Contributors

Marvin Adelson	(pp. 433–462)
Samuel Aroni	(pp. 433–462)
Klaus Brockhoff	(pp. 291–321)
J. Douglas Carroll	(pp. 402–431)
Norman C. Dalkey	(pp. 236–261, 327–337, 387–401)
Lawrence H. Day	(pp. 168–194)
Alex Ducanis	(pp. 288–290)
Selwyn Enzer	(pp. 195–209)
Nancy H. Goldstein	(pp. 210–226)
Olaf Helmer	(pp.xix-xx)
Irene Anne Jillson	(pp. 124–159)
Robert Johansen	(pp. 517–534, 550–562)
Chester G. Jones	(pp. 160–167)
Julius Kane	(pp. 369–382)
Harold A. Linstone	(pp. 3–12, 15–16, 75–83, 229–235, 325–326, 385–386, 489–496, 573–586, 589–614)
John Ludlow	(pp. 102–123)
Richard H. Miller	(pp. 517–534)
Ian I. Mitroff	(pp. 17–36)
Norman W. Mulgrave	(pp. 288–290)
Charlton R. Price	(pp. 497–516)
D. Sam Scheele	(pp. 37–71)
M. Scheibe	(pp. 262–287)
J. Schofer	(pp. 262–287)
James A. Schuyler	(pp. 550–562)
Thomas B. Sheridan	(pp. 535–549)
M. Skutsch	(pp. 262–287)
John W. Sutherland	(pp. 463–486)
Murray Turoff	(pp. 3–12, 15–16, 17–36, 75–83, 84–101, 229–235, 325–326, 338–368, 385–386, 489–496, 563–569, 589–614)
Jacques F. Vallee	(pp. 517–534)
Myron Wish	(pp. 402–431)

The Delphi Method
Techniques and Applications

Edited by

Harold A. Linstone
Portland State University

Murray Turoff
New Jersey Institute of Technology

With a Foreword by
Olaf Helmer
University of Southern California

1975

Addison-Wesley Publishing Company
Advanced Book Program
Reading, Massachusetts

London·Amsterdam·Don Mills, Ontario·Sydney·Tokyo

Delphi — The Tholos, Courtesy of the Greek National Tourist Office
(End papers for hardbound edition only.)

The illustration on the cover is taken from a silver stater (about 350 B.C.) formerly in the collection of John Pierpont Morgan. It depicts the Greek god Apollo Pythios seated on the sacred *omphalos* at Delphi. Apollo, son of Zeus and Leto, became master of Delphi upon slaying the dragon Pythos there. He was renowned not only for his youth and perfect beauty, but even more for his ability to foresee the future. The omphalos, located in the Temple of Apollo, was a conically shaped stone representing the navel of the earth and the center of the universe. Thus Delphi assumed the status of the most revered oracular site in Greece.

Library of Congress Cataloging in Publication Data
Main entry under title:

The Delphi method.

"Delphi bibliography": p.
Includes index.
1. Delphi method. I. Linstone, Harold A.
II. Turoff, Murray.
T174.D44 — 001.4'33 — 75-25650
ISBN 0-201-04294-0
ISBN 0-201-04293-2 pbk.

ABCDEFGHIJ-MA-798765

Manufactured in the United States of America

Contents

Biographical Data xiii
The Editors xiii
The Contributors xiv
Foreword by *Olaf Helmer* xix

I. INTRODUCTION 1

Harold A. Linstone and *Murray Turoff*
INTRODUCTION 3
General Remarks 3
Characteristics of the Delphi 5
The Evolution of Delphi 10

II. PHILOSOPHY 13

A. *Harold A. Linstone* and *Murray Turoff*
INTRODUCTION 15

B. *Ian A. Mitroff* and *Murray Turoff*
PHILOSOPHICAL AND METHODOLOGICAL FOUNDATIONS OF DELPHI 17
Introduction 17
Inquiring Systems (IS) 20
Lockean IS 21
Leibnizian IS 23
Kantian IS 25
Hegelian, or Dialectical, IS 29
Singerian–Churchmanian IS 33
Concluding Remarks . . . . . 3[illegible]

C. *D. Sam Scheele*
REALITY CONSTRUCTION AS A PRODUCT OF DELPHI INTERACTION 37
Concepts of Reality 40
Mental Climates and Styles of Thinking 48
Group Influence on Reality Construction 56
What I Meant to Say 64
Summary 64
Some Design Considerations 64

III. GENERAL APPLICATIONS 73

A. *Harold A. Linstone* and *Murray Turoff*
INTRODUCTION 75
Government Planning 75
Business and Industry 76
Comments on Other Studies 78

B.1. *Murray Turoff*
THE POLICY DELPHI 84
Introduction 84
The Committee and the Delphi Process 85
The Mechanics of a Policy Delphi 87
Examples of Policy Delphis 95
The Problems of a Policy Delphi 100

B.2. *John Ludlow*
DELPHI INQUIRIES AND KNOWLEDGE UTILIZATION 102
Introduction 102
Outline of the Procedures: Social, Political, and Economic Trends 104
Sources of Pollution 108
Recommended Waste-Water Treatment and Disposal Systems 110
Regional Opportunities, Problems, and Planning Strategies 112
Evaluation of the Methodology 114
Summary of Findings and Recommendations 119

B.3. *Irene Anne Jillson*
THE NATIONAL DRUG-ABUSE POLICY DELPHI: PROGRESS REPORT AND FINDINGS TO DATE 124
Introduction 124
The Questionnaires 127
Results 134
Preliminary Epilogue 158

B.4. *Chester G. Jones*
A DELPHI EVALUATION OF AGREEMENT BETWEEN ORGANIZATIONS 160
Introduction 160
Background 160
Convergence 162
Correlation between Questions 163
Bias by Time Period 164

Laboratory Bias 165
Summary 167

C.1. *Lawrence H. Day*
DELPHI RESEARCH IN THE CORPORATE ENVIRONMENT 168
Introduction 168
Examples of Corporate Delphi Research 168
Delphi and Bell Canada 171
Delphi in the Corporate Environment 187
Future Use of Delphi in Corporations 191

C.2. *Selwyn Enzer*
PLASTICS AND COMPETING MATERIALS BY 1985: A DELPHI FORECASTING STUDY 195
Anticipated Changes in Properties of Existing Materials 196
Other Materials Suggested by the Panel as Likely to Become Important by 1985 202
Overall Forecasts of U.S. Plastic Production 205

C.3. *Nancy H. Goldstein*
A DELPHI ON THE FUTURE OF THE STEEL AND FERROALLOY INDUSTRIES 210
Introduction 210
Round One 212
Round Two 216
Round Three 220
Comparison of Delphi and Panel Studies 222
Advice to Future Delphi Monitors and Designers 226

IV. EVALUATION 227

A. *Harold A. Linstone* and *Murray Turoff*
INTRODUCTION 229
Dispersion as a Function of Remoteness of Estimate 229
Distribution of Responses 230
Optimism-Pessimism Consistency 230
Accuracy of Forecasts 231
Delphi Statements 232
Basis for Respondents' Intuitive Forecast 233
Self-Rating of Experts 233

B. *Norman C. Dalkey*
TOWARD A THEORY OF GROUP ESTIMATION 236

Theory of Errors 249
Probabilistic Approach 251
Axiomatic Approach 257

C. *M. Scheibe, M. Skutsch,* and *J. Schofer*
EXPERIMENTS IN DELPHI METHODOLOGY 262
Introduction 262
Description of Procedure 263
An Experiment on the Effect of Feedback 267
Comparison of Scaling Techniques 272
The Effect of Personal Variables on Participant Behavior 273
Opinion Stability as a Method of Consensus Measurement 277
Delphi Service Program 281
Closure 282

D. *Norman W. Mulgrave* and *Alex J. Ducanis*
PROPENSITY TO CHANGE RESPONSES IN A DELPHI ROUND AS A FUNCTION OF DOGMATISM 288
Method 288
Results 289
Discussion 290

E. *Klaus Brockhoff*
THE PERFORMANCE OF FORECASTING GROUPS IN COMPUTER DIALOGUE AND FACE-TO-FACE DISCUSSION . . . 291
Group Performance and Communication System 297
Group Performance and Type of Task: Fact-Finding Questions and Forecasting 299
The Relationship between the Hypotheses 300
Participants, Group Formation, Place, and Time 301
The Organization of the Delphi Groups and the Face-to-Face Groups 302
The Questions 305
The Volume of the Tests 306
The Distribution of Expertise 307
The Significance of the Self-Ratings 310
The Performance of the Delphi Groups 312
The Size of the Groups 315
Hypothesis *D* 318
Hypothesis *F* 319

V. CROSS-IMPACT ANALYSIS 323

A. *Harold A. Linstone* and *Murray Turoff*
INTRODUCTION 325

B. *Norman C. Dalkey*
AN ELEMENTARY CROSS-IMPACT MODEL 327
Introduction 327
Consistent Systems of Probabilities 329
Resolution of Inconsistencies 333
Scenarios 335
Conclusion 336

C. *Murray Turoff*
AN ALTERNATIVE APPROACH TO CROSS IMPACT ANALYSIS . . . 338
Introduction 338
Theory 341
Example 357
Applications 361
Annotated Bibliography 365

D. *Julius Kane*
A PRIMER FOR A NEW CROSS-IMPACT LANGUAGE —KSIM (WITH EXAMPLES SHOWN FROM TRANSPORTATION POLICY) 369
Introduction 369
Aircraft Competition 370
A Metropolitan Transit Model 375
Conclusions 380

VI. SPECIALIZED TECHNIQUES 383

A. *Harold A. Linstone* and *Murray Turoff*
INTRODUCTION 385

B. *Norman C. Dalkey*
A DELPHI STUDY OF FACTORS AFFECTING THE QUALITY OF LIFE 387
Earlier Related Studies 387
Models of the Quality of Life 388
Procedure 390
Results 391

Additional Analyses 396
Discussion 398

C. *J. Douglas Carroll* and *Myron Wish*
MULTIDIMENSIONAL SCALING:
MODELS, METHODS, AND RELATIONS TO DELPHI 402
Ordinary or "Two-Way"
Multidimensional Scaling 409
Individual Differences in Perception and the
INDSCAL Model 414
Annotated Bibliography 426

D. *Marvin Adelson* and *Samuel Aroni*
DIFFERENTIAL IMAGES OF THE FUTURE 433
Introduction 433
Purpose 434
Background 434
Hypotheses 436
Procedure 436
Results 437
Conclusions and Plans 439
Appendix I: Questionnaires for Images of the
Future Study 441
Appendix II: Some Significant Correlations 447
Appendix III: Expectations II: A Survey
about the Future 448

E. *John W. Sutherland*
ARCHITECTING THE FUTURE:
A DELPHI-BASED
PARADIGM FOR NORMATIVE SYSTEM-BUILDING 463
Prelude 463
The Scenario-Building Logic 464
An Amplification of the Scenario-Building Technology . . . 470
Constructing the Problem Network 477
Generation of Action Proposals (Prescriptive Policies) . . . 480
The "Learning" Aspect 482
Summary 484

VII. COMPUTERS AND THE FUTURE OF DELPHI 487

A. *Harold A. Linstone* and *Murray Turoff*
INTRODUCTION 489

B. *Charlton R. Price*
CONFERENCING VIA COMPUTER: COST EFFECTIVE
COMMUNICATION FOR THE ERA OF FORCED CHOICE 497
Introduction 497
What Is Computer Conferencing? 498
Costs and Benefits, Financial and Otherwise 502
Origins of Computer Conferencing 505
Some Possible Futures for Computer Conferencing 509

C. *Robert Johansen, Richard H. Miller,*
and *Jacques Vallee*
GROUP COMMUNICATION THROUGH ELECTRONIC
MEDIA: FUNDAMENTAL CHOICES AND SOCIAL EFFECTS 517
Current Research Approaches 520
Isolating Key Elements in Mediated Group
Communication 526
Conclusion 532

D. *Thomas B. Sheridan*
TECHNOLOGY FOR GROUP DIALOGUE
AND SOCIAL CHOICE 535
Introduction 535
The Formal Social Choice Problem 537
Potential Contributions of Electronics 539
The Steps in a Group Feedback Session 540
Experiments in Progress 544
Extending the Meeting in Space and Time 548

E. *Robert Johansen* and *James A. Schuyler*
COMPUTERIZED CONFERENCING IN AN
EDUCATIONAL SYSTEM: A SHORT-RANGE SCENARIO 550
Citizen Sampling on Current Events and Futures 552
Course and Curricular Evaluation 554
Committee Work and Long-Range Planning 555
Conflict Management and Diagnosis 557
Interface with Computer-Based Learning 558
The Computer Learns to Talk 559
Possible Misuses of ORACLE 560

