s(x_{\max}, \lambda_i) = 1.0$. Then compute the compaction values relative to the m^{th} ranked alternative:

$$cr_{ij} \leftarrow cr_{ij} - cr_{im}$$

Note that this normalization is not sensitive to the lower endpoint because once the upper endpoint is chosen all the lower endpoint does is fix the additive constant which cancels when relative values are computed. It does make \underline{cr}_{ij} very sensitive to \underline{cr}_{im} which could imply a strong influence by irrelevant alternatives.

Share of relative value: Compute the share relative to the m^{th} ranked alternative:

$$\underline{cr}_{ij} \leftarrow \frac{\underline{cr}_{ij} - \underline{cr}_{im}}{\sum_{\ell} (\underline{cr}_{i\ell} - \underline{cr}_{im})}$$

Note that this normalization is sensitive to neither the upper or lower endpoints but is extremely sensitive to the characteristics of the m^{th} ranked alternative.

First ranked alternative: It is intuitively more satisfying to have a heuristic normalization that is sensitive to the alternative with the largest scalar measure of goodness than to have one that is sensitive to the alternative with the smallest value. This heuristic uses share of relative deviaton from the first ranked alternative as a measure of dissatisfaction. A negative linear transformation is then taken to regain the property of preference:

$$\underline{cr}_{ij} \leftarrow 1 - \frac{\underline{cr}_{il} - \underline{cr}_{ij}}{\sum_{\ell} (\underline{cr}_{il} - \underline{cr}_{i\ell})}$$

Note that this normalization is sensitive to neither the upper or lower endpoints but is of course sensitive to the characteristics of the first ranked alternative.

Choosing among heuristics: This list is not exhaustive but hopefully is sufficient to indicate the issues. Since these are heuristics, none can be selected a priori as best. Instead some should be eliminated based on the judgement of the analyst and the remainder tested with the testing technique of Chapter 10. The compaction assessment for the MIT HMO (Section 6.4) uses common endpoints. The empirical example in section 7.3.5 uses share with fixed lower endpoint.

7.3.4 Two Alternative Logit Equations as an Empirical Bayesian Model

The popular logit model is usually derived as a random utility model (section 7.2.2.2) but it can also be derived as a special case of the empirical Bayesian model. Note though that the Bayesian model deals with the rank order events whereas the logit model can be alternative specific, none-the-less under very special assumptions the two-alternative logit equation emerges as a special case. (This derivation is similar to that of Quarmby [118]). Continue with the notation of section 7.3.1.

Assumption 1: Heuristic normalization via relative value:

$$cr_i = cr_{i1} - cr_{i2}$$

Note: $p(e_2 | cr_i) = 1 - p(e_1 | cr_i)$.

Assumption 2: $p(\text{cr}_i | e_1)$ and $p(\text{cr}_i | e_2)$ are normally distributed with means $\bar{\text{cr}}_1$ and $\bar{\text{cr}}_2$ respectively.

Assumption 3: Homoscedasticity, i.e., $p(\text{cr}_i | e_1)$ and $p(\text{cr}_i | e_2)$ have the same associated variance, σ^2 .

Using the Bayesian equation and substituting the normal density for $p(\text{cr}_i | e_j)$ yields:

$$\begin{aligned}
 p(e_1 | \text{cr}_i) &= \frac{n_1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\text{cr}_i - \bar{\text{cr}}_1)^2/2\sigma^2}}{n_1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\text{cr}_i - \bar{\text{cr}}_1)^2/2\sigma^2} + n_2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\text{cr}_i - \bar{\text{cr}}_2)^2/2\sigma^2}} \\
 &= \frac{e^{\log n_1} e^{-\text{cr}_i^2/2\sigma^2} e^{(2\text{cr}_i \bar{\text{cr}}_1 + \bar{\text{cr}}_1^2)/2\sigma^2}}{e^{\log n_1} e^{-\text{cr}_i^2/2\sigma^2} e^{(2\text{cr}_i \bar{\text{cr}}_1 + \bar{\text{cr}}_1^2)/2\sigma^2} + e^{\log n_2} e^{-\text{cr}_i^2/2\sigma^2} e^{(2\text{cr}_i \bar{\text{cr}}_2 + \bar{\text{cr}}_2^2)/2\sigma^2}} \\
 &= \frac{1}{1 + e^{a-b\text{cr}_i}} \quad = \quad \frac{1}{1 + e^{a-b(\text{cr}_{i1} - \text{cr}_{i2})}}
 \end{aligned}$$

where

$$a = \log \frac{n_2}{n_1} + (\bar{\text{cr}}_2^2 - \bar{\text{cr}}_1^2)/2\sigma^2$$

$$b = (\bar{\text{cr}}_1 - \bar{\text{cr}}_2)/\sigma^2, \quad \text{note } \bar{\text{cr}}_1 \geq \bar{\text{cr}}_2$$

Examining the assumptions necessary to obtain this result we see that the two strongest are that the relative values matter and that the distributions $p(\underline{cr}_i | e_j)$ are homoscedastic. The Bayesian model allows both to be relaxed and therefore allows more flexible tuning.

A reasonable extension to the multiple choice case is to assume that $p(\underline{cr}_i | e_j)$ is multivariate normal and homoscedastic for all j and that $\text{cov}(\underline{cr}_{ij}, \underline{cr}_{il} | e_k) = 0$ for all k , for all j , and for all $l \neq j$ with $\text{var}(\underline{cr}_{ij} | e_k) = \sigma^2$ for all j and k . Substituting these assumptions into the Bayesian model does not yield a simple multinomial logit formulation but instead yields:

$$p(e_j | \underline{cr}_i) = \frac{1}{1 + \sum_{l \neq j} e^{\gamma_{lj}}}$$

with

$$\gamma_{lj} = \ln(n_l/n_j) - (1/\sigma^2) \sum_k \underline{cr}_{ik} (\bar{cr}_{jl} - \bar{cr}_{jk}) + \sum_k (\bar{cr}_{lk}^2 - \bar{cr}_{jk}^2)$$

where $\bar{cr}_{lk} = E[\underline{cr}_{ik} | e_l]$

7.3.5 An Empirical Example: Aerosol Deodorants

The empirical Bayesian model was formulated to be compatible with direct assessment of compaction functions, but it can be used with other cardinal measures. In this example the data is obtained from constant sum paired comparison questions where the individual was asked to allocate 11 chips between two alternatives according to his preferences for those

alternatives. This data was obtained for every possible pair of alternatives in each individual's evoked set and was statistically (Torgenson [141]) reduced to shares of chips. I.e., estimates of the share of chips an individual would have placed on each alternative had he allocated them all simultaneously. The choice situation was actual choice in a simulated retail store environment. Thus the Bayesian model is to link the constant sum preferences (chips) to actual choice. (This data was collected by Glen Urban and Alvin Silk and is part of the data used in ASSESSOR [132].)

7.3.5.1 Calibration

Step 1: Estimate $p(e_j)$. Out of 178 observations 147 choose their first preference, 27 choose their second preference, and 4 choose their third preference. No one chose any higher than third preference. (Ties in chips were allocated 50-50 or 33-33-33 between categories.) Thus $p(e_1) = .83$, $p(e_2) = .15$, $p(e_3) = .02$, and $p(e_j) = 0$ for $j > 3$. Therefore $m = 3$.

Step 2: Calibrate separate models for the following two cases: (1) evoked set size equals two, (2) evoked set size is greater than two. What is reported here is the second case, evoked set size greater than two, the first case being similar but simpler and has a smaller sample size.

Step 3: Since $m = 3$ assume that those products ranked greater than 3 have no effect and renormalize the share of chips as if only the first,

second, and third ranked alternatives are in the evoked set. I.e.:

$$f_{i1} = cr_{i1}/(cr_{i1} + cr_{i2} + cr_{i3})$$

$$f_{i2} = cr_{i2}/(cr_{i1} + cr_{i2} + cr_{i3})$$

$$f_{i3} = cr_{i3}/(cr_{i1} + cr_{i2} + cr_{i3})$$

Note that to calculate $p(e_j | cr_i)$ it is sufficient to know $p(e_j | f_{i1}, f_{i2})$ since $m = 3$ and $f_{i3} = 1 - f_{i1} - f_{i2}$. This normalization implicitly makes some behavioral assumptions. See section 7.3.3.

Step 4: Enter an exploratory plotting stage to estimate the correct forms for $p(f_{i1}, f_{i2} | e_j)$. In this case the range of (f_{i1}, f_{i2}) was discretized and histograms of various marginal and conditional frequency distributions were plotted. Exponential smoothing was used to recover smooth densities from the histograms. Plotting marginal distributions for $p(f_{i1} | e_j)$ and conditional distributions for $p(f_{i2} | e_j, f_{i1})$ revealed that the conditional distribution for f_{i2} was not independent of f_{i1} . An alternative behavioral assumption is that choice depends on both the share for the largest alternative, f_{i1} , and the share for the remaining chips that are allocated to the second largest alternative, i.e., $f_{i2}/(f_{i2} + f_{i3}) = f_{i2}/(1 - f_{i1})$. When the conditional distributions for $p[f_{i2}/(1 - f_{i1}) | e_j, f_{i1}]$ were plotted, they seemed to indicate that these were indeed roughly independent effects. Preliminary histograms of $p(f_{i1} | e_j)$ appear in Figure 7.5; smoothed histograms of $p(f_{i1} | e_j)$ appear in Figure 7.6; and smoothed histograms of $p[f_{i2}/(1 - f_{i1}) | e_j, f_{i1}]$ appear in Figure 7.7 for two

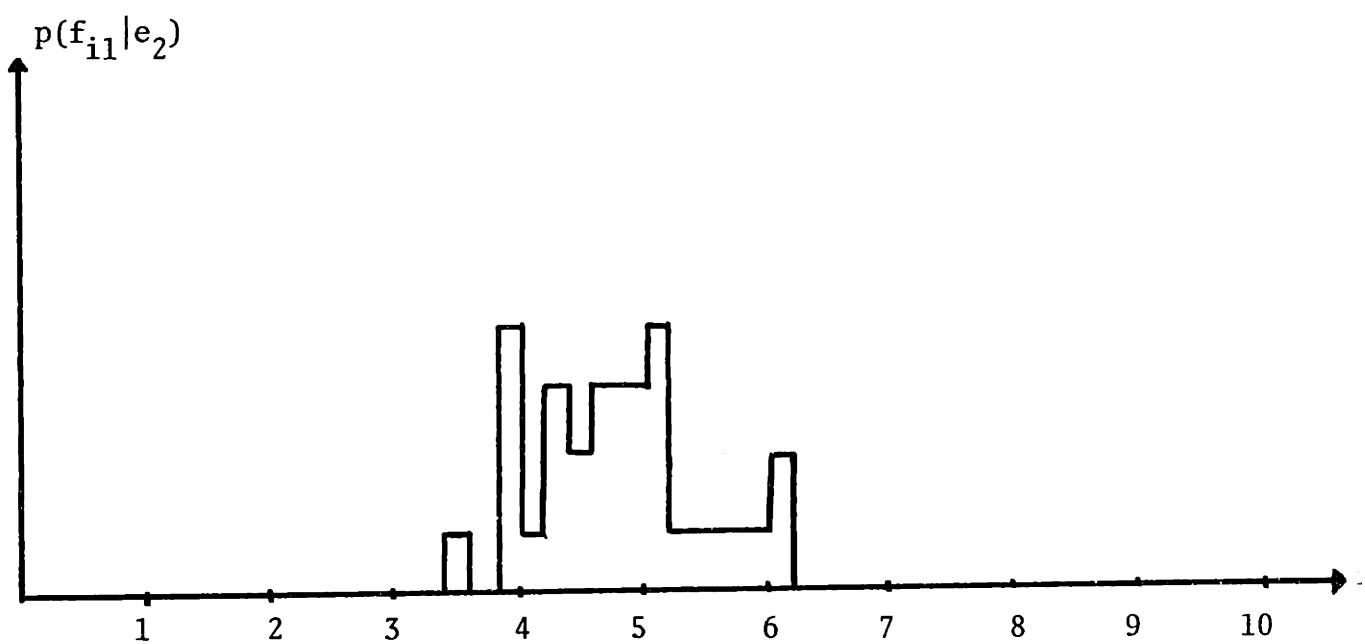
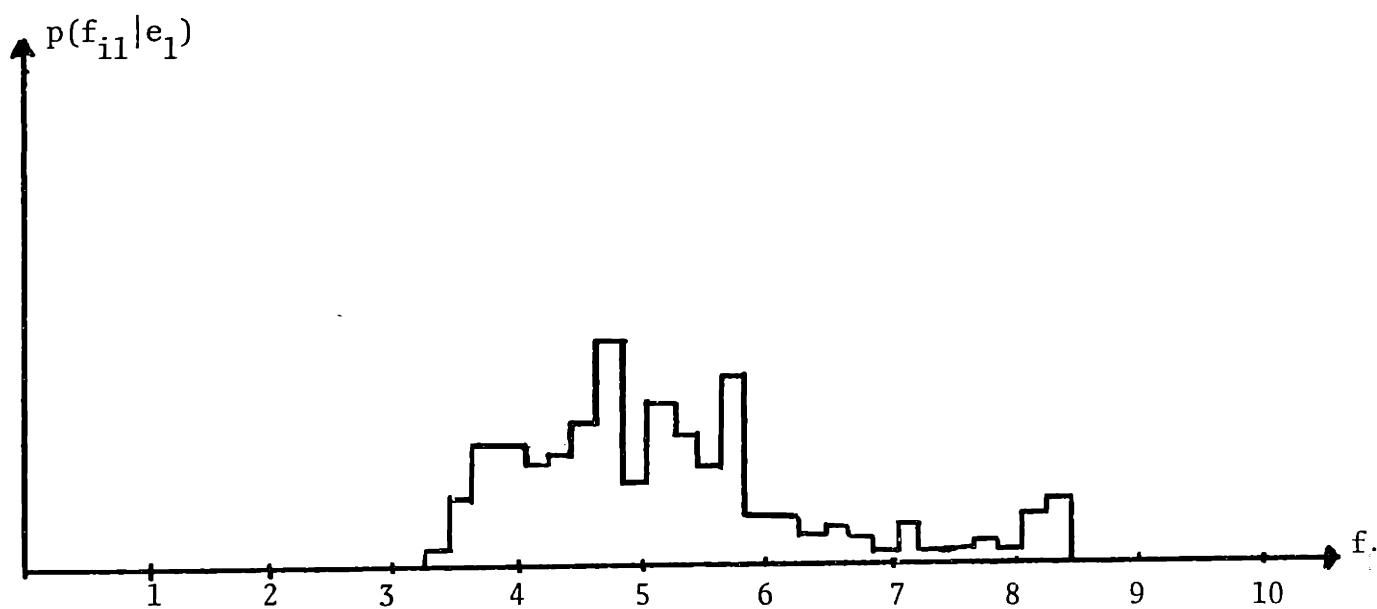


Figure 7.5: Histograms of Chip Distribution

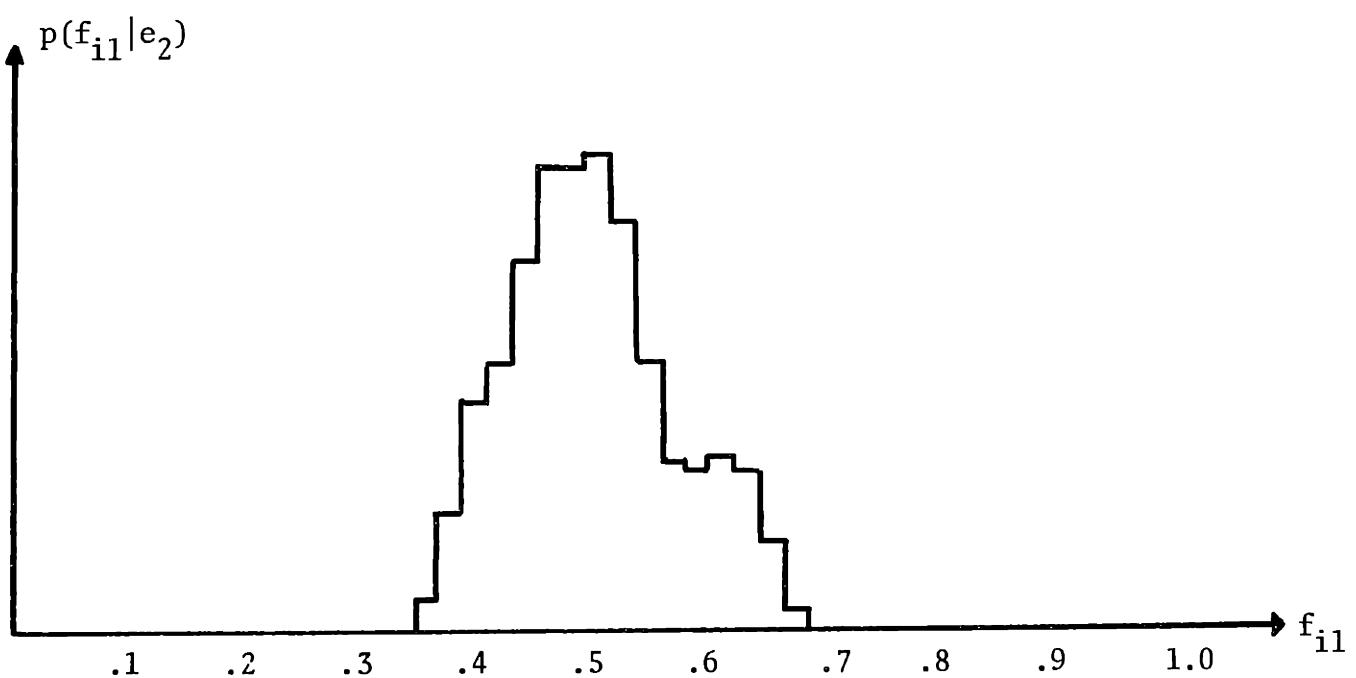
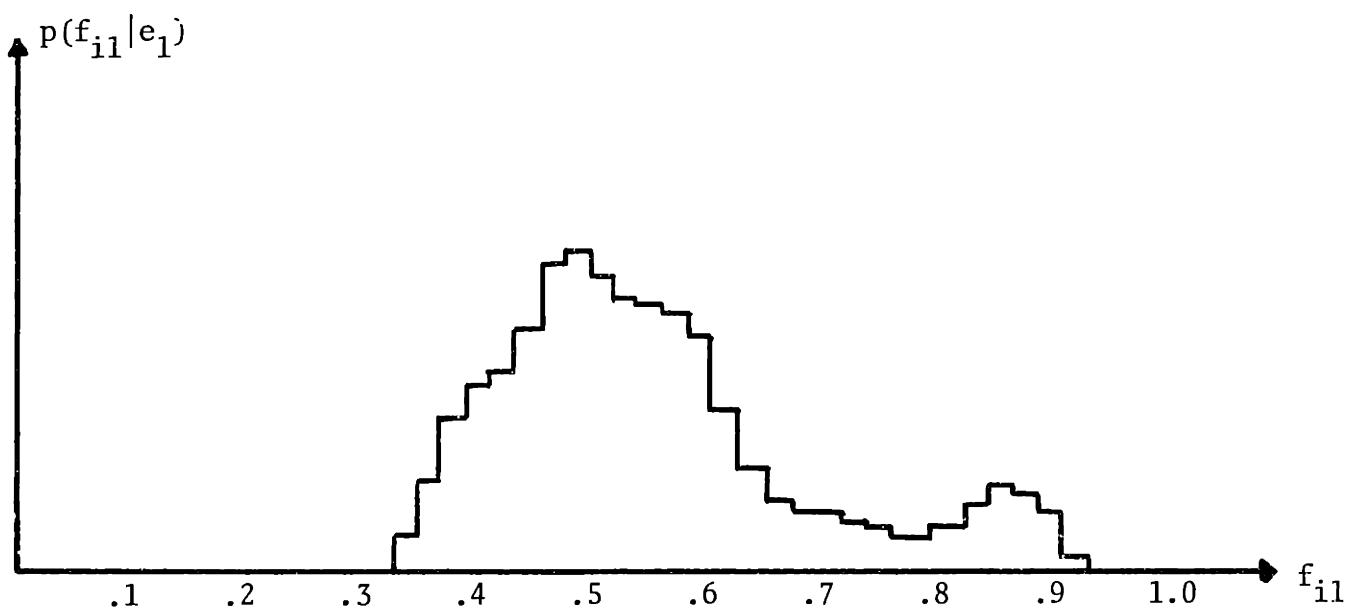
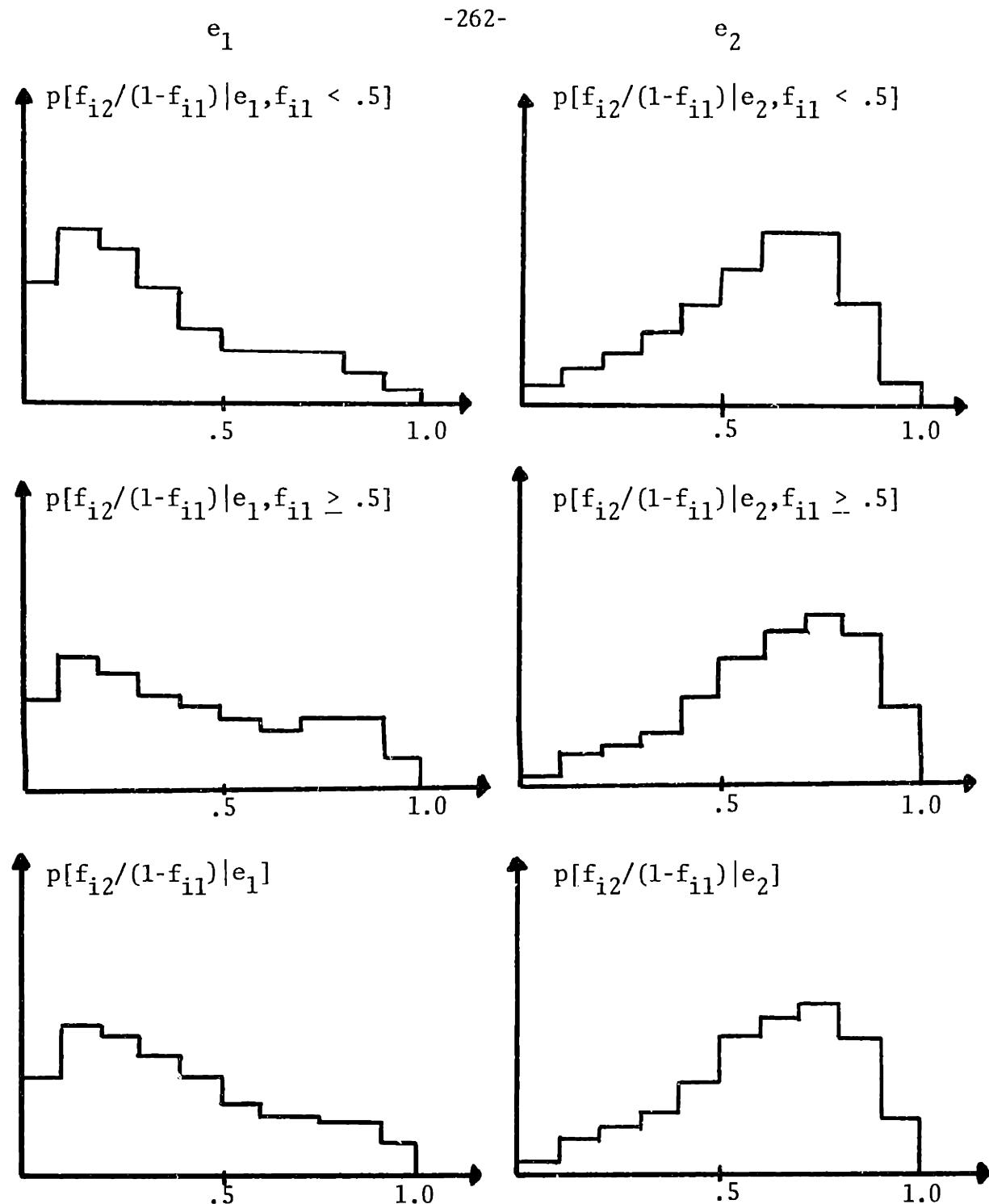


Figure 7.6: Smoothed Histograms of Chip Distribution



abscissa = $f_{i2}/(1-f_{i1})$

Figure 7.7: Smoothed Histograms of Marginal Distributions

separate ranges of f_{i1} .

This identification of independent effects allows representation of the joint density of $p(f_{i1}, f_{i2} | e_j)$ as a product of independent marginals:

$$p(f_{i1}, f_{i2} | e_j) = p(f_{i1} | e_j) \cdot p(\frac{f_{i2}}{1-f_{i1}} | e_j)$$

This independent representation allows easier parameter fitting.

There were not enough observations to realistically fit $p(f_{i1}, f_{i2} | e_3)$, thus since $p(e_3)$ was extremely small, the joint density was approximated with little error by a uniform density.

Examination of Figure 7.6 reveals that empirically if $f_{i1} \geq .66$ then an individual will always choose the first ranked alternative. Then conditioned on $f_{i1} < .66$, $p(f_{i1} | e_1, f_{i1} \leq .66)$ appears to be reasonably approximated by a beta distribution. Note that $f_{i1} \geq .33$ since it is the share of the largest of three values, and similarly the range of $f_{i2}/(1-f_{i1})$ is corrected because $(1-f_{i1})/2 \leq f_{i2} \leq \min(f_{i1}, 1-f_{i1})$. Finally $p(\frac{f_{i2}}{1-f_{i1}} | e_2)$ also appears to be reasonably approximated by a beta distribution. Analytically this results in the following model (i subscript suppressed):

$$p(e_1 | f_1, f_2) = \begin{cases} 1.0 & \text{if } f_1 \geq .66 \\ \frac{.83 p(f_1, f_2 | e_1)}{.83 p(f_1, f_2 | e_1) + .15 p(f_1, f_2 | e_2) + .02} & \text{otherwise} \end{cases}$$

where for $j=1,2$

$$p(f_1, f_2 | e_j) = \frac{\Gamma(\sigma_{ij} + v_{ij})}{\Gamma(\sigma_{ij}) \Gamma(v_{ij})} \left(\frac{\delta_{ij}^{-1}}{3f_1 - 1} \right)^{\sigma_{ij}-1} \left(\frac{v_{ij}^{-1}}{2 - 3f_1} \right)^{v_{ij}-1}$$

$$\cdot \frac{\Gamma(r_{2j} + v_{2j})}{\Gamma(\sigma_{2j}) \Gamma(v_{2j})} \left(\frac{2f_2}{1-f_1} - 1 \right) \left(2 - \frac{2f_2}{1-f_1} \right)^{v_{2j}-1}$$

if $\frac{1}{2} \leq f_1 \leq \frac{2}{3}, \quad \frac{1-f_1}{2} \leq f_2 \leq 1-f_1$

$$= \frac{\Gamma(\sigma_{ij} + v_{ij})}{\Gamma(\sigma_{ij}) \Gamma(v_{ij})} \left(\frac{\delta_{ij}^{-1}}{3f_1 - 1} \right)^{\sigma_{ij}-1} \left(\frac{v_{ij}^{-1}}{2 - 3f_1} \right)^{v_{ij}-1}$$

$$\cdot \frac{\Gamma(\sigma_{2j} + v_{2j})}{\Gamma(\sigma_{2j}) \Gamma(v_{2j})} \left[\frac{2f_2 + f_1 - 1}{3f_1 - 1} \right]^{\sigma_{ij}-1} \left[\frac{2(f_1 - f_2)}{3f_1 - 1} \right]^{v_{2j}-1}$$

if $\frac{1}{3} \leq f_1 \leq \frac{1}{2}, \quad \frac{1-f_1}{2} \leq f_2 \leq f_1$

Step 5: Determine the parameters of the distributions for $p(f_1, f_2 | e_j)$.

Since these distributions are beta with normalized variates, the parameters $\{\sigma_{1j}, v_{1j}, \sigma_{2j}, v_{2j} : j=1,2\}$ were determined by calculating the sample mean and variance for the normalized variates and choosing the appropriate beta

parameters. These parameters are given in Table 7.2. A chi-squared goodness of fit test was used to test the histograms against the beta distributions. The fit was just significant at the 10% level.

	σ_{1j}	ν_{1j}	σ_{2j}	ν_{2j}
j=1	3.40	3.57	1.0	1.69
j=2	4.40	4.69	2.14	1.50

Table 7.2: Calibration of Bayesian Model

7.3.5.3 Comment on Empirical Example

We see from this example that the Bayesian model is indeed a tuning model. Overall there was a strong rank order effect - 83% choose their "preferred" product - but knowing the full cardinal vector c_i did provide "tuned" probabilistic estimates of choice. Although no major design diagnostics were uncovered, some interesting observations can be made. (1) There is a class of people who felt strongly about preference, they said so, i.e., $f_{i1} > .66$, and they acted accordingly. (2) A significant proportion of people chose their second "preferred" product but very few chose their third and none chose lower than third. Thus a manufacturer really should try to be number one or two in this particular product category. (3) Although analytically complex due to the normalization induced by the rank order effect, the basic model was simply calibrated, and the empirical distributions were reasonably approximated with analytically "nice" distributions. Future research will

compare this somewhat complex model to other rank order tuning models and to non-rank order probability models. One such tuning model is in ASSESSOR [131] where a binary logit model based on cr_{i1}/cr_{i2} is used to tune rank order effects. The advantage of the ASSESSOR model is that calibration of the binary logit model is based on optimizing fit, whereas the Bayesian model is still strongly based on intuition and heuristics.

This empirical example represents a first test of a new model. Based on this test it appears that the model has potential for uncovering behavioral effects, but requires intuitive interpretation by the analyst.

7.3.6 Conclusion: Empirical Bayesian Model

Section 7.2 showed that it was extremely difficult to explicitly model the various sources of error which cause uncertainty in predicting choice. Pending further investigation of explicit modeling, this section presented an empirical alternative which implicitly models the error.

Starting with the assumptions that the probability model is symmetric (i.e., the rank order transformation is valid) and that it encompasses the population (two individuals with the same ranked compaction vector, cr_i , have the same ranked probabilities, pr_i) the empirical Bayesian model was derived. It was then shown that the model has a number of desirable properties, one of which was that its calibration was insensitive to the data distribution, $p(cr_i)$.

Finally an empirical example was given of the application of this model to constant sum paired comparison data.

Future research will compare Bayesian models, other tuning models such as Silk and Urban's [131] logit formulation, and non-tuning models such as the multinomial logit formulation. The testing criterion will be the information test developed in Chapter 10.

7.4 Explicit Modeling of Similarities Among Alternatives

The choice axiom, axiom 6.4, used in deriving compaction functions assumed that the preference ordering among two alternatives does not change if a third alternative is added to the choice set. We have seen that in the case of hierarchical choice this assumption may not hold, but if the hierarchy is explicitly modeled compaction theory can still be used and the probability models presented in sections 7.2 and 7.3 can be readily adapted to this choice process.

Another assumption in compaction was that once the compaction values were known, the choice probabilities could be computed, but these probabilities do not explicitly depend upon similarities of the alternatives. (They do depend implicitly upon the similarities because with idiosyncratic compaction functions alternatives perceived as similar will have similar compaction values.) An important future research area is to explicitly include similarity measures in the probability model.

One possible research direction is to adapt Tversky's [143] Elimination-by-Aspects (EBA) to compaction theory. EBA describes choice by a stochastic sequential elimination process. An individual stochastically selects one aspect and eliminates some alternatives;

he then stochastically selects another aspect and eliminates some more alternatives. This continues until one alternative remains. This model is similar to lexicographic decision rules (Fishburn [32], Allaire [2]) but differs in that the order in which aspects are considered is stochastic.

At present this model requires direct estimation of certain scale values and is oriented toward describing behavior with respect to existing stimuli rather than predicting and controlling behavior with respect to new stimuli. Before EBA can be used in the methodology, techniques must be developed to relate the scale values to instrumental performance measures so that predictions of choice behavior can be linked to managerial design decisions.

7.5 Simultaneous, Independent, and Sequential Choice

(This section is tangential to the development and test of the methodology. It is included because it relates some of the theoretical issues in compaction and probability of choice to the very real prediction problem of modeling the interactions between mode choice and destination choice in transportation demand prediction.)

In transportation individuals often make choices from interdependent sets. For example, an individual might choose his destination and then his travel mode. This is a sequential decision. He might choose them independently, or he might choose them simultaneously. This problem has been studied extensively in the transportation demand literature, see for example Ben-Akiva [10], but how can it be handled in

the framework of the methodology?

7.5.1 Simultaneous Models

One option is to model the decisions as completely simultaneous. Consider two disjoint choice sets A and B, e.g., modes of travel and destinations, and suppose that an individual is to make choices among atoms of the form (a,b) , i.e., the true choice set is $A \times B$. Then a simultaneous model would determine a compaction function over the set of performance measures, X , where X equals the union of the performance measures describing A with those describing B. I.e., $X = X'_A \cup X'_B$ where $X'_A \cap X'_B$ is not necessarily empty. The probability model would then be of the form $p(\{a_j, b_\ell\} | \underline{c}_i)$ where \underline{c}_i is now a matrix of the form $\underline{c}_i = ||c_{ij\ell} = c_s(x_j^a, x_\ell^b; \lambda_i) ||$ where $x_j^a \in X'_A$, $x_\ell^b \in X'_B$. All of the previous theory then applies directly to this model.

This model is theoretically sound, but runs into problems if the choice set gets large. For example if there are 10 elements in A and 10 in B, then there are 100 elements in $A \times B$. With so many options in the choice set the inherent random noise in the compaction function and probability model can become large compared to the predicted probabilities. (Notice with 100 items in the choice set, the "equally likely" probability is .01.) There is sometimes a need to simplify the choice process.

7.5.2 Independent Models

A complete simplification of the problem of simultaneous choice is to assume that the choices are independent, i.e., that $p(\{a, b\} \text{ from } A \times B) = p(a \text{ from } A) \cdot p(b \text{ from } B)$. In the vocabulary of Chapter 6 this says that characteristics and stochastic preference orderings among the choice set A are independent of the characteristics and stochastic preference orderings among the choice set B. Before presenting some intuitively obvious sufficient conditions for independence, let us make the following notational definitions. (1) $X_{AB} = X_A' \cap X_B'$, i.e., performance measures common to both A and B; (2) $X_A = X_A' - X_A' \cap X_B'$ i.e., performance measures specific to A, and (3) $X_B = X_B' - X_A' \cap X_B'$, i.e., performance measures specific to B. Note that these definitions partition X into mutually exclusive and collectively exhaustive sets, $X = X_A \cup X_B \cup X_{AB}$.

The first obvious sufficient condition is that $X_{AB} = \emptyset$ and that X_A and X_B are mutually utility dependent. (If only deterministic choices were being made then preferential independence would be sufficient, but we want cardinality for the probability model and also want to be able to consider uncertain alternatives, i.e., alternatives characterized by implicit lotteries.) Note that $X_{AB} = \emptyset$ would not be sufficient because that would leave open the possibility that preference orderings among A^* (lotteries over A) could be dependent upon the choice from B and visa versa. This restriction once stated is obvious but be careful, it is easy to overlook.

It could be that $X_{AB} \neq \emptyset$ but still an independent model applies. The following less restrictive conditions are also sufficient.

(1) $x_{a'b} \sim_{\gamma} x_{a''b}$, $x_{ab} \sim_{\gamma} x_{a'b''}$ where $\gamma = A^* \times B^*$, $x_{ab} \in X_{AB}$ for all $a, a' \in A, b, b' \in B$, and (2) X_A utility independent of $X_B \cup X_{AB}$ and X_B utility independent of $X_A \cup X_{AB}$. Note that if $X_{AB} = \emptyset$ this reduces to the previous condition. Essentially what this says is that there is some overlap in the performance measures but that for any fixed element of one set, say $a \in A$, all elements of the other set, B , are indifferent with respect to X_{AB} . Coupling this with the appropriate utility independence conditions yields the above.

There are other very special sufficient conditions, but the above two are the most likely to occur in practice. The proofs consist simply of unfolding the definitions of utility independence and for brevity will not be stated here.

7.5.3 Sequential Models

Sometimes choices may not be independent but their interdependence may be quite asymmetric, that is one choice, say mode of travel, may depend upon another, say destination, but not visa versa. In this case a sequential model would be appropriate, i.e., $p(\{a,b\} \text{ from } A \times B) = p(\{a,b\} \text{ from } A \times \{b\}) \cdot p(b \text{ from } B)$. Note that although this looks like conditional probability, which would be true by definition, there are various behavioral independence assumptions in the equation. Again sufficient conditions can be stated for a sequential model. They are simply the asymmetric form of the conditions

for the independent model, i.e., (1) $x_{ab}' \sim_{\gamma} x_{ab}''$ for all $a \in A$, $b', b'' \in B$ and (2) X_B utility independent of $X_A \cup X_{AB}$. The interpretation is similar.

7.5.4 Quasi-separable Models

The idea behind quasi-separable models is to add some robustness to simultaneous models by explicitly considering certain segments of the population with limitations on one of the choice sets, A or B. For example consider a simple city where the only modes of travel are auto and bus. Some people may not own an auto and hence their simultaneous choice of mode and destination is limited to the options of $\{\text{bus}, d_{\text{bus}}\}$ where $\{d_{\text{bus}}\}$ are the destinations served by bus. Similarly some people may be auto loyal. To model this effect we introduce the concept of a restriction rule:

Definition 7.7: Suppose an individual is simultaneously choosing from two disjoint choice sets, A and B. Then a restriction rule $f: B \rightarrow A$, is a rule which limits choice to atoms of the form $(f(b), b)$. For a given f , call the set of all atoms the restricted set, B_f . I.e., $B_f = \{(f(b), b): b \in B\}$. Notice $B_f \subseteq A \times B$.

Consider the following example of a restriction rule: Suppose you are at MIT and going shopping at either (1) the Tech Coop (on campus), (2) downtown's Jordan Marsh, or (3) Hingham's general store. Now rather than considering all modes for all destinations, a restriction rule would limit choice to the following sets (walk, Tech Coop), (subway, Jordan

Marsh), and (auto, Hingham).

The decision can then be modeled as sequential in the sense that first a restriction rule, f , is chosen and then second an atom of the form $(f(b), b)$ is chosen. Note that for a given rule the number of alternatives for the second choice is the same as the number of elements of B . The catch is that the number of decision rules is quite large. For example, if there are two modes for each of ten destinations, then there are $2^{10} = 1024$ restriction rules. Thus a quasi-separable model is useful only if the set of restriction rules can be limited a priori.

One extreme case where this idea is useful is when significant portions of the population are either captive-transit or auto-loyal. In this case one would first estimate whether the restriction rule is (1) transit only, (2) auto only, or (3) other. In the latter case analysis would revert to a simultaneous model, but for significant numbers of cases, i.e., (1) and (2), the simple restriction model would be used (See Figure 7.8.).

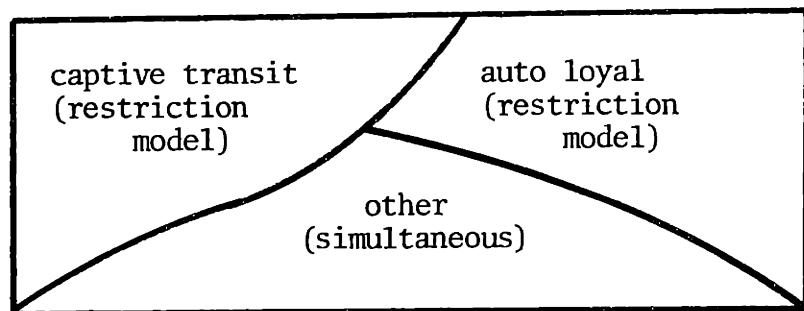


Figure 7.8 : Quasi-separability: Restriction and Simultaneous Models

Robustness is gained because the simultaneous model, with its potential danger of relatively large noise is only used for a smaller portion of the

population.

Another use of this idea is to assume that there exist a relatively small number of a priori specified decision rules, for example deterministic utility functions. The individual then stochastically chooses decision rules, which, once chosen, specifies his restriction rule, f . Then he chooses from the restricted set B_f .

As always, the merits of simultaneous, independent, sequential, and quasi-separable models are highly dependent on the choice situation being modeled. This section attempted to raise some of the issues without particularly resolving any, but hopefully it can act as a guide to an analyst facing the task of modeling "simultaneous" choice from two or more disjoint choice sets.

7.6 Stability Over Time

In transportation there is often a need to project 15 years into the future to 1990, or in health there is a need to project 10 years into the future to decide whether a new health care facility should be built. How stable are predictions from the methodology? Could it have predicted the switch from large to small cars in 1974 resulting from the increased price of fuel? This section looks briefly at sources of temporal instability and how they affect the methodology.

New performance measures: Suppose a new performance measure, such as degree of automatic control in automobiles, becomes important in describing

the alternatives in the evoked set. If this new measure, X_n , is utility independent of the previously applicable performance measures, X , and X is utility independent of X_n then by Theorem 6.7, $c_s(x_n, x; \lambda_i) = k c_s(x; \lambda_i) + k_n c_s(x_n; \lambda_i) + (1-k-k_n) c_s(x; \lambda_i) c_s(x_n; \lambda_i)$ where $c_s(x; \lambda_i)$ is the compaction function over the previously applicable performance measures and $c_s(x_n; \lambda_i)$ is the conditional compaction function for X_n .

Thus all that need be done to incorporate a new performance measure is to assess the uni-attribute scale $c_s(x_n; \lambda_i)$ and the scaling constants k and k_n . The probability model still applies and new choice probabilities can be calculated. Of course if independence properties cannot be verified then complete reassessment of the compaction function may be necessary. If statistical compaction procedures are used, recalibration may be necessary.

Changes in saliency: Suppose by advertising or by general shifts in the population's preferences one or more performance measures becomes relatively more important. If X_ℓ was utility independent of $X_{\bar{\ell}}$ and remains so, then the compaction function remains quasi-additive. If further the risk characteristics of all performance measures remain unchanged then all that need be done is to reassess the scaling constants. If the risk characteristics of any one performance measure changes then all scaling constants affecting it must be reassessed, and if the appropriate independence properties do not hold the entire compaction function must be reassessed. The probability model still applies and

new choice probabilities can be calculated. If statistical compaction procedures are used, recalibration may be necessary.

Changes in the values of the performance measures: There are primarily two ways that the values of the performance measures can change, actual change and perceived change. The former can result from external effects such as the fuel crisis or from design changes. The latter can result from advertising or from good and bad experience. In either case neither the compaction function nor the probability model changes and the choice probabilities can be re-computed. In fact it is this particular case that the methodology is best suited to handle.

New Alternatives: If a new alternative is offered and the probability model is extendable then neither the compaction function nor the probability model changes and the new choice probabilities can be readily calculated.

Comments: Note that many of the causes of temporal instability can be explicitly handled with the methodology. Thus although it is doubtful that projections to 1990 could be made with great confidence, trend indications could be made. The methodology would be then used to track the predictions and explicitly tune them in response to external (unpredicted) events which cause changes in preferences, perceptions, and alternatives.

7.7 Conclusion of Probability of Choice

Chapter 6 identified a function which produced a scalar measure of goodness for each alternative in an individual's evoked set. This chapter then recognized that there would be certain errors in prediction and discussed models to explicitly or implicitly consider these errors in transforming the vector of scalar measures of goodness into a vector of choice probabilities.

Section 7.1 began a formal development by identifying desirable properties for a probability model and then discussed dependency on the choice set. Next the relative merits of revealed preference and proxy choice were discussed with the conclusion that an appealing strategy might be to measure perceptions on real alternatives, but to measure preferences with respect to proxy choice and tune them with a probability model calibrated on revealed choice.

Section 7.2 introduced models based on the postulate of some (perhaps partly unknown) utility function which consumers maximize when making a choice. First the general integral equations were presented, but for any reasonable compaction function these were too complex to solve analytically. The more feasible "random utility models" were then presented and discussed.

Section 7.3 presented an alternative to explicit error modeling. It introduced a new empirical Bayesian model which implicitly incorporates error. This model makes strong use of the rank order nature of choice relative to the scalar measures of goodness and tunes the

predictions with the cardinal values of these measures. It is shown that under the assumption that any two individuals with the same compaction vector have the same choice probabilities it is possible to empirically calibrate such a model and that if the number of observations is sufficiently large and varied then the calibrated parameters are independent of the distribution of the observed data. The use of this model was illustrated by its application to empirical data on aerosol deodorants.

Section 7.4 briefly discussed explicit modeling of similarities among alternatives, section 7.5 presented a discussion of the issues of simultaneous choice from two disjoint choice sets, and finally section 7.6 discussed the stability over time of the predictions based on the methodology.

The next chapter will discuss how to aggregate the individual choice probabilities to predict group response.

Chapter 8

AGGREGATION

Reduction, abstraction, compaction, and probability of choice all deal with the individual choice process. They provide useful diagnostic information to help the design team understand and control choice, but the design team needs estimates of group response before a design strategy can be evaluated and a GO/NO GO decision made. Aggregation provides these estimates by combining the individual choice probabilities to produce estimates of the mean and variance of the total number of people choosing each alternative. (See figure 8.1.) In doing this, corrections must be made for awareness, availability, and evoking of the alternative, for frequency of choice, and if appropriate for the trial/repeat nature of the choice process.

This chapter indicates how to "aggregate" the individual choice probabilities, how to use the individual choice models to predict the response to design decisions, and how to correct for evoking phenomena. In addition it discusses the issues of trial/repeat, and frequency of choice, and suggests existing models which can be used to complement this methodology in dealing with these issues.

Care was taken in abstraction to identify homogeneous segments not only to enhance the accuracy of the choice models but also because of the possibility of differentially targeted alternatives. Thus in predicting group response, aggregation must be performed separately within each of the identified segments. Keeping this in mind for the remainder

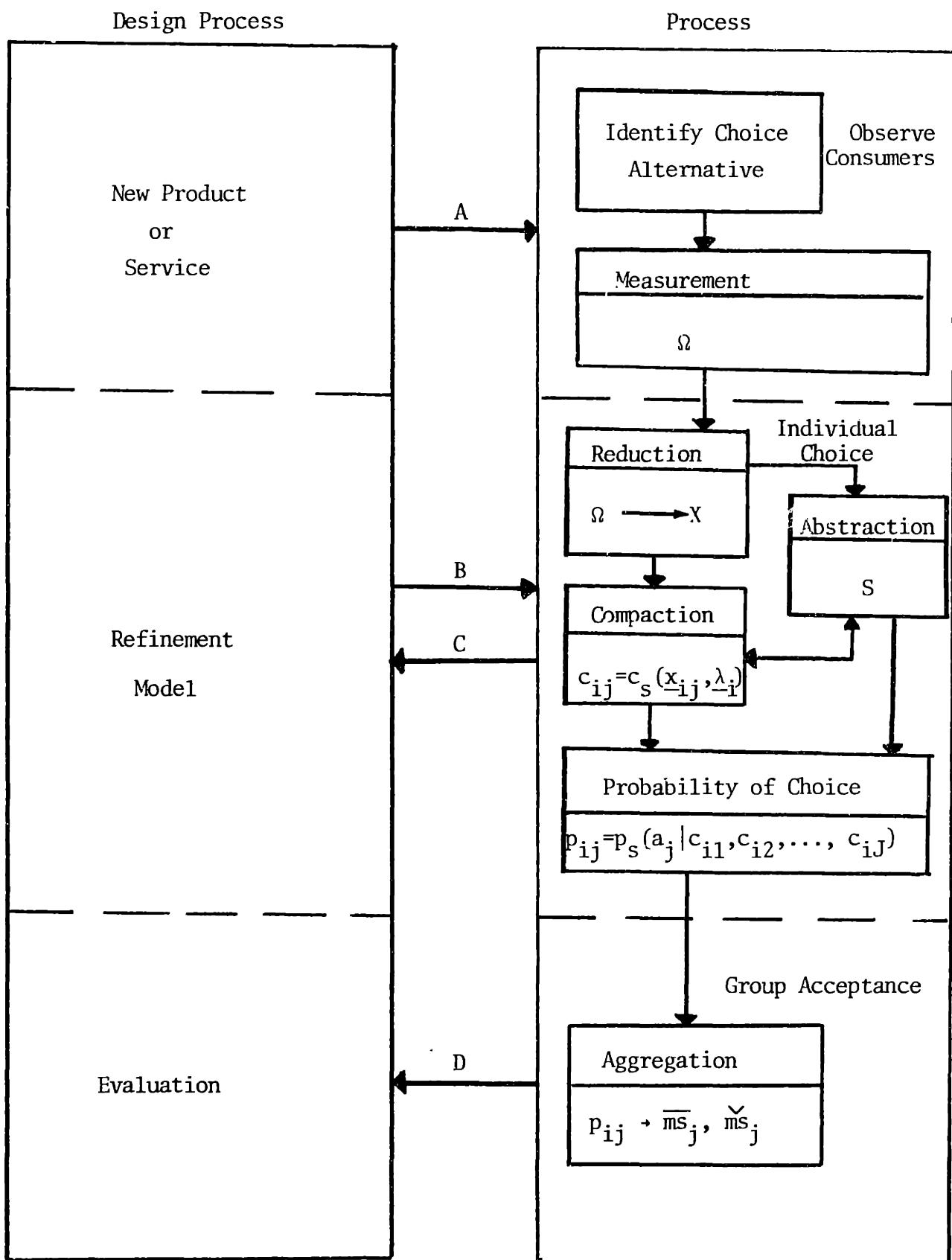


Figure 8.1: Relationship of Aggregation to the Methodology

of this chapter we will assume that we are dealing with a single population segment.

8.1 Combining Individual Choice Probabilities to Predict Group Response

For the moment neglect modeling evoking phenomena and imagine that we are given for each individual, i , and for each alternative, a_j , the probability, p_{ij} , that individual i will choose alternative a_j , where $p_{ij} \geq 0$ if a_j is not in individual i 's evoked set. Let alternative a_0 be the null alternative (no choice is made) and thus $\sum_{j=0}^J p_{ij} = 1$ for all i . Notationally define $\underline{p_i} = (p_{i0}, p_{i1}, p_{i2}, \dots, p_{iJ})$. Let N_j be the total number of people choosing alternative, a_j .

To calculate N_j we make use of a state indicator random variable:

Let $\delta_{ij} = \begin{cases} 1 & \text{if individual } i \text{ chooses alternative } a_j \\ 0 & \text{otherwise} \end{cases}$

then $p(\delta_{ij} = 1) = p_{ij}$

$$p(\delta_{ij} = 0) = 1 - p_{ij}$$

Now $N_j = \sum_i \delta_{ij}$

thus $\bar{N}_j = E(N_j) = \sum_i E(\delta_{ij})$

where we have assumed that individuals are independent. But $E(\delta_{ij}) = 1 \cdot p_{ij} + 0 \cdot (1 - p_{ij}) = p_{ij}$, thus:

$$\bar{N}_j = \sum_i p_{ij}$$

Similarly:

$$\begin{aligned}
 N_j &= \text{var}(N_j) = \sum_i \text{var}(\delta_{ij}) \\
 &= \sum_i [E(\delta_{ij}^2) - E(\delta_{ij})^2] \\
 &= \sum_i [p_{ij} - p_{ij}^2] = \sum_i p_{ij}(1-p_{ij})
 \end{aligned}$$

Finally if $\text{cov}(N_j, N_k)$ is needed it is given by

$$\text{cov}(N_j, N_k) = -\sum_i p_{ij}p_{ik}$$

Similar results can be obtained for share since the market share of a_j is given by $ms_j = N_j/N$.

We see by the above that given p_i for all i it is conceptually easy to obtain the mean and variance of group choice. Note also that N_j is given by the sum of independent Bernoulli probabilities most of which are not degenerate (i.e., $p_{ij} < 1$ for all j) thus when $\bar{N}_j > 3\sqrt{N_j}$ and i is sufficiently large then by the Central Limit Theorem (Drake [29]) the joint distribution of $\underline{N} = (N_0, N_1, \dots, N_J)$ is multivariate normal with means, variances, and covariances as calculated above. When $\bar{N}_j < 3\sqrt{N_j}$ the Poisson approximation can be used.

The above results hold when the discrete summation over individuals is replaced by an integration over a distribution of p_i . Unfortunately if there is great variation in perceptions (or actual values) of the performance measures across the population, and if this population has many segments, the integration which is easy to write down becomes tedious and time consuming. Thus, occasionally approximations in the integration

must be made. This issue is important in a major urban transportation studies where predictions are made for every zonal interchange. For a complete treatment of the problems of calculating \bar{N}_j when the multinomial logit model is used for p_i see Koppleman [80].

8.2 Prediction by Changing Attributes, Performance Measures, or Preference Parameters

This methodology was formulated for more than just describing choice behavior, it also produces a first order prediction of the number of people choosing a new alternative or the gain(loss) in number of people due to a design decision effecting an existing alternative. Again temporarily ignoring evoking and frequency phenomena, design decisions can be simulated (arrow A in figure 8.1) by changes in (1) the attributes, Y, (2) the performance measures, X, or (3) the preference parameters, Λ .

Attributes: If the attributes were chosen to be instrumental (section 3.1.1.1) then certain design decisions will be reflected directly in changes in the attributes.¹ If the reduction is then computable (definition 4.1) a new distribution of performance measures can be determined. Then proceed as if the changes were made directly in the performance measures.

Performance measures: The design team can effect changes in the performance measures by changing the attributes, but they may also wish to stimulate creativity by simulating a direct change in the performance measures or they may wish to simulate an advertising strategy which changes

peoples' perceptions. In any case we are given a new set of values for the performance measures, \underline{x}_{ij} , for all i, j . (Most often the change will only be along one performance measure for a single alternative. Assume all measures not directly changed remain unchanged.)

The aggregation technique is quite simple. Using the sample population as representative of the target population, (remember segmentation is corrected for), compute for each individual his new compaction vector, \underline{c}_i , and then his new choice probability vector, \underline{p}_i . Total choice is given by the method of section 8.1. In actuality the design team may not change all the individual performance measures but instead will specify a shift in the mean. In this case assume the shape of the distribution of perceptions remains the same and simply calculate the new perceptions by adding a constant value (new mean-old mean) to each individual's performance measure for alternative a_j .

In some cases especially transportation it is convenient to represent the population by a joint probability distribution for the performance measures and the preference parameters and integrate to find the share of choice. For example, given that the performance measures, \underline{x}_j , for each alternative and the preference parameters, $\underline{\lambda}$, are distributed across the population with joint probability distribution $p_s(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_J; \underline{\lambda})$ the predicted share for alternative 1 of two alternatives is:

$$ms_j = \int p[a_j | c(\underline{x}_1, \underline{\lambda}), c(\underline{x}_2, \underline{\lambda})] p(\underline{x}_1, \underline{x}_2, \underline{\lambda}) d\underline{x}_1 d\underline{x}_2 d\underline{\lambda}$$

Preference parameters: The design team may decide to effect choice by changing the saliency of certain performance measures through advertising. One method to simulate this change is to predict a shift in the values of the preference parameters. Aggregate estimates are then obtained in the same way as they were for shifts in the performance measures.

Sensitivity: Predictions made by direct shifts in the performance measures or preference parameters are exploratory by nature. It is imperative that sensitivity analyses be performed by trying a number of shifts rather than just one.

8.3 Correction for Evoked Set

In observing consumers we were careful to make observations relative to the evoked set and in some cases, especially for new alternatives, we forced evoking via concept statements, mockups, or the real alternative. In prediction we must correct for evoking. Three cases will be discussed:

- (1) no new alternative, no change in evoking, (2) no new alternative, design change effects evoking, and (3) new alternative introduced.

No new alternative, no change in evoking: This is the base case where the design change does not effect the consumer's evoking of the alternatives, only their perceptions and preferences. Aggregation is as discussed in sections 8.1 and 8.2.

No new alternatives, design change effects evoking: One of the possible diagnostics from the individual choice model is that the product

is good but no one is aware of it. Thus some design decisions may be directly aimed at increasing the percent of people evoking an alternative. For example, more advertising or sampling (free one day pass on dial-a ride) may be aimed at awareness. To be completely theoretically sound this change in evoking should be modeled at the individual level, but no practical models to do this are currently available. A reasonable aggregate strategy is to predict the new percentage, E_j^n , evoking alternative a_j and assume that the increase is equally likely to occur for all individuals not now evoking a_j . Mathematically let I_j be the set of all individuals currently evoking a_j , let \bar{I}_j be its complements, and let E_j^0 be the old percentage evoking. ($E_j^0 = (\text{number in } I_j) / (\text{number in } I_j + \bar{I}_j)$).

Step 1: Simulate the design change over the set I_j to get a prediction of choice, $\underline{N}^0 = (N_1^0, N_2^0, \dots, N_J^0)$.

Step 2: Simulate the design change over the set \bar{I}_j with no change in evoking to get a prediction of choice, $\underline{\bar{N}}^0 = (\bar{N}_1^0, \bar{N}_2^0, \dots, \bar{N}_J^0)$.

Step 3: Simulate the design change over the set \bar{I}_j with everyone evoking alternative a_j to get a prediction of choice, $\underline{\bar{N}}^n = (\bar{N}_1^n, \bar{N}_2^n, \dots, \bar{N}_J^n)$.