F. *Murray Turoff*
MEETING OF THE COUNCIL ON CYBERNETIC STABILITY:
A SCENARIO 563

VIII. EIGHT BASIC PITFALLS: A CHECKLIST 571

Harold A. Linstone
EIGHT BASIC PITFALLS: A CHECKLIST 573
Discounting the Future 574
The Prediction Urge 578
The Simplification Urge 579
Illusory Expertise 581
Sloppy Execution 582
Optimism-Pessimism Bias 584
Overselling 584
Deception 585

APPENDIX: DELPHI BIBLIOGRAPHY 587

Harold A. Linstone and *Murray Turoff*
APPENDIX: DELPHI BIBLIOGRAPHY 589
Distribution of Bibliography 590

SUBJECT INDEX 615

Biographical Data

THE EDITORS

Harold A. Linstone received his B. S. degree in 1944 from the City College of New York and his Ph. D. in Mathematics from the University of Southern California in 1954. He has been a member of the RAND Corporation, a Senior Scientist at Hughes Aircraft Company, and Associate Director of Corporate Development Planning at Lockheed Aircraft Corporation. From 1965 to 1973, he was also Adjunct Professor of Industrial and Systems Engineering at the University of Southern California. There he introduced courses in "Technological Forecasting" and "Planning Alternative Futures." The latter won USC's Dart Award for Innovation in Teaching in 1970. He has presented numerous seminars on Technological Forecasting and Long-Range Planning in the United States and Europe and started the Center for Technological and Interdisciplinary Forecasting at Tel-Aviv University. Since 1970, he has been Professor of Systems Science at Portland State University and Director of its new interdisciplinary Ph. D. Program in this field. He is Senior Editor of the international journal, ***Technological Forecasting and Social Change.***

Murray Turoff received his B. S. degree in 1958 from the University of California at Berkeley in Physics and Mathematics, and his Ph. D. in Physics from Brandeis University in 1965. He has been a Systems Engineer for IBM, a member of the Science and Technology Division of the Institute for Defense Analyses and a Senior Operations Research Analyst for the Office of Emergency Preparedness in the Executive Offices of the President. In 1972-73 he introduced and taught a course in "Technological Forecasting" at American University. Since 1973 he has been an Associate Professor of Computer and Information Science at

I. Introduction

HAROLD A. LINSTONE and MURRAY TUROFF

General Remarks

It is common, in a book of this kind, to begin with a detailed and explicit definition of the subject—the Delphi technique. However, if we were to attempt this, the reader would no doubt encounter at least one contribution to this collection which would violate our definition. There is in addition a philosophical perspective that when something has attained a point at which it is explicitly definable, then progress has stopped; such is the view we hold with respect to Delphi.

In 1969 the number of Delphi studies that had been done could be counted in three digits; today, in 1974, the figure may have already reached four digits. The technique and its application are in a period of evolution, both with respect to how it is applied and to what it is applied. It is the objective of this book to expose the richness of what may be viewed as an evolving field of human endeavor. The reader will encounter in these pages many different perspectives on the Delphi method and an exceedingly diverse range of applications.

For a technique that can be considered to be in its infancy, it would be presumptuous of us to present Delphi in the cloak of a neatly wrapped package, sitting on the shelf and ready to use. Rather, we have adopted the approach, through our selection of contributions, of exhibiting a number of different objects having the Delphi label and inviting you to sculpt from these examples your own view and assessment of the technique. For, if anything is "true" about Delphi today, it is that in its design and use Delphi is more of an art than a science.

However, as editors, we would be remiss if there were not some common thread underlying the articles brought together in this volume. As long as we restrict ourselves to a very general view, it is not difficult to present an acceptable definition of the Delphi technique which can be taken as underlying the contributions to this book:

> Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem.

To accomplish this "structured communication" there is provided: some feedback of individual contributions of information and knowledge; some assessment of the group judgment or view; some opportunity for individuals to revise views; and some degree of anonymity for the individual responses. As the reader will discover, there are many different views on what are the "proper," "appropriate," "best," and/or "useful" procedures for accomplishing the various specific aspects of Delphi. We hope that the reader will find this book a

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)

rich menu of procedures from which he may select his own repast if he should seek to employ the Delphi technique.

When viewed as communication processes, there are few areas of human endeavor which are not candidates for application of Delphi. While many people label Delphi a forecasting procedure because of its significant use in that area, there is a surprising variety of other application areas. Among those already developed we find:

- Gathering current and historical data not accurately known or available
- Examining the significance of historical events
- Evaluating possible budget allocations
- Exploring urban and regional planning options
- Planning university campus and curriculum development
- Putting together the structure of a model
- Delineating the pros and cons associated with potential policy options
- Developing causal relationships in complex economic or social phenomena
- Distinguishing and clarifying real and perceived human motivations
- Exposing priorities of personal values, social goals

It is not, however, the explicit nature of the application which determines the appropriateness of utilizing Delphi; rather, it is the particular circumstances surrounding the necessarily associated group communication process: "Who is it that should communicate about the problem, what alternative mechanisms are available for that communication, and what can we expect to obtain with these alternatives?" When these questions are addressed, one can then decide if Delphi is the desirable choice. Usually, one or more of the following properties of the application leads to the need for employing Delphi:

- The problem does not lend itself to precise analytical techniques but can benefit from subjective judgments on a collective basis
- The individuals needed to contribute to the examination of a broad or complex problem have no history of adequate communication and may represent diverse backgrounds with respect to experience or expertise
- More individuals are needed than can effectively interact in a face-to-face exchange
- Time and cost make frequent group meetings infeasible
- The efficiency of face-to-face meetings can be increased by a supplemental group communication process
- Disagreements among individuals are so severe or politically unpalatable that the communication process must be refereed and/or anonymity assured
- The heterogeneity of the participants must be preserved to assure validity of the results, i.e., avoidance of domination by quantity or by strength of personality ("bandwagon effect")

Hence, for the application papers in this book the emphasis is not on the

used and how it was implemented. From this the reader may be able to transpose the considerations to his own area of endeavor and to evaluate the applicability of Delphi to his own problems.

Those who seek to utilize Delphi usually recognize a need to structure a group communication process in order to obtain a useful result for their objective. Underlying this is a deeper question: "Is it possible, via structured communications, to create any sort of collective human intelligence[1] capability?" This is an issue associated with the utility of Delphi that has not as yet received the attention it deserves and the reader will only find it addressed here indirectly. It will, therefore, be a subjective evaluation on his part to determine if the material in this book represents a small, but initial, first step in the long-term development of collective human intelligence processes.

Characteristics of the Delphi

The Delphi process today exists in two distinct forms. The most common is the paper-and-pencil version which is commonly referred to as a "Delphi Exercise." In this situation a small monitor team designs a questionnaire which is sent to a larger respondent group. After the questionnaire is returned the monitor team summarizes the results and, based upon the results, develops a new questionnaire for the respondent group. The respondent group is usually given at least one opportunity to reevaluate its original answers based upon examination of the group response. To a degree, this form of Delphi is a combination of a polling procedure and a conference procedure which attempts to shift a significant portion of the effort needed for individuals to communicate from the larger respondent group to the smaller monitor team. We shall denote this form *conventional Delphi.*

A newer form, sometimes called a "Delphi Conference," replaces the monitor team to a large degree by a computer which has been programmed to carry out the compilation of the group results. This latter approach has the advantage of eliminating the delay caused in summarizing each round of Delphi, thereby turning the process into a real-time communications system. However, it does require that the characteristics of the communication be well defined before Delphi is undertaken, whereas in a paper-and-pencil Delphi exercise the monitor team can adjust these characteristics as a function of the group responses. This latter form shall be labeled *real-time Delphi.*

Usually Delphi, whether it be conventional or real-time, undergoes four distinct phases. The first phase is characterized by exploration of the subject under discussion, wherein each individual contributes additional information he feels is pertinent to the issue. The second phase involves the process of

[1] We refer to "intelligence" in this context as including attitudes and feelings which are part of the process of human motivation and action.

reaching an understanding of how the group views the issue (i.e., where the members agree or disagree and what they mean by relative terms such as importance, desirability, or feasibility). If there is significant disagreement, then that disagreement is explored in the third phase to bring out the underlying reasons for the differences and possibly to evaluate them. The last phase, a final evaluation, occurs when all previously gathered information has been initially analyzed and the evaluations have been fed back for consideration.

On the surface, Delphi seems like a very simple concept that can easily be employed. Because of this, many individuals have jumped at trying the procedure without carefully considering the problems involved in carrying out such an exercise. There are perhaps as many individuals who have had disappointing experiences with a Delphi as there are users who have had successes. Some of the common reasons for the failure of a Delphi are:

- Imposing monitor views and preconceptions of a problem upon the respondent group by overspecifying the structure of the Delphi and not allowing for the contribution of other perspectives related to the problem
- Assuming that Delphi can be a surrogate for all other human communications in a given situation
- Poor techniques of summarizing and presenting the group response and ensuring common interpretations of the evaluation scales utilized in the exercise
- Ignoring and not exploring disagreements, so that discouraged dissenters drop out and an artificial consensus is generated
- Underestimating the demanding nature of a Delphi and the fact that the respondents should be recognized as consultants and properly compensated for their time if the Delphi is not an integral part of their job function

In addition to the latter problems associated with the Delphi technique another class of criticisms directed at Delphi is often raised in the literature. These are the "virtual" problems that do not in themselves affect the utility of the technique.[2] Typical of these is the question of how to choose a "good" respondent group. This is, in fact, a problem for the formation of any group activity—committees, panels, study groups, etc. One has this problem no matter what communication mode is used; therefore, while it is a real and significant problem, it is not a problem unique to Delphi. However, the nature of certain applications does, in fact, dictate special consideration of this problem and it is discussed in a number of articles. Another virtual problem frequently arises when a particular Delphi design for a particular application is taken as representative of all Delphis, whereupon it is then observed that this design does not work for some other application. The problem here is that of making too explicit and restrictive a definition for Delphi. A third virtual

[2]See, for example, Gordon Welty, "A Critique of the Delphi Technique," *Proceedings of the*

problem is the honesty of the monitor team, and it is of the same concern as the honesty of any study or analysis group. In fact, there is probably more likelihood in most instances of exposure of misrepresentation in a Delphi summary than in a typical group study report. Finally, misunderstandings may arise from differences in language and logic if participants come from diverse cultural backgrounds. Since we consider these virtual issues to be somewhat irrelevant to Delphi per se, we have made no attempt to give them special attention within this book. Other problems will be discussed in Chapter VIII.

It is quite clear that in any one application it is impossible to eliminate all problems associated with Delphi. There is, for example, a natural conflict in the goal of allowing a wide latitude in the contribution of information and the goal of keeping the communication process efficient. It is the task of the Delphi designer to minimize these problems as much as possible and to balance the various communication "goals" within the context of the objective of the particular Delphi and the nature of the participants. Arriving at a balanced design for the communication structure is still very much an art, even though there is considerable experience on how to ask and summarize various types of questions.

It can be expected that the use of Delphi will continue to grow. As a result of this, one can observe that a body of knowledge is developing on how to structure the human communication process for particular classes of problems. The abuse, as well as the use, of the technique is contributing to the development of this design methodology.

Table 1 compares the properties of normal group communication modes and the Delphi conventional and real-time modes. The major differences lie in such areas as the ability of participants to interact with the group at their own convenience (i.e., random as opposed to coincident), the capacity to handle large groups, and the capability to structure the communication. With respect to time considerations, there is a certain degree of similarity between a committee and a conventional Delphi process, since delays between committee meetings and Delphi rounds are unavoidable. Also, the real-time Delphi is conceptually somewhat analogous to a randomly occurring conference call with a written record automatically produced. It is interesting to observe that within the context of the normal operation of these communication modes in the typical organization—government, academic, or industrial—the Delphi process appears to provide the individual with the greatest degree of individuality or freedom from restrictions on his expressions. The items highlighted in the table will be discussed in more detail in many of the articles in this book.

While the written word allows for emotional content, the Delphi process does tend to minimize the feelings and information normally communicated in such manners as the tone of a voice, the gesture of a hand, or the look of an eye. In many instances these are a vital and highly informative part of a communication process. Our categorization of group communication processes is not meant to imply that the choice for a particular objective is limited, necessarily, to one

TABLE 1
Group Communication Techniques

	Conference Telephone Call	Committee Meeting	Formal Conference or Seminar	Conventional Delphi	Real-Time Delphi
Effective Group Size	Small	Small to medium	Small to large	Small to large	Small to large
Occurrence of Interaction by Individual	Coincident with group	Coincident with group	Coincident with group	Random	Random
Length of Interaction	Short	Medium to long	Long	Short to medium	Short
Number of Interactions	Multiple, as required by group	Multiple, necessary time delays between	Single	Multiple, necessary time delays between	Multiple, as required by individual
Normal Mode Range	Equality to chairman control (flexible)	Equality to chairman control (flexible)	Presentation (directed)	Equality to monitor control (structured)	Equality to monitor control or group control and no monitor (structured)

TABLE 1 (continued)

Group Communication Techniques

	Conference Telephone Call	Committee Meeting	Formal Conference or Seminar	Conventional Delphi	Real-Time Delphi
Principal Costs	Communications	—Travel —Individuals' time	—Travel —Individuals' time —Fees	—Monitor time —Clerical —Secretarial	—Communications —Computer usage
Other Characteristics	Time-urgent considerations	Forced delays		Forced delays	Time-urgent considerations
	—Equal flow of information to and from all —Can maximize psychological effects		—Efficient flow of information from few to many	—Equal flow of information to and from all —Can minimize psychological effects —Can minimize time demanded of respondents or conferees	

communication mode. As the readers will see from some of the contributions to this book, there are instances where it is desirable to use a mix of these approaches.