Step 4: Now the change in evoking will cause everyone in I_j to evoke a_j and $(E_j^n - E_j^0) / (1 - E_j^0)$ of the individuals in set \bar{I}_j to evoke a_j . (See figure 8.2.) Thus under the assumption of equally likely evoking over \bar{I}_j :

$$N_k = N_k^0 + \left[\frac{E_j^n - E_j^0}{1 - E_j^0} \right] \cdot \bar{N}_k^n + \left[1 - \frac{E_j^n - E_j^0}{1 - E_j^0} \right] \cdot \bar{N}_k^0$$

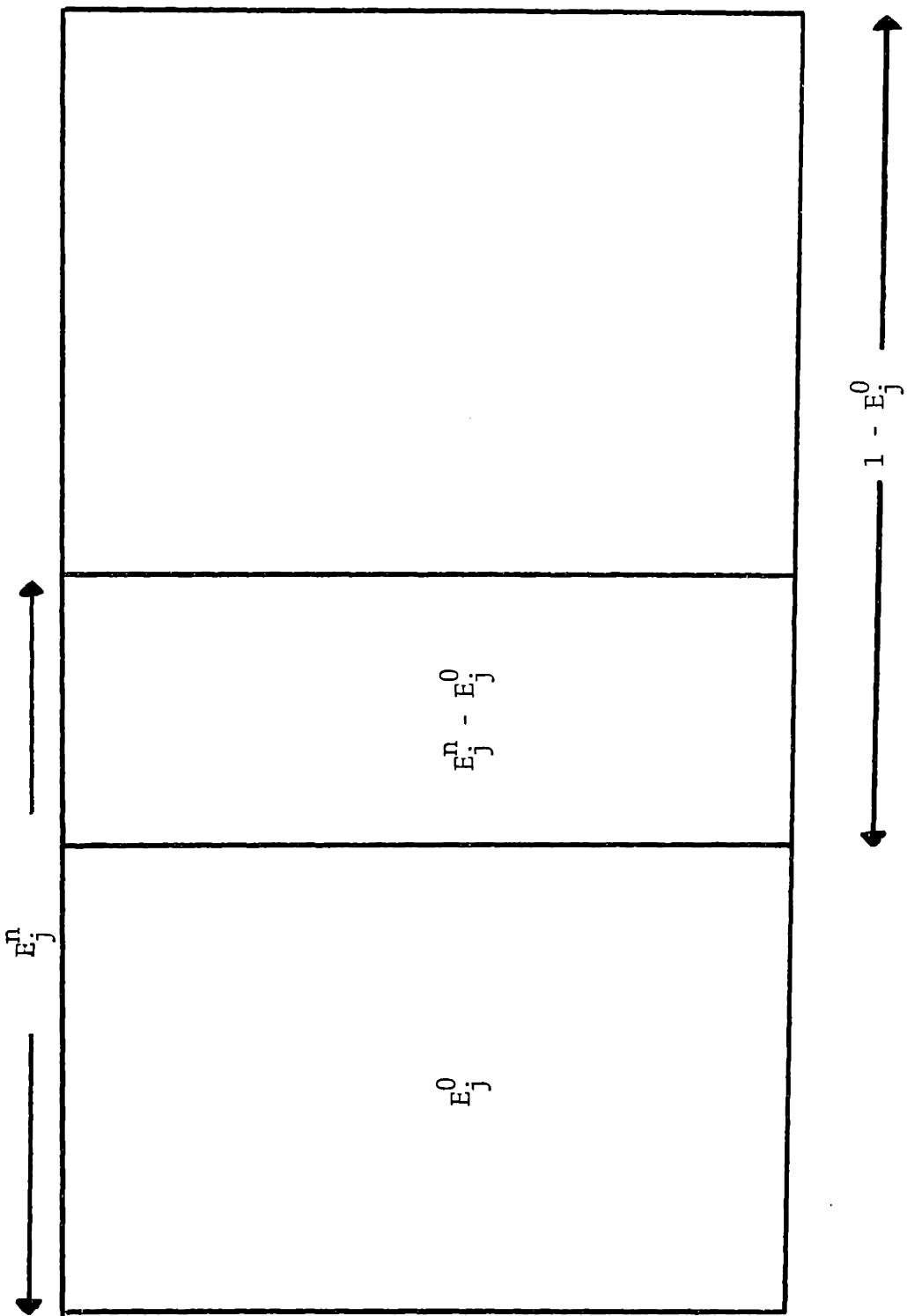


Figure 8.2: Change in the Evoked Set

Where the above equation is true for all alternatives, a_k , not just a_j .

New alternative introduced: When a new alternative is introduced part of its design is a strategy to cause some percentage of the population to evoke the new alternative. The correction for evoking proceeds as above except that $E_j^0 = 0$, $I_j = \phi$, and \bar{I}_j is now the entire population. (As always corrections for evoking are done within segments.)

8.4 Dynamics: Trial, Repeat, and Frequency

When possible it is preferable to observe consumers making choices from sets of actual alternatives, but, especially with new products or services, observation is sometimes made with respect to proxy alternatives such as concept statements. If possible, this effect must be corrected for in aggregation because choices from proxy alternatives give trial probabilities. Once an alternative is tried, perceptions can change and the repeat probabilities can be different from the trial probabilities.

Even if trial and repeat probabilities were known for each individual, there is still a frequency phenomenon and a time dependent diffusion process. For example the number of dial-a-ride trips per month would be much larger if everyone used it daily than if everyone used it twice monthly. Furthermore, steady state usage would not be reached on opening day.

One technique to handle frequency is to treat it as a choice made simultaneously with the choice of alternatives. (Section 7.5 discusses simultaneous choice models.) Ben Akiva [10] uses a variation of this technique in transportation demand modeling. Another technique is to build a macro-flow tracking model in which consumers with different choice

(or usage) frequencies are handled separately as identifiable segments. This latter technique has the advantage that it can also explicitly handle trial/repeat phenomena.

Urban in SPRINTER MOD III [147] has developed a fairly complete macro-flow model which can simultaneously handle a diversity of dynamic effects. Rather than give a complete description of Urban's model we will instead give a naive example of a macro-flow model.

Suppose we are only tracking the choice of one alternative and suppose that 1000 consumers can be classified as being in one of five states: (1) initial pool, (2) aware of the alternative, its available, but have not tried, (3) are trying the alternative for the first time, (4) recent bad experience and (5) recent good experience. (See figure 8.3.) Suppose from our methodology we know the trial probability, p_T , for people in state (2) and the repeat probability for people in state (4) and state (5), p_b and p_g . We are given time dependent probabilities of awareness, p_{at} , from our advertising agency and we know that the probability of a good experience with our alternative is p_G . P_1 percent of the consumers are frequent users (every time period) and P_2 percent are occasional users (every other time period). The process is memoryless (transitions depend only on state) and we enter the process at a random time but with everyone in state (1).

The macro-flow concept is simply to simulate the flow of consumers through this process and count the number of choices per time period. An example of this is given in table 8.1, and the resulting plot of usage over time is given in figure 8.4.

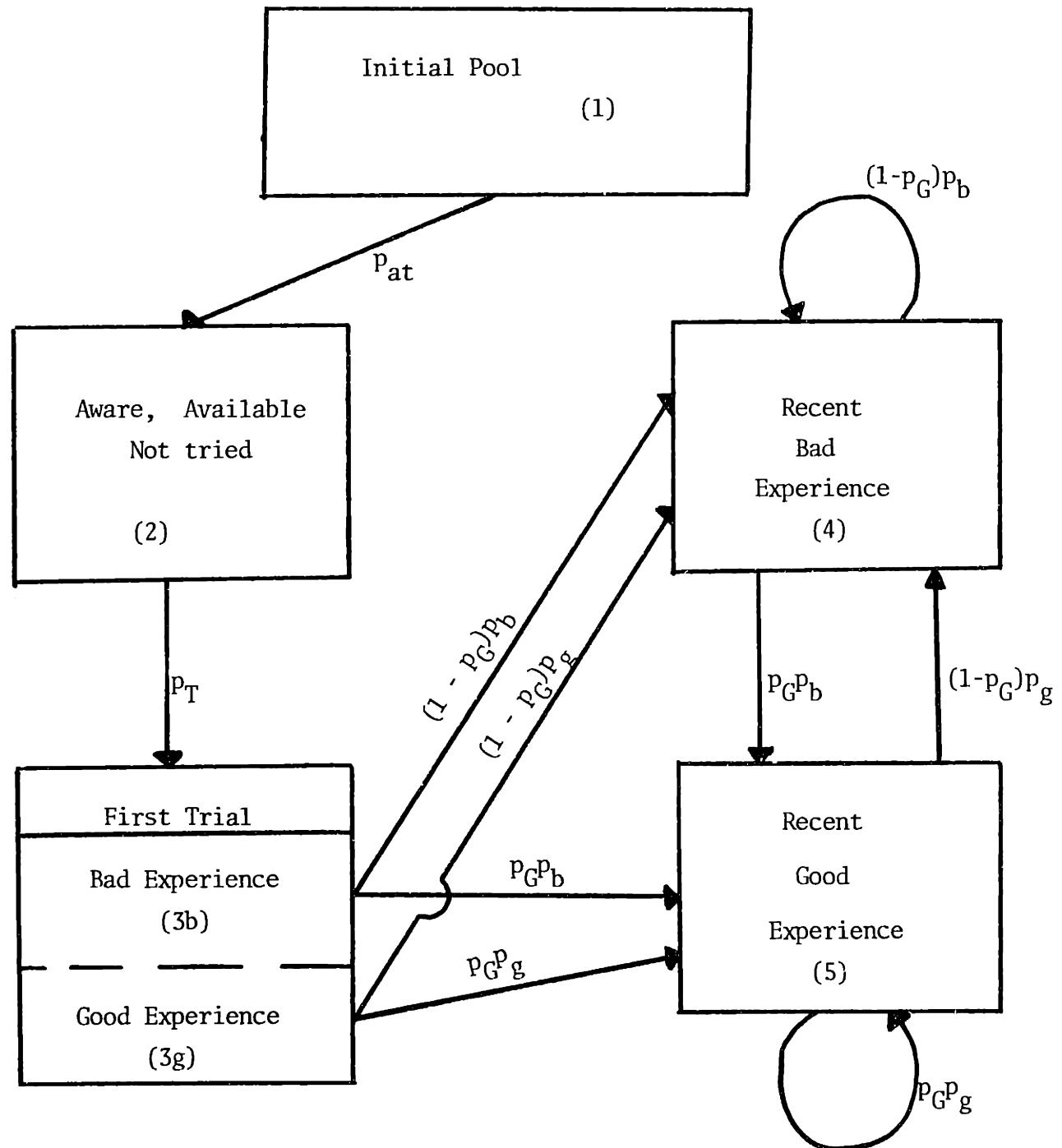


Figure 8.3: Dynamics-Macro Flow Model

Table 8.1: Dynamics-Macro Flow Model (page 1 of 3)

Time	Frequent Users					Occasional Users (Odd days)						
	(1)	(2)	(3b)	(3g)	(4)	(5)	(1)	(2)	(3b)	(3g)	(4)	(5)
t=0	400	-	-	-	-	-	300	-	-	-	-	-
t=1	200	200	-	-	-	-	150	150	-	-	-	-
t=2	120	200	32	48	-	-	150	150	-	-	-	-
t=3	84	156	61	58	16	25	105	135	24	36	-	-
t=4	67	111	80	49	43	50	105	135	24	36	-	-
t=5	60	74	90	36	75	65	95	91	44	39	13	18
t=6	60	44	93	25	107	71	95	91	44	39	13	18
t=7	60	26	91	16	134	73	95	55	54	30	31	35
t=8	60	16	86	9	158	71	95	55	54	30	31	35
t=9	60	10	81	6	178	65	95	34	57	19	52	43
t=10	60	6	75	3	193	63	95	34	57	19	52	43
t=∞	60	0	0	0	206	54	95	0	0	0	173	52

Table 8.1: Dynamics-Macro Flow Model (page 2 of 3)

Occasional Users (Even Days)						Totals						
Time	(1)	(2)	3b)	(3g)	(4)	(5)	(1)	(2)	(3b)	(3g)	(4)	(5)
t=0	300	-	-	-	-	-	1000	-	-	-	-	-
t=1	300	-	-	-	-	-	650	350	-	-	-	-
t=2	180	120	-	-	-	-	450	470	32	48	-	-
t=3	180	120	-	-	-	-	369	411	85	94	16	25
t=4	144	108	19	29	-	-	316	354	123	114	43	50
t=5	144	108	19	29	-	-	299	273	153	104	88	83
t=6	144	65	34	32	10	15	299	200	171	96	130	104
t=7	144	65	34	32	10	15	299	146	179	78	175	123
t=8	144	39	41	22	26	28	299	110	181	61	215	134
t=9	144	39	41	22	26	28	299	83	179	47	256	136
t=10	144	23	43	14	42	34	299	63	175	36	287	140
t=∞	144	0	0	0	131	25	299	0	0	0	590	111

Table 8.1: Dynamics-Macro Flow Model (page 3 of 3)

Time	Usage	PARAMETERS
$t=0$	-	Awareness
$t=1$	-	$p_{at} = (.5, .4, .3, .2, .1)$
$t=2$	80	
$t=3$	181	Trial
$t=4$	184	$p_T = .4$
$t=5$	220	
$t=6$	195	Repeat
$t=7$	201	$p_g = .8$
$t=8$	171	$p_b = .1$
$t=9$	175	Experience
$t=10$	150	$p_G = .6$
$t=\infty$	148	Frequency
		$p_1 = .4$
		$p_2 = .6$

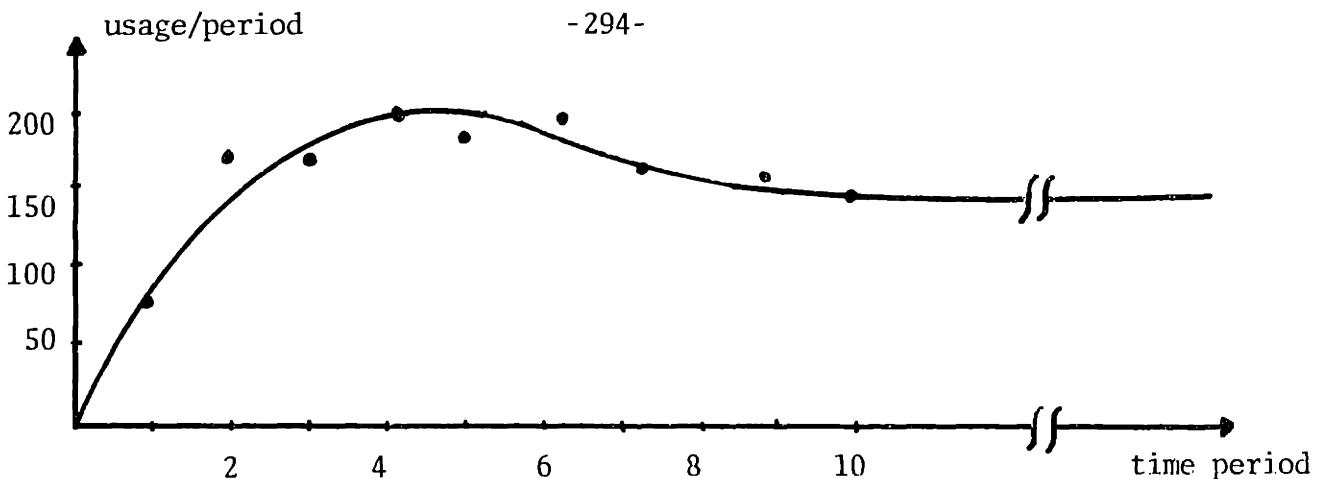


Figure 8.4: Dynamics—Usage Over Time

Notice in the above example that when advertising stops the usage begins to decay. But this decay is not the fault of the advertising! The macro flow model shows that what is really happening is that users are having bad experiences with the alternative and are thus less likely to reuse it. If we change the probability of a good experience from .6 to .8 then the steady state usage more than doubles (usage = 306) even if there is no more advertising.

From this example we can see the importance of complementing the methodology's predictions with external dynamic models.

8.6 Empirical Example: HMO Study

A simple aggregation example is to use the statistical compaction technique of preference regression and deterministic choice.

Base Case: Preference regression was used to determine a statistical compaction function for the performance measures, quality, personalness, convenience, and value, relative to the choice of health care delivery systems. (See Section 6.4.1.) Scores on these performance measures for each individual for each of four health care plans were determined

by factor analysis. (See Section 4.2.1.3.) These scores for individuals not in the M.I.T. HMO pilot program were substituted in the compaction function and it was assumed that everyone would choose the maximum valued plan if he was aware of it. Finally awareness was externally estimated and the total values calculated. (See figure 8.5.) Re-enrollment for those already in the pilot program was calculated from the estimated repeat rate (92.5%) and estimates of migration out of the M.I.T. community. This resulted in a forecast of 3600 families which is just financially sufficient to maintain the HMO. Considering the inherent risk involved in any new service venture, the decision to expand the existing HMO could not be supported based on the initial design.

Prediction: One diagnostic from reduction was that changing hospital affiliation would improve the perception of quality. Another was that active media advertising could move the perceptions of personalness and value for the concept up to the levels of the people in the M.I.T. HMO pilot program. These actions were simulated by increasing the mean of the M.I.T. HMO concept quality perception one-half the distance between the Harvard Community Health Plan (HCHP) and M.I.T., and the means of value and personalness one half the distance to that of the M.I.T. HMO pilot. Awareness under an improved marketing strategy was estimated at 85%. These results are given in figure 8.5. Sensitivity analyses were done by varying the degree improvement and varying competition from HCHP. The most likely forecast was based on HCHP being offered with an improved M.I.T. plan. This resulted in a forecast of 4950 family enrollments and was sufficient to make a positive recommendation to M.I.T. based on

CASE I - EXISTING DESIGN

New Enrollment:

<u>Groups</u>	<u>Number not now in pilot HMO</u>		<u>Enrollment if aware</u>		<u>Estimated awareness</u>		<u>Estimated Enrollment Number</u>
Students	8000	x	33%	x	70%	=	1848
Faculty	3800	x	15%	x	70%	=	399
Staff	<u>3400</u>	x	<u>22%</u>	x	<u>70%</u>	=	<u>523</u>
Total	17,200		23%		70%		2,770

Re-enrollment:

Existing HMO Subscribers	Repeat Rate	Estimated to Remain at MIT
1067	x 92.5%	x 86.3% = 852

Total Enrollment = 3622

CASE II - IMPROVED DESIGN

New Enrollment:

<u>Groups</u>	<u>Number not now in pilot program</u>		<u>Enrollment if aware</u>		<u>Estimated awareness</u>		<u>Estimated Enrollment Number</u>
Students	8000	x	42%	x	85%	=	2856
Faculty	3800	x	25%	x	85%	=	808
Staff	<u>3400</u>	x	<u>30%</u>	x	<u>85%</u>	=	<u>867</u>
Total	17,200		31%		85%		4,531

Re-enrollment:

Existing HMO Subscribers	Repeat Rate	Estimated to Remain at MIT
1067	x 95%	x 86.3% = 874

Total Enrollment = 5,405

Figure 8.5: Forecast of M.I.T. Enrollment

the response of consumers to revised communication and design strategies.

A more complete discussion of the M.I.T. HMO case appears in chapter 11.

8.7 Conclusion of Aggregation

Aggregation is the act of combining the individual probabilities to predict group response. This chapter presented the mathematics of aggregation, indicated how to correct for evoking, and discussed an external dynamic model to complement the methodology.

The output of aggregation is the mean and variance of number of people choosing each alternative as a function of design decisions. This information is crucial in strategy evaluation, but it is not the only information needed. The next chapter discusses the interaction between the analytic process and the design process and indicates some evaluation techniques.

Chapter 9

INTERACTION WITH THE DESIGN PROCESS

The first interaction of the consumer model with the managerial process is in evaluation (arrows A + D in figure 9.1). The prediction of the mean and variance of the number of consumers choosing each alternative are critical evaluation inputs, but other factors such as investment, cost, risk, or political considerations must be integrated in a model that results in a decision to terminate (NO), or to continue development (ON), or to introduce to the market (GO).

It is rare that a GO decision is reached on a first cycle through the methodology, more likely an ON decision is made and the design is subjected to a refinement effort. One criteria for the methodology was that it facilitate successful innovation by eliciting and focusing creativity. The methodology does this in the refinement stage by providing insight and diagnostics (arrows B + C in figure 9.1) on four levels (1) reduction identifies how the attributes combine to form performance measures and pictorially represents the alternatives in the perceptual space of performance measures, (2) abstraction identifies strategically relevant segments homogeneous with respect to perception and/or preference, (3) compaction explicitly identifies the relative importance of the performance measures, how strongly they interact, and how important risk characteristics are in the choice process, (4) probability of choice gives numerical implications of the scalar measures of goodness.

Design Process

Consumer-Response
Process

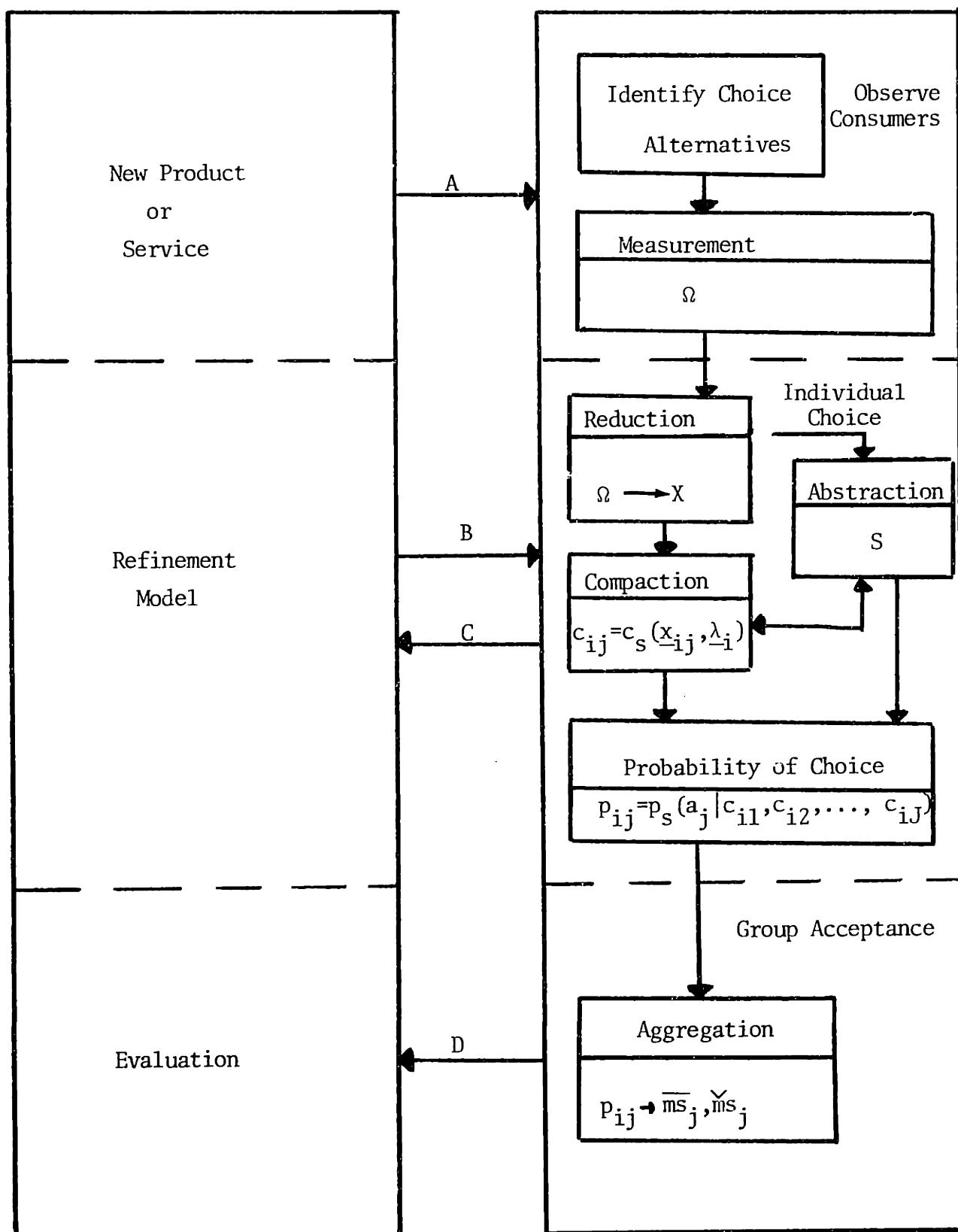


Figure 9.1: Interaction with the Design Process

Based on the insight gained from the individual choice models, the attributes of the alternative and the "marketing" strategy are refined. The analytic model is then used to simulate the new "alternative." This process is iterated until an acceptable design is found. If necessary the analyst can then take new measurements on the refined alternative and rerun the methodology to evaluate it.

This chapter begins by indicating the process of evaluation and suggesting evaluation models. It then goes on to indicate the process of refinement and reviews the various types of diagnostic input available from the individual choice models.

9.1 Evaluation/Simulation Mode of Interaction

9.1.1 Process of Design and Strategy Changes

New designs and strategies can be simulated in a number of ways:

Attribute changes: Changes in the characteristics of the offered alternative can be tested by changing the values of the attribute measures (judgementally if rating scales rather than natural measures are used). The calibrated reduction, compaction, probability, and aggregation models are then used to simulate the effects of the change on consumer choice within each segment. For example a specific refinement such as changing the hospital affiliation of an M.I.T. HMO can be tested by changing the average rating on a scale of hospital quality to be equivalent to that of an HMO which already has the better hospital affiliation (e.g., Harvard Community Health Plan).

Performance measure shifts: The design team may want an indication of more general changes such as the overall quality of the HMO. These changes are modeled as judgemental shifts in the values of the performance measures and are input directly to the compaction model. The calibrated compaction, probability, and aggregation models are then used to simulate the effects of the shift on consumer choice within each segment. Such simulated shifts are extremely useful in guiding the refinement effort. For example suppose that a slight shift in quality produces a dramatic change in choice. The design team would then concentrate on creative ideas to shift the actual or perceived quality. Another example of a shift in the values of the performance measures might be to simulate an advertising strategy which would bring the perception of the performance measures up to their directly calculated values.

Evoking changes: The alternative may be fantastic, but if no one is aware of it, or if it is not available to anyone, or if those aware of it can not easily get more detailed information about it, it may be a failure in the "marketplace." Advertising, sampling, distribution and other "marketing" strategies can be tested by simulating their effects on the evoking rates in the aggregation model. Similarly different segmentation strategies can be tested by simulating differentially targeted designs and strategies.

The outputs of the above changes (attributes, performance measures, and evoking) are diagnostic from each of the individual choice models as well as evaluative based on the output of aggregation.

9.1.2 Data Available for Evaluation

Evaluation data is available from the analytic model, but additional "supply" data is necessary from other models.

Demand data: For each combination of design and strategy decisions the analytic model provides estimates of the mean and variance of the new or changed alternative. It also indicates cannibalization (taking share from existing alternatives) by providing before and after estimates of the share of these alternatives. Note that because the compaction model is individual specific more share will be taken from those alternatives similar to the new alternative. Thus, one strategy may be preferred to another even if they capture the same share because they may capture that share by cannibalizing different alternatives. This becomes very important if one of these alternatives is another product or service given by the same organization or if competitive effects are important.

Supply data: The prediction of demand is a crucial evaluation input, but other factors are also needed. Separate models or estimates of the cost of providing the alternative, the facilities and investment required, the political considerations, and possibly competitive response should be coupled with the analytic model to provide the inputs necessary for evaluation. These models are external to the dissertation and will not be described here.

9.1.3 Models for Evaluation

Decision Analysis: Decision analysis as presented by Raiffa [120] and used by Keeney [67] and others is a formal structuring of a decision problem through the use of decisions trees and prescriptive utility theory. Basically the problem is represented by a tree with decision nodes (Δ) and chance nodes (0). (See figure 9.2.) At each decision node the design team is faced with a decision among a number of alternative courses of action represented as branches in the tree. At each chance node a number of outcomes are possible. A probability of occurrence is associated with each branch. At the end of the tree the outcomes, 0_ℓ , are represented by a vector of performance measures (not to be confused with the reduced space perceptions of chapter 4) which represent the attributes of that outcome. A multiattributed utility function is assessed to represent the design team's preferences for the levels of the performance measures. (The theory underlying this assessment is the theory described in chapter 6 and is meant here to be prescriptive rather than descriptive.) Using a technique described in Raiffa [120], the tree is "folded back" by calculating the expected utility at each chance node and choosing the decision at each decision node to maximize expected utility.

This evaluation model is particularly useful for new service innovations such as the design of a health maintenance organization. The performance measures might be (1) the predicted number of members, (2) the net cost (or profit) of supplying service, (3) the fixed facilities required, (4) a political index and any other measures deemed appropriate.

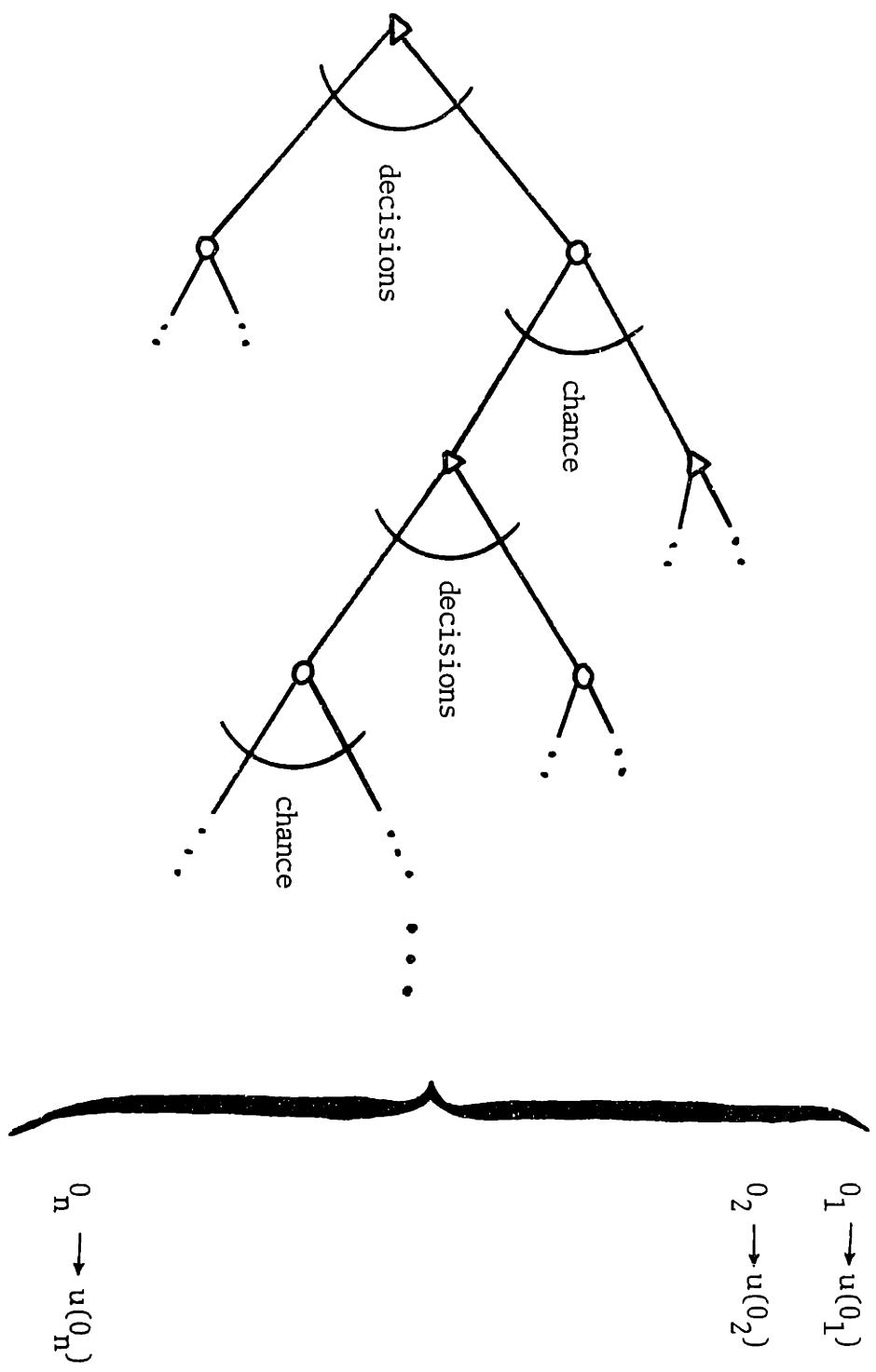


Figure 9.2: Decision Tree

Notice that in this case the predicted number of members is included in addition to the resulting revenue.

Decision Quadrant: Sometimes the design team may not wish to quantify its preferences in the form of a utility function but instead have a pictoral representation of the outcome to act as a guide to their decisions. Urban [145] suggests such a pictoral representation of uncertainty in profit.

Suppose external models combine the demand estimates produced by the methodology to produce estimates of the monetary benefit of the innovation under consideration. Then Urban's criteria for a GO decision is that the probability of obtaining a target discounted rate of return must be greater than a specified level p_{go} , and his criteria for a NO decision is that it be less than some specified level, p_{no} . Let b_j be this monetary benefit discounted at the target rate of return and let I_j be the investment required for alternative a_j . Mathematically Urban's criteria are stated as:

$$\text{Prob}(b_j \geq I_j) \geq p_{go} \rightarrow \text{GO}$$

$$\text{Prob}(b_j \geq I_j) \leq p_{no} \rightarrow \text{NO}$$

$$\text{otherwise} \rightarrow \text{ON}$$

Suppose that the methodology produces the mean, \bar{b}_j , and standard deviation, σ_j , of the discounted benefit, then the GO decision can be written as:

$$\text{Prob} \left\{ \frac{b_j - \bar{b}_j}{\sigma_j} \geq \frac{I_j - \bar{b}_j}{\sigma_j} \right\} \geq p_{go} \rightarrow \text{GO}$$

This is equivalent to

$$(\bar{b}_j - I_j)/\sigma_j \geq -t_{go} \longrightarrow GO$$

where t_{go} is the $(p_{go})^{\text{th}}$ fractile of the normalized (zero mean, unity variance) distribution of b_j . This is shown in figure 9.3 for a normal distribution where the shaded area represents p_{go} .

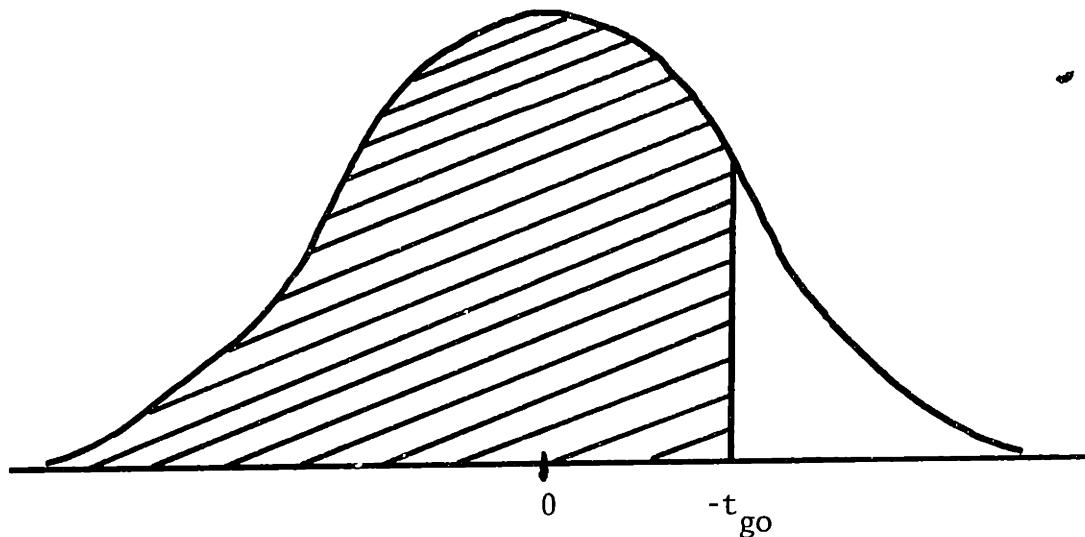


Figure 9.3: GO fractile

Similarly for the NO criterion:

$$(\bar{b}_j - I_j)/\sigma_j \leq -t_{no} \longrightarrow NO$$

These inequality constraints are straight lines and are used to partition the decision quadrant (expected value-standard deviation space) into mutually exclusive collectively exhaustive GO/ON/NO areas. (See figure 9.4.)

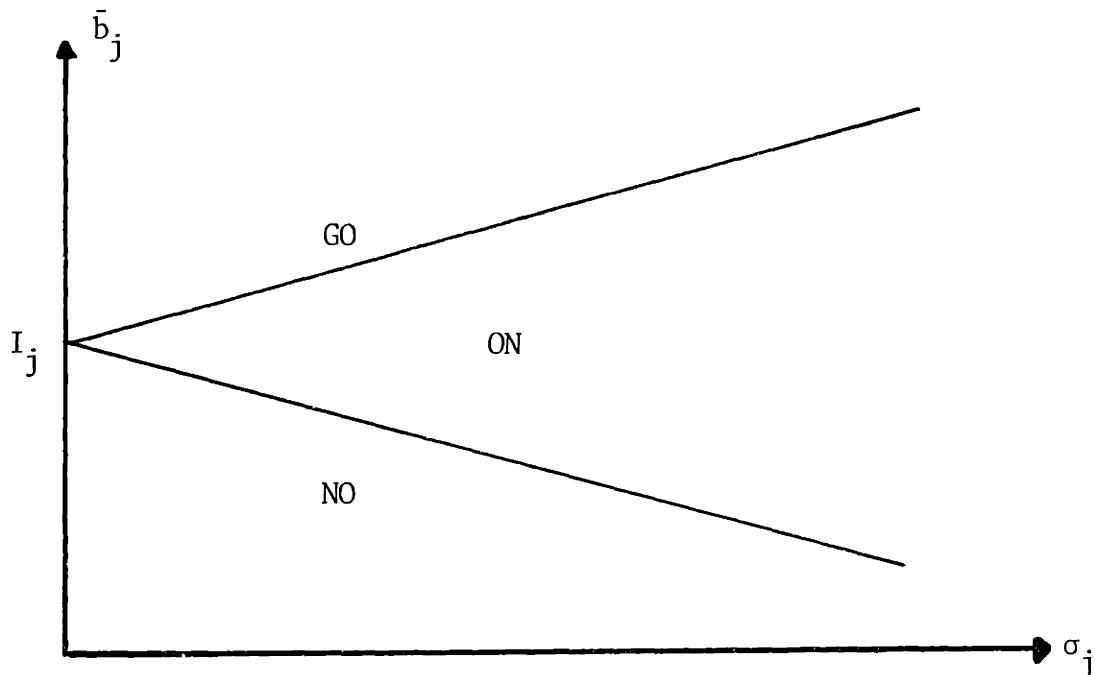


Figure 9.4: Decision Quadrant

Urban goes on to suggest ways to select among two or more GO alternatives, but what is most useful here is the very simple pictorial representation of uncertainty that can be used as an initial GO/ON/NO screen. This representation can prove very useful in communicating with managers. Note that the derivation here is for discounted monetary benefit, but it could just as easily have been modified for some other measure such as a target enrollment in an HMO.

9.1.4 Summary of Evaluation

In early screening when the choice process is not well understood the decision quadrant approach is very useful because of its ease of application and because it does not require detailed structuring of a manager's decision criteria. Later after the initial ideas are refined, the consumer choice process is better understood, and the manager's goals are better structured, decision theory and utility theory are very powerful tools to select optimal innovation strategies.

Early in the innovation process the uncertainty as represented by σ_j in figure 9.4 will be high and an ON decision will be likely. The next section discusses how the methodology guides refinement of the innovation strategy.

9.2 Refinement

The methodology is useful for early screening of ideas and later testing of alternatives, but its greatest contribution is that it focuses creativity by providing insight and understanding of the choice process through a evolutionary series of diagnostic information.

9.2.1 Refinement is Guided by the Individual Choice Models

By its very nature, refinement is a creative process. Rather than replace or circumvent the experience, insight, and creativity of the design team, the methodology works with and guides the creative process. Although all data is processed at the individual level, each model in the individual choice process provides aggregate numerical and pictoral indicators which help the design team understand (1) how consumers perceive the

alternatives, (2) what their preferences are relative to the way they perceive, (3) what segments are relevant, and (4) how all of this effects choice.

The design team couples this understanding with its knowledge (or models) of the "supply" side, i.e., cost, investment, political considerations, etc., to effectively refine the alternatives. These are then simulated to test and update the intuition of the design team. Further data, such as actual choice rather than stated preference, can be collected and more complete models calibrated, if the information they provide becomes necessary for further understanding of the choice process and strategic refinement of the innovations. This sequential evolutionary nature of the methodology insures that optimal use is made of collected data and that new data is collected only when the potential benefits are worth the added cost.

9.2.2 Diagnostics Available from the Individual Choice Models

Measurement: Although the attributes are too numerous for clear understanding and internalization, the identification of a complete set of attributes is useful to insure that the design team has not neglected what could be an important effect. Measurement of average perceptions of the attributes provides a check to see if any one or more are perceived as particularly bad or particularly good.

Reduction: A parsimonious set of performance measures are identified, named, and measured. Their average values, presented to the decision maker as perceptual maps, see figures 4.2 and 4.8, provide him with

understanding of the consumer perception process which he can easily internalize. In addition the representations of existing alternatives and the new alternative in average perceptual space provide useful insight on how the new alternative is perceived relative to its competition.

Abstraction: Potentially a number of alternative segmentation strategies are identified, each partitioning the population into groups homogeneous with respect either perception or preference. These segmentations prove useful in suggesting potential differentially targeted alternatives. Abstraction prevents the design of an average alternative which satisfies no one and encourages fitting new alternatives to promising segments.

Compaction: The primary purpose of compaction is to identify and structure consumers' preference. Both statistical compaction and direct assessment identify the relative importance (saliency) of the various performance measures. This identification allows innovation effort to be concentrated where it can best increase the "goodness measure," hence the choice probability, and ultimately the market share. In addition direct assessment identifies the risk characteristics of the performance measures giving the manager an idea of the relative (and absolute) importance of reliability for each performance measure and for the overall alternative. Compaction also identifies interaction effects, i.e., whether the performance measures act as substitutes or complements. This is important in design strategy because it indicates the effectiveness of concentrating on one performance measure versus simultaneously increasing two or more. Direct assessment provides separate measures for

these effects. (importances, risk characteristics, and interactions) but the manager also needs a representation of the simultaneous action of all the effects. The indifference curves of figure 6.16 provide a pictoral representation of this which is easy to comprehend.

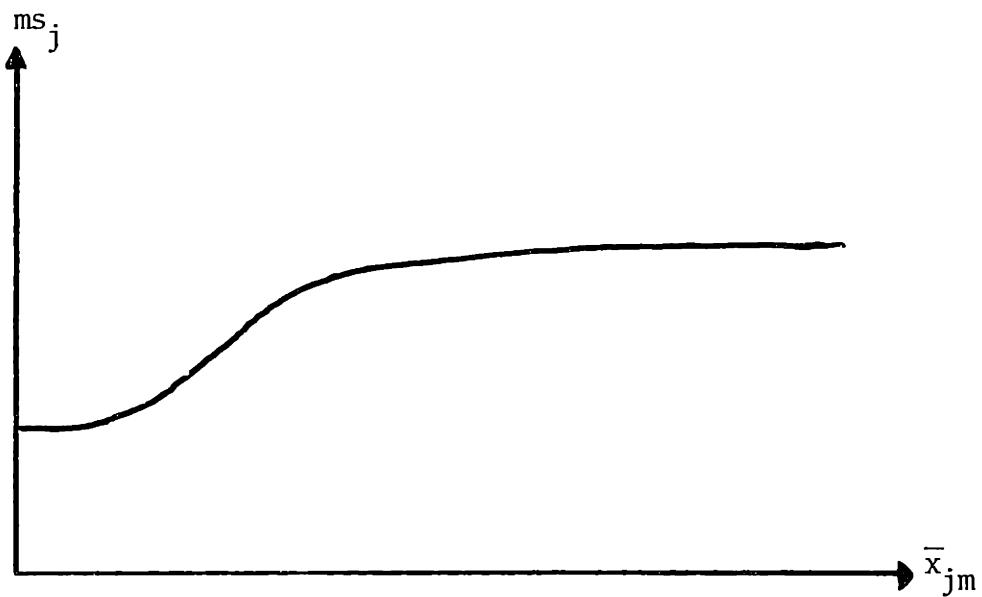
Probability of Choice: The main purpose of probability of choice is to transform the goodness measures into predicted numerical indications of choice, but diagnostically it does provide an indication of the strength of the compaction function.

Aggregation: As mentioned in evaluation, the outputs of aggregation help update a manager's intuition by providing estimates of the number of choices for a given simulated design. An alternative strategy would be to vary the mean or variance of one or more performance measures to produce curves which indicate aggregate share as a function of the level and/or reliability of a given performance measure. (Assuming competition remains constant.) See for example figure 9.5.

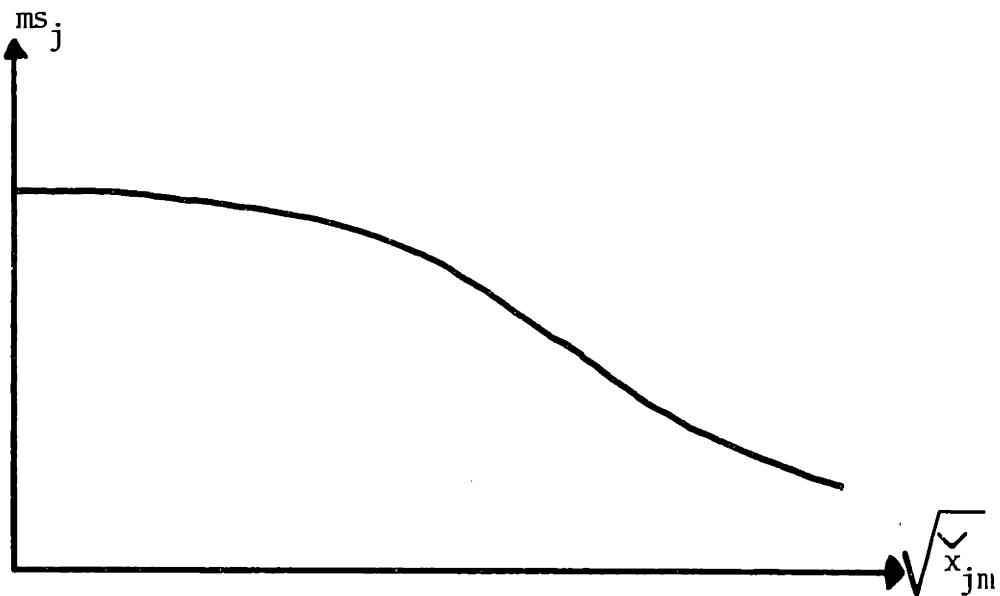
9.3 Conclusion of Interaction with the Design Process

In summary, the methodology facilitates testing and screening of alternatives by the evaluation step which uses the outputs of aggregation. But perhaps even more important, it facilitates successful innovation through interactions between the refinement models and the various submodels of the individual choice process.

This completes the micro-description of the methodology. The next chapter discusses formal tests which are applicable to the various submodels in the methodology.



$(x_k: k \neq j, x_{jm}, \bar{x}_{jm} \text{ constant})$



$(x_k: k \neq j, x_{jm}, \bar{x}_{jm} \text{ constant})$

Figure 9.5: Market Share Sensitivity

Chapter 10

FORMAL TESTS

The proposed methodology has many steps each of which allows the analyst to choose among many potential submodels. To make this choice some tests are needed to indicate the relative accuracy and usefulness of competing combinations of submodels. This chapter presents two types of tests.

Section 10.1 presents aggregate tests which act as indicators of the accuracy of the "aggregate" predictions but do not directly indicate the accuracy of the disaggregate, individual specific, predictions. Section 10.2 then presents a test which explicitly compares the predicted individual probabilities with the individual choice outcomes. This test, based on honest reward and information theory, provides useful benchmarks with which to measure both the accuracy and the usefulness of a particular model.

10.1 Aggregate Tests

There are basically two types of aggregate tests, those which test the rank order recovery due to compaction and those which compare predicted group response (share) with actual group response.

10.1.1 Rank Order Recovery due to Compaction

First preference: If compaction is successful then the stochastic preference model should closely approximate the certainty rule. Thus one

indicator of the accuracy of compaction is to count up the number of times an individual actually chooses the alternative which maximizes his scalar measure of goodness. The percent of individuals doing this is called the first preference recovery factor.

Rank order recovery: For a particular individual a compaction model may switch first and second preference but correctly predict the remaining rank order of preference, this is better than complete reversal of rank order. The first preference test is not sensitive to this effect. (Note that for this test to make any sense the observed choice must be rank ordering of proxy choices rather than actual choice.) One indicator of rank order recovery is to count the number of times an individual's first preference (from rank order) is the alternative with the largest scalar measure of goodness, plus the number of times second preference is the alternative with the second largest measure, and so on for the complete ranking of alternatives. The percent of times this occurs is the rank order recovery factor.

Use: Clearly if revealed preference is used the rank order recovery factor is inappropriate since the observed data is insufficient, but suppose that we have rank order data. Which test is preferred? Most likely there will be other factors influencing choice, thus it is suggested that both factors be calculated and if no model dominates with respect to both measures professional judgement and other tests should be used.

An example: Both measures can be readily computed from a match recovery table. Table 10.1 is the match recovery table for the statistical compaction described in section 6.4.

Table 10.1: Rank Order Recovery with Statistical Model

		Predicted Rank Order			
Actual Rank Order		1 st	2 nd	3 rd	4 th
		1 st	2 nd	3 rd	4 th
1 st		90	43	22	18
2 nd		56	65	34	18
3 rd		17	44	76	36
4 th		10	21	41	101

Chi-squared: One weak measure of accuracy is to test rank order recovery against a random model. A random model based on completely random rank ordering would result (for four alternatives) in the match recover table 10.2. Table 10.1 can be viewed as a chi-squared table and tested against the model for table 10.2.

I.e., let n = number of individuals

m = number of alternatives which are rank ordered

$n_{j\ell}$ = number of individuals who ranked an alternative as the j^{th} alternative but for whom the compaction model ranked as ℓ^{th} .

Table 10.2: Rank Order Recovery with Random Model

		Predicted Rank Order			
		1 st	2 nd	3 rd	4 th
Actual Rank Order	1 st	.25	.25	.25	.25
	2 nd	.25	.25	.25	.25
	3 rd	.25	.25	.25	.25
	4 th	.25	.25	.25	.25
1 st preference (fraction)		rank order (fraction)			
matches:		(.25)		(.25)	

then

$$\chi^2 = \sum_{j=1}^m \sum_{\ell=1}^m \frac{(n_{j\ell} - nm/m^2)^2}{nm/m^2}$$

where χ^2 is approximately Chi-squared distributed with $m^2 - 2m + 1$ degrees of freedom. Use a right-hand tail test, i.e., reject random hypothesis if χ^2 is large.

10.1.2 Tests on Predicted Share

Certainly a necessary, but far from sufficient, test of a combination of submodels is that they accurately predict aggregate (market) share. To see that such a test is not sufficient consider the naive probability model which assigns to each individual the observed market share. The expected value of the market share of such a model will always reproduce the observed share, but such a model could hardly be considered to improve

our understanding of the choice process. None the less such aggregate tests do serve the useful purpose of a sieve to remove bad models.

Least squares: For aggregate shares both the predictions and observations are real numbers in the interval $[0,1]$. Thus, the traditional least squared error test is not inappropriate, i.e.,

$$LS = \sum_{j=1}^J ms_j^0 \cdot |ms_j^0 - \hat{ms}_j|^2$$

where ms_j^0 = observed share of alternative a_j

\hat{ms}_j = predicted share of alternative a_j

Of course one may or may not wish to weight the squared terms by ms_j^0 and one may wish to use some power other than squared, e.g., absolute deviation.

Chi-squared: The predicted market share can be viewed as a hypothesized frequency distribution and compared against the model of observed choice being drawings from a Bernoulli population with probabilities equal to the predicted market share. I.e., let

n = number of individuals

n_j = number of individuals choosing alternative a_j

ms_j = predicted share of alternative a_j , $ms_j \in [0,1]$

then

$$\chi_s^2 = \sum_{j=1}^m \frac{(n_j - n \cdot ms_j)^2}{n \cdot ms_j}$$

where χ_s^2 is approximately chi-squared distributed with $m-1$ degrees of freedom. Use a left of threshold test, i.e., do not reject hypothesis if χ_s^2 is small.

10.2 Disaggregate Tests

The aggregate tests serve the useful purpose of screening out particularly bad models, but they are not consistent with the fundamental disaggregate philosophy of the methodology. What we need is a test which is sensitive-on an individual level-to the explicit comparison of the predicted choice probabilities with the observed outcomes which are 0-1 events. (I.e., an individual either chooses alternative a_j or he does not.) Still since competing submodels are to be compared, the test must produce a single measure by which to rank models.

This section begins by showing the inappropriateness of traditional tests. Based on these shortcomings a number of formal criteria are identified which imply a unique set of tests with easily interpretable "benchmarks."

10.2.1 Traditional Tests

Least squares: Perhaps the most traditional test is the least squared error test. In this case, unlike in section 10.1, it would be applied to individual predictions, i.e.,

let $\delta_{ij} = \begin{cases} 1 & \text{if individual } i \text{ chooses alternative } a_j \\ 0 & \text{otherwise} \end{cases}$

$$p_{ij} = p_s(a_j | c_i)$$

n = number of individuals in the sample

then

$$\ell s = \frac{1}{n} \sum_{i=1}^n \ell s_i = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J (\delta_{ij} - p_{ij})^2$$

Although $\sum_j \delta_{ij} = \sum_j p_{ij} = 1$ for all i , the data are of different types;

δ_{ij} is discrete 0 or 1 while p_{ij} is continuous between 0 and 1, i.e.,

$(0 < p_{ij} < 1)$. Thus we begin to wonder whether such a formulation can accurately test individual outcomes against predicted probabilities.

Let us examine¹ the test in terms of a television weatherman, whom we shall call Dylan. Now Dylan was trained at a great eastern technological school and can come up with probabilities, $p_i = (p_{i1}, p_{i2}, \dots, p_{iJ})$, which he feels truly reflect his best predictions for the J states of tomorrow's weather. Dylan's employer knows Dylan is risk neutral with respect to salary and plans to pay Dylan a bonus. This bonus is to be based on the probabilities Dylan announces on the 6:00 news, $q_i = (q_{i1}, q_{i2}, \dots, q_{iJ})$, and the next day's weather, $\delta_{ij} = 1$ if state j occurs, 0 otherwise. If the bonus is $(J - \ell s_i)$, should Dylan announce $q_i = p_i$?

The answer is no! He should not announce the values he believes! Consider the following.

Since Dylan is risk neutral, his appropriate criterion is to select q_i to minimize the expected value of ℓs_i . For the least squares test this is:

$$E[\ell s_i] = \sum_{j=1}^J p_{ij} \cdot [(1 - q_{ij})^2 + \sum_{k \neq j} (0 - q_{ik})^2]$$

By explicitly writing out this expression, adding the constraint $\sum_j q_{ij} = 1$ with a Lagrange multiplier, and by differentiating $E(\ell s_i) + \gamma [\sum q_{ij} - 1]$ with respect to q_{ik} for all k , it is possible to show that Dylan's choice of q_i should be to set $q_{ij} = 1.0$ for the weather state with the highest probability ($\max_k \{p_{ik}\}$) and $q_{ij} = 0.0$ for all other states. Since this test causes Dylan to give dishonest answers when the criterion is to optimize the expected value of the test, we will call ℓs dishonest.²

Maximum score: Another intuitive test is to simply count up the number of correct predictions. That is count up the number of times an individual chooses the outcome which was predicted as having the highest probability. This test, which is equivalent to the first preference aggregate test (since probabilities are monotonic in compaction values), is unsatisfactory for a disaggregate test since it has poor discriminability, i.e., it gives exactly the same "reward" if the predicted probability was .5001 or .9999.

Chi-squared aggregate tests: Section 10.1 proposed two aggregate tests based on the chi-squared distribution. Both tests are theoretically sound and are useful screening tests, but neither test is sensitive to the comparison of individual probabilities and outcomes and neither test provides particularly strong measures with which to rank competing models.

Thus, we see that some test must be designed which is honest in the expected reward sense, discriminates well, and is sensitive to individual predictions and outcomes. The next subsection formalizes these

criteria.

10.2.2 Criteria for Disaggregate Tests

(The first five criteria stated in this section were formalized by Raiffa [119], they are restated here in terms of our notation.)

Criterion 1, disaggregate: Clearly the test must be sensitive to the individual predicted probabilities \underline{p}_i and the individual outcomes, $\underline{\delta}_i = (\delta_{i1}, \delta_{i2}, \dots, \delta_{iJ})$. We will consider tests which give a "reward", $r(\underline{p}_i, \underline{\delta}_i)$, to an each individual choice event and sum these over the population, i.e.,

$$R = \sum_{i=1}^n r(\underline{p}_i, \underline{\delta}_i)$$

Criterion 2, honesty: As stated earlier the probabilities the model gives, \underline{q}_i , must be equal to the probabilities it believes, \underline{p}_i . In other word the believed expected reward, $E(R)$, must be maximized when $\underline{q}_i \equiv \underline{p}_i$ where:

$$E(R) = \sum_i \sum_j p_{ij} \cdot r(\underline{q}_i, a_j \text{ chosen})$$

Criterion 3, relevance: If a_j is chosen and that is all that is observed then the reward given to that outcome should only depend on p_{ij} not on the entire vector of probabilities, \underline{p}_i . Formally:

$$r(\underline{p}_i, a_j \text{ chosen}) = r_j(p_{ij}, a_j \text{ chosen})$$

Criterion 4, invariance: One of the desirable properties of the probability model was that it be symmetric (section 7.1.1). It is desirable that the reward function exhibit a similar property, that is, if $p_{ij} = p_{ik}$ then the reward should be the same whether a_j or a_k is chosen. Formally:

$$r_j(p_{ij}, a_j \text{ chosen}) = r_k(p_{ik}, a_k \text{ chosen})$$

or in other words drop the subscript on r :

$$r_j(p_{ij}, a_j \text{ chosen}) = r(p_{ij}, a_j \text{ chosen})$$

Criterion 5, strong discriminability: If our weather man, Dylan, tried to predict the exact temperature and got it correct he would expect a higher reward than if he just tried for the range. Similarly suppose a person had a hierarchical decision process, that is he first chose the type of health care delivery and then the plan or doctor. We would want to give a reward first for prediction of the type of delivery system and then for the exact plan or doctor. Furthermore we would want a reward which is higher for the composite prediction (delivery and plan) than the sum of the rewards for the parts of the prediction. Stated mathematically $r(p_{1j}, \delta_{1j} = 1) + r(p_{2k}, \delta_{2k} = 1)$ should be monotonically increasing in the probability of the composite event (Think of the subscript $i = 1, 2$ as indexing days for Dylan, individuals for us). In the case of an independent model (section 7.5) the probability of this composite event is $p_{1j}p_{2j}$. Formally:

$$r(p_{ij}, \delta_i) + r(p_{hk}, \delta_h)$$

is monotone increasing in $p_{ij}p_{hk}$.