The Evolution of Delphi

The Delphi concept may be viewed as one of the spinoffs of defense research. "Project Delphi" was the name given to an Air Force-sponsored Rand Corporation study, starting in the early 1950's, concerning the use of expert opinion.[3] The objective of the original study was to "obtain the most reliable consensus of opinion of a group of experts ... by a series of intensive questionnaires interspersed with controlled opinion feedback."

It may be a surprise to some that the subject of this first study was the application of "expert opinion to the selection, from the point of view of a Soviet strategic planner, of an optimal U. S. industrial target system and to the estimation of the number of A-bombs required to reduce the munitions output by a prescribed amount." It is interesting to note that the alternative method of handling this problem at that time would have involved a very extensive and costly data-collection process and the programming and execution of computer models of a size almost prohibitive on the computers available in the early fifties. Even if this alternative approach had been taken, a great many subjective estimates on Soviet intelligence and policies would still have dominated the results of the model. Therefore, the original justifications for this first Delphi study are still valid for many Delphi applications today, when accurate information is unavailable or expensive to obtain, or evaluation models require subjective inputs to the point where they become the dominating parameters. A good example of this is in the "health care" evaluation area, which currently has a number of Delphi practitioners.

However, because of the topic of this first notable Delphi study, it took a later effort to bring Delphi to the attention of individuals outside the defense community. This was the "Report on a Long-Range Forecasting Study," by T. J. Gordon and Olaf Helmer, published as a Rand paper in 1964.[4] Its aim was to assess "the direction of long-range trends, with special emphasis on science and technology, and their probable effects on our society and our world." "Long-range" was defined as the span of ten to fifty years. The study was done to explore both the methodological aspects of the technique and to obtain substantive results. The authors found themselves in "a near-vacuum as far as tested techniques of long-range forecasting are concerned." The study covered

[3]N. Dalkey and O. Helmer, "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science* 9, No. 3 (April 1963), p. 458.

[4]Rand Paper P-2982. Most of the study was later incorporated into Helmer's *Social Technology*, Basic Books, New York, 1966.

six topics: scientific breakthroughs; population control; automation; space progress; war prevention; weapon systems. Individual respondents were asked to suggest future possible developments, and then the group was to estimate the year by which there would be a 50 percent chance of the development occurring. Many of the techniques utilized in that Delphi are still common to the pure forecasting Delphis being done today. That study, together with an excellent related philosophical paper providing a Lockean-type justification for the Delphi technique,[5] formed the foundation in the early and mid-sixties for a number of individuals to begin experimentation with Delphi in nondefense areas.

At the same time that Delphi was beginning to appear in the open literature, further interest was generated in the defense area: aerospace corporations and the armed services. The rapid pace of aerospace and electronics technologies and the large expenditures devoted to research and development leading to new systems in these areas placed a great burden on industry and defense planners. Forecasts were vital to the preparation of plans as well as the allocation of R&D (research and development) resources, and trend extrapolations were clearly inadequate. As a result, the Delphi technique has become a fundamental tool for those in the area of technological forecasting and is used today in many technologically oriented corporations. Even in the area of "classical" management science and operations research there is a growing recognition of the need to incorporate subjective information (e.g., risk analysis) directly into evaluation models dealing with the more complex problems facing society: environment, health, transportation, etc. Because of this, Delphi is now finding application in these fields as well.

From America, Delphi has spread in the past nine years to Western Europe, Eastern Europe, and the Far East. With characteristic vigor the largest Delphi undertaken to date is a Japanese study. Starting in a nonprofit organization, Delphi has found its way into government, industry, and finally academe. This explosive rate of growth in utilization in recent years seems, on the surface, incompatible with the limited amount of controlled experimentation or academic investigation that has taken place. It is, however, responding to a demand for improved communications among larger and/or geographically dispersed groups which cannot be satisfied by other available techniques. As illustrated by the articles in this book, aside from some of the Rand studies by Dalkey, most "evaluations" of the technique have been secondary efforts associated with some application which was the primary interest. It is hoped that in coming years experimental psychologists and others in related academic areas will take a more active interest in exploring the Delphi technique.

While many of the early articles on Delphi are quite significant and liberally mentioned in references throughout this book, we have chosen to concentrate

[5]O. Helmer and N. Rescher, "On the Epistemology of the Inexact Sciences," Project Rand Report R-353, February 1960.

on work that has taken place in the past five years and which represents a cross section of diverse applications.

Although the majority of the Delphi efforts are still in the pure forecasting area, that application provides only a small part of the contents of this volume.

Chapters II and III of this book consist of articles which provide an overview of the Delphi technique, its utility, the underlying philosophy, and broad classes of application.

Chapter IV takes up recent studies in the evaluation of the technique. Precision and accuracy are considered in this context. Between the reviews, articles, and associated references, the reader should obtain a good perspective on the state of the art with respect to experimentation.

Chapters V and VI describe some of the specialized techniques that have evolved for asking questions and evaluating responses. Foremost among them is cross-impact analysis (Chapter V). This concept reflects recognition of the complexity of the systems dealt with in most Delphi activities, systems where "everything interacts with everything." In essence, these sections explore the quantitative techniques available for deeper analysis of the subjective judgments gathered through the Delphi mechanism.

The effect computers can have on Delphi and speculations on the future of the technique itself are discussed in Chapter VII. The book concludes with a summary of pitfalls which can serve the practitioner as a continuing checklist (Chapter VIII).

We have striven to avoid making this volume a palimpsest of previously published papers: all but four of the articles have been especially prepared for this work. The four reprinted articles were selected from the journal *Technological Forecasting and Social Change*, a rich lode of material on Delphi. The extensive bibliography in the Appendix provides a guide to those who wish to probe the subject further. It is thus our hope that this volume will serve the reader as a useful reference work on Delphi for a number of years.

II. Philosophy

II.A. Introduction

HAROLD A. LINSTONE and MURRAY TUROFF

Any human endeavor which seeks recognition as a professional or scientific activity must clearly define the axioms upon which it rests. The foundation of a discipline, as the foundations of a house, serves as a guide and basis for the placement of the building blocks of knowledge gathered through research and development activities. It is the definition, exposure, and investigation of the philosophical foundation that distinguishes a scientific profession from other endeavors.

In a well-established scientific endeavor, the foundation is made explicit so that one is able to recognize when the resulting structure can no longer be properly supported and a reexamination of the fundamentals is in order. A classic example of this was the impact of quantum mechanics on the foundations of physics. With respect to new disciplines, such as the investigation of Delphi methodology, the situation is one where not enough of the structure has been blueprinted to discriminate which of many possible foundations supply the "best" underpinnings.

The early attempt by Helmer and Rescher in their classic paper "On the Epistemology of the Inexact Sciences" proposed one foundation, largely of a Lockean nature, which was very adequate for the typical technological forecasting applications for which Delphi has been popular. However, in recent years extensions to Delphi methodology have demonstrated a need for a broader basis. Certainly the theme of this book, which largely views Delphi as the process of structuring human communications, further enhances this position.

The first article by Mitroff and Turoff, examines what the various classic or "pure mode" epistemologies of Western philosophy have to offer for insight into the Delphi process. The philosophies covered are those represented by Locke, Leibniz, Kant, Hegel, and Singer. It largely follows the morphological structure of philosophical inquiry first proposed by C. West Churchman in his "Design of Inquiring Systems." As with any young discipline, it should not come as a surprise that such a rich diversity of foundation axioms may be used to give form and shape to Delphi. In a sense this is an expression of the yet untapped potential for future development of the technique.

The second article, by Scheele, illustrates how a user of Delphi may compose for his own view and application of Delphi a very specific philosophical foundation. The author, being primarily concerned in many of his applications with the perceptions of individuals as they may relate to marketing problems, adapts elements of the Lockean, Kantian, and Singerian philosophies and merges them with the existentialist concept of subjective or negotiated reality. The result is a foundation for a design precisely matched to the user's unique needs.

Throughout the book one will find in the various articles explicit or implicit support for a mode or manner of applying Delphi which rests on the philosophies brought out in these two papers. It is interesting to note that a recent sociological perspective views Delphi as a ritual.[1] Primitive man always approached the future ritualistically, with ceremonies involving utensils, liturgies, managers, and participants. The Buckminster Fuller World Game, Barbara Hubbard's SYNCON, as well as Delphi, can be considered as modern participatory rituals. The committee-free environment and anonymity of Delphi stimulate reflection and imagination, facilitating a personal futures orientation. Thus, the modern Delphi is indeed related to its famous Greek namesake.

[1] A. Wilson and D. Wilson, "The Four Faces of the Future," New York, Grove Press, 1974.

II. B. Philosophical and Methodological Foundations of Delphi

IAN I. MITROFF and MURRAY TUROFF

It takes two of us to create a truth, one to
utter it and one to understand it.
Kahlil Gibran

Introduction

The purpose of this article is to show that underlying *any* scientific technique, theory, or hypothesis there is always *some* philosophical basis or theory about the nature of the world upon which that technique, theory, or hypothesis fundamentally rests or depends. We also wish to show that there is more than one fundamental basis which can underlie any technique, or in other words, that there is no one "best" or even "unique" philosophical basis which underlies any scientific procedure or theory. Depending upon the basis which is presumed, there results a radically different developmental and application history of a technique. Thus in this sense, the particular basis upon which a scienfific procedure depends is of fundamental practical importance and not just of philosophical interest.

We human beings seem to have a basic talent for disguising through phraseology the fundamental similarities that exist between common methodologies of a different name. As a result, we often bicker and quarrel about such superficial matters as whether this or that name is appropriate for a certain technique when the real issue is whether the philosophical basis or system of inquiry that underlies a proposed technique or methodology is sound and appropriate. We are indeed the prisoners of our basic images of reality. Not only are we generally unaware of the different philosophical images that underlie our various technical models, but each of us has a fundamental image of reality that runs so deep that often we are the last to know that we hold it. As a result we disagree with our fellow man and we experience inner conflict without really knowing why. What's worse, we ensure this ignorance and conflict by hiding behind catchwords and fancy names for techniques. The field of endeavor subsumed under the name of Delphi is no less remiss in this respect than many other disciplines. Its characteristic vocabulary more often obscures the issues than illuminates them.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

One of the basic purposes of our discussion is to bring these fundamental differences and conflicts of methodology up to the surface for conscious examination so that, one hopes, we can be in a better position to choose explicitly the approach we wish to adopt. In order to accomplish this we consider a number of fundamental historical stances that men have taken toward the problem of establishing the "truth content" of a system of communication signals or acts. More precisely, the purpose of this article is to examine the variety of ways and mechanisms in which men have chosen to locate the criteria which would supposedly "guarantee" our "true and accurate understanding" of the "content" of a communication act or acts. We will also show that every one of these fundamental ways differs sharply from the others and that each of them has major strengths as well as major weaknesses. The moral of this discussion will be that there is *no one "single best way"* for ensuring our understanding of the content of a set of communication acts or for ascribing validity to a communication. The reason is that there is no one mode of ensuring understanding or for prescribing the validity of a communication that possesses all of the desired characteristics that one would like any preferred mode to possess. As we wish to illustrate, this awareness itself constitutes a kind of strength. To show that there is no one mode that can satisfy our every requirement, i.e., that there is no one mode that is *best* in all senses and for all circumstances, is not to say that each of these modes does not appear to be "better suited" for some special set of circumstances.

Since these various modes or characteristic models for ensuring validity basically derive from the history of Western philosophy, another objective of this article is also to show what philosophy and, especially, what the philosophy of science specifically and concretely has to offer the field of Delphi design. For example, one of the things we wish to show is which among these various philosophical modes have been utilized to date (and how) and which have been neglected. When there has been little or no utilization of a particular philosophical basis then we may infer existing gaps in the development of the Delphi to date.

Before we describe each of thesse philosophical modes or systems more fully, we can rather easily and simply convey the general spirit of each of them by means of the following exercise. Suppose we are given a set of statements or propositions by some individual or group which pretend to describe some alleged "truth." Then each of our philosophical systems (hereafter referred to as an Inquiring System, or IS) can be simply differentiated from one another in terms of the kind of characteristic question(s) that each would address either to the statement itself or to the individual (group) making the statement or assertion. Each question in effect embodies the major philosophical criterion that would have to be met before that Inquiring System would accept the propositions as valid or as true.