Criterion 6, benchmarks: To Raiffa's formal criteria add the following informal criterion: "The test should provide reasonable rules of thumb, or benchmarks, with which to get an intuitive feel for the relative accuracy and usefulness of a probability model, $p(a_j | c_i)$."

10.2.3 Formal Tests: Theory

Honest reward functions: Based on the first five criteria, Raiffa [119] identifies a test which satisfies them and he shows that this test is unique to a positive linear transformation. Formally:

Theorem 10.1 [Raiffa]: There exists a unique reward function satisfying criteria 1,2,3,4, and 5. It is given by

$$r(p_i, a_j \text{ chosen}) = A \log p_{ij} + B$$

where A,B are constants and $A > 0$.

All that remains is to chose the constants A and B to satisfy criterion 6, that is to provide useful benchmarks.

Information Theory: The probability model, $p(a_j | c_i)$ for all j, can be viewed as an information system. In other words the "observable occurrence," i.e., the "outcome" or exact value of c_i , is providing information about "unobservable events," i.e., about the choice outcome, $a_j \in A$. Thus we can use the information measure, $I(a_j, c_i)$, (Gallagher [42]) to quantify the information provided by c_i . Formally:

$$I(a_j, c_i) = \log \frac{p(a_j | c_i)}{p(a_j)}$$

where $p(a_j)$ is the prior likelihood of the outcome, a_j chosen.

Now the information criteria does satisfy criterion 6 in that it provides benchmarks. The first benchmark is the expected information provided by the model, $I(A, C)$, where:

$$I(A, C) = \sum_{c_i \in C} \sum_j p(a_j, c_i) \log \frac{p(a_j | c_i)}{p(a_j)}$$

with $p(a_j, c_i)$ the joint probability of an "observation" of c_i and an "event," a_j chosen.

Another benchmark is the total uncertainty in the system which is measured by the prior entropy, $H(A)$, where:

$$H(A) = - \sum_j p(a_j) \log p(a_j)$$

The prior entropy measures the uncertainty before "observing" c_i . After observing c_i the uncertainty is reduced to the posterior entropy, $H(A|C)$, where

$$H(A|C) = - \sum_{c_i \in C} \sum_j p(a_j, c_i) \log p(a_j | c_i)$$

It can be shown (Gallagher [42]) that the amount by which c_i reduces the entropy is exactly the expected information provided by C , i.e., $H(A) - H(A|C) = I(A, C)$. Thus, even before the empirical reward is calculated the "usefulness" of the model can be calculated by comparing the expected information, $I(A, C)$, to the entropy, $H(A)$.

Note that for a sample of n individuals we can use $p(a_j, c_i) \equiv p(a_j | c_i)p(c_i)$ by setting $p(c_i) = (\# \text{ of times } c_i \text{ occurs})/n$ and by setting $p(a_j)$ either equal to the observed market share fraction, ms_j (naive

model), or equal to $1/(\# \text{ alternatives})$ (equally likely model), or to any other prior belief on $p(a_j)$. For comparing $p(a_j | c_i)$ against the naive model:

$$I(A, C) = \sum_i \sum_j (1/n) p(a_j | c_i) \log \frac{p(a_j | c_i)}{m_s_j} \quad \text{equation 10.1}$$

and

$$H(A) = -\sum_j m_s_j \cdot \log (m_s_j) \quad \text{equation 10.2}$$

Note that since $0 \leq m_s_j \leq 1$, $H(A)$ is positive.

The accuracy of the model is then calculated by comparing the empirical information, $I(\text{observed})$, with the expected information. To compute the empirical reward use the δ_{ij} notation where $\delta_{ij} = 1.0$ when individual i chooses alternatives a_j and $\delta_{ij} = 0$ otherwise. Then

$$I(\text{observed}) = (1/n) \sum_i \sum_j \delta_{ij} \log \frac{p(a_j | c_i)}{m_s_j} \quad \text{equation 10.3}$$

Equivalence: So far the honest reward function satisfies criteria 1-5 and the intuitively pleasing information measure satisfies criterion 6. Which one is better? Fortunately the decision need not be made since the observed information, $I(\text{observed})$, is an honest reward function. This is shown by the following simple theorem:

Theorem 10.2: The observed information measure, $I(\text{observed})$ is an honest reward function as specified by theorem 10.1.

Proof: For an honest reward function measuring the accuracy of the probability model, $p(a_j | c_i)$, the group reward, R , is given by:

$$R = \sum_i \sum_j \delta_{ij} [A \log p(a_j | c_i) + B]$$

choose $A = 1/n$ and $B = H(A) = -\sum_k m s_k \log m s_k$

Then:

$$R = (1/n) \sum_i \sum_j \delta_{ij} \log p(a_j | c_i) - \sum_i \sum_j \delta_{ij} \sum_k m s_k \log m s_k$$

but

$$\begin{aligned} \sum_i \sum_j \delta_{ij} \sum_k m s_k \log m s_k &= \sum_i \sum_k m s_k \log m s_k \\ &= \sum_i \sum_k (1/n) \delta_{ik} \log m s_k \end{aligned}$$

thus switching the index k back to j and collecting terms gives:

$$R = (1/n) \sum_i \sum_j \delta_{ij} (\log p(a_j | c_i) - \log m s_j)$$

which is just $I(\text{observed})$ since $\log p(a_j | c_i) - \log m s_j$ is equal to

$$\log \frac{p(a_j | c_i)}{m s_j} .$$

Monotonic in likelihood ratio: An additional useful property of $I(\text{observed})$ is that it is monotonic in the likelihood ratio of the probability model to the naive model. To see this notice that $I(\text{observed})$

can be rewritten as:

$$I(\text{observed}) = (1/n) \log \left[\frac{\prod_{i=1}^n \prod_{j=1}^J p(a_j | c_i) \delta_{ij}}{\prod_{i=1}^n \prod_{j=1}^J p(a_j) \delta_{ij}} \right]$$

Notice that $X^{\delta_{ij}}$ equals X if $\delta_{ij} = 1$ and equals 1 (and thus drops out of the product) if $\delta_{ij} = 0$. Thus the quantity in brackets is indeed one form of a likelihood ratio.³

10.2.4 Formal Test: Use

Step 1, calculate the entropy: This first step sets up a goal because not only is the entropy, (equation 10.2), a measure of our prior uncertainty but also a measure of how well a perfect model would perform. A perfect model would assign $p_{ij} = 1$ if a_j is chosen and $p_{ij} = 0$ otherwise. In other words deterministic prediction. To see this substitute $p(a_j | c_i) = \delta_{ij}$ into equation 10.3:

$$I(\text{observed}) = (1/n) \sum_i \sum_j \delta_{ij} \log \frac{\delta_{ij}}{m s_j}$$

$$= (1/n) \sum_j (\# \text{ times } a_j \text{ chosen}) (-\log m s_j)$$

$$+ (1/n) \sum_j (\# \text{ times } a_j \text{ not chosen}) \cdot 0 \cdot \log \frac{0}{m s_j}$$

$$= -\sum_j m s_j \log m s_j = H(A)$$

Where the last substitution is made recognizing $\lim_{x \rightarrow 0} (x \log x) = 0$

and $ms_j = (\# \text{ times } a_j \text{ chosen})/n$. Note that in calculating the entropy the analyst has a choice of using (1) prior observed market share, (2) posterior observed market share, (3) prior belief on the market share, (4) predicted market share, or (5) equally likely share, i.e., $ms_j = 1/J$. The interpretation of the test then depends on the analyst's choice of ms_j .

Step 2; calculate the expected information: This second step tests the usefulness of the model because the expected information, (equation 10.1), measures how much uncertainty the model $p(a_j | c_i)$ reduces. Thus a comparison of $I(A,C)$ with $H(A)$ gives the percent of uncertainty reduced or, in another interpretation, the percent of information provided.

Step 3, calculate the empirical information: This third step tests the accuracy of the model, because if $p(a_j | c_i)$ is the best measure of choice based on c_i then $I(A,C)$ is the true expected value of $I(\text{observed})$. Thus for a large sample by the weak law of large numbers (Drake [29]), we expect $I(\text{observed})$ to be close to its expected value.⁴

Step 4, pictoral representation: A one-dimensional plot as in figure 10.1 is useful in visually assessing the usefulness and accuracy of the model. Notice that the 0-point is explicitly plotted for comparison. This is to provide a visual comparison of the ratio of $I(A,C)$ to $H(A)$, and because any model with $I(\text{observed})$ less than zero is counterproductive in the sense that it does worse than the naive model.

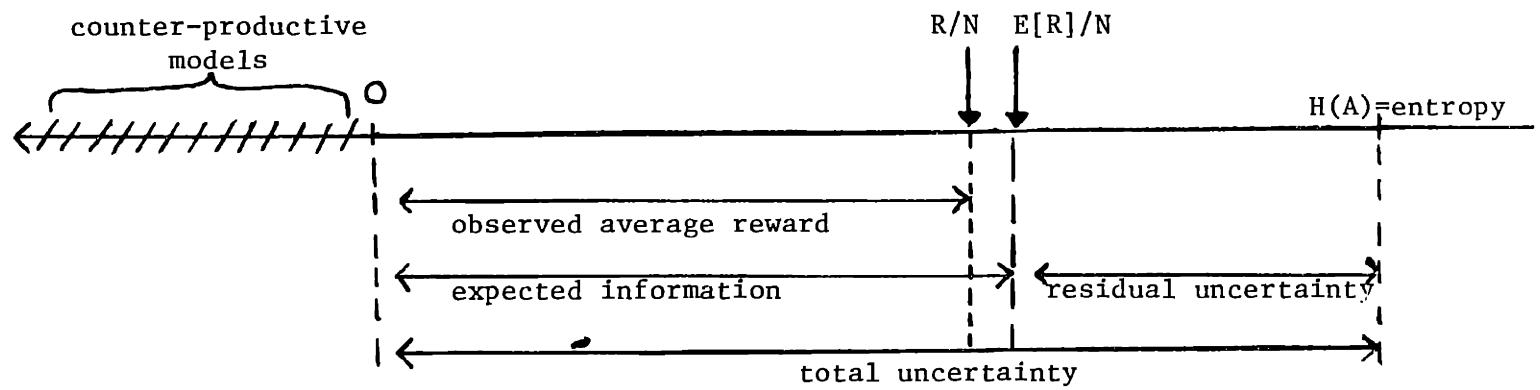


Figure 10.1: Schematic of the Information Test

Step 5, comparing competing models: The model with the larger $I(A,C)$ is the more useful model, as long as $I(\text{observed})$ is close to $I(A,C)$ for the model. (Note: depending on his interpretation the analyst may wish to use $I(\text{observed})$ as the measure with which to rank the model.) Comparing $I(A,C)$ to $H(A)$ is analogous to comparing R^2 to 1.0 in regression, except that information and entropy are "natural" measures and satisfy the six criteria of section 10.2.2. Comparing $I(\text{observed})$ to $I(A,C)$ is somewhat analogous to an F-test in regression.

10.3 Conclusion of Testing

This chapter presented two basic types of tests. The aggregate tests screen out particularly bad models and give managers confidence in good models, while the disaggregate test considers the fundamental individual specific prediction of the methodology. In any application these tests should be used to complement each other and act as inputs to the choice of models. The ultimate choice is left to the judgement of the analyst who can implicitly consider external effects which no test can statistically measure.

Chapter 11

CASE STUDY: THE DESIGN OF A HEALTH MAINTENANCE ORGANIZATION

The emphasis of the preceding chapters has been with the technical detail of the various models in the methodology. In each chapter some of these models, especially the new models, were illustrated on data collected to aid in the design of a health maintenance organization (HMO) for the MIT Medical Department. These empirical examples are summarized in this chapter. Presenting them together illustrates the interrelationships of the various modules in the methodology and indicates how they combine to form a useful consumer response model. Section 11.1 summarizes the empirical experience on two separate paths through the methodology and section 11.2 relates the empirical experience to the managerial issues of the case.

11.1 Summary of Empirical Experience

The empirical studies done for the M.I.T. HMO case can be summarized by two parallel paths. See figure 11.1. Path 1 is a typical implementation based on statistical compaction across individuals. Path 2 is a typical implementation based on an individual specific compaction (assessment)¹ technique. Both paths use the same consumer observation models.

¹This empirical study was jointly done by Glen Urban and John Hauser. This chapter is similar to that presented in Hauser and Urban [56].

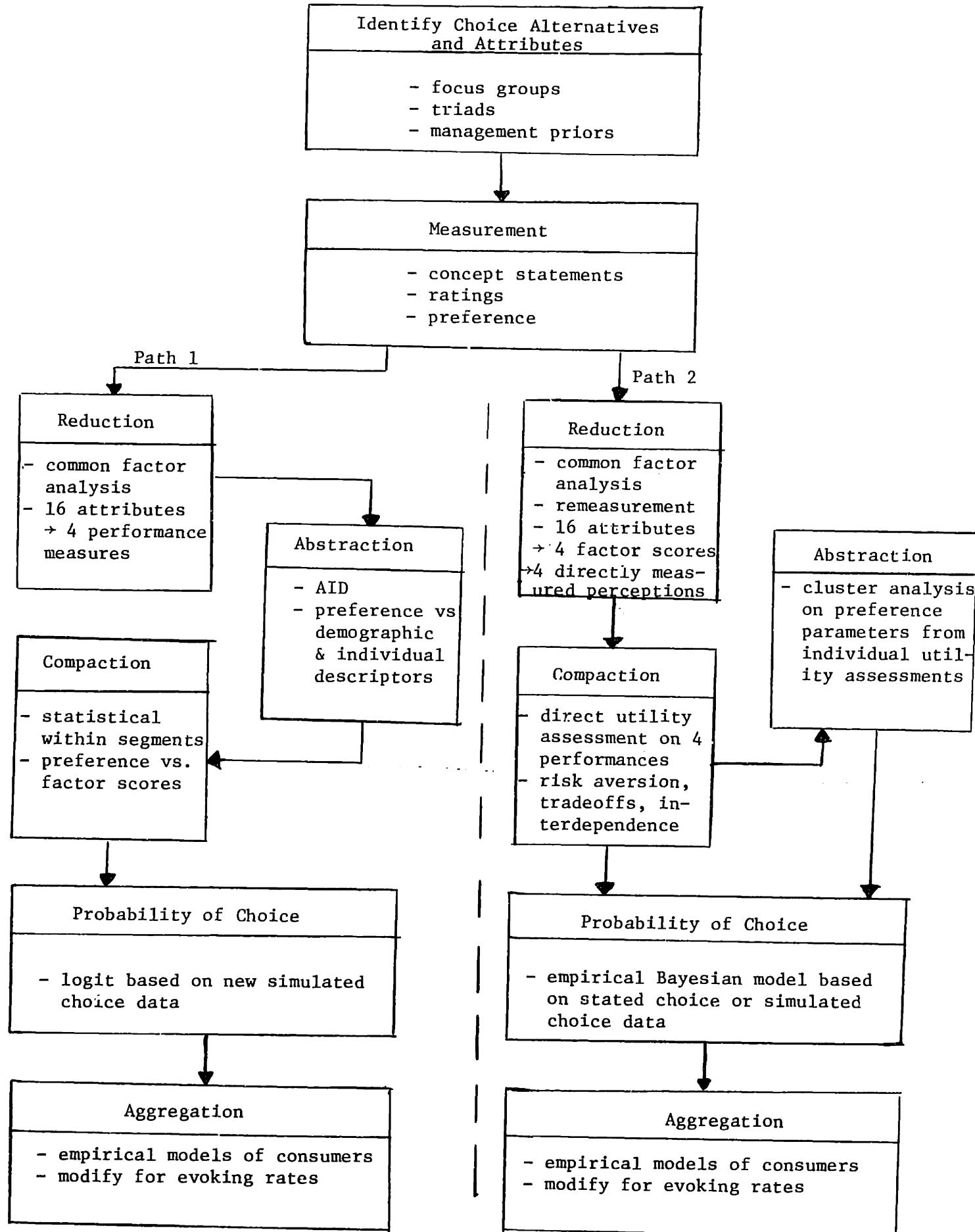


Figure 11.1: Alternative Empirical Analyses of HMO Data

Common Models

Consumer observation (Chapter 3): The first step was to identify the choice alternatives and the attributes consumers use to describe those alternatives. Preliminary open-ended surveys, focus groups and triads were used in the HMO case. The evoked set consisted of three concept statements (figure 3.2), a new M.I.T. HMO, the Harvard Community Health Plan, and a mythical Massachusetts Health Foundation, as well as the existing pattern of care. The attributes were represented by 16 five-point agree/disagree scales identified by consumers as relevant to their choice of health plan (figure 3.6). The consumers rated each plan on the 16 scales, stated rank order preference, and answered demographic and pattern of care questions (mailed survey in appendix 1).

Path 1: Statistical Compaction and Complementary Models

Reduction (Section 4.2.1.3): The 16 scales were too numerous to give any useful insight on perceptions or preference, thus these attributes were reduced to a set of underlying performance measures. Common factor analysis (figure 4.7) identified four performance measures as being relevant to health care: quality, value, personalness, and convenience. For input into later models, estimates of these measures were determined for each individual for each plan, but to communicate with management perceptual maps (figure 4.8) representing the position of each plan were determined based on the average factor scores. These maps proved useful in identifying needed improvements in hospital affiliation and in communication strategy.

Abstraction (Section 5.2): There is a fundamental paradox in abstracting segments when statistical compaction is used. One would like to abstract based on the preference parameters, but the preference parameters can not be determined until statistical compaction is done within segments. Thus segments must be abstracted by indirect means. In path 1 two techniques were tried, cluster analysis on perceptions and AID analysis with preference vs. demographics. Cluster analysis (Section 5.2.1) was unsuccessful in explaining more variance on the empirical sample than it could on a random sample. AID indicated that an individual's previous pattern of care effected his choice of plans, but the results were not very strong. A less formal abstraction method is prior beliefs and strategic relevance. In the HMO case segmentation by student-staff-faculty was tried and the results of statistical compaction did not dispel these beliefs. (Statistical compaction was also done with segments based on pattern of care. The results were not significantly stronger than those based on student-staff-faculty.)

Compaction (Section 6.4.1): The statistical technique used in the HMO case was preference regression with the dependent variable rank order preference and the explanatory variables the factor scores of quality, value, personalness, and convenience for each individual for each plan (tables 6.1 and 6.2). Overall, quality and value were found to be most important but they did not dominate personalness and convenience. When the regressions were done within segments it was found that significant variation did occur with faculty most concerned with quality, students with value, and staff with personalness. This suggested management might consider offering a variety of plans or using different

communication strategies within segments. Together the segment regressions had a first preference recovery of 48% and a rank order recovery of 52%.

Probability of Choice (Section 7.2): As of this writing the statistical compaction values have not been formally linked to probability of choice. Potential models for this are the recursive binary logit formulation, the multinomial logit model, or the empirical Bayesian model. An interim approach is to use rules of thumb such as using the compaction values to rank order the plans for each individual and assigning probabilities based on rank order alone. These are obtained by comparing stated rank order preference to predicted rank order preference.

Aggregation (Section 8.6): Aggregation was performed in the HMO case by using the preference regression models with corrections for awareness and with corrections for comparison of the base case predictions with actual choice (figure 8.4). Judgemental shifts in the mean values of the performance measures predicted that an M.I.T. HMO with the existing design would just break even, but if hospital affiliation and communication strategy were improved enough enrollment could be expected to justify the risks. This completes path 1; the managerial issues of the case will be discussed in section 11.2.

Path 2: Compaction by Direct Assessment and Complementary Models

Reduction (Section 4.2.1.3): Path 2 also begins with factor analysis, but directly measures each individual's perceptions on the reduced performance measures. This is necessary because the individual

specific compaction functions are assessed with respect to direct perceptions. The factor scores can then be correlated to the directly measured perceptions.

Compaction (Section 6.4.2): Using a personal interview questionnaire (appendix 2) based on the axiomization of section 6.2, individual specific compaction functions were directly assessed for a random sample of M.I.T. students. Quality and value were again the most important but with personalness and convenience also significant (table 6.3). The students viewed the performance measures as substitutes rather than complements and they were risk averse with respect to all of the measures (table 6.3). Although the preference parameters are kept individual specific for prediction, their mean values were used to produce indifference curves to indicate how much quality and convenience are necessary to support a given price level (figure 6.16). The assumptions necessary for the multiplicative form were satisfied in 39-64% of the cases and approximately satisfied in 66-89% of the cases (table 6.4). The first preference recovery was 49.5% and the rank order recovery was 47.4% (table 6.5). This first experiment indicates the potential feasibility of direct assessment and the appropriateness of a stochastic definition of preference.

Abstraction (Section 5.2.1): Individual specific compaction makes it possible to directly abstract segments based on homogeneity of preference parameters. These segments can be abstracted by using cluster analysis on the preference parameters. This has not yet been done.

Probability of Choice (Section 7.3): The empirical Bayesian model was formulated especially to be compatible with direct assessment of

compaction functions. This model has not yet been calibrated on the HMO data, but section 7.3.5 presents an empirical example calibrated on constant sum paired comparision data collected prior to introduction of a nationally known aerosol deoderant. This calibration indicates the feasibility of the technique and illustrates that significant information about choice is carried in the full vector of compaction values rather than just in their rank order.

Aggregation (Section 8.1, 8.2, and 8.3): When the choice probability model is calibrated for the HMO case, aggregation will be done using the central limit theorem with corrections for evoking. New alternatives will be tested by using the individual compaction functions as a simulated consumer pool.

Comparison of Paths 1 and 2

Path 2 has the relative advantages of axiomatic identification of functional forms, individual specific preference parameters, and abstraction based on the individual parameters. Its main disadvantages are that it requires extensive individual specific measurement and has no structure to explicitly deal with measurement error. Path 1 has the relative advantage of easier measurement and a proven method of compaction within a segment, but it does not provide individual specific parameters and is limited to abstraction by indirect means. Both paths provide useful insight to the managerial design team.

11.2 Managerial Issues

A health maintenance organization is a very appealing concept. For the consumer it could mean more convenient, higher quality care with less financial risk. (It features preventative care, attracts good doctors, has specialists available, guarantees service for a fixed monthly fee, has a central facility, makes it easier to get the right doctor, and is open long hours.) For the staff it could mean guaranteed income with shorter hours, more peer contact, and freedom from minor administrative chores. For the organization itself, the stable revenues enable centralized manpower and financial planning. Despite these advantages an HMO represents considerable risk to the organization because it requires significant investment and long term planning and because of the uncertainty in consumer enrollment. Thus, an organization needs consumer response models to predict enrollment. Furthermore an HMO is not a simple product, it is a complex service which may be a success or failure depending upon how well it is designed. To develop a "best" design an organization needs a diagnostic consumer model. Thus, both predictive and diagnostic models were needed to analyze a decision by M.I.T. to expand its HMO from a pilot program of roughly 1000 families.

The methodology in its evaluation mode is a predictive model. New enrollments were estimated based on stated preference and intent and upon estimated awareness. These together with estimates of re-enrollment and estimates of migration out of the M.I.T. community forecast a total enrollment of 3600 families (see figure 8.4). Since this was just financially sufficient to maintain an HMO, the inherent risk indicated that

a decision to expand the existing pilot HMO could not be supported based on the existing design.

However, the research demonstrates clearly the existing pilot was not the best design. The individual choice models provided useful diagnostics which suggested design changes. Reduction identified the need for aggressive communication to close the gap between perceptions and performance and for a change in hospital affiliation. Compaction emphasized the importance of quality but cautioned that value, convenience, and personalness should not be neglected. Abstraction suggested either a variety of plans or different communication strategies within segments. The combined models also showed that students as well as staff and faculty represented a potential market and that most likely students were needed to ensure sufficient enrollment for financial viability.

The individual choice models were then used to simulate these design changes. These simulations indicated that an improved design would attract sufficient enrollment even under competitive pressure from Harvard Community Health Plan (See Section 8.6).

M.I.T. is now expanding its HMO to meet the indicated need and more aggressively marketing its HMO to faculty, students, and staff as facilities become available. These decisions were aided by the model but were also the result of other considerations. For example, M.I.T. could not simply change its hospital affiliation because of the need to have its doctors accredited at the hospitals. Besides the M.I.T. hospitals are quality hospitals, just not perceived as such. Thus the strategy is to make more visible use of the better Cambridge hospital and actively communicate its quality. Other considerations such as financial risk were

also important in the decision to expand at the existing facility rather than build a new facility.

Based upon this initial application it appears that the methodology is relevant to the management of innovation and can be useful in improving designs of new products and services.

Chapter 12

CONCLUSION

An important strategic problem faced by both the private and the public sectors is how to design and introduce innovative products and services. Such innovation is linked to increased effectiveness and productivity, but often represents a high risk to the organization since the success or failure of innovation is dependent upon consumer response. A methodology which models consumer response and is compatible with the managerial decision process is necessary to increase the rate of success in innovation.

12.1 Summary

This dissertation has presented a model based methodology to improve the effectiveness of the design and implementation of innovative products and services. The methodology couples a consumer response process based on state-of-the-art knowledge in the fields of psychometrics, utility theory, and stochastic choice theory, with a parallel managerial design process which encourages effective integration of model outputs into creative design efforts. The design process consists of idea generation, evaluation, and refinement, while the consumer response process is based on consumer measurement, models of the individual choice process, and aggregation of individual choice predictions. The consumer models aid idea generation and refinement by providing diagnostics on consumer perceptions, preferences, choice, and segmentation. They aid

evaluation by providing numerical predictions of consumer choice.

The preceding chapters have discussed in detail the various models in the methodology by presenting a formal development of the issues and a review and exposition of relevant existing models. In addition new models were introduced which promise to prove useful in modeling consumer response. Many of the models were illustrated by application to consumer data collected to support the design of a new health maintenance organization (HMO) at M.I.T. After a formal discussion of testing techniques for the models, the managerial design implications were illustrated by application to the M.I.T. HMO case.

12.2 Contribution of the Research

The development of this methodology which is presented in Hauser and Urban [56] and repeated in this document is a first test of a set of complementary models which are oriented towards aiding the creative design and implementation of innovation. Besides the methodology itself, which is a useful communication device to bring together diverse models in psychometrics, utility theory, and stochastic choice theory to solve an important problem, this joint research has produced (1) new measurement instruments to allow mass direct assessment of consumers' compaction functions, (2) the use of psychometrics to get a complete and parsimonious set of performance measures to make mass assessment feasible, (3) a method of directly abstracting segments based on preference, (4) a stochastic interpretation of preference and (5) an empirical Bayesian model which computes choice probabilities based on compaction values. Perhaps most important, these new techniques were demonstrated

by application to the empirical problem of design a new HMO.

This dissertation contains much of the formal theoretic development necessary to support the new models in the methodology. Chapter 6 defines a form of stochastic preference and identifies a psychological choice axiom which together with a restatement of the von Neumann-Morgenstern axioms establishes an isomorphism between utility theory and compaction theory and makes possible the application of prescriptive choice theorems to descriptive choice. These "utility" theorems establish the existence and uniqueness of a cardinal compaction function, identify relatively simple functional forms, and establish measurement techniques to assess the parameters of these functions and to verify the assumptions necessary for these forms. Chapter 7 presents the formal development of the empirical Bayesian model based on rank order events and "tuning" of rank order probabilities. The model is shown to be consistent and thus asymptotically independent of any peculiarities in a data set. Chapter 8 shows that with little additional effort the central limit theorem can be used to determine variances and covariances of total choices in addition to the means. Chapter 10 presents the formal development of an information theoretic/honest reward test for the methodology. The advantage of this test is that it compares predicted probabilities and observed outcomes on an individual level and provides useful benchmarks with which to judge the model being tested.

This dissertation also briefly summarizes existing models in psychometrics, utility theory, and stochastic choice theory. The formal definitions in chapters 3 through 8 and the common notation used in the presentation of the issues and mathematics of the models enable

practitioners to compare the models and to choose that set of complementary models which is best suited for the problem to be solved.

12.3 Suggestions for Future Research

Further work is suggested in theoretical development, empirical comparison of models, and in application.