The Leibnizian analyst or IS would ask something like:

> How can one independently of any empirical or personal considerations give a *purely rational* justification of the proposed proposition or assertion? Can one build or demonstrate a rational model which underlies the proposition or assertion? How was the result deduced; is it precise, certain?

The Lockean analyst or IS would ask something like:

> Since for me data are always prior to the development of formal theory, how can one independently of any formal model justify the assertion by means of some objective data or the consensus of some group of expert judges that bears on the subject matter of the assertions? What are the supporting "statistics"? What is the "probability" that one is right? Are the assertions a good "estimate" of the true empirical state of affairs?

The Kantian analyst or IS would ask something like:

> Since data and theory (models) *always* exist side by side, does there exist *some combination* of *data* or expert judgment *plus* underlying *theoretical justification* for the data that would justify the propositions? What *alternative* sets of propositions exist and which best satisfy my objectives and offer the strongest combination of data plus model?

The Hegelian (Dialectical) analyst or IS would ask something like:

> Since every set of propositions is a reflection of a more general theory or *plan* about the nature of the world as a *whole system*, i.e., *a world-view*, does there exist some alternate sharply differing world-view that would permit the serious consideration of a completely opposite set of propositions? Why is this opposing view not true or more desirable? Further, does this conflict between the plan and the counterplan allow a third plan or world-view to emerge that is a *creative synthesis* of the original plan and counterplan?

Finally, the Singerian analyst or IS would ask:

> Have we taken a broad enough perspective of the basic problem? Have we from the very beginning asked the right question? Have we focused on the right objectives? To what extent are the questions and models of each inquirer a reflection of the unique *personality* of each inquirer as much as they are felt to be a "natural" characteristic or property of the "real" world?"

Even at this point in the discussion, it should be apparent that as a body these are very different kinds of questions and that each of them is indicative of a fundamentally different way of ascribing content to a communication. It should also be apparent, and it should really go without saying, that these do not exhaust the universe of potential questions. There are many more philosophical positions and approaches to "validity" than we could possibly hope to deal with in this article. These positions do represent, however, some of the

most significant basic approaches and, in a sense, pure-modes from which others can be constructed.

The plan of the rest of this article is briefly as follows: first, we shall describe each inquirer in turn and in general terms, but we hope in enough detail to give the reader more of a feel for each system; second, along with the description of each inquirer, we shall attempt to point out the influence or lack of influence each philosophy of inquiry has had on the Delphi technique; and third, we shall attempt to point out some general conclusions regarding the nature and future of the Delphi technique as a result of this analysis.

It should be borne in mind as we proceed that the question of concern is not how we can determine or agree on the meaning of "truth" with "perfect or complete certainty," for put in this form, the answer is clearly that we cannot know anything with "perfect certainty." We cannot even know with "perfect certainty" that "we cannot know anything with 'perfect certainty.'" The real question is *what* can we know and, even more to the point, how we can *justify* what we think we can know. It is on this very issue that the difference between the various Inquiring Systems arises and the utility or value of the Delphi technique depends.

Inquiring Systems (IS)

The process of inquiry, whether it be for a single individual or a group of individuals, may be "represented" by a very general system. We start with some *assumed* "external event" or "raw data set" which for the moment we *consider* to be a characteristic property of the "real world," i.e., we *assume* the data set "exists" in the "external world." (As we shall see in a moment, this amounts to assuming a Lockean IS beginning. The point is that we can't even begin to describe the "world" and our "knowledge" of "it" without having to invoke some "conceptualization," i.e., some Inquiring System characterization, of "it.") Next we apply some transformation and/or filter to the "raw data" in order to get it into the "right form" for input to some model. The model, which may be *any sort* of structured process, is represented by a set of rules which may be either in the form of an algorithm or a set of heuristic principles. The model acts on the input to transform it from the state of "input data" to the state of "output information." This output information may in turn be passed through another filter or transform to put it in the right form so that a decisionmaker can recognize and utilize it as "information" or as a "policy recommendation." In terms of this general configuration, the various IS can be differentiated from one another with respect to (1) the priority assigned to the various systems components, i.e., which components are regarded as more important or fundamental by one IS than by another, and (2) the degree of interdependence assigned to the various systems components by each IS.

Our objective in the following discussion will be to draw a sufficient distinction between the philosophical Inquirying System (IS) concepts so that we can place alternative Delphi design methodologies into this perspective.

Lockean IS

As first pioneered by Dalkey, Helmer, and Rescher at Rand, the Delphi technique represents a prime example of Lockean inquiry. Indeed, one would be hard pressed to find a better contemporary example of a Lockean inquirer than the Delphi.

The philosophical mood underlying the major part of *empirical science* is that of Locke. The sense of Lockean IS can be rather quickly and generally grasped in terms of the following characteristics:

(1) Truth is *experiential,* i.e., the truth content of a system (or communication) is associated *entirely* with its empirical content. A model of a system is an *empirical model* and the truth of the model is measured in terms of *our* ability (a) to reduce every complex proposition down to its simple empirical referents (i.e., simple observations) and (b) to ensure the validity of each of the simple referents by means of the widespread, freely obtained *agreement* between different human observers.

(2) A corollary to (1) is that the truth of the model does not rest upon *any* theoretical considerations, i.e., upon the prior assumption of any theory (this is the equivalent of Locke's *tabula rasa*). Lockean inquirers are opposed to the prior presumption of theory, since in their view this exactly reverses the justifiable order of things. Data are that which are prior to and justify theory, not the other way around. The only general propositions which are accepted are those which can be justified through "direct observation" or have already been so justified previously. In sum, *the data input sector is not only prior to the formal model or theory sector but it is separate from it as well. The whole of the Lockean IS is built up from the data input sector.*

In brief, Lockean IS are the epitome of *experimental, consensual* systems. On any problem, they will build an empirical, inductive representation of it. They start from a set of elementary empirical judgments ("raw data," observations, sensations) and from these build up a network of ever expanding, increasingly more general networks of factual propositions. Where in the Leibnizian IS to be discussed shortly the networks are theoretically, deductively derived, in a Lockean IS they are empirically, inductively derived. The guarantor of such systems has traditionally been the function of human agreement, i.e., an empirical generalization (or communication) is judged "objective," "true," or "factual" if there is "sufficient widespread agreement" on it by a group of "experts." The final information content of a Lockean IS is identified almost exclusively with its empirical content.

A prime methodological example of Lockean thinking can be found in the field of statistics. Although statistics is heavily Leibnizian in the sense that it devotes a considerable proportion of its energies to the formal treatment of data and to the the development of formal statistical models, there is a strong if not almost pure Lockean component as well. The pure Lockean component manifests itself in the attitude that although statistical methods may "transform" the "basic raw data" and "represent" "it" differently, statistical methods themselves are presumed not to create the "basic raw data." In this sense, the "raw data" are presumed to be *prior to* and *independent of* the formal (theoretical) statistical treatment of the data. The "raw data" are granted a prior existential status. Another way to put this is to say that there is little or no match between the theory that the observer of the raw data has actually used (and has had to use) in order to collect his "raw data" in the first place and the theory (statistics) he has used to analyze it in the second place A typical Lockean point of view is the assertion that one doesn't need any theory in order to collect data first, only to analyze it subsequently.

As mentioned at the beginning of this section, the Delphi, at least as it was originally developed, is a classic example of a Lockean inquirer. Furthermore, *the Lockean basis of Delphi still remains the prime philosophical basis of the technique to date.*

As defined earlier Delphi is a procedure for structuring a communication process among a large group of individuals. In assessing the potential development of, say, a technical area, a large group (typically in the tens or hundreds) are asked to "vote" on when they think certain events will occur. One of the major premises underlying the whole approach is the assumption that a large number of "expert" judgments is required in order to "treat adequately" any issue. As a result, a face-to-face exchange among the group members would be inefficient or impossible because of the cost and time that would be involved in bringing all the parties together. The procedure is about as pure and perfect a Lockean procedure as one could ever hope to find because, first, the "raw data inputs" are the opinions or judgments of the experts; second, *the validity of the resulting judgment of the entire group is typically measured in terms of the explicit "degree of consensus" among the experts.* What distinguishes the Delphi from an ordinary polling procedure is the feedback of the information gathered from the group and the opportunity of the individuals to modify or refine their judgments based upon their reaction to the collective views of the group. Secondary characteristics are various degrees of anonymity imposed on the individual and collective responses to avoid undesirable psychological effects.

The problems associated with Delphi illustrate the problems associated with Lockean inquiry in general. The judgments that typically survive a Delphi procedure may not be the "best" judgments but, rather, the compromise position. As a result, the surviving judgments may lack the significance that extreme or conflicting positions may possess.

The strength of Lockean IS lies in their ability to sweep in rich sources of experiential data. In general, the sources are so rich that they literally overwhelm the current analytical capabilities of most Leibnizian (analytical) systems. The weaknesses, on the other hand, are those that beset all empirical systems. While experience is undoubtedly rich, it can also be extremely fallible and misleading. Further, the "raw data," "facts," or "simple observables" of the empiricist have *always* on deeper scientific and philosophical analysis proved to be exceedingly complex and hence further divisible into other entities themselves thought to be indivisible or simple, ad infinitum. More troublesome still is the almost extreme and unreflective reliance on agreement as the sole or major principle for producing information and even truth out of raw data. The trouble with agreement is that its costs can become too prohibitive and agreement itself can become too imposing. It is not that agreement has nothing to recommend it. It is just that agreement is merely one of the many philosophical ways for producing "truth" out of experiential data. The danger with agreement is that it may stifle conflict and debate when they are needed most. As a result, Lockean IS are best suited for working on well-structured problem situations for which there exists a *strong consensual* position on "the nature of the problem situation." If these conditions or assumptions cannot be met or justified by the decisionmaker—for example, if it seems too risky to base projections of what, say, the future will be like on the judgments of experts—then no matter how strong the agreement between them is, some alternate system of inquiry may be called for.

While the consensus-oriented Delphi may be appropriate to technological forecasting it may be somewhat inappropriate for such things as technology assessment, objective or policy formulation, strategic planning, and resource allocation analyses. These latter applications of Delphi often or should involve the necessity to explore or generate *alternatives,* which is very different from the generation of consensus.

Leibnizian IS

The philosophical mood underlying the major part of theoretical science is that of Leibniz. The sense of Leibnizian inquiry can be rather quickly and generally captured in terms of the following characteristics:

(1) Truth is *analytic*; i.e., the truth content of a system is associated *entirely* with its formal content. A model of a system is a formal model and the truth of the model is measured in terms of *its* ability to offer a theoretical explanation of a wide range of general phenomena and in terms of *our* ability as model-builders to state clearly the formal conditions under which the model holds.

(2) A corollary to (1) is that the truth of the model does not rest upon *any external considerations,* i.e., upon the raw data of the external world. Leibnizian inquirers regard empirical data as an inherently risky base upon which to

found universal conclusions of any kind, since from a finite data set one is never justified in inferring any general proposition. The only general propositions which are accepted are those that can be justified through purely rational models and/or arguments. Through a series of similar arguments, *Leibnizian IS not only regard the formal model component as separate from the data input component but prior to it as well.* Another way to put this is to say that *the whole of the Leibnizian IS is contained in the formal sector and thus it has priority over all the other components.*

In short, Leibnizian IS are the epitome of formal, symbolic systems. For *any* problem, they will characteristically strive to reduce it to a formal mathematical or symbolic representation. They start from a set of elementary, primitive "formal truths" and from these build up a network of ever expanding, increasingly more general, formal propositional truths. The *guarantor* of such systems has traditionally been the precise specification of what shall count as a proof for a derived theorem or proposition; other guarantor notions are those of internal consistency, completeness, comprehensiveness, etc. The final information content of Leibnizian IS is identified almost exclusively with its symbolic content.

A prime example of Leibnizian inquiry is the field of Operations Research (OR) in the sense that the major energies of the profession have been almost exclusively directed toward the construction and exploration of highly sophisticated formal models. OR is a prime example of Leibnizian inquiry not because there is no utilization of external data whatsoever in OR models but because in the training of Operations Researchers significantly more attention is paid to teaching students how to build sophisticated models than in teaching them equally sophisticated methods of data collection and analyses. There is the implication that the two activities are separable, i.e., that data can be collected independently of formal methods of analysis.

Delphi by itself is not a Leibnizian inquirer and is better viewed from the perspective of some of the alternative Inquiring Systems. However, many of the views and assertions made with respect to the Delphi technique involve Leibnizian arguments. Delphi has, for example, been accused of being very "unscientific." When assertions of this type are examined one usually finds the underlying proposition rests on equating what is "scientific" to what is "Leibnizian." This is a common misconception that has also affected other endeavors in the social, or so-called soft, sciences where it is felt that the development of a discipline into a science must follow some preordained path leading to the situation where all the results of the discipline can be expressed in Leibnizian "laws." We have today in such areas as economics, sociology, etc., schools of research dedicated to the construction of formal models as ends in themselves.

In Delphi we find a similar phenomenon taking place where models are constructed for the purpose of describing the Delphi process and for determining *the* "truth" content of a given Delphi. (See, for example, the articles in Chapter IV.) One model hypothesizes that the truth content of a Delphi result (often measured as the error) increases as the size of the Delphi group increases.

This concept is often used to guide the choice of the size of the participant group in a Delphi. Other formal models have been proposed to measure an individual's "expertise" as a function of the quantity of information supplied and the length of associated questions. All such models, which are *independent of the content* of what is being communicated but look for structured relationships in the process of the communication, are attempts to ascribe Leibnizian properties to the Delphi process. The existence of such models in certain circumstances do not in themselves make the Delphi technique any more or less "scientific." They are certainly useful in furthering our understanding of the technique and should be encouraged. However, they are based upon assumptions, such as the superiority of theory over data and the general applicability of formal methods of reasoning, which are quite suspect with respect to the scope of application of the Delphi technique and the relative experimental bases upon which most of these models currently rest. The utility of Delphi, at least in the near future, does not appear to rest upon making Delphi appear or be more Leibnizian but, rather, in the recognition of what all the IS models can contribute to the development of the Delphi methodology. Our current understanding of human thought and decision processes is probably still too rudimentary to expect generally valid formal models of the Delphi process at this time.

For which kinds of problem situations are Leibnizian analyses most appropriate? First of all, the situations must be sufficiently "well understood" and "simple enough" so that they can be modeled. Thus, Leibnizian IS are best suited for working on clearly definable (i.e., well-structured) problems for which there exists an analytic formulation as well as solution. Second, the modeler must have strong reasons for believing in the assumptions which underlie Leibnizian inquiry, e.g., that the model is universally and continually applicable. In a basic sense, the fundamental guarantor of Leibnizian inquiry is the "understanding" of the model-builder; i.e., he must have enough faith in his understanding of the situation to believe he has "accurately" and "faithfully" represented it.

Note that there is no sure way to prove or justify the assumptions underlying Leibnizian inquiry. The same is true of all the other IS. But then this is not the point. The point is to show the kinds of assumptions we are required to make if we wish to employ Leibnizian inquiry so that if the decisionmaker or modeler is unwilling to live with these assumptions he will know that another IS may possibly be called for.

Kantian IS

The preceding two sections illustrate the difficulties that arise from emphasizing one of the components of a tightly coupled system of inquiry to the detriment of other components. Leibnizian inquiry emphasizes theory to the detriment of data. Lockean inquiry emphasizes data to the detriment of theory. When these attitudes are translated into professional practice, what often

sults is the development of highly sophisticated models with little or no ncern for the difficult problems associated with the collection of data or the emingly endless proliferation of data with little regard for the dictates of rrently existing models.

The recent controversy surrounding the attempts of Forrester and Meadows[1] build a "World Model" is a good illustration of the strong differences tween these two points of view. In our opinion, the work of Forrester and leadows represents an almost pure Leibnizian approach to the modeling of rge, complicated systems. The Forrester and Meadows model is, in effect, ta independent. One can criticize the model on pure Leibnizian grounds, g., whether the internal theory and structure of the model are sound with spect to current economic and social theory, and some of the critics have osen to do this. However, it would seem to us that more often than not the itics have chosen to offer a Lockean critique, i.e., that some other way, say ing accurate statistical data, is a better way to build a sound forecast model the world. While this is a legitimate method of criticism, to a large extent it nly further exacerbates the differences between the two approaches and hence isses the real point. To us the real point is not whether the Forrester-leadows approach is *the* correct Leibnizian approach, or whether there is *a* rrect Lockean approach, but rather, whether *any* Leibnizian or Lockean proach acting independently of the other could ever possibly be "correct." rrester and Meadows seek to justify (guarantee?) their approach through the obustness and richness of their *model*, and their Lockean critics attempt to tablish the validity of their approach through the priority and "regularity" of statistical *data* to which they appeal. Perhaps if the debate proves anything, aises the serious question as to whether an advanced modern society can ntinue to rely on purely Leibnizian or Lockean efforts for its planning. In rder to evaluate the relative merits of separate Leibnizian or Lockean in-uirers, it is necessary to go to a competing philosophy which incorporates th, such as the Kantian inquirer.

The sense of Kantian inquiry can be rather quickly grasped through the llowing set of general characteristics:

(1) Truth is *synthetic;* i.e., the truth content of a system is not located in either s theoretical or its empirical components, but in *both.* A model of a system is a nthetic model in the sense that the truth of the model is measured in terms of model's ability (a) to associate every theoretical term of the model with ne empirical referent and (b) to show that (how) underlying the collection of ery empirical observation related to the phenomenon under investigation re is an associated theoretical referent.

(2) A corollary to (1) is that neither the data input sector nor the theory ctor have priority over one another. Theories or general propositions are built

[1]Meadows, Dennis "Limits to Growth" 1972 Universe Books.

up from data, and in this sense theories are dependent on data, but data cannot be collected without the prior presumption of some theory of data collection (i.e., a theory of "how to make observations," "what to observe," etc.), and in this sense data are dependent on theories. *Theory and data are inseparable.* In other words, Kantian IS require some coordinated image or plan of the system as a whole before any sector of the system can be worked on or function properly.

These hardly begin to exhaust all the features we identify with Kantian inquiry. A more complete description would read as follows: Kantian IS are the epitome of multimodel, synthetic systems. On any problem, they will build at least two alternate representations or models of it. (If the alternate representations are complementary, we have a Kantian IS; if they are antithetical, we have a Hegelian IS, as described in the next section.) The representations are partly Leibnizian and partly Lockean; i.e., Kantian IS make explicit the strong interaction between scientific theory and data. They show that in order to collect some scientific data on a problem a posteriori one *always* has had to presuppose the existence of some scientific theory a priori, no matter how implicit and informal that theory may be. Kantian IS presuppose at least two alternate scientific theories (this is their Leibnizian component) on any problem or phenomenon. From these alternate Leibnizian bases, they then build up at least two alternate Lockean fact nets. The hope is that out of these alternate fact nets, or representations of a decisionmaker's or client's problem, there will be one that is "best" for representing his problem. The defect of Leibnizian and Lockean IS is that they tend to give only one explicit view of a problem situation. Kantian IS attempt to give *many* explicit views. The guarantor of such systems is the degree of fit or match between the underlying theory (theoretical predictions) and the data collected under the presumption of that theory plus the "deeper insight" and "greater confidence" a decisionmaker feels he has as a result of witnessing many different views of his problem.

The reason Kantian IS place such a heavy emphasis on alternate models is that in dealing with problems like planning for the future, the real concern is how to get as many perspectives on the nature of the problem as possible. Problems which involve the future cannot be formulated and solved in the same way that one solves problems in arithmetic, i.e., via a single, well-structured approach. There seems to be something fundamentally different about the class to which planning problems belong. In dealing with the future, we are not dealing with the concrete realities of human existence, but, if only in part, with the hopes, the dreams, the plans, and the aspirations of men. Since different men rarely share the same aspirations, it seems that the best way to "analyze" aspirations is to compare as many of them against one another as we can. If the future is 99 percent aspiration or plan, it would seem that the best way to get a handle on the future is *to draw forth explicitly as many different aspirations or plans for the future as possible.* In short, we want to examine as many different alternate futures as we can.

In the field of planning, Normative Forecasting, Planning Programming Budgeting Systems (PPBS), and Cost-Effectiveness or Cost-Benefit Analyses are all examples of Kantian inquiry, although these are such low-level Kantian inquirers as to be almost more Leibnizian in nature than Kantian. The Kantian element that these various approaches share is the fact that they are all concerned with *alternate paths*, or *methods*, of getting from a present state to a future state characterized by certain objectives, needs, or goals. When these various planning vehicles have failed, it is not just because we are dealing with an inherently fuzzy problem—indeed that is the basic nature of the problem—but because we have failed to produce alternatives that are true alternatives and to show that the data, models, and objectives cannot be separated for purposes of planning.

In recent years, there have been a number of Delphi studies which in contrast to the original Lockean-based consensus Delphis begin "to take on" more actively the characteristics of Kantian inquiry. The initial Delphis were characterized by a strong emphasis on the use of consensus by a group of "experts" as *the* means to converge on a single model or position on some issue. In contrast, the explicit purpose of a Kantian Delphi is to elicit alternatives so that a comprehensive overview of the issue can take place. In terms of communication processes, while a "consensus," or Lockean, Delphi is better suited to setting up a communication structure among an already informed group that possesses the same general core of knowledge, a Kantian, or "contributory," Delphi attempts to design a structure which allows many "informed" individuals in different disciplines or specialties to contribute information or judgments to a problem area which is much broader in scope than the knowledge that any one of the individuals possesses. This type of Delphi has been applied to the conceptualization of such problems as: (1) the definition of a structural model for material flows in the steel industry (see Chapter III, C, 3); (2) the examination of the current and the potential role of the mentally retarded in society (see Chapter VI, D); (3) the forecasting of the future characteristics of recreation and leisure (see Chapter VI, D); and (4) the examination of the past history of the internal combustion engine[1] for a clue to significant events possibly affecting its future. While all of these Delphis had specific forecasting objectives, none of them could be achieved if all the parties to the Delphi were drawn from the same specialized interest group. The problems were broader than that which could be encompassed by any single discipline or mode of thinking. For example, the examination of the role of the mentally retarded in our society is neither the exclusive problem nor the sole province of any special group. Educators, psychiatrists, parents, and teachers all have different and valid perspectives to contribute to the definition of the "problem." Consensus on a single definition is not the goal, at least not in the

[1]proprietary Delphi in 1969 by Kenneth Craver of Monsanto Company.

initial stages, but rather, the eliciting of many diverse points of view and potential aspects of the problem. In essence, the objective is establishing how to fit the pieces of a jigsaw puzzle together, and even determining if it is one or many puzzles.

The problem of conceptualizing goals and objectives is not an explicit part of the three inquiry processes we have discussed so far. That is, the Leibnizian and Lockean IS are not explicitly goal directed. For example, Leibnizian IS assume that the same rational model is applicable no matter what the problem and the objectives of the decisionmaker or who it is that has the problem. In contrast, the Kantian IS is explicitly goal oriented, i.e., it hopes by presenting a decisionmaker with several alternative models of his problem to better clarify both the problem and the nature of the objectives, which after all are part of the "problem."

Kantian inquiry is best suited to problems which are inherently ill-structured, i.e., the kinds of problems which are inherently difficult to formulate in pure Leibnizian or Lockean terms because the nature of the problem does not admit of a clear consensus or a simple analytic attack. On the other hand, the Kantian inquiry is not especially suited for the kinds of problems which admit of a single clear-cut formulation because here the proliferation of alternate models may not only be costly but time consuming as well. Kantian inquiry may also overwhelm those who are used to "the single best model" approach to any problem. Of course this in itself is not necessarily bad if it helps to teach those who hold this belief that there are some kinds of problems for which there is no one best approach. Social problems inherently seem to be of this kind and thus to call for Kantian approach. The concept of "technology assessment" as a vehicle for determining the relationships between technology and social consequences would also seem to imply the necessity of at least a Kantian approach. Many efforts which fall under the heading of "assessments" have proved to be inadequate because they were conducted on pure Leibnizian or Lockean bases.

Hegelian, or Dialectical, IS

The idea of the Hegelian, or Dialectical, IS can be conveyed as follows:

(1) Truth is *conflictual*; i.e., the truth content of a system is the result of a highly complicated process which depends on the existence of a *plan* and a *diametrically opposed counterplan*. The plan and the counterplan represent strongly divergent and opposing conceptions of the whole system. The function of the plan and the counterplan is to engage each other in an unremitting debate over the "true" nature of the whole system, in order to draw forth a new plan that will, one hopes, reconcile (synthesize, encompass) the plan and the counterplan.

(2) A corollary to (1) is that by itself the data input sector is totally

meaningless and only becomes meaningful, i.e., "information," by being coupled to the plan and the counterplan. Further, it is postulated that there is a particular input data set which can be shown to be consistent with both the plan and the counterplan; i.e., by itself this data set supports neither naturally, but there is an interpretation of the data such that it is consistent with both the plan and the counterplan. It is also postulated that without both the plan and the counterplan the meaning of the data is incomplete, i.e., partial. Thus, under this system of inquiry, the plan and the counterplan which constitute the theory sector are prior to the input sector and indeed constitute opposing conceptions of the whole system. Finally, it is also assumed that *on* every *issue of importance, there can be found or constructed a plan and a counterplan; i.e., a dialectical debate can be formulated with respect to* any *issue*. On any issue of importance there will be an intense division of opinion or feeling.