Theoretical Development

Perhaps the most crucial theoretic development needed is an explicit modeling of error sources in direct utility assessment to feasibly handle complex functional forms such as multiplicative or quasi-additive with concave (risk averse) conditional utility functions. It is possible to model the error sources with integral equations (chapter 8) and under strong assumptions develop analytic models (McFadden [93]), but for complex utility functions these equations become too complex. Furthermore, the direct assessment techniques development in chapter 6 exactly specify the utility function and leave no degrees of freedom for estimation. Maximum likelihood techniques to estimate these complex functions based on feasible redundant measurement would be extremely useful and most likely increase the accuracy of the methodology.

Two serendipitous results of direct assessment in the M.I.T. HMO case were that importances seemed to be correlated with risk aversion and that individuals were more risk averse with respect to X_{λ} if the levels of $X_{\bar{\lambda}}$ were reduced. Both of these are intuitive results but current utility theory has no way to incorporate these into a utility

function which can be feasibly assessed.

The von Neumann-Morgenstern axioms and the psychological axiom together imply weak simple scalability. This means that hierarchical choices must explicitly be modeled and that the effect on choice of similarities among alternatives in the choice set is not explicitly modeled. Because utilities and probabilities are modeled on the individual level, aggregate choice will be sensitive to similarities, but nonetheless an explicit modeling on the individual level could provide managerial insight. A useful direction to follow is Tversky's [143] elimination by aspects or McFadden's [94] random utility model in which the distribution of the disturbance term is dependent upon the performance measures of the alternatives. (See also Manksi [99].)

The HMO case used common factor analysis across individuals and stimuli as a first reduction technique, followed by individual specific remeasurement of perceptions. Linkages between the factor scores and direct perceptions required heuristic functions. A completely idiosyncratic reduction technique, compatible with direct compaction assessment and based on feasible individual measurement would prove useful. At present it is possible to use Ting's [140] theory of value functions to define parsimonious sets of individual specific performance measures, but the measurement costs are still too high for use in mass assessment.

Empirical Comparison of Models for the Methodology

The methodology was formulated to be modular to allow the interchange of various models for each step. Since each model has its relative advantages and disadvantages, direct comparisons of the models are

needed to guide the analyst in choosing the model appropriate for the problem. Data is being collected¹ which will enable the comparison of compaction by utility theory to that done by conjoint analysis, preference regression, and multinominal logit theory. At the probability of choice phase, the information theoretic/honest reward test will be used to compare the empirical Bayesian, the binary logit, and the multinominal logit probability models. A test of aggregation techniques with respect to the multinominal logit model is contained in Koppleman [80].

Applications

The methodology is being tested and elaborated on in several real decision environments. It is currently being used to study the consumer response to possible repositioning of the master of science program at M.I.T.'s Sloan School of Management, to the design of financial service packages, and the positioning of new frequently purchased consumer products (antacids, personal care products, and pain relievers). Other high potential uses are being explored in the design of banking services, the improvement and understanding of consumer reaction to new transportation modes such as Dial-a-Ride, and the marketing of preventive health services. Unexplored uses exist in the design of prepaid legal aid, the design of unified services for the elderly such as food, recreation, and medical programs, and in the design of day care centers. Perhaps parts of the methodology could also be adapted to industrial marketing.

Empirical experience to date indicates that the theory presented in this dissertation can be applied through the normative methodology and that this methodology can help managers to create new products and services, to understand consumer response to innovation, and to reduce the risks of failure.

BIBLIOGRAPHY

1. Ahl, David H., "New Product Forecasting Using Consumer Panels," Journal of Marketing Research, May 1970, pp. 160-167.
2. Allaire, Yvan, "The Measurement of Heterogeneous Semantic, Perceptual, and Preference Structures," unpublished PhD Dissertation, M.I.T., 1973.
3. Amstutz, A.E., Computer Simulation of Competitive Market Response, The M.I.T. Press, Cambridge, MA, 1967.
4. Armstrong, J.S. and A.C. Shapiro, "Analyzing Quantitative Models: Think Snow?", Wharton School, University of Pennsylvania, January 1973.
5. Arrow, K.J., Social Choice and Individual Values, 2nd Edition, Wiley and Sons, New York, 1963.
6. Ashton, Winifred D., The Logit Transformation, Griffin Co., London, 1972.
7. Bass, F., "A New Product Growth for Model Consumer Durables," Management Science 15, No. 5, January 1969.
8. Bass, F., "The Theory of Stochastic Preference and Brand Switching," Journal of Marketing Research XI, February 1974, pp. 1-20.
9. Bass, F. and W.W. Talarzyk, "An Attitude Model for the Study of Brand Preference," Journal of Marketing Research 9, February 1972, pp. 93-96.
10. Ben-Akiva, M.E., "Structure of Passenger Travel Demand Model," PhD Thesis, Department of Civil Engineering, M.I.T., 1973.
11. Bodily, S.E., "The Utilization of Frozen Red Cells in Blood Banking Systems: A Decision Theoretic Approach," Technical Report No. 94, Operations Research Center, M.I.T., May 1974.
12. Bohrnstedt, G.W., "Reliability and Validity Assessment in Attitude Measurement," Chapter 3 in Attitude Measurement, G.F. Summers, ed., Rand-McNally, Chicago, 1970.
13. Booz, Allen, and Hamilton, Management of New Products, 1971.
14. Boyle, B., "The Decision to Grant Credit," PhD Thesis, Department of Electrical Engineering, M.I.T., May 1974.

15. Burger, P., "COMP: A New Product Forecasting System," Northwest University Report No. 123-72.
16. Butler, D.H. and B.F. Butler, Hendrodynamics: Fundamental Laws of Consumer Dynamics, The Hendry Corporation, August 1966.
17. Cacoullos, T., Discriminant Analysis and Applications, Academic Press, New York, 1973.
18. Campbell, D. and D.W. Fiske, "Convergent and Discriminant Validation by the Multitrait, Multimethod Matrix," Psychological Bulletin 56, No. 2, March 1959.
19. Campbell, D.T. and J.C. Stanley, Experimental and Quasi-experimental Designs for Research, Rand McNally College Publishing Co, Chicago, 1963.
20. Cannell, C.F. and R.L. Kahn, "The Collection of Data by Interviewing," in Research Methods in Social Relations, L. Festinger and D. Katz, eds., Holt, Rinehart and Winston, New York, 1953.
21. Carroll, J.D., "A General Method for Preference Mapping of Perceptual Space," Bulletin of the Operations Research Society of America 16, 1968, pp. 227-228.
22. Carroll, J.D., "Individual Differences and Multidimensional Scaling in Multidimensional Scaling: Theory and Application in the Behavioral Sciences, R.N. Shepard, A.K. Romney and S. Nerlove, eds. Academic Press, New York, 1972, pp. 105-157.
23. Carroll, J.D. and J.J. Chang, "Analysis of Individual Difference in Multidimensional Scaling Via N-Way Generalization of 'Ekart-Young' Decomposition," Psychometrika 35, September 1970, pp. 283-319.
24. Carroll J.D. and J.J. Chang, "Relating Preference Data to Multidimensional Scaling Solutions via a Generalization of Coombs' Unfolding Model," Bell Telephone Laboratories, Murray Hill, NJ, 1967 (mimeographed).
25. Claycamp, H. and L.E. Liddy, "Prediction of New Product Performance An Analytical Approach," Journal of Marketing Research VI, November 1969, pp. 414-420.
26. Davis, O. and J.U. Farley, "Allocating Sales Force Effort with Commissions and Quotas," Management Science 18, No. 4, December 1971.
27. Domencich, T.A., G. Kraft and J.P. Valette, "Estimation of Urban Passenger Travel Behavior: An Economic Demand Model," Highway Research Record No. 238, 1968, pp. 64-78.

28. Douglas, A. and R. Lewis, "Trip Generation Techniques," Traffic Engineering and Control, November, December 1970, January 1971 (three part series).
29. Drake, A.W., Fundamentals of Applied Probability, McGraw-Hill, New York, 1967.
30. Ellis, H.M., "The Application of Decision Analysis to the Problem of Choosing an Air Pollution Control Problem for N.Y.C., Doctoral Dissertation, Harvard Business School, 1970.
31. Ellis, H.M. and R.L. Keeney, "A Rational Approach for Government Decisions Concerning Air Pollution," in Analysis of Public Systems, A.W. Drake, R.L. Keeney and P.M. Morse, eds., The M.I.T. Press, Cambridge, MA, 1972.
32. Farley, J. and L.W. Ring, "An Empirical Test of the Howard-Sheth Model of Buyer Behavior," Journal of Marketing Research VII, November 1970, pp. 427-438.
33. Fishbein, M., "Attitudes and the Prediction of Behavior," in Readings in Attitude Theory and Measurement, John Wiley and Sons, New York, 1967.
34. Fishbein, M. and L. Anderson, "Prediction of Attitude from the Number Strength and Evaluative Aspect of Beliefs about the Attitude Object: A Comparison of Summation and Congruity Theories," in Readings in Attitude Theory and Measurement, M. Fishbein, ed., John Wiley and Sons, New York, 1967.
35. Fishburn, P.C., Decision and Value Theory, John Wiley and Sons, New York, 1964.
36. Fishburn, P.C., "Lexicographic Orders, Utilization, and Decision Rules: A Survey," Management Science 20, No. 11, July 1974.
37. Fishburn, P.C., Mathematics of Decision Theory, UNESCO, 1972.
38. Fishburn, P.C., "von-Neumann-Morgenstern Utility Functions on Two Attributes," Operations Research 22, No. 1, January 1974, pp. 35-45.
39. Fishburn, P.C. and R.L. Keeney, "Generalized Utility Functions and Some Implications," Technical Report No. 93, Operations Research Center, M.I.T., April 1974.
40. Fleet, C.R. and S.R. Robertson, "Trip Generation in the Transportation Planning Process," Highway Research Record No. 240, pp. 11-26.

41. Frank, R.E., W. Massy and Y. Wind, Market Segmentation, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1972.
42. Gallagher, R., Information Theory and Reliable Communication, John Wiley and Sons, New York, 1968.
43. Golob, T., E. Canty and R. Gustafson, "An Analysis of Consumer Preferences for a Public Transportation System," Research Laboratories of General Motors Corporation, Warren, MI, 1970.
44. Golob, T.F. and R.L. Gustafson, "Economic Analysis of a Demand-Responsive Public Transportation System," Research Laboratories of the General Motors Corporation, Warren, MI, 1971.
45. Green, B., F. Mosteller and R.R. Bush, "Selected Quantitative Techniques and Attitude Management," reprinted from the Handbook of Social Psychology, Addison-Wesley, Reading, MA, 1954.
46. Green, P.E., "On the Robustness of Multidimensional Scaling Techniques," Journal of Marketing Research XII, February 1975, pp. 73-81.
47. Green, P. and F.J. Carmone, "Multidimensional Scaling: An Introduction and Comparison of Nonmetric Unfolding Techniques," Journal of Marketing Research VI, August 1969, pp. 330-341.
48. Green, P. and V. Rao, Applied Multidimensional Scaling, Holt, Rinehart and Winston, Inc., New York, 1972.
49. Green, P. and Y. Wind, Multiattribute Decisions in Marketing, The Dryden Press, Hinsdale, IL, 1973.
50. Greer, P.S. and E.M. Suuberg, "Segmentation Issues in Marketing and HMO," SM Thesis, Sloan School of Management, M.I.T., 1975.
51. Gustafson, R.L., H.N. Curd and T.F. Golob, "User Preference for a Demand-Responsive Transportation System: A Case Study Report," Research Laboratories of the General Motors Corporation, Warren, MI, 1971.
52. Harman, H.H., Modern Factor Analysis, University of Chicago Press, Chicago, IL, 1967.
53. Hatch, T. and G.L. Urban, "New Products: Maximizing Creativity and Minimizing Risk," Working Paper 752-74, Sloan School of Management, M.I.T., December 1974.
54. Hauser, J.R., "An Efficient Model for Planning Bus Routes in Communities with Populations Between 20,000 and 250,000," Working Paper OR 029-73, Operations Research Center, M.I.T., 1973.

55. Hauser, J.R., "Modeling Decisions of Choice Among Finite Alternatives: Applications to Marketing and Transportation Demand Theory," Working Paper OR 038-74, Operations Research Center, M.I.T., October 1974.
56. Hauser, J.R. and G.L. Urban, "A Normative Methodology for Modeling Consumer Response to Innovation," Technical Report No. 109, Operations Research Center, M.I.T., May 1975.
57. Herniter, J., "A Comparison of the Entropy and the Hendry Model," Journal of Marketing Research XI, February 1974, pp. 21-29.
58. Herniter, J., "An Entropy Model of Brand Purchase Behavior," Journal of Marketing Research X, November 1973, pp. 361-375.
59. Herniter, J., "A Probabilistic Market Model of Purchase Timing and Brand Selection," Management Science 18, No. 4, December 1974.
60. Herniter, J. and V.J. Cook, "Microsimulation Evaluation of Advertising Budget Strategies," presented at the International TIMS meeting, March 27, 1969.
61. Hertz, D.B., "Risk Analysis in Capital Investment," Harvard Business Review, January-February 1974, pp. 95-106.
62. Highway Research Board, "Urban Travel Demand Forecasting," HRB Special Report 143, 1973.
63. Hill, D.M. and H.G. VonCube, "Development of a Model for Forecasting Travel Mode Choice in Urban Areas," Highway Research Record No. 38, 1963, pp. 78-96.
64. Horgan, D., "A Decision Analysis for Sewage Sludge Disposal Alternatives for Boston Harbor," SM Thesis, Operations Research Center M.I.T., September 1972.
65. Howard, J. and J.N. Sheth, The Theory of Buyer Behavior, John Wiley and Sons, New York, 1969, pp. 83-114.
66. Kaufman, G.M., "Multi-Attributed Preferences," Class Notes, Sloan School of Management, M.I.T., December 1972.
67. Keeney, R.L., "A Decision Analysis with Multiple Objectives: The Mexico City Airport," Bell Journal of Economics and Management Science 4, 1973, pp. 101-117.
68. Keeney, R.L., "Decomposition of Multiattribute Utility Functions," Technical Report No. 80, Operations Research Center, M.I.T., May 1973.

69. Keeney, R.L., "A Group Preference Axiomization with Cardinal Utility," presented at the 46th National ORSA/TIMS Meeting, San Juan, Puerto Rico, October 1974.
70. Keeney, R.L., "Multidimensional Utility Functions: Theory, Assessment, and Application," Technical Report No. 43, Operations Research Center, M.I.T., October 1969.
71. Keeney, R.L., "Multiplicative Utility Functions," Operations Research 22, No. 1, January 1974, pp. 22-33.
72. Keeney, R.L., "Utility Functions for Multiattributed Consequences," Management Science 18, 1972, pp. 276-287.
73. Keeney, R.L., "Utility Independence and Preference for Multiattributed Consequences," Operations Research 19, 1971, pp. 875-893.
74. Keeney, R.L. and C.W. Kirkwood, "Grouping Decision Making Using Cardinal Social Welfare Functions," Technical Report No. 83, Operations Research Center, M.I.T., October 1973.
- 74a. Keeney, R.L. and G.L. Lilien, "A Utility Model for Product Positioning," Working Paper 777-75, Sloan School of Management, M.I.T., March 1975.
75. Keeney, R.L. and K. Nair, "Decision Analysis for Siting of Nuclear Power Plants: The Relevance of Multiattribute Utility Theory," Technical Report No. 96, Operations Research Center, M.I.T., June 1974.
76. Keeney, R.L. and H. Raiffa, Decision Analysis with Multiple Conflicting Objectives, John Wiley and Sons, New York (forthcoming).
77. Kelly, G.A., The Psychology of Personal Constructs, Volume 1, Norton Wiley, New York, 1955.
78. Kendall, M.G., A Course in Multivariate Analysis, Hafner Publishing Co., New York, 1961.
79. Kirkwood, C.W., "Decision Analysis Incorporating Preferences of Groups," Technical Report No. 74, Operations Research Center, M.I.T., June 1972.
80. Koppleman, F., "Prediction of Travel Behavior with Disaggregate Choice Models," PhD Thesis, Department of Civil Engineering, M.I.T. June 1975.
81. Koppleman, F., M. Ben-Akiva, and P. Roberts, "Development of an Aggregate Model of Urbanized Area Travel Behavior," Technical Proposal submitted to the U.S. Department of Transportation, RFP DOT-50001, September 6, 1974.

82. Kraft, G. and T.A. Domencich, "Free Transit," presented at Transportation and Poverty Conference, American Academy of Arts and Sciences, Brookline, MA, June 7, 1968.
83. Krantz, D.H., "The Scaling of Small and Large Color Differences," PhD Thesis, University of Pennsylvania, (University Microfilms, Ann Arbor, MI, 1964.) As referenced by Tversky [143].
84. Krischer, J.P., "An Analysis of Patient Management Decisions as Applied to Cleft Palate, Technical Report 12-74, Center for Research in Computing Technology, Harvard University, 1974.
85. Kruskal, J.B., "Multidimensional Scaling by Optimizing Goodness of Fit to a Non-Metric Hypothesis," Psychometrika 24, 1964, pp 1-27.
86. Lave, C.A., "A Behavioral Approach to Model Split Forecasting," Transportation Research 3, 1969, pp. 463-480.
87. Lehrman, S., "A Disaggregate Behavioral Model of Urban Mobility Decisions," PhD Thesis, Department of Civil Engineering, M.I.T. June 1975.
88. Levinson, H.S. and F.H. Wynn, "Some Aspects of Future Transportation in Urban Areas," Highway Research Board Bulletin 326, Urban Transportation: Demand and Coordination, 1962.
89. Little, J.D.C., "BRANDAID: An On-Line Marketing-Mix Model," Working Paper 586-72, Sloan School of Management, M.I.T., February 1972.
90. Little, J.D.C., "Models and Managers: The Concept of a Decision Calculus," Management Science, May 1970.
91. Lodish, L.M., "CALLPLAN, An Interactive Salesman's Call Planning System," University of Pennsylvania.
92. Luce, R.D., Individual Choice Behavior, John Wiley and Sons, New York, 1959.
93. McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers in Econometrics, P. Zarenblea, ed., Academic Press, 1970, pp. 105-142.
94. McFadden, D., "Quanta Choice Analysis: A Survey," NSF-NBER Conference on Decision Roles and Uncertainty, University of California, Berkeley, CA, March 22-23, 1974.
95. McLynn, J.M., A.J. Goldman, P.R. Meyer and R.H. Watkins, "Analysis of a Market Split Model," Journal of Research of the National Bureau of Standards 72B, No. 1, 1968.

96. McLynn, J.M. and T. Woronka, "Passenger Demand and Modal Split Models," prepared for Northeast Corridor Transportation Project, United States Department of Transportation, Report No. NECTP-730, Bethesda, MD, 1969.
97. Manheim, M., "Practical Implications of Some Fundamental Properties of Travel Demand Models," Department of Civil Engineering Report R-72-63, M.I.T.
98. Manski, C., "Maximum Score Estimation of the Stochastic Utility Model of Choice," Working Paper, Carnegie-Mellon University, Pittsburgh, PA, January 1974.
99. Manski, C., "The Structure of Random Utility Models," Working Paper, Carnegie-Mellon University, Pittsburgh, PA, February 1975.
100. Martin, B.V., F.W. Memmott, III, and A.J. Bone, "Principles and Techniques of Predicting Future Demand for Urban Area Transportation," M.I.T. Report No. 3, The M.I.T. Press, Cambridge, MA, June 1961.
101. Massy, W.F., "Forecasting the Demand for New Convenience Products," presented to the Educator's Conference of the American Marketing Association, Denver, CO, August 1968.
102. Massy, W.F., D.B. Montgomery and D.G. Morrison, Stochastic Models of Buying Behavior, The M.I.T. Press, Cambridge, MA, 1970.
103. Memmott, F.W., III, "Home Interview Survey and Data Collection Procedures," Highway Research Record No. 41, 1963, pp. 7-12.
104. Meyer, J.R., J.F. Kain and M. Wohl, The Urban Transportation Problem, Harvard University Press, Cambridge, MA, 1965.
105. Meyer, R.F., "On the Relationship Among the Utility of Assets, the Utility of Consumption, and Investment Strategy in an Uncertain, but Time Invariant World," proceedings of the Fifth International Conference on Operations Research, Travistock Publications. As referenced in Kaufman [66].
106. Montgomery, D. and A.J. Silk, "Estimating Dynamic Effects of Market Communications Expenditures," Management Science 18, No. 10, June 1972.
107. Montgomery, D., A.J. Silk and C.E. Zavagoza, "A Multiple Product Sales Force Allocation Model," Management Science 18, No. 4, December 1971.
108. Montgomery, D.B. and G.L. Urban, Management Science in Marketing, Prentice-Hall, Englewood Cliffs, NJ, 1969.

109. Mood, A.M. and F.A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill Book Company, New York, 1963.
110. Morrison, D., Multivariate Statistical Methods, McGraw-Hill Book Company, New York, 1967.
111. Nerlove, M. and S.J. Press, "Univariate and Multivariate Log-Linear and Logistic Models," Rand Report R-1306-EOA/NIH, December 1973.
112. Payne, S.L., The Art of Asking Questions, Princeton University Press, Princeton, NJ, 1951.
113. Pessemier, E.A., STRATOP: A Model for Designing Effective Product and Communication Strategies, Working Paper No. 470, Purdue University, West Lafayette, IN, August 1974.
114. Plourde, R.P., "Development of a Behavioral Model of Travel Mode Choice, PhD Thesis, Department of Civil Engineering, M.I.T., 1971.
115. Pollack, R.A., "Additive von-Neumann-Morgenstern Utility Functions," Econometrics 35, 1967, pp. 485-595.
116. Quandt, R.E. and W.J. Baumol, "The Demand for Abstract Transportation Modes: Theory and Management," Journal of Regional Science 6, No. 2, 1966.
117. Quandt, R.E. and K.H. Young, "Cross-Sectional Travel Demand Models: Estimate and Tests," Journal of Regional Science 9, No. 2, 1969.
118. Quarmby, D.A., "Choice of Travel Mode for the Journey to Work," Journal of Transportation Economics and Policy 1, No. 3, September 1967, pp. 273-314.
119. Raiffa, H., "Assessments of Probabilities," unpublished manuscript, January 1969.
120. Raiffa, H., Decision Analysis: Introducing Lectures on Choices Under Uncertainty, Addison-Wesley, Reading, MA, 1970.
121. Raiffa, H. and R. Schlaifer, Applied Statistical Decision Theory, Harvard University Press, Cambridge, MA, 1961.
122. Rassum, P.R., R.H. Ellis and J.C. Bennett, "The N-Dimensional Logit Model: Development and Application," paper presented to the 50th Annual Highway Research Board meeting, January 1971.
123. Rogers, E. and J.D. Stanfield, "Adoption and Diffusion of New Products: Emerging Generalizations and Hypotheses," in Applications of the Sciences in Marketing Management, F. Bass, et al., eds., Wiley and Sons, New York, 1968.

124. Rosenburg, M.J., "Cognitive Structure and Attitude Effect," Journal of Abnormal and Social Psychology 53, 1956, pp. 367-372.
125. Ruiter, E.R., "Toward a Better Understanding of the Intervening Opportunities Model," Transportation Research, 1967, pp. 47-56.
126. Rummel, R.J., Applied Factor Analysis, Northwestern University Press, Evanston, IL, 1970.
127. Schneider, L.M., Marketing Urban Mass Transit: A Comparative Study of Management Strategies, Harvard University Press, Cambridge, MA 1965.
128. Shepard, R., A. Romney and S. Nerlove, Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Seminar Press, New York, 1972.
129. Sheth, J.N. and W.W. Talarzyk, "Perceived Instrumentality and Value Importance as Determinants of Attitudes," Journal of Marketing Research 9, February 1972, pp. 6-9.
130. Sicherman, A., "An Interactive Computer Program for Assessing and Using Multi-Attributed Utility Functions," SM Thesis, Operations Research Center, M.I.T., June 1975.
131. Silk, A.J., "Preference and Perception Measures in New Product Development: An Exposition and Review," Industrial Management Review, Sloan School of Management, M.I.T., Cambridge, MA, Fall 1969.
132. Silk, A.J. and G.L. Urban, "ASSESSOR: A Pretest Market Evaluation Model," Working Paper, Sloan School of Management, M.I.T., April 1975.
133. Smith, S.A., "A Derivation of Entropy and the Maximum Entropy Criterion in the Context of Decision Problems," IEEE Transactions on Systems, Man and Cybernetics SMC-4, No. 2, March 1974.
134. Sonquist, J.A., E.L. Baker and J.N. Morgan, "Searching for Structure (Alias AID III)," Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 1971.
135. Stefflre, V. "Market Structure Studies: New Products for Old Markets and New Markets (Foreign) for Old Products," in Applications of the Science in Marketing Management, F. Bass, C. King, and E. Pessimier, eds., John Wiley and Sons, New York, 1968, pp. 251-268.
136. Stopher, P.R., "A Probability Model of Travel Mode Choice for the Work Journey," Highway Research Record No. 283, pp. 57-65.

137. Stopher, P.R. and T.E. Lisco, "Modelling Travel Demand: A Disaggregate Behavioral Approach - Issues and Applications," Transportation Research Forum Proceedings, 1970, pp. 195-214
138. Systems Analysis and Research Corporation (SARC), "Demand for In-territory Passenger Travel in the Washington-Boston Corridor," Report to the U.S. Department of Commerce, 1963.
139. Thurstone, L., "A Law of Comparative Judgment," Psychological Review 34, 1927, pp. 273-286.
140. Ting, H.M., "Aggregation of Attributes for Multiattributed Utility Assessment," Technical Report No. 66, Operations Research Center, M.I.T., August 1971.
141. Torgerson, W.S., Methods of Scaling, John Wiley and Sons, New York, pp. 105-112.
142. Tucker, L.R. and S. Messick, "An Individual Differences Model for Multidimensional Scaling," Psychometrika 28, 1963, pp. 333-367.
143. Tversky, A., "Elimination by Aspects: A Theory of Choice," Psychological Review 79, No. 4, 1972, pp. 281-299.
144. Tversky, A., "A General Theory of Polynomial Conjoint Measurement," Journal of Mathematical Psychology 4, 1967, pp. 1-20.
145. Urban, G.L., "A New Product Analysis and Decision Model," Management Science 14, No. 8, April 1968.
146. Urban, G.L., "PERCEPTOR: A Model for Product Position," Management Science VIII, April 1975, pp. 858-871.
147. Urban, G.L., "SPRINTER mod III: A Model for the Analysis of New Frequency Purchase Consumer Products," Operations Research 18 September-October 1970, pp. 805-853.
148. U.S. Department of Transportation, Federal Highway Administration, Bureau of Public Roads, "Calibrating and Testing a Gravity Model for Any Size Urban Area," 1968.
149. Utterback, J.M., "Innovation in Industry and the Diffusion of Technology," Science 183, February 1974, pp. 620-625.
150. Vith, J.E., H.J. Bauer, E.T. Carty and T.F. Golob, "Determining the Importance of User-Related Attributes for a Demand-Responsive Transportation System," Purdue University, Lafayette, IN, 1969.
151. von Neumann, J. and O. Morgenstern, The Theory of Games and Economic Behavior, 2nd ed., Princeton University Press, Princeton, NJ, 1947.

152. Wilson, A.G., "Travel Demand Forecasting: Achievements and Problems," Highway Research Board Special Report 143, 1973.
153. Witheford, D.K., "Comparison of Trip Distribution by Opportunity Model and Gravity Model," presented to the Origin and Destination Committee, Highway Research Board, January 9, 1961.
154. Wynn, F.H. and H.S. Levinson, "Some Considerations in Appraising Bus Transit Potentials," Highway Research Record No. 197, 1967, pp. 1-24.
155. Young, F.W. and W.S. Torgerson, "TORSCA, A Fortran IV Program for Shepard-Kruskal Multidimensional Scaling Analysis," Behavioral Science 12, 1967, pp. 498.
156. Zeleny, M., "The Attribute-Dynamic Attitude Model," Working Paper, Graduate School of Business, Columbia University, New York, May 1974.

FOOTNOTES

Chapter 1

1. In practice, the "utility" function used by choice theorists contains socio-economic variables. This allows it to vary across segments even if its parameters do not.
2. The methodology has been previously published in Hauser and Urban [56].

Chapter 2

no footnotes

Chapter 3

1. A follow-up mailing was used to encourage respondents to complete the questionnaire. Statistical comparison indicated there was no significant difference between the first wave and the following wave.

Chapter 4

1. Since the evoked set varies by individual, there will be $N \cdot \bar{J}$ columns where \bar{J} is the average evoked set size.
2. Commonalities were defined by iteration with the BIOMED statistical routines. 40% of the variance was common. Principal components analysis yielded a similar interpretation of factors and factored 55% of the total variance into the four factors.

Chapter 5

no footnotes

Chapter 6

1. The redundant notation, $c(\{y, z\})$ is changed to $c(y, z)$ when the meaning is clear from the context.
2. This term is due to Keeney [70].
3. Parts of this section are strongly influenced by Kaufman [66].
4. This is the general quasi-additive form.
5. This term is due to Keeney [70].

6. This is possible since $\{X_1, X_j\}$ p.i. $X_{\overline{1j}}$. The performance measures have been renumbered appropriately.
7. If $\sum k_j > 1$ then $K \in (-1, 0)$, if $\sum k_j = 1$ then $K = 0$, and if $\sum k_j < 1$ then $K \in (0, \infty)$. Restricted to these ranges, the normalization equation has only one real root.
8. Fits in the range .1 to .2 are reported in Sheth and Talarzyk [129], but Bass and Talarzyk [9] report correct predictions of first preference by linear attitude mode in the range of .63 to .75.
9. Tversky [144] defines a set of regular extensions $\overset{e}{>}_x$ of $>_x$. The axiom then states that there exists at least one regular extension such that $a \overset{e}{>}_x a$ does not hold for at least one a .
10. To produce a preferential compaction function, first preference was set equal to 4.0, second preference 3.0, third preference 2.0, and fourth preference 1.0. Ties were handled by averaging e.g., a first-second tie received 3.5 for each.
11. To account for individual scale differences, factor scores were normalized across plans for each individual.
12. The functional form is flexible. If $p_{il} > .5$ then $r_{il} > 0$ and the individual is risk averse. If $p_{il} < .5$ then $r_{il} < 0$ and the individual is risk prone. If $p_{il} = .5$ then the utility function is linear in X_l and the individual is risk neutral.
13. Points were generated with a computer program developed by Sicherman [130].

Chapter 7

1. A diary of choices kept by a consumer which is useful in monitoring choice behavior over time. (Massy, Montgomery, Morrison [102]).
2. Depending upon interpretation, perhaps perceptions of the performance measures.
3. The integrals in this section expand upon those in Keeney and Lille [74a] and in McFadden [93].
4. Even if a few consumers choose more than one alternative the observed probabilities will most likely be arbitrarily set to 1/2, 1/3, etc.

5. The number of parameters required by many distributions is not linear in the number of variates, v . For example the multivariate normal distribution requires v parameters for the means and v for the variances, but $(1/2)v(v - 1)$ for the covariances. In many cases, orthogonalization of effects reduces the multivariate conditional distribution. Such a product is linear in the number of its parameters. See, for example, the empirical calibration of section 7.3.3.
6. For example, the beta distribution used in section 7.3.3.

Chapter 8

1. If the design changes are significant, remeasurement may be necessary.

Chapter 9

no footnotes

Chapter 10

1. This is a variation of an example due to Raiffa [119].
2. Term due to Raiffa [119].
3. Although it is not a generalized likelihood ratio. See Mood and Graybill [109] pages 286, 298.
4. $I(\text{observed})$ is the average value of the information measure. The weak law of large numbers states that a sample mean converges to its expected value. This can be shown by the Chebyshev inequality. See Drake [29] pages 204-207.

Chapter 11

1. Conjoint analysis can also be individual specific.

Chapter 12

1. This data is being collected as part of a study for possible repositioning of the master of science program at M.I.T.'s Sloan School of Management.
2. The applications to financial service and frequently purchased consumer products are being done by Management Decision Systems, Inc.

APPENDIX ONE

First Questionnaire

Mail Survey of 1000

M.I.T. Faculty, Students, and Staff

(actual survey was printed on one side of the page only)

April, 1974

Dear Member of the M.I.T. Community:

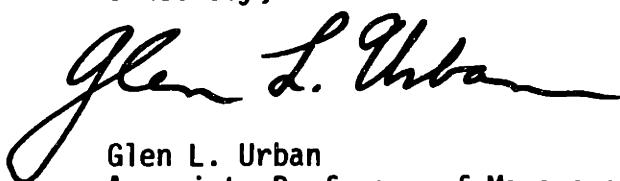
Good health is important to all of us. Today many changes are taking place to improve medical care. One of the most important is the increasing role of consumers in the planning for health service delivery. In order to better reflect patients' needs, new ways for consumers to express their feelings are being developed. The enclosed questionnaire is part of a research effort to determine health care requirements and desires of consumers such as you.

Your name has been selected at random to be part of this survey of health care attitudes and preferences in the M.I.T. community. The number of people being asked to participate is small, so your answers are very important. The results of the survey will also be used by the M.I.T. Medical Department to remain as responsive as possible to the needs of the M.I.T. community.

You are asked not to put your name on the questionnaire and there is no way to link your answers to your name. The questionnaire requests information about the health services now available to you and asks for your opinion on several new methods of delivering health care. The last questions relate to some demographic characteristics that are important in projecting the responses from this small survey to the M.I.T. community as a whole. The questionnaire takes about 20-30 minutes to complete. Most people find it easy and interesting to answer the questions and they think it is important to make their feelings known to those who provide health care. A summary of the survey can be made available to you after the project is completed.

I hope you will be able to take a few minutes in the next day or two to fill out this questionnaire. It is important to M.I.T. in planning for your needs and it is important for our research project to improve understanding of consumer response to health services. If you have any questions about this survey, please feel free to call me at 253-6615.

Sincerely,



Glen L. Urban
Associate Professor of Management

We are pleased that you have decided to fill out this questionnaire. It gives you a chance to make your feelings known so better health services can be provided.

1. Do you feel that most people in Massachusetts get good medical care? (Please check one of the boxes below)

— YES
— NO
— NOT SURE

2. How do you feel about the prices doctors charge their patients?

- _____ outrageously high
- _____ too high
- _____ reasonable considering the services
- _____ a little low
- _____ too low

3. We would like to ask where you go or would go to receive health service. Several medical problems will be described and you are asked to indicate where you would most likely go to receive health care. There are no right or wrong answers, so just think about what you would actually do in each situation.

A. If you woke up in the morning with a headache, a temperature over 102°, and stomach pains, where would you go first or whom would you contact first? **CHECK ONE** of the following:

— HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

— PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING: Do you have the name of a specific doctor in mind? YES NO

Is the doctor a: General Practitioner Specialist - what type _____

— M.I.T. MEDICAL DEPARTMENT

— OTHER SOURCE OF HEALTH CARE: Please describe _____

— WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

— I DO NOT KNOW WHAT I WOULD DO

B. If you wanted to get a routine physical examination and health checkup, where would you go? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

C. If you experienced frequent and sometimes severe chest pains for several days, where would you go first or whom would you contact? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

D. If you had a severe toothache, where would you go? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

PRIVATE DENTIST - Do you have the name of a specific dentist in mind? YES NO

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

E. If you fell down the stairs at home and felt that you may have broken your ankle, where would you go first or whom would you call? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

F. If you felt extremely nervous and tense all the time due to serious personal problems, where would you ^{first} go or whom would you contact? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

G. IF YOU ARE FEMALE, please answer this question, OTHERWISE GO DIRECTLY TO QUESTION 4.

If you wanted a PAP smear test (female cancer check up), where would you go? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

H. IF YOU ARE FEMALE AND BETWEEN THE AGES OF 16 and 45, please answer the following question, OTHERWISE GO DIRECTLY TO QUESTION 4.

If you thought you were pregnant, where would you go for medical care? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

4. IF YOU ARE MARRIED, please answer the following questions about your spouse (husband or wife), OTHERWISE GO TO QUESTION 5.

A. If your spouse experienced frequent and severe chest pains for several days, where would he or she go first or whom would he or she contact? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner
 Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

B. If your spouse wanted a routine physical examination or health check up, where would he or she go? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner
 Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

C. IF YOUR SPOUSE IS FEMALE, please answer the following question, OTHERWISE GO TO QUESTION 5.

If your wife wanted a PAP smear test (female cancer test), where would she go? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:
Do you have the name of a specific doctor in mind?
YES NO
Is the doctor a: General Practitioner
 Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

D. IF YOUR WIFE IS BETWEEN THE AGES OF 16 and 45, please answer the following question, OTHERWISE GO TO QUESTION 5.

If your wife thought she was pregnant where would she go for medical care?

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:

Do you have the name of a specific doctor in mind?

YES NO

Is the doctor a: General Practitioner

Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

5. IF YOU HAVE CHILDREN, please answer the following questions, OTHERWISE GO DIRECTLY TO QUESTION 6.

A. If your child woke up in the morning with a headache, a temperature over 102°, and stomach pains, where would you go first or whom would you contact? Check one.

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:

Do you have the name of a specific doctor in mind?

YES NO

Is the doctor a: General Practitioner

Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

B. If your child needed a routine physical examination and health check up, where would you take him or her?

HOSPITAL - IF YOU CHECKED THIS BOX INDICATE WHICH AREA OF THE HOSPITAL: emergency room, front desk, outpatient clinic, other, what is the name of the hospital? _____

PRIVATE DOCTOR - IF YOU CHECKED THIS BOX ANSWER THE FOLLOWING:

Do you have the name of a specific doctor in mind?

YES NO

Is the doctor a: General Practitioner

Specialist - what type _____

M.I.T. MEDICAL DEPARTMENT

OTHER SOURCE OF HEALTH CARE: Please describe _____

WOULD NOT THINK IT NECESSARY TO DO ANY OF THE ABOVE

I DO NOT KNOW WHAT I WOULD DO

6. You have just described the health services you would use if certain problems occurred. We would now like to learn how you feel about the health care you, your spouse, and your children are now receiving. On the following page you will find a list of various statements about health services. Please consider each statement and place an X in the space that you feel most accurately reflects your agreement or disagreement with the statement. Consider the overall health services you now receive. For example, consider the following statement:

The waiting rooms are beautiful.

strongly X agree	agree	neither agree nor disagree	disagree	strongly disagree
a	b	c	d	e

If you agree that the waiting rooms you use are beautiful, you would put an X in box a or b, depending upon whether you strongly agree (box a) or just agree (box b). If you feel the waiting rooms you use are ugly, you would put an X in box d or e, depending upon how ugly you think they are. If you have no feeling on whether the waiting rooms are beautiful or ugly, you would mark box c, neither agree nor disagree. In the example shown here, the respondent has placed an X in box a, indicating he strongly feels the waiting rooms are beautiful.

Now consider the statements on the following page and place an X in the box which indicates your OVERALL FEELINGS about the current health services you (and your family) use.

YOUR EXISTING HEALTH SERVICES

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
--	-------------------	-------	----------------------------------	----------	----------------------

1. I can get medical service and advice easily any time of the day and night.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

2. I have to wait a long time to get service.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

3. I can trust that I am getting really good medical care.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

4. My health services are inconveniently located and are difficult to get to.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

5. I pay too much for my required medical services.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

6. I get a friendly, warm, and personal approach to my medical problems.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

7. The health services I get help me prevent medical problems before they occur.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

8. I can easily find a good doctor.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

9. My health services offer modern, up-to-date treatment methods.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

10. No one has access to my medical record except medical personnel.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

11. My health services do not provide continuing interest in my health care.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

12. My health services use the best possible hospital.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

13. Too much work is done by nurses and assistants rather than doctors.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

14. My health services represent an organized and complete medical service for me and my family.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

15. My health services have a lot of redtape and bureaucratic hassle.

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

16. My health services have highly competent doctors and specialists to

<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
----------	----------	----------	----------	----------

7. In general, please indicate how satisfied you are with your current health care. Check one box.

- extremely satisfied
- very satisfied
- satisfied
- indifferent
- dissatisfied
- very dissatisfied
- extremely dissatisfied

8. In the space below please describe any other feelings you have concerning your existing health services. For example, you might comment on what you like or dislike about your health care. If you do not want to say anything more, go on to Question 9.

9. Are you covered by health insurance (or prepaid health coverage)? YES NO

10. IF YOU ARE A STUDENT, please answer the following questions, otherwise go directly to Question 11.

A. Do you have hospitalization insurance in addition to the student health fee coverage: YES NO

If yes, is it the hospital insurance offered by M.I.T.? YES NO

B. Are you covered by your parents' hospitalization insurance? YES NO

11. Which of the following statements best reflects your feeling about your health insurance:

- covers everything
- very adequate coverage
- satisfactory coverage
- not satisfactory coverage
- very inadequate coverage
- covers almost nothing
- do not know

12. Have you heard of the Harvard Community Health Plan? YES NO

If yes, which statement reflects your level of information about the Harvard Community Health Plan.

- I know almost everything about it.
- I know a lot about it.
- I know a little about it.
- I know almost nothing about it.

13. Have you heard of the M.I.T. Health Plan? YES NO

If yes, which statement reflects your level of information about the M.I.T. Health Plan?

- I know almost everything about it.
- I know a lot about it.
- I know a little about it.
- I know almost nothing about it.

14. Now we would like to learn what your reactions are to three new health care delivery systems. Some of them are real systems, others are hypothetical. Each new system is described on one page. Then you are asked to rate that system on the same basis as you rated your existing system. You will be asked to consider the same list of statements about medical care and then to indicate if you agree or disagree with each statement. We are interested in finding out how you feel about the new systems, based on the one page description. There are no right or wrong answers, so just indicate how you think the new service would be if you actually were part of the new health delivery system.

After reading the description below of the M.I.T. HEALTH PLAN, please rate how you feel about it (as if you were a member) on the following page. (If you are now a member of the M.I.T. HEALTH PLAN, we are still interested in your reactions to this description.)

DESCRIPTION OF THE M.I.T. HEALTH PLAN

M.I.T. announces a new health care plan for YOU AND YOUR FAMILY. By joining the M.I.T. HEALTH PLAN you can get comprehensive health care at a low, fixed monthly charge. Virtually all your medical needs will be met. You will not have to face unexpected doctor or hospital bills and you will not have to worry about finding a good doctor for you or your family.

The cost of joining the M.I.T. HEALTH PLAN is only a little more than regular Blue Cross/Blue Shield health insurance, but you get more services and comprehensive care. There are no charges for doctor visits, nursing and laboratory services, or hospital services. Women in the plan pay nothing extra for prenatal, delivery, or maternity care. The services are comprehensive and include mental health care and emergency services.

The costs are kept low by the utilization of preventive care to keep you well. The plan succeeds by keeping you and your family well and out of the hospital. In addition, the use of trained paramedics and technology helps reduce costs while maintaining the quality of care.

You choose your own personal doctor (specialist in internal medicine for yourself and a pediatrician for your children) from our staff of physicians. Your doctor supervises your total health care at the health center and in the hospital. He will be sure you get the highest quality of care. When you are a member of the M.I.T. HEALTH PLAN you can be sure of getting health care around the clock from the staff of physicians, nurses, social workers and allied health personnel.

The M.I.T. HEALTH PLAN delivers its services from the Homberg Memorial Building on the M.I.T. campus. Parking is available during patient visits. Hospital services are provided by the Mount Auburn and Cambridge City Hospitals. Maternity and gynecological care are provided through the resources of the Boston Hospital for Women. For emergencies outside the Boston area, local hospitals can be used.

You can become a member of the plan by paying \$1.50 per month more than your Blue Cross/Blue Shield coverage if you are single and \$4.00 more per month if you are married. If you are a single student and do not have hospital insurance, the cost is \$8.25/month more than the student health fee you are currently paying; if you are a married student, the cost is \$20.00/month more than the student health fee. These fees cover all of your medical costs except: the first \$50 and 20% of the balance of prescription charges and the excess of \$10 per visit for psychotherapy (over \$5 per visit for group therapy). The plan does not include eye glasses, hearing aids, cosmetic surgery, custodial confinement, or dental care done outside a hospital. If you join the plan, you must remain a member for one year.

The M.I.T. HEALTH PLAN is designed to make comprehensive, high quality health care available to you and your family at a low cost.

YOUR RATING OF THE M.I.T. HEALTH PLAN

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1. I would be able to get medical service and advice easily any time of the day and night.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
2. I would have to wait a long time to get service.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
3. I could trust that I am getting really good medical care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
4. The health services would be inconveniently located and would be difficult to get to.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
5. I would be paying too much for my required medical services.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
6. I would get a friendly, warm, and personal approach to my medical problems.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
7. The plan would help me prevent medical problems before they occurred.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
8. I could easily find a good doctor.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
9. The service would use modern, up-to-date treatment methods.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
10. No one has access to my medical record except medical personnel.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
11. There would not be a high continuing interest in my health care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
12. The services would use the best possible hospitals.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
13. Too much work would be done by nurses and assistants rather than doctors.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
14. It would be an organized and complete medical service for me and my family.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
15. There would be much redtape and bureaucratic hassle.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
16. Highly competent doctors and specialists would be available to serve me.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>

After reading the description below of the MASSACHUSETTS HEALTH FOUNDATION, please rate how you feel about it (as if you were a member) on the following page.

DESCRIPTION OF THE MASSACHUSETTS HEALTH FOUNDATION

The Massachusetts Medical Association announces a new way to finance your family's health care. If you join the MASSACHUSETTS HEALTH FOUNDATION you pay only a fixed monthly charge for comprehensive health services. You pay no hospital or doctors' bills.

In the MASSACHUSETTS HEALTH FOUNDATION you can keep your relationship with an individual, private physician. You visit him in his neighborhood office and you are cared for at his hospital. Almost all physicians in Massachusetts are affiliated with the MASSACHUSETTS HEALTH FOUNDATION so probably you can keep your current doctor and hospital. You do not have to change physicians or travel to a new place to receive your health services.

Your physician will be responsible for the quality of your total health care and will provide you with the kind of individual attention you expect from a private, personal physician. In addition, a group of doctors from the Foundation periodically reviews the quality of care individual patients receive. This is an additional guarantee that you are receiving the best possible health care and that excessive costs will not occur.

The cost of joining the MASSACHUSETTS HEALTH FOUNDATION is only a little more than the cost of regular Blue Cross/Blue Shield health insurance, but for the small additional cost you will receive more comprehensive services. There are no charges for doctor's visits, nursing or laboratory services, or hospital care. The benefits include all maternal services (prenatal, delivery, and postpartum care). You can become a member of the plan by paying \$6.50 per month more than your Blue Cross/Blue Shield coverage if you are single and 20.00 more per month if you are married. If you are a single student and do not have hospital insurance, the cost is \$13.25 month more than the student health fee you are now paying; if you are a married student, the cost is \$36.00/month more than the student health fee. The fees do not include prescriptions, eye glasses, hearing aids, cosmetic surgery, custodial confinement, or dental care done outside a hospital. If you join the plan you must remain a member for one year.

MASSACHUSETTS HEALTH FOUNDATION is designed to give you the opportunity to pay only a fixed monthly charge for comprehensive individual health care. If you join the Foundation and your doctor is affiliated with the Foundation you can continue to see your current doctor in his office and receive care at his hospital as you do now.

YOUR RATING OF THE MASSACHUSETTS HEALTH FOUNDATION

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1. I would be able to get medical service and advice easily any time of the day and night.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
2. I would have to wait a long time to get service.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
3. I could trust that I am getting really good medical care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
4. The health services would be inconveniently located and would be difficult to get to.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
5. I would be paying too much for my required medical services.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
6. I would get a friendly, warm, and personal approach to my medical problems.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
7. The plan would help me prevent medical problems before they occurred.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
8. I could easily find a good doctor.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
9. The service would use modern, up-to-date treatment methods.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
10. No one has access to my medical record except medical personnel.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
11. There would not be a high continuing interest in my health care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
12. The services would use the best possible hospitals.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
13. Too much work would be done by nurses and assistants rather than doctors.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
14. It would be an organized and complete medical service for me and my family.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
15. There would be much redtape and bureaucratic hassle.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
16. Highly competent doctors and specialists would be available to serve me.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>

After reading the description below of the HARVARD COMMUNITY HEALTH PLAN, please rate how you feel about it (as if you were a member) on the following page.

DESCRIPTION OF THE HARVARD COMMUNITY HEALTH PLAN

Blue Cross announces a new health care plan for YOU AND YOUR FAMILY. By joining the HARVARD COMMUNITY HEALTH PLAN you can get comprehensive health care at a prepaid monthly charge. Almost all your medical needs will be met. You will be assured that you can find a good doctor for yourself and your family and you will not have unexpected doctor or hospital bills.

The cost of joining the HARVARD COMMUNITY HEALTH PLAN is just a little greater than regular Blue Cross/Blue Shield health insurance, but you get more services and comprehensive care. There is only a \$1.00 charge for doctor visits and no charges for nursing, laboratory, and hospital services. Women in the plan pay no extra charges for prenatal, delivery, and maternity care. The services are comprehensive and include emergency services and mental health care.

The HARVARD COMMUNITY HEALTH PLAN provides its care at either the Health Center in Kenmore Square or Inman Square. Parking is available at the health centers. Hospital services are provided in Boston at either Beth Israel, Boston Hospital for Women, Peter Bent Brigham Hospital, or the Children's Medical Center. In Cambridge, services are supplied at the Cambridge City Hospital. For emergencies outside the Boston area, local hospitals may be used.

You choose your own personal doctor from our staff of participating physicians. You will have a specialist in internal medicine for yourself and a pediatrician for your children. Your doctor directs your total care at the Health Center and in the hospital. He will be sure you get the highest quality of care. When you are a member of the HARVARD COMMUNITY HEALTH PLAN you can be sure of getting health care 24 hours a day from the staff of physicians, nurses, social workers, and allied health personnel.

The plan succeeds by keeping you and your family well and out of the hospital. The plan utilizes preventive care. The plan employs trained paramedics and technology to lower costs while maintaining the quality of care and control costs.

You can become a member of the plan by paying \$6.50 per month more than your Blue Cross/Blue Shield coverage if you are single or \$20.00 more per month if you are married. If you are a single student and do not have hospital insurance, the cost is \$13.25/month more than the student health fee you are now paying; if you are a married student, the cost is \$36.00/month more than the student health fee. These fees cover all of your medical costs except: \$1.00 per doctor visit (\$5. for necessary home visit and \$10. per visit for mental health services (if more than 15 visits are made). Prescriptions are available at a low cost. The plan does not include eye glasses, hearing aids, cosmetic surgery, custodial confinement or dental care done outside a hospital. If you join the plan you are committed to it for one year.

THE HARVARD COMMUNITY HEALTH PLAN is designed to make comprehensive health care available to you and your family at a low cost and high quality.

YOUR RATING OF THE HARVARD COMMUNITY HEALTH PLAN

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1. I would be able to get medical service and advice easily any time of the day and night.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
2. I would have to wait a long time to get service.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
3. I could trust that I am getting really good medical care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
4. The health services would be inconveniently located and would be difficult to get to.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
5. I would be paying too much for my required medical services.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
6. I would get a friendly, warm, and personal approach to my medical problems.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
7. The plan would help me prevent medical problems before they occurred.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
8. I could easily find a good doctor.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
9. The service would use modern, up-to-date treatment methods.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
10. No one has access to my medical record except medical personnel.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
11. There would not be a high continuing interest in my health care.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
12. The services would use the best possible hospitals.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
13. Too much work would be done by nurses and assistants rather than doctors.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
14. It would be an organized and complete medical service for me and my family.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
15. There would be much redtape and bureaucratic hassle.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
16. Highly competent doctors and specialists would be available to serve me.	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>

15. You have now read and rated three new health plans in addition to your existing medical service. We are now interested in your preference for these alternatives. Below are listed the alternatives. Place a "1" next to the one which would be your first choice. Place a "2" next to your second choice. Place a "3" next to your third choice. Place a "4" next to your last choice.

- Your existing health services
- HARVARD COMMUNITY HEALTH PLAN
- M.I.T. HEALTH PLAN
- MASSACHUSETTS HEALTH FOUNDATION

16. If only the M.I.T. HEALTH PLAN and your existing health service were available which would you actually choose. Check one.

- Your existing health services.
- M.I.T. HEALTH PLAN

16a. If you selected the M.I.T. HEALTH PLAN, which of the following statements reflects how you feel about your choice: Check one. (Otherwise go to Question 17.)

- I definitely would select the M.I.T. HEALTH PLAN
- I probably would select the M.I.T. HEALTH PLAN
- I might select the M.I.T. HEALTH PLAN

17. If only the M.I.T. HEALTH PLAN and THE MASSACHUSETTS HEALTH FOUNDATION and your existing health system were available, which would you choose? Check one.

- Your existing health services
- M.I.T. HEALTH PLAN
- MASSACHUSETTS HEALTH FOUNDATION

17a. Which of the following statements reflects how you feel about your choice: Check one.

- I definitely would select this alternative
- I probably would select this alternative
- I might select this alternative

18. If you were actually considering joining the M.I.T. HEALTH PLAN would you take actions to get more information?

YES NO

If YES, what would you do?

GO TO QUESTION 21.

21-39. The final section of this questionnaire concerns several important facts about you. These are important in understanding your responses and projecting the results to the M.I.T. community. As you recall your answers are not identified with your name and all responses are confidential. These are important questions and you should find them easy to answer.

21. Indicate your relationship to M.I.T.

<input type="checkbox"/> student	(<input type="checkbox"/> undergraduate) (<input type="checkbox"/> graduate)	<input type="checkbox"/> spouse of student	(<input type="checkbox"/> undergraduate) (<input type="checkbox"/> graduate)
<input type="checkbox"/> staff	(<input type="checkbox"/> hourly) (<input type="checkbox"/> biweekly) (<input type="checkbox"/> monthly) (<input type="checkbox"/> instructor) (<input type="checkbox"/> other)	<input type="checkbox"/> spouse of staff	(<input type="checkbox"/> hourly) (<input type="checkbox"/> biweekly) (<input type="checkbox"/> monthly) (<input type="checkbox"/> instructor) (<input type="checkbox"/> other)
<input type="checkbox"/> faculty	(<input type="checkbox"/> asst. prof.) (<input type="checkbox"/> assoc. prof.) (<input type="checkbox"/> full prof.)	<input type="checkbox"/> spouse of faculty	(<input type="checkbox"/> asst. prof.) (<input type="checkbox"/> assoc. prof.) (<input type="checkbox"/> full prof.)

22. Are you now-a member of the M.I.T. pilot Health Plan? YES NO

23. How many years have you been in the M.I.T. community?

1 year, 2 years, 3 years, 4 years, more than 5 years.

24. Age

15-20 years, 21-24 years, 25-29 years, 30-34 years, 35-39 years,
 40-44 years, 45-49 years, 50-54 years, 55-59 years, over 60.

25. What is your sex?

male female

26. Where do you live?

<input type="checkbox"/> M.I.T. Campus	<input type="checkbox"/> Brighton
<input type="checkbox"/> Cambridge (not on campus)	<input type="checkbox"/> Watertown
<input type="checkbox"/> Boston	<input type="checkbox"/> Belmont
<input type="checkbox"/> Somerville	<input type="checkbox"/> Newton
<input type="checkbox"/> Arlington	<input type="checkbox"/> Charlestown
<input type="checkbox"/> Brookline	<input type="checkbox"/> suburban city or town: how many miles from M.I.T.?

27. How much do you pay for health insurance (how much do you actually pay out of pocket)?

\$5-\$10/month, \$10-\$15/month, \$15-\$20/month, \$20-\$30/month,
 \$30-\$50/month, greater than \$50/month, Do not know.

28. Your insurance may not have paid for all of your health expenses (medical and dental). In the last year, how much did you have to pay for doctors, hospitals, and prescriptions, over and above what was covered by your insurance? If you do not know exactly, make your best estimate.

0-\$50, \$50-\$100, \$100-\$200, \$200-\$300, \$300-\$400, \$400-\$500,
 \$500-\$750, \$750-\$1000, greater than \$1000.

29. How many times have you, your spouse, or children (living at home) visited a medical doctor in the last year?

once, 2-3 times, 4-5 times, 6-12 times, 12-24 times,
 24-50 times, more than 50 times.

30. Have you or your spouse or children (living at home) been hospitalized in the last year?
 NO, go on to question 31.

YES, total number of days you or members of your family spent in the hospital last year.

one day, 2-3 days, 4-5 days, 6-7 days,
 7-14 days, 21-50 days, greater than 50 days.

31. When did you last visit the dentist?

last month, 6 months ago, 6-12 months ago, 12-24 months ago,
 more than 24 months ago.

31a. What did he do?

clean teeth, fill cavities (how many were filled?),
 extract teeth, other (please describe _____)

32. Are you married? NO YES (if no children, go on to Question 33.)

32a. How many children? Ages: , , , , , ...

32b. Are you/is your wife pregnant? YES NO

32c. Do you plan to have a child in the next year? YES NO

32d. How many more children would you definitely like to have?

none, one, two, three or more.

33. During the last year how many days were you unable to work due to medical problems?

none, 1-3 days, 4-7 days, 7-14 days, 14-21 days, more than 21 days.

34. Are you currently being treated by a doctor for a continuing illness?

NO

YES, how many times do you see him?

about once a week, less than once a week,
 more than once a week.

35. How would you rate your overall health?

extremely good, good, OK, not so good, poor

36. If you are married, rate your spouse's health.

extremely good, good, OK, not so good, poor

37. If you have children, how would you rate your children's overall health?

extremely good, good, OK, not so good, poor

38. What is your most important health problem?

39. In the space below we would welcome any other comments you would like to make.

We sincerely want to thank you for completing this questionnaire - THANKS!

APPENDIX TWO

Second Questionnaire

Personal Survey of 80 M.I.T. Students

(actual survey was printed on one side of the page only)

attempt to contact.