Hegelian, or Dialectical, IS are the epitome of *conflictual*, synthetic systems. On any problem, they build *at least two, completely antithetical,* representations of it. Hegelian IS start with either the prior existence (identification) of or the creation of two strongly opposing (*contrary*) Leibnizian models of a problem. These opposing representations constitute the contrary underlying assumptions regarding the theoretical nature of the problem. Both of these Leibnizian representations are then applied to the *same* Lockean data set in order to demonstrate the crucial nature of the underlying theoretical assumptions, i.e., that the same data set can be used to support either theoretical model. The point is that data are not information; *information is that which results from the interpretation of data*. It is intended that out of a dialectical confrontation between opposing interpretations (e.g., the opposing "expert" views of a situation), the underlying assumptions of both Leibnizian models (or opposing policy experts) will be brought up to the surface for conscious examination by the decision-maker who is dependent upon his experts for advice. It is also hoped that as a result of witnessing the dialectical confrontation between experts or models, the decisionmaker will be in a better position to form his own view (i.e., build his own model or become his own expert) on the problem that is a "creative synthesis" of the two opposing views.

In considering the resource allocation and decision processes which govern our society and institutions, the role of the "expert" has become somewhat confused and clouded. In a historical perspective the emergence of systems analysts, efficiency or productivity experts, and operation researchers can be viewed as the establishment of a new group of advocates. They advocate decisions, policies, and actions which may optimize certain unique measures such as benefits, costs, efficiency, etc. However, their training does not enable them to reflect on all the factors which the decisionmaker must account for in the process of reaching a decision. Perhaps part of the problem we have had in the past is a misconception that the "expert" has the only view pertinent to the

decision and our error in our not attempting to balance and place in perspective the views arising from political, sociological, psychological and ethical considerations which may advocate alternative options. Perhaps "experts" can be better used by the decision processes if they are viewed from the perspective of the Hegelian inquirer as just *one component* of the decision analysis process. This view of the use of expertise underlies concepts such as the Policy Delphi.

Whereas, in the Lockean IS the guarantor of the validity of a proposition is agreement, in the Hegelian it is intense conflict, i.e., the presumption that conflict will expose the assumptions underlying an expert's point of view that are often obscured *precisely because of* the agreement between experts. Hegelian IS are best suited for studying "wickedly" ill-structured problems. These are the problems that, precisely because of their ill-structured nature, will produce intense debate over the "true" nature of the problem. Conversely, Hegelian IS are extremely unsuited to well-structured, clear-cut problems because here conflict may be a time-consuming nuisance.

Except for the Policy Delphi concept of Turoff (see Chapter III, B, 1,3), the use of conflict as a methodology is conspicuously absent in the field of technological forecasting or in Delphi studies in general. In the Policy Delphi the communication process is designed to produce the best underlying pro or con arguments associated with various policy or resource allocation alternatives. In a non-Delphi mode of communication (e.g., face to face), one of the most interesting applications can be found in the activity of corporate or strategic planning. In an important case study, Mason[2] literally pioneered the development of what may be termed the Dialectical Policy inquirer. The situation encountered by Mason was one in which the nature of the problem prevented traditional well-structured technical approaches to planning (i.e., Leibnizian and Lockean methods) from being used.

In the company situation studied by Mason, there were two strongly opposing groups of top executives who had almost completely contrary views about the fundamental nature and management of their organization. They were faced with a crucial decision concerning the future of their company. It was literally a life-and-death situation, since the decision would have strong repercussions throughout all of their company's activities. The two groups each offered fundamentally differing plans as to how to cope with the situation. Neither of the plans could be proved or "checked out" by performing any technical study, since each plan rested on a host of assumptions, many of them unstated, that could probably never be verified in their entirety even if time to do this were available, which it wasn't. Indeed, if the executives wanted to be around in the future to check on how well their assumptions turned out, they

[2]Richard Mason "A Dialectical Approach to Strategic Planning," *Management Science* 15, No. 8 (April 1969).

had to make a decision in the present. It was at this point that the company agreed to let Mason try the Dialectical Policy inquirer to see if it could help resolve the impasse and suggest a way out.

After careful study and extensive interviews with both sides, Mason assembled both groups of executives and made the following presentation to them: First, he laid out side by side on opposite halves of a display board what he took to be the underlying assumptions on which the two groups were divided. Thus, for every assumption on the one side there was an opposing assumption for the other side. It is important to appreciate that this had never been done before. Prior to Mason's contact, both groups had never fully developed their underlying positions. They were divided, to be sure, but they didn't know precisely how and why. In this sense Mason informed both groups about what they "believed" individually. Next, Mason took a typical set of characteristic operating data on the present state of the company (profit, rate of return on investment, etc.) and showed that every piece of data could be used to support either the plan or the counterplan; i.e., there was an interpretation of the data that was consistent with both plans. Hence, the real debate was never really over the surface data, as the executives had previously thought, but over the underlying assumptions. Finally, as a result of witnessing this, both groups of executives were asked if they, not Mason, could now formulate a new plan that encompassed their old plans. Fortunately in this case they could and because of the intense and heated debate that took place, both groups of executives felt that they had achieved a better examination of their proposed course of action than normally occurred in such situations.

Of course, it should be noted that such a procedure does not guarantee an optimal solution. But then, the DIS (Dialectical Inquiring System) is most applicable for those situations in which the problem cannot be formulated in pure Leibnizian terms for which a unique optimal solution can be derived. DIS are most appropriate for precisely those situations in which there is no better tool to rely on than the opinions of opposing experts. If the future is 99 percent opinion and assumption, or at least in those cases where it is, then the DIS may be the most appropriate methodology for the "prediction" and "assessment" of the future.

It is important to appreciate that the DIS and Policy Delphis differ fundamentally from other techniques and procedures that make use of conflict. In particular, they differ greatly from an ordinary courtroom debate or adversary procedure. In an ordinary courtroom debate, both sides are free to introduce whatever supporting data and opposing arguments they wish. Thus, the two are often confounded. In a DIS, Hegelian inquirer or Policy Delphi, the opposing arguments are kept strictly apart from the data so that the crucial function of the opposing arguments can be explicitly demonstrated. This introduces an element of artificiality that real debates do not have, but it also introduces a strong element of structure and clarity that makes this use of conflict much

more controlled and systematic. In essence, the Hegelian Inquiry process dictates a conceptual communication structure which relates the conflict to the data and the objectives. Under this conception of inquiry, conflict is no longer antithetical to Western science's preoccupation with objectivity; indeed, conflict actually serves objectivity in this case. This will perhaps be puzzling to those who have been brought up on the idea that objectivity is that upon which men can agree and not on what they disagree. While the Hegelian inquirer does not always lead to a new agreement (i.e., a new plan), the resulting synthesis or new agreement, when it occurs, is likely to be stronger than that obtained by the other inquirers.

Singerian-Churchmanian IS

Singerian IS are the most complicated of all the inquirers encountered thus far and hence the most difficult to describe fully. Nevertheless, we can still give a brief indication of their main features as follows:

(1) Truth is *pragmatic*; i.e., the truth content of a system is relative to the overall goals and objectives of the inquiry. A model of a system is *teleological*, or explicitly goal-oriented, in the sense that the "truth" of the model is measured with respect to its ability to define (articulate) certain systems objectives, to propose (create) several alternate means for securing these objectives, and finally, at the "end" of the inquiry, to specify new goals (discovered only as a result of the inquiry) that remain to be accomplished by some future inquiry. Singerian inquiry is thus in a very fundamental sense nonterminating though it is response oriented at any particular point in time; i.e., Singerian inquirers never give final answers to any question although at any point they seek to give a highly refined and specific response.

(2) As a corollary to (1), Singerian IS are the most strongly coupled of all the inquirers. No single aspect of the system has any fundamental priority over any of the other aspects. The system forms an inseparable whole. Indeed, Singerian IS take *holistic thinking* so seriously that they constantly attempt to sweep in new variables and additional components to broaden their base of concern. For example, it is an explicit postulate of Singerian inquiry that the *systems designer is a fundamental part of the system*, and as a result, he must be explicitly considered in the systems representation, i.e., as one of the system components. *The designer's psychology and sociology are inseparable from the system's physical representation.*

Singerian inquirers are the epitome of synthetic multimodel, *interdisciplinary* systems. In effect, Singerian IS are meta-IS, i.e., they constitute a theory about all the other IS (Leibnizian, Lockean, Kantian, Hegelian). Singerian IS include all the previous IS as submodels in their design. Hence, Singerian inquiry is a theory about how to manage the application of all the other IS. In effect, Singerian inquiry has been illustrated throughout this chapter in our

descriptions of the inquirers, for example, in our previous representations of the inquirers and in our discussions of which kinds of problems the inquirers are best-suited to study. A different theory of inquiry would have described each of the preceding inquirers differently.

Singerian IS contain some rather distinctive features which none of the other IS possess. One of their most distinctive features is that they speak almost exclusively in the language of commands, for example, "*Take* this model of the system as the "true" model (or the true model within some error limits $\pm \epsilon$)." The point is that all of the models, laws, and facts of science are only approximations. All of the "hard facts" and "firm laws" of science, no matter how "well-confirmed" they are, are only hypotheses, i.e., they are only "facts" and "laws" providing we are willing to accept or make certain strong assumptions about the nature of the reality underlying the measurement of the facts and the operation of the laws. The thing that serves to legitimize these assumptions is the command, in whichever form it is expressed, to *take* them seriously, e.g., "*Take* this is as the true model underlying the phenomenon in question so that with this model as a background we can do such-and-such experiments." Thus, for example, the Bohr model of the atom is not a "factually real description of the atom," but if we regard it as such, i.e.,. if we *take* it as "true," then we can perform certain experiments and make certain theoretical predictions that we would be unable to do without the model. What Singerian inquirers do is to draw these hidden commands out of every system so that the analyst is, he hopes, in a better position to choose carefully the commands he wishes to postulate. Although it is beyond the scope of this chapter, it can be shown how this notion leads to an interesting reconciliation between the scientist's world of facts (the language of "is") and the ethicist's world of values (the language of "ought"). In effect, *Singerian inquiry shows how it is possible to sweep ethics into the design of every system.* If a command underlies every system, it can be shown that *behind every technical-scientific system is a set of ethical presuppositions.*

Another distinctive feature is that Singerian IS greatly expand on the potential set of systems designers and users. In the extreme, the set is broadened to include all of mankind, since in an age of larger and larger systems nearly everyone is affected by, or affects, every other system. While the space is not available here to discuss the full implications of this proposition, it can be shown that *every Singerian IS is dependent upon the future for its complete elucidation.* If the set of potential users for which a system exists is broadened to include all of mankind, then this implies that every system must be designed to satisfy not only the objectives of the present but also the objectives of the future. Thus, a Singerian theory of inquiry is explicitly concerned with the future and is by definition involved with the forecasting of the future. Singerian IS attempt to base their forecast of the future on the projections of as many diverse disciplines, professions, and types of personalities as possible.

Singerian inquiry has been conspicuously absent from the field of Delphi design; hence, unfortunately, we cannot talk about any current applications of Singerian IS to Delphi. There are hints of Singerian overtones in those few Delphis that ask people for the contrast in their real views and the views they would state publicly. However, none of these has ever explored the underlying values and psychology to the extent of warranting a Singerian label. Nevertheless, we can say something about what a Singerian Delphi would look like.

Of all the many features that Singerian inquiry could potentially add to Delphi design, one of the primary ones would be a general broadening of the class of designers. That is, at some point the participants should not merely participate in a Delphi but be swept into its design as well. In a Singerian Delphi, one of the prime features of the exercise would not only be to add to our "substantive knowledge" of the subject matter under investigation, but just as much, to add to the participants' knowledge of themselves. How do the participants change as the result of participating in a Delphi? Are their conceptions of policy formation, and of who and what constitutes an "expert," the same afterwards as before? How is it possible to sweep the participants more actively and more consciously into the design of the Delphi? What are the values and/or psychology that led me and my fellow respondents to answer with this view? These are only a very few of the many issues with which a Singerian-designed Delphi would be concerned, and as a result, would thus act to build into the design of the Delphi the potential for pursuing these questions systematically. In short, a Singerian-based Delphi is concerned with raising and building explicitly into the design of the technique *the self-reflective question: How do I learn about myself in the act of studying others and the world?* Why is it that some minds think they can best learn about the world and the contents of other minds (i.e., their communications) by formal models only? Why do others believe they can best learn through empirical consensual means, and others still, through multiple synthetic or conflictual means? And finally, why do Singerians want to spend so much time studying the others? What kind of mind is it that studies others? Perhaps, perverse; most certainly, reflective—the very spirit that moved the first pioneers of the Delphi technique to want to study how and under which circumstances a *group* of reflective minds was better than one.

Concluding Remarks

In many ways a brief commentary on the strengths and weaknesses of Singerian inquiry provides the most fitting summary to this chapter.

The strength of Singerian inquiry is that it gives the broadest possible modeling of any inquirer on any problem. The weakness is the potentially prohibitive cost involved in comprehensive modeling efforts. However, given

the increased fear and concern with our environment, we may no longer have the choice but to pay the price. We may no longer be able to afford the continued "luxury" of building large-scale Leibnizian and Lockean technological models that are devoid of serious and explicit ethical considerations and which fail to raise the self-reflective question. We certainly no longer seem able to afford the faulty assumption that there is only one philosophical base upon which a technique can rest if it is to be "scientific." Indeed if our conception of inquiry is "fruitful" (notice, not "true" or "false" but "productive") then to be "scientific" would demand that we study something (model it, collect data on it, argue about it, etc.) from as many diverse points of view as possible. In this sense strict Leibnizian and Lockean modes of inquiry are "unscientific" because they inhibit this effort, a conclusion which we are sure most of our "scientific" colleagues would be surprised to find and even more reluctant to accept. But then, believing in conflict as we do, we might have a good debate on the matter. If one were to design a Delphi to investigate the matter, which Delphi inquirer design do *you* think we (you) *ought* to use?

References

The references listed below are intended to provide the reader with general reviews, further background, and some specific examples of topics covered in the article. On the subject of Inquiring Systems the best place to seek further explanation would be:

C. West Churchman, *The Design of Inquiring Systems*, Basic Books, New York, 1971.

Those interested in attempts to construct formal mathematical representations of Inquiring Systems are directed to the following three articles:

Ian I. Mitroff, "A Communication Model of Dialectical Inquiring Systems—A Strategy for Strategic Planning," *Management Science*, 17, No. 10 (June 1971), pp. B634-B648.

Ian I. Mitroff and Frederick Betz, "Dialectical Decision Theory: A Meta-Theory of Decision Making," *Management Science* 19, No. 1 (September 1972), pp. 11-24.

Ian I. Mitroff, "Epistemology as a Basis for Building a Generalized Model of General Policy-Sciences Models," *Management Science*, special issue on "The Philosophy of Science of Management Science," to appear.

This chapter is, in large part, a specialization of an earlier more general article:

Ian I. Mitroff, and Murray Turoff, "Technological Forecasting and Assessment: Science and/or Mythology?" *Technological Forecasting and Social Change*, 5, No. 1 (1973).[3]

[3]A condensed version of the above directed to an engineering audience appeared in the March 1973 issue of *Spectrum*, which is the magazine of the Institute of Electronic and Electrical Engineers.

II. C. Reality Construction as a Product of Delphi Interaction

D. SAM SCHEELE

> The mind is but a barren soil;
> a soil which is soon exhausted, and
> will produce no crop, or only one,
> unless it be continually
> fertilized and enriched
> with foreign matter.
>
> Sir Joshua Reynolds

Reality is a name we give our collections of tacit assumptions about what is. We bring along these realities to give meaning to our interactions. Each of us maintains several of these realities—at least one for every significant set of others in our lives. We have domestic realities, parental-family realities, professional realities, sexual realities, organizational realities, stylistic realities... Since this article is about Delphi design, the important thing is not how many different realities we each have, but that one important product of each Delphi panel is the reality that is defined through its interaction.

Realities can be described as presumed agreements which give meaning to our thoughts and make reasonable our actions in each setting. Most of these agreements about reality are implicit, and are merely confirmed and elaborated by our acts and conversations. Sometimes our interactions subtly modify these realities. Occasionally, a group's reality is actively renegotiated or even constructed *de novo* for a new situation. Delphi inquiries might produce any of these results. The purpose of this essay is to suggest ways of managing Delphi interactions in order to create intentionally a reality that will prompt the appropriate kinds of active interventions.

I believe with others that there is nothing more practical than good theory. Much of what is related in this article is theory, but theory in search of application, by the reader, out of his understanding, to produce results in his specific contexts. I have used a number of examples to illustrate how realities are asserted, modified, reconceptualized during a Delphi interaction. Each example will require detailed consideration. This is painful if the reader is interested merely in an overview. Therefore, I have tried to write the discursive text so that it makes sense even if the reader skip the examples. Each example is a freeze-dried caricature of a set of rich interactions. To reconstitute, the

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

reader must supply the cerebral juices and attention, such connective interpretation being necessary to make static diagrams into what might pass for interaction.

Another problem with examples is that they represent only a small fraction of the myriad of potential applications for this approach to Delphi inquiry. Clearly it would be incestuous to build a design rationale solely on generalizations drawn from available applications. Taking an expository approach based on cases would foster an already widespread predilection—method and technique in search of applications. This would be the antithesis of my primary recommendation: that the particular qualities of the circumstances that prompt and define the inquiry be used as a basis for the Delphi design. Further, let me suggest that the results of a Delphi be seen as the product of a carefully designed and managed interaction and not answers to a set of abstract questions that are obtained by following prescribed methods. Hence, a slogan for this essay: Concepts from doing.

This paper might have been called: What to think about when considering, designing, and managing (even interpreting) a Delphi. The reader will find many propositions asserted that require a reflection and reinterpretation for application to his particular undertaking. The extensive illustrations are intended to enable the reader to develop a feel for the importance of details of style and tone in presenting materials to panelists. These illustrations should not be thought of as "cases" to emulate, but as necessary to describe the more general pedagogic points about the importance of self-conscious presentation of information in suggestive, but open-ended, frameworks to facilitate the negotiation of realities. Most of the illustrations are based on Delphis we have conducted. In several cases the substantive content of the illustration has been changed because of the proprietary nature of the inquiry or the possibility that our intent would be misinterpreted if the material were to be seen out of context. Also included is some of our thinking which occurred when we did not do a Delphi.

The italic text provides a setting for some of the illustrative examples. Each illustration depicts a synthesis of the interaction between the panelists with summarization, juxtaposition, interpretation, reconjecture by the Delphi monitor. The role of the diagrammatic presentation of the examples is described in the illustration below (Fig. 1). The intention is not to create order or to impose a unique conceptualization. Neither are the diagrams supposed to be balanced, "well-designed," synthesizing abstractions, or even documentations. In some of the Delphis, the major part of the panelists' comments were sent back on tape cassettes. The emphasis is a personal verisimilitude with the process of undertaking conceptual forays. Most of the panelists' thinking processes cannot be directly shared, so we have attempted to depict for the group some typical points of view out of which to define a reality of relevance.

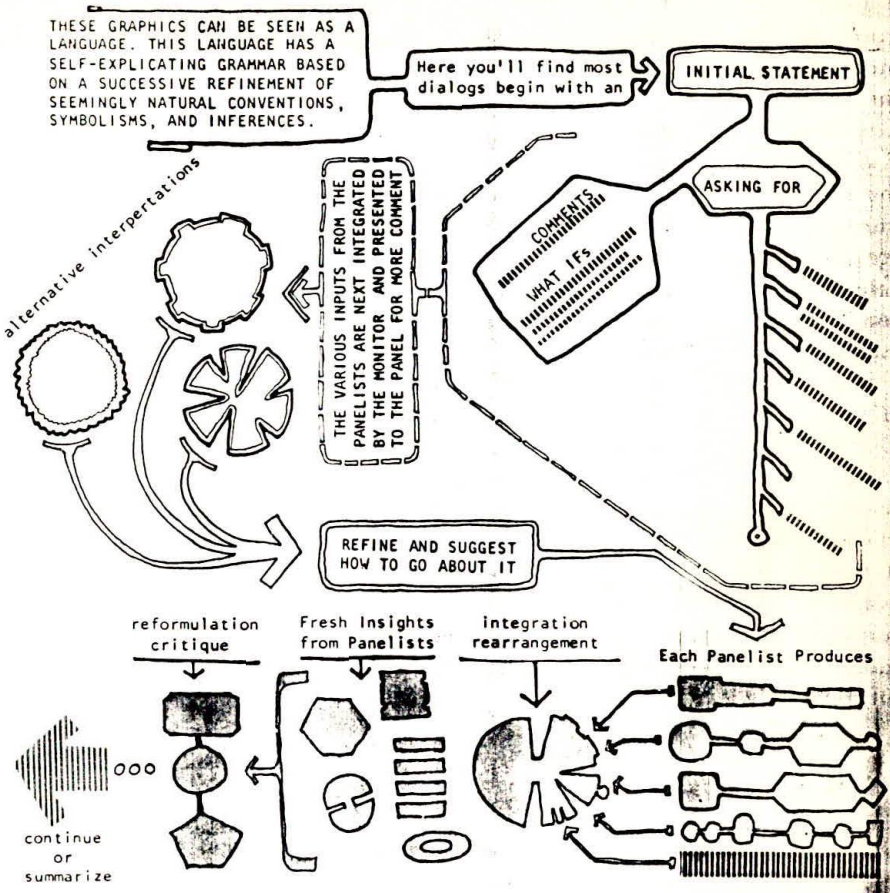


Fig. 1. Role of the diagrammatic presentation of examples.

The particular graphic style adopted in this paper is a personal one. It developed out of use. It is intended to support the process of thinking and not simply to represent completed conceptualizations. Graphic aids can be useful in stimulating your own thinking and organization of ideas. Start by trying your hand at whatever seems cogent to you and make adjustments as you go. There are a rationale and some techniques that have evolved in the use of graphics in thinking, but their explanation would make another book. The quality of the

drawings is intended to convey the liveliness of the concepts and their important properties effectively and quickly. This is difficult to do in writing or with more formal diagrams. Also this drawing style encourages participation, and the organization of the diagrams usually readily admits of modification or extension. In addition to aiding individual thinking, common graphic constructions, or explicit group memories, are useful in moderating discussions. Here the shorthand of positional relationships and the insightfulness of successive interpretations and alterations prove very productive. Since Delphi inquiries are group processes, these kinds of graphic representations have proved equally valuable in Delphi applications. The figures may not mean the same thing to you as to me, but your explanation is accessible, and therefore should be more useful to you than mine would be.

Some of the points may seem trivial—like, "use bright colors" or "state ideas in emotive language"—but their impact is significant. Others—like "provide a concrete situational context" or "depict an explicit theoretical framework"—have resulted from trying to de-abstract Delphi inquiries. It is important to deal with the different assumptions of panelists, monitor,[1] and sponsor which, when left alone, limit the potential fruitfulness of the Delphi interaction. Most of the methodological insights suggested here have resulted from efforts at designing other kinds of subjective information processes, such as diffusion of innovations, learning by doing, technology transfer, policy development, management of creativity, and design of service systems. Look before you leap, but eventually leap.

Concepts of Reality

The impact of one conceptualization of a situation upon others and the influence of the various constructions of reality assumed by the Delphi panelists generate what I believe are the most significant results from any Delphi inquiry. Panelists can be made aware of these seemingly subtle differences in the nature of the realities they presume in the course of the interactions. The panel can then produce a common reality for the situation at hand as a result of their participation.[2] How this comes about is determined by the monitor of the Delphi interaction. If panelists are reluctant to make specific contributions or if a very wide, almost unrelatable, array of conjectures is produced, one suspects that there are great differences in the meaning each panelist is attributing to the way the inquiries are stated in the materials provided to stimulate response and the way panelists expect the results to be used. This

[1]Monitor is a term I use for the person or group conducting the Delphi inquiry, i.e., preparing the materials, interpreting the responses, integrating the insights, presenting the results, etc.