POTENTIAL RESPONDENT'S NAME _____

CONTACTED BY: _____ DATE: _____ HOUR: _____

IF APPOINTMENT REJECTED, REASON GIVEN: no answer _____ other _____

IF APPOINTMENT ACCEPTED, APPOINTMENT DATE: _____

APPOINTMENT HOUR: _____

MEETING PLACE: _____

RESPONDENT'S PHONE NUMBER: _____
office

home

DELIVERY ADDRESS FOR WRITTEN QUESTIONNAIRE:

.....

CROSS VALIDATION CHECK BY: _____

DATE: _____ HOUR: _____

Sample Telephone Call

Hello, my name is _____, I am involved in a research effort to determine health care requirements and desires of consumers such as you. This study is sponsored by the Sloan School of Management in cooperation with the M.I.T. health department.

Your name has been selected at random to be part of this survey of health care attitudes and preferences in the M.I.T. community. The number of people being asked to participate is small, so your answers are very important. The results of the survey will also be used by the M.I.T. Medical Department to remain as responsive as possible to the needs of the M.I.T. community.

We are asking for help in completing a questionnaire. The questionnaire requests information about the health services now available to you and asks for your opinion on several new methods of delivering health care. The last questions relate to some demographic characteristics that are important in projecting the responses from this small survey to the M.I.T. community as a whole. The questionnaire takes about 20-30 minutes to complete. Most people find it easy and interesting to answer the questions and they think it is important to make their feelings known to those who provide health care. After filling out the questionnaire I would like to ask you some additional questions in person.

It is important to M.I.T. in planning for your needs and it is important for our research project to improve understanding of consumer response to health services. Could I deliver this questionnaire?

Interviewer: Get respondents address and set up a time for a personal interview.

PERSONAL INTERVIEW ON HEALTH CARE PREFERENCES

1. You will recall you read and rated various aspects of three new health care plans and your existing services. For example, the new M.I.T. Health Plan.

Interviewer: Open the mailed questionnaire to the description of the M.I.T. Health Plan and allow respondent to review it if necessary.

What did you like or dislike about this plan?

Interviewer: Allow a minute or two of open ended response on each plan including if appropriate, existing care, writing down key phrases on likes and dislikes. This question serves the purpose of reviewing the concepts with the respondent, understanding consumer semantics which may help later in the interview, and gaining valuable critiques of the plans.

2. Now we would like you to evaluate the complete plans. Imagine you are given 11 chips to allocate between two plans according to how much you like each one. For example if you liked your existing care just a little more than the Massachusetts Health Foundation, you might allocate 6 chips to your existing care and 5 chips to the Massachusetts Health Foundation:

Existing care: Massachusetts Health Foundation

If you prefer your existing care 10 times as much as Massachusetts Health Foundation you would put 10 chips in the box for existing care and 1 chip in the box for Massachusetts Health Foundation.

Interviewer: Hand evaluation sheet to respondent and have him fill it out.

3. In question 9 of the mailed survey you described your feelings about some specific characteristics of the health care you are now receiving. We would also like to know how you feel in general about this care. Please consider the following factors which describe your health care: (1) quality of care, (2) personalness of service, (3) convenience of obtaining care, and (4) value for the price paid. To express your feelings consider carefully each scale and rate your (and your family's) existing care. For example:

Interviewer: Hand example rating sheet to respondent.

If you feel that the overall quality of your existing plan is poor, you would put an X in box 3. If you feel that the quality is excellent you would put an X in box 7.

Interviewer: If respondent wishes to rate a plan as 5½ allow him to put an X between box 5 and box 6.

Now consider the four scales on the next page and place an X in the box which indicates your general feelings about the current health care you (and your family) use. Remember that these are your feelings and that there are no right or wrong answers.

Interviewer: Hand rating sheet for existing health services to respondent and have him fill it out. Point out that the scale for quality has a different range than the other scales.

Interviewer: Allow respondent to relate to the scales in any way he feels comfortable. Just so long as he remains consistent throughout the survey.

4. Now we would like to learn about your general feelings on the three new health care delivery systems described earlier in this questionnaire. You are asked to rate each system on the same basis as you rated your existing system. We are interested in your overall response to each system based on the one page description. There are no right or wrong answers so just indicate how you think the new service would be if you actually were part of the new health care delivery system.

Interviewer: Hand rating sheets to respondent one at a time for M.I.T., Massachusetts Health Plan, and Harvard Community Health Plan.

5. The next set of questions allows you to indicate your feelings on the relative importance of each of the four factors: quality, personalness, convenience, and value. To do this you are given two hypothetical health care plans. The first plan is rated on each of the four scales. The second plan is rated on 3 of the 4 scales with the scale value of the fourth factor left blank. You are asked to set the level of the fourth factor in such a way as to make the two plans equal in your eyes. In other words, how would the second plan have to rank on the scale that has been left blank in order for both plans to look equally attractive to you as a potential health plan member.

Perhaps an example would better explain this type of question.

Interviewer: Hand example of comparison question to respondent.

Notice that quality and personalness are satisfactory for both plans. Suppose for a moment that the value of plan B were extremely poor (rated 1). Would you prefer plan A or plan B? _____

Interviewer: Respondent should prefer A to B. If not, make sure he understands question.

Suppose the value of plan B were excellent (rated 7). Would you prefer plan A or plan B? _____

Interviewer: Most respondents will now switch their preference and prefer B to A. A few will not. If they do not this means that convenience is much more important to them than value. If they do not switch make sure this is their reason and that they understand the question.

What you are asked to do is choose some value, say 2 (very poor) such that you neither prefer plan A or plan B.

Interviewer: Allow respondent to consider setting the scale for value. After he sets it, allow him to consider the choice between plan A and plan B with his setting. If he prefers either one, he has set the scale wrong. Iterate to get the correct value and to make sure he understands the question.

Now consider the plans on the next page and choose the level of the fourth factor in such a way that you are indifferent between the two plans.

Interviewer: Hand plan comparison sheet 1 to respondent and have him fill it out. (You may wish to iterate, giving him the choice once the scale is set and making sure he has indeed given you an indifference value.)

Consider the pair C-D again. If quality and convenience were both poor (rated 3) you would probably be less satisfied with both plans. We would like to know if this would change your answers. Remember to express your true feelings, there are again no right or wrong answers. You may refer to your original answers again if you wish.

Interviewer: Hand comparison sheet 2 to the respondent and have him fill it out. Give him or her sheet 1 for comparison. Emphasize that it is perfectly okay to have the same answers on both sheet 1 and sheet 2. It is also perfectly okay to change.

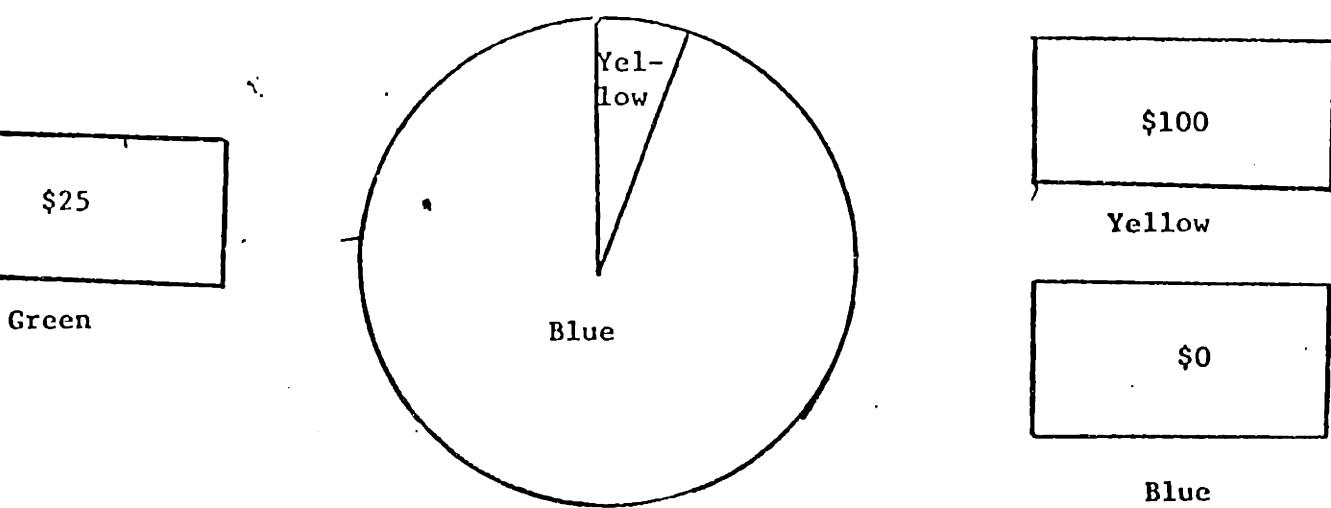
6: The next set of questions allows you to express how important you feel it is that you can be certain of the characteristics of your health care plan. Most people find these questions difficult but interesting to answer, but feel it is important to express their feelings on this aspect of health care.

To better understand this question, let us first consider a little game of chance. Imagine that someone is going to spin this wheel.

Interviewer: Show respondent the probability wheel with the yellow area set to 1/10.

If it comes up yellow (chances are 1 in 10) you win \$100; if it comes up blue you don't win anything.

Interviewer: Lay out the probability wheel and the yellow and blue 3x5 cards marked \$100 and \$0 respectively. I.e.,



Interviewer: Now lay out the green card marked \$25 and ask:
"Would you accept \$25 instead of playing this game?"

YES

NO

Most respondents would accept \$25 rather than play the game. Make sure to emphasize that the game is a once and done game and that if chosen they can not switch after seeing the consequences. Also you may wish to point out that the "expected monetary value" of the game is \$10.

A few respondents will still rather play the game. This is okay if they truly understand the question and have a valid reason.

Now assume that the odds are improved to 9 in 10.

Interviewer: Flip the probability wheel over such that the yellow area is now 9/10 of the wheel. The props in front of the respondent should now look like those in figure 2.

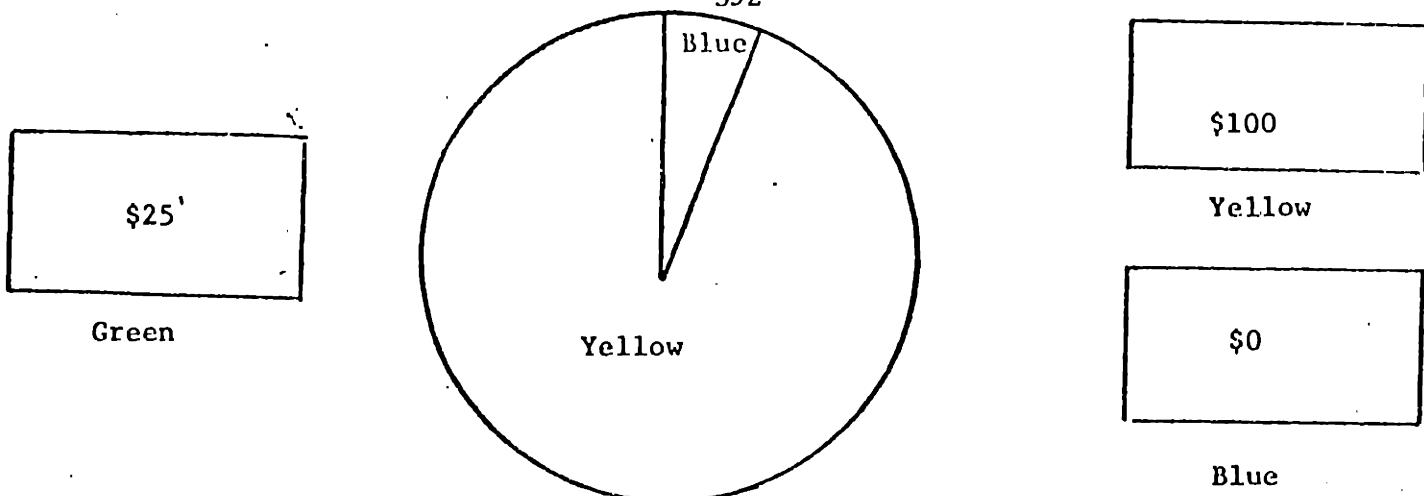


FIGURE 2

If the wheel comes up yellow you win \$100, blue means no win. Would you accept \$25 rather than play this game?

YES

NO

Interviewer: Most respondents would now rather play the game, although a few may prefer \$25. Make sure their reasons are valid and that they truly understand the game. You may wish to point out that the "expected monetary value" of the game is \$90.

Imagine now this mysterious person is going to ask you to set the odds. In other words, set the size of the yellow area.

Interviewer: Change the size of the yellow area of the probability wheel to give the respondent an idea of how the odds can change. Then ask him:

Try to find some "indifference" setting such that you would play the game if the setting were larger, but you would not play the game if the setting were smaller.

In other words, at what setting of the odds would you be indifferent between playing and not playing the game?

Interviewer: Either allow the respondent to play with the wheel and select his own odds, or change it for him in an interactive way such that together you arrive at an indifference setting.

Interviewer: Present him with the lottery again using the odds he has selected. If he strongly prefers either the \$25 or the lottery he has set the wrong odds. If so, iterate through again to get the correct odds. Some respondents will answer 25%. This means that they are "risk neutral". Watch out here that they are not "playing the expected value game". Try to make the lottery a real situation to them and emphasize there are no right or wrong answers and that they should express their true feelings.

The next 4 questions concern attributes of health care plans, and allow you to express how you feel about guaranteed service.

6a. Imagine you can only choose between two health plans, plan 1 and plan 2. In both plans personalness, convenience, and value are good (rated 5). You are familiar with plan 1 and know that quality is satisfactory plus (rated 4). You are not sure of the quality of plan 2. In fact it is as if you were playing the game with the spinning wheel. That is, if you choose plan 2, then the wheel is spun and the quality you will experience for the entire year depends on the outcome of the wheel. If it comes up yellow, the quality is very good (rated 6) and if it comes up blue the quality is just adequate (rated 2). Graphically this is stated:

Interviewer: Set out the 3x5 cards for the plan 1-plan 2 lottery as shown in figure 3.

Interviewer: If respondent is unsure of the meanings of the scales allow him to look at scales for existing health care.

Interviewer: Give the probability wheel to the respondent and ask the following question:

At what setting of the odds (size of the yellow area) would you be indifferent between plan 1 and plan 2?

Interviewer: Allow respondent to set the odds on the probability wheel. Record answer on Interviewer's Answer Sheet under heading "1st time". Try to get a specific value not a range of odds.

Interviewer: Most respondents will be "risk adverse" and set the yellow area at greater than 50%. As always, if they set it lower this is okay as long as they truly understand the question.

Interviewer: Present him with the lottery again using the odds he has selected. If he strongly prefers either the reliable plan or the lottery he has set the wrong odds. If so iterate through again to get the correct odds.

Interviewer: Now change the 3x5 cards such that they appear as in figure 4. This is simply accomplished by flipping the cards for each plan over. Now ask the following question:

Plan 1

P = 5	Q = 4
C = 5	(Quality)
V = 5	

GREEN

Plan 2

P = 5	Q = 6
C = 5	(Quality)
V = 5	

YELLOW

Plan 2

P = 5	Q = 2
C = 5	(Quality)
V = 5	

BLUE

RULES

- wheel is spun after you make your decision
- you must accept the consequences and can not switch

PINK

FIGURE 3

Plan 1	
P = 3	Q = 4 (Quality)
C = 3	
V = 3	

GREEN

Plan 2	
P = 3	Q = 6
C = 3	(Quality)
V = 3	

YELLOW

Plan 2	
P = 3	Q = 2
C = 3	(Quality)
V = 3	

BLUE

RULES	
- wheel is spun <u>after</u> you	make your decision
- you must accept the	consequences and can not
	switch

PINK

FIGURE 4

If in both plans personalness, convenience, and value were all poor (rated 3) you would be less satisfied with each plan. Would this now change your feeling about the reliability of quality?

Interviewer: If he answers "no I would leave the odds the same", then record the same odds on Interviewer's Answer Sheet under heading "2nd time".

Interviewer: Emphasize that it is perfectly okay to have the same setting of the wheel. There are no right or wrong answers and he should try to express his true feelings. It is also perfectly okay to change. You must be very careful to be objective and not to influence the respondent so he feels he should (or should not) change the setting. Record setting.

6b. Imagine you are faced with another choice, but this time the unreliable factor is convenience. Graphically this is stated:

Interviewer: Use same technique for questions 5b, 5c, and 5d, that you used for 5a. Of course changing the cards for appropriate plan comparisons. Make sure to ask the change question as well as the indifference questions.

6c. You are given still another choice, but this time with personalness.

6d. One final time with value:

7. You have just expressed how important reliability is in your personal health care. If you will bear with us, we would like to ask one more question to determine how interdependent the various aspects of health care are. To do this we would like you to play the game with the spinning wheel 1 more time. Again you are offered two health plans.

The first plan, plan A, has very good (rated 6) quality but has very poor (rated 2) personalness, convenience and value.

The second plan, plan B is uncertain in every regard. In fact, it is as if someone spins the wheel. If it comes up yellow, quality, personalness, convenience, and value are all very good (rated 6). If it comes up blue, personalness, convenience, and value are all poor (rated 2), and quality is just adequate.

Interviewer: Lay out the 5x7 cards as shown in figure 5.

Plan A

Quality	= 6 (very good)
Personalness	= 2
Convenience	= 2 } (very poor)
Value	= 2

GREEN

Plan B

Quality	= 6
Personalness	= 6
Convenience	= 6 } (very good)
Value	= 6

YELLOW

Plan B

Quality	= 2 (just adequate)
Personalness	= 2 } (very poor)
Convenience	= 2 }
Value	= 2

RULES

- wheel is spun after you make your decision
- you must accept the consequences and can not switch

PINK

FIGURE 5

How large should the yellow area be to make you indifferent between the two plans.

Interviewer: Give the respondent the probability wheel and allow him to set the odds.

Interviewer: Finally present him with the lottery again using the odds he has selected. If he strongly prefers either the reliable plan or the lottery he has set the wrong odds. If so, iterate through and get the correct odds.

Interviewer: Record the setting of the yellow area on Interviewer's Answer Sheet.

Interviewer: Collect all materials making sure that all questions have been answered. You should have:

- 1) mailed questionnaire
- 2) key phrases on likes and dislikes
- 3) evaluation sheet
- 4) 4 rating sheets (existing, MIT HCHP, MHF)
- 5) plan comparison sheets 1 and 2
- 6) interviewer's answer sheet
- 7) all props and examples

I sincerely want to thank you for helping with this project. Your answers will be held in strict confidentiality and will be valuable input to the planning of health care delivery systems.

PROPS REQUIRED

- (1) Probability wheel blue/yellow calibrated in 1/10's plus lines for 1/20's.
- (2) money cards. Use format in figure 1.
- (3) 3x5 cards for health care. Use the format in figures 3 and 4.

The following 12 cards are needed:

<u>Front</u>	<u>Color</u>	<u>Back</u>
1. Plan 1 P, C, V = 5 Q = 4	green	Plan 1 P, C, V = 3 Q = 4
2. Plan 2 P, C, V = 5 Q = 6	yellow	Plan 2 P, C, V = 3, Q = 6
3. Plan 2 P, C, V, = 5 Q = 2	blue	Plan 2 P, C, V = 3 Q = 2
4. Plan 3 Q, P, V = 5 C = 4	green	Plan 3 Q, P, V = 3, C = 4
5. Plan 4 Q, P, V = 5 C = 6	yellow	Plan 4 Q, P, V = 3 C = 6
6. Plan 4 Q, P, V = 5 C = 2	blue	Plan 4 Q, P, V = 3 C = 2
7. Plan 7 Q, V, C = 5 P = 4	green	Plan 7 Q, V, C = 3 P = 4

8.	Plan 8 Q, V, C = 5 P = 6	yellow	Plan 8 Q, V, C = 3 P = 6
9.	Plan 8 Q, V, C = 5 P = 2	blue	Plan 8 Q, V, C = 3 P = 2
10.	Plan 5 Q, P, C = 5 V = 4	green	Plan 5 Q, P, C = 3 V = 4
11.	Plan 6 Q, P, C = 5 V = 6	yellow	Plan 6 Q, P, C = 3 V = 6
12.	Plan 6 Q, P, C = 5 V = 2	blue	Plan 6 Q, P, C = 3 V = 2

(4) 3x5 rules card. Any color except blue, green, or yellow. Use format in figure 3.

(5) 5x7 corner pt cards. Use format in figure 5.

MATERIALS FOR PROPS

8-1/2 x 11 sheets	1 each	4, B
3x5 cards	5 each	4, B, G
5x7 cards	1 each	4, B, G

KEY PHRASES ON LIKES AND DISLIKES

NEW MIT HEALTH PLAN

HARVARD COMMUNITY HEALTH PLAN

MASSACHUSETTS HEALTH FOUNDATION

EXISTING CARE

Interviewer's Initials

Form 1

EVALUATION SHEET

New MIT Health Plan

--	--

Harvard Community Health Plan

Existing

--	--

New MIT Health Plan

Harvard Community Health Plan

--	--

Massachusetts Health Foundation

Massachusetts Health Foundation

--	--

New MIT Health Plan

Existing

--	--

Harvard Community Health Plan

Massachusetts Health Foundation

--	--

Existing

EXAMPLE RATING SHEET

Personalness (warm, friendly, personal approach, doctors not assistants, no red tape or bureaucratic hassle.)

Extremely poor poor satisfactory good very good excellent

1 2 3 4 5 6 7

YOUR EXISTING HEALTH SERVICES

ality

(trust in doctor, easily find a good doctor, up-to-date treatment, good hospitals, competent doctors and specialists).

very poor	just adequate	Satisfactory	Satisfactory plus	good	very good	excellent
1	2	3	4	5	6	7

rsonalness

(warm, friendly, personal approach, doctors not assistants, no red tape or bureaucratic hassle).

Extremely poor	very poor	poor	satisfactory	good	very good	excellent
1	2	3	4	5	6	7

onvenience

(availability of service, waiting time, location, open hours)

Extremely poor	very poor	poor	satisfactory	good	very good	excellent
1	2	3	4	5	6	7

alue

(good buy for the money paid)

Extremely poor	very poor	poor	satisfactory	good	very good	excellent
1	2	3	4	5	6	7

HARVARD COMMUNITY HEALTH PLAN

(trust in doctor, easily find a good doctor, up-to-date treatment, good hospitals, competent doctors and specialists).

very poor	just adequate	Satisfactory	Satisfactory plus	good	very good	excellent
—	—	—	—	—	—	—
1	2	3	4	5	6	7

(warm, friendly, personal approach, doctors not assistants, no red tape or bureaucratic hassle).

Extremely poor	very poor	poor	satisfactory	good	very good	excellent
—	—	—	—	—	—	—
1	2	3	4	5	6	7

(availability of service, waiting time, location, open hours)

Extremely poor	very poor	poor	satisfactory	good	very good	excellent
—	—	—	—	—	—	—
1	2	3	4	5	6	7

(good buy for the money paid)

Extremely poor	very poor	poor	satisfactory	good	very good	excellent
—	—	—	—	—	—	—
1	2	3	4	5	6	7

MASSACHUSETTS HEALTH FOUNDATION

Quality

(trust in doctor, easily find a good doctor, up-to-date treatment, good hospitals, competent doctors and specialists).

very poor just Satisfac- Satisfac- good very good excellent
adequate tory tory plus

1 2 3 4 5 6 7

Personalness

(warm, friendly, personal approach, doctors not assistants, no red tape or bureaucratic hassle).

Extremely very poor poor satisfac- good very good excellent
poor tory

1 2 3 4 5 6 7

Convenience

(availability of service, waiting time, location, open hours)

Extremely very poor poor satisfac- good very good excellent
poor tory

1 2 3 4 5 6 7

Value

(good buy for the money paid)

Extremely very poor poor satisfac- good very good excellent
poor tory

1 2 3 4 5 6 7

NEW MIT HEALTH PLAN

Quality

(trust in doctor, easily find a good doctor, up-to-date treatment, good hospitals, competent doctors and specialists).

very poor just Satisfac- Satisfac- good very good excellent
adequate tory tory plus

1 2 3 4 5 6 7

Personalness

(warm, friendly, personal approach, doctors not assistants, no red tape or bureaucratic hassle).

Extremely very poor poor satisfac- good very good excellent
poor tory

1 2 3 4 5 6 7

Convenience

(availability of service, waiting time, location, open hours)

Extremely very poor poor satisfac- good very good excellent
poor tory

1 2 3 4 5 6 7

Value

(good buy for the money paid)

Extremely very poor poor satisfac- good very good excellent
poor tory

1 2 3 4 5 6 7

EXAMPLE COMPARISON QUESTION

Plan A

Quality = 4 (satisfactory plus)

Personalness = 4 (satisfactory)

Convenience = 5 (good)

Value = 1 (extremely poor)

Plan B

Quality = 4 (satisfactory plus)

Personalness = 4 (satisfactory)

Convenience = 2 (very poor)

Value =

Excellent
Very good
Good
Satisfactory
Poor
Very poor
Extremely poor

PLAN COMPARISONS: SHEET 1

Plan C

Value = 5 (good)
Convenience = 5 (good)
Personalness = 6 (very good)
Quality = 2 (just adequate)

Plan D

Value = 5 (good)
Convenience = 5 (good)
Personalness = 2 (very poor)

Quality =

Very poor	Just adequate	Satisfactory	Good	Very good	Excellent
1	2	3	4	5	6

Plan E

Value = 5 (good)
Personalness = 5 (good)
Convenience = 6 (very good)
Quality = 2 (just adequate)

Plan F

Value = 5 (good)
Personalness = 5 (good)
Convenience = 2 (very poor)
Quality =

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Plan G

Convenience = 5 (good)
Personalness = 5 (good)
Value = 5 (very good)
Quality = 2 (just adequate)

Plan H

Convenience = 5 (good)
Personalness = 5 (good)
Value = 2 (very poor)
Quality =

1	2	3	4	5	6	7
---	---	---	---	---	---	---

PLAN COMPARISONS: SHEET 2

Plan J

Value = 3 (poor)
Convenience = 3 (poor)
Personalness = 6 (very good)
Quality = 2 (just adequate)

Plan K

Value = 3 (poor)
Convenience = 3 (poor)
Personalness = 2 (very poor)
Quality = 1 2 3 4 5 6 7

Plan M

Value = 3 (poor)
Personalness = 3 (poor)
Convenience = 6 (very good)
Quality = 2 (just adequate)

Plan N

Value = 3 (poor)
Personalness = 3 (poor)
Convenience = 2 (very poor)
Quality = 1 2 3 4 5 6 7

Plan P

Convenience = 3 (poor)
Personalness = 3 (poor)
Value = 6 (very good)
Quality = 2 (just adequate)

Plan Q

Convenience = 3 (poor)
Personalness = 3 (poor)
Value = 2 (very poor)
Quality = 1 2 3 4 5 6 7

Excellent
Very good
Good
Satisfactory plus
Satisfactory
Just adequate
Very poor

INTERVIEWER'S ANSWER SHEET

Indifference Questions

<u>Uncertain Factor</u>	<u>Setting of Wheel</u>	
	<u>First Time</u>	<u>Second Time</u>
Money	_____ % yellow	_____
Personalness	_____ % yellow	_____
Convenience	_____ % yellow	_____
Quality	_____ % yellow	_____
Value	_____ % yellow	_____

Corner Point Question

Yellow area = _____ %

Comments on interview (if any)

BIOGRAPHICAL NOTE

John R. Hauser earned an S.M. degree in Civil Engineering (Transportation Systems) and S.M. and S.B. degrees in Electrical Engineering from M.I.T. in February 1973. He is currently associated with the Operations Research Center at M.I.T. and has been supported by a National Science Foundation Fellowship and by research grants through the Operations Research Center, the Civil Engineering Department and at the Sloan School of Management, all at M.I.T.

As a research assistant, Mr. Hauser has worked on modeling consumer response to health care delivery systems and to management education programs. In addition he has worked on the design and implementation of Dial-a-Ride in Rochester, N.Y., the design of computer algorithms to integrate Dial-a-Ride with fixed route transit, and the use of consumer utility as an objective function in Dial-a-Ride routing algorithms. He has been a teaching assistant in a marketing seminar and a course on management science in public systems. Earlier he was a consultant to Volunteers in Technical Assistance (VITA), an organization which assists social agencies and small communities in the design of public transportation services, and was on the summer faculty of Ryder Technical Institute in Allentown, PA.

Mr. Hauser's professional papers include "A Normative Methodology for Modeling Consumer Response to Innovation" (with G.L. Urban), 'Modeling Decisions of Choice Among Finite Alternatives: Applications to Marketing and Transportation Demand Theory,' "An Efficient Model for Planning Bus Routes in Communities with Populations Between 20,000 and 250,000," "A Cost Model for RTS (Rochester, N.Y.) Conventional Bus Routes," and "The Chemung County Transit Survey" (with S. Lehrman, G. Harvey, M. McFadden and J. Harvey).

In September 1975, Mr. Hauser will join the faculty of the Graduate School of Management, Northwestern University as an Assistant Professor of Marketing and Transportation.