[2]For a longer exposition of, and details about, this concept see S. Lyman and M. B. Scott, *A Sociology of the Absurd*, Appleton-Century-Crofts, New York, 1970, and H. N. Lee, *Percepts, Concepts, and Theoretical Knowledge*, Memphis State University Press, Memphis 1973.

ambiguity often might be what you want—productive of interesting premises. An array of differently bracketed realities that include a particular object-event-concept is often useful. For example:

Four young adults who are retarded enter a restaurant with older couple. Panelists were asked to select likely responses for restaurant manager, waiters, and other patrons from a list provided. Many panelists added their own. The panel included parents of, and professionals who work with the retarded, as well as individuals to simulate response of general community. The responses of the panelists could be mapped:

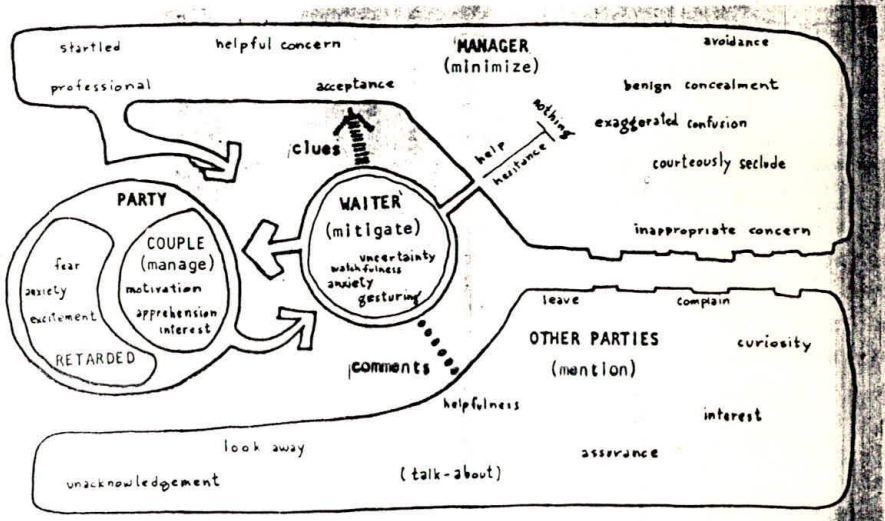


Fig. 2-A

Later, panelists were asked how repeated contact with the retarded would affect the key actors:

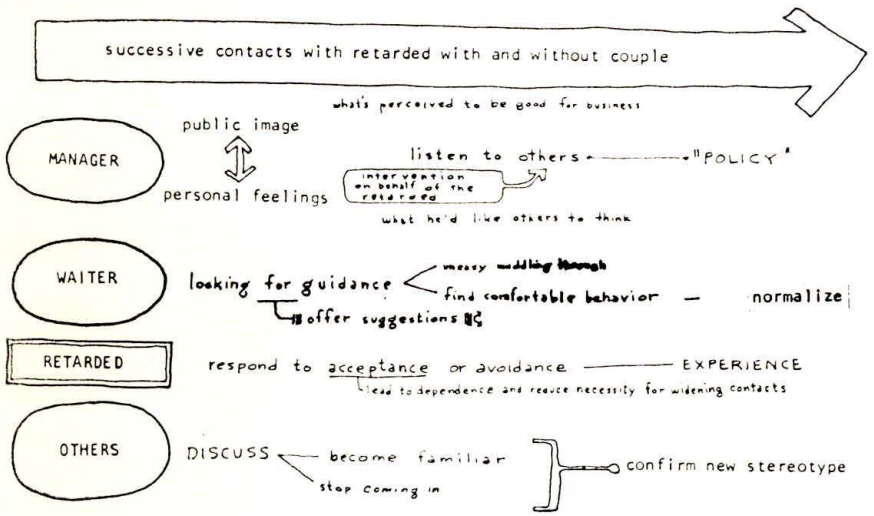


Fig. 2. What is going on in the restaurant.

On the other hand, you may want to create greater focus and consensus. If this is so you can begin with examples for interpretation instead of general questions. This enables you to direct attention in subsequent rounds to contrasts between the assumptions imbedded in the initial situations and panelists' contributions. Differing reality constructs can produce divergence from even seemingly unambiguous statements. Focusing attention on differences in the reality constructs will usually yield either a more refined and widely agreed-upon definition of the appropriate context or clearer and more precise distinctions between competing contexts—possibly leading to an estimation of the relative probabilities of each, or a search for present options that could influence the circumstances.

In the preceding discussion the notion of socially constructed or intentionally negotiated realities was employed. This concept grew out of the work of

Husserl,[3] Merleau-Ponty,[4] Heidegger,[5] and others, which led to the formulation of a phenomenological epistemology that is now being applied by neo-symbolic interactionalists and ethnomethodologists. The concept of a negotiated reality can be related to Mitroff and Turoff's discussion of the philosophical bases for inquiry systems in the preceding article. This discussion describes a range of inquiry systems (IS) using as differentiating labels the name of the principal philosopher whose concepts undergird each approach. The array includes the Leibnitzian, Lockean, Kantian, Hegelian, and Singerian IS's. Since these categories are well defined there, the philosophical premise for an IS based on a view of reality as a context-specific product of interaction will be described in relation to this framework. First, to select a label consistent with the others will be slightly misleading, because any one name tends to obscure the contributions of others and imply that the ideas are largely set. Dubbing this territory of philosophic exploration after Merleau-Ponty seemed the least misleading. To make a contrast with the Singerian analyst, the Merleau-Pontyean is concerned with the particular reality created by the "bracketing" of an event or idea out of the great din of experience, rather than explicating a pragmatic reality that can be used to define possible actions. Truth to the Merleau-Pontyean is agreement that enables action by confirming or altering "what is normal" or to be expected. By contrast, the Singerian views truth as an external articulation of systems to define goals and options for action. Reality is viewed by the Merleau-Pontyean as the product created out of intentions and actions instead of an external basis for intelligent actions. To reiterate Mitroff and Turoff, the importance is not which philosophy is "correct," but which is appropriate to the kinds of situations one is attempting to impact. The Merleau-Pontyean inquiry system seems applicable to situations either where a redefinition of contextual reality can facilitate the generation of new options, or where the acceptance of a new reality must be negotiated to create the impetus for technical or social change—as, for example, in defining as "progress" a reduced or more limited consumption that would permit reallocations instead of "progress" as lower unit production cost to support increased demand. This philosophical point of view leads to viewing the future as a situation where both the dominant reality and the technology are invented as well as inherited, and where culture is transformed as well as transmitted.

Merleau-Ponty and others suggest reality be viewed in a new way: as a currently prevailing shared assumption about a specific situation. This implies that reality is the product of our experience and not external to it. In a

[3]Edmund Husserl, *Ideas: General Introduction into Phenomenology* (trans. W. R. Boyce), Allen & Unwin, London, 1931.

[4]Remy C. Kant, *Merleau-Ponty and Phenomenology*, in *Phenomenology* (Kockelmans, ed.), Doubleday, Garden City, N. Y., 1967.

[5]Calvin O. Schrag, *Phenomenology, Ontology and History in the Philosophy of Martin Heidegger*, Revue Internationale de Philosophie, vol 2 (1858).

commonsense view, reality is a collection of observable things and occurrences which is animated by a society of individuals. Although we are not usually aware of these distinctions, our everyday realities can be seen as created by us out of the meanings we give things and events.[6] Since we do not exist alone, we are continuously asserting and having validated or challenged our definition of "what's going on" or "what it's all about." Our collection of situational definitions constitutes our reality. We select realities from our repertoire that seem appropriate in order to know how to act, attribute meaning, and interpret behavior. This means that instead of continuously discovering more of an external verity—"the reality out there"—we are, wittingly or not, continuously adding, verifying, or revising our "shelf-stock" or versions of what is normal or to be expected in particular circumstances. We each have a shelf-stock of realities that have been produced by our interactions.

Earlier the basic philosophic question had been "What is the structure of social reality?" Now phenomonological insights have transformed this question into "What realities have been or are being socially constructed?"

What does this mean for conducting Delphis? *First*, since the results of a Delphi are produced by interaction, albeit highly structured, the results can be said to constitute a reality construct for the group.[7] Because processes of successive refinement, like the Delphi, strongly tend to induce convergence and agreement, the monitor of a Delphi should purposely introduce ambiguities, even disruptions. These might take the form of "angle" items[8] to challenge and redefine reality as well as "quirk" items[9] to act as catalysts to explore the limits of the reality. For example:

Mass transit could compete with private vehicles by offering more than lower cost—particularly in enhancing the use of commuting time by offering:

Round 1: *What types of services?*

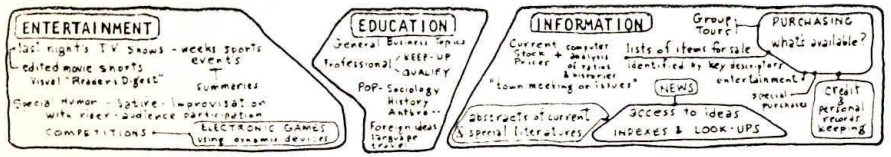


Fig. 3-A

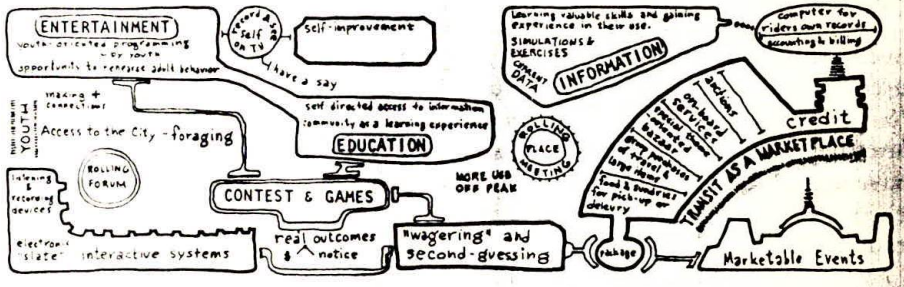
[6]Jack Douglas, *Understanding Everyday Life*, Aldine, New York, 1970.

[7]Hans Peter Drietzel, *Recent Sociology No. 2—Patterns of Communicative Behavior*, Macmillan, New York, 1970.

[8]Example: "Suppose you had invented the better mousetrap and people had beaten a path to your door—would they buy?"

[9]Example: the famous rejoinder, "And how does that relate to the Jewish question?" or Stephen Potter's functional equivalent, "...but what of the South?"

Round 2: *Can't money change hands—transit as market place?* (*angle*)—*Relate services to attracting and serving youth?* (*quirk*)



Round 3: *How might the service be organized and supplied?*

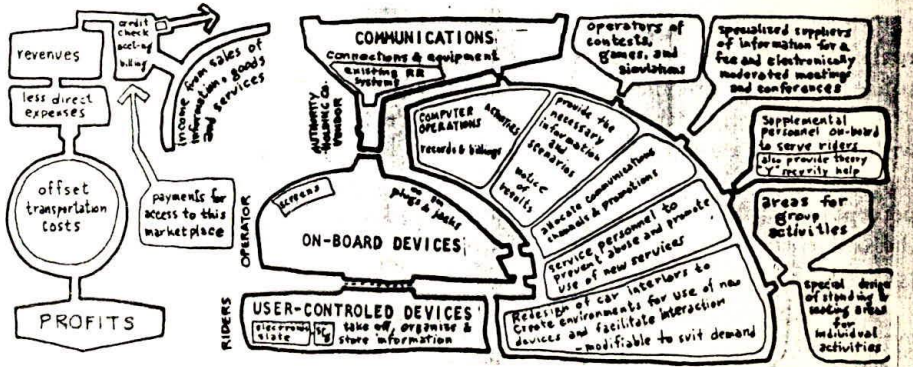


Fig. 3. Activities to enrich mass transit experience.

Second, since the knowable reality is in competition with the other conceptions, including the idea of reality as a negotiable construct, the unknown or unexplained cannot simply be attributed to greater degrees of complexity (i.e., the "more data and better instruments" gambit). This means that further efforts to obtain information, such as a Delphi, must go beyond attempts to unravel what has often been assumed as merely additional complexities. Instead, information should be sought that can shape reality, such as identifying new considerations or introducing new options. This means that the systems

ing described are viewed as indeterminate, arbitrary, delimited, multiplistic, ven convenient fictions if this facilitates discourse, but not as complex artesian clockworks. In conducting a Delphi then, "what if" and "why not" ms might be introduced or highlighted if suggested by a panelist to prompt nsideration of new alternatives.

Third, the reality we construct can be expected to be different by at least as ich in the future as our technology will be advanced or our society restruc- ured. This is almost always overlooked by forecasters and other futurists. edictions may well occur as forecast, but their occurrence will not necessarily an the same thing then as it now seems that it would. You can note people's aive understanding of this in their response to the prediction of new occur- ices with the statement, "...but I guess they [the people of the future] will be dy for it by then."

Fourth, expect reality to continue to be negotiated. This means that the kinds realities within which occurrences will be given meaning and be understood ll vary from those prevailing at present. To a large extent changes in reality nape the kind of attention, consideration, and effort that will be spent on veloping a new idea. They also determine whether the new concept seems lausible, desirable, and feasible. Further, both interest in, and advocacy of, a ew concept, along with precipitous events that come to be associated with it, n shape reality to the extent that a new concept becomes one of the ways that ality is defined. Our present notion of the "urban crisis" is an example of a ncept that has become imbedded in our realities.

A Delphi inquiry, then, might explore two sides of the negotiation of reality with regard to a specific occurrence: (1) how alternative realities might affect e meaning of the occurrence and how likely each is and (2) how public reeptions of the occurrence, the interests and activities of its proponents, and he concepts and ideology that come to be associated with it will shape the ality that is negotiated.

In many cases it also may be useful to consider the possibility of precipitating vents. These events (seen from hindsight) have often dramatically altered the llective awareness of consciousness of the society. For example, context- ecific realities were shaped by the assassinations of the Kennedys and King or gun-control measures, Nader and "Unsafe at Any Speed" for auto safety mpaigns, and the Soviet Sputnik for the U.S. space program and educational direction.

Let us look at how these considerations can be handled as part of a Delphi quiry. To grasp how different prevailing realities might affect a particular pic, one can posit several alternative "reality gestalts." Encourage partici- pants to supply their own realities. But I believe examples are necessary to arify this notion initially—even if these proto-realities are overly "cari- tured." The example below illustrates this kind of inquiry.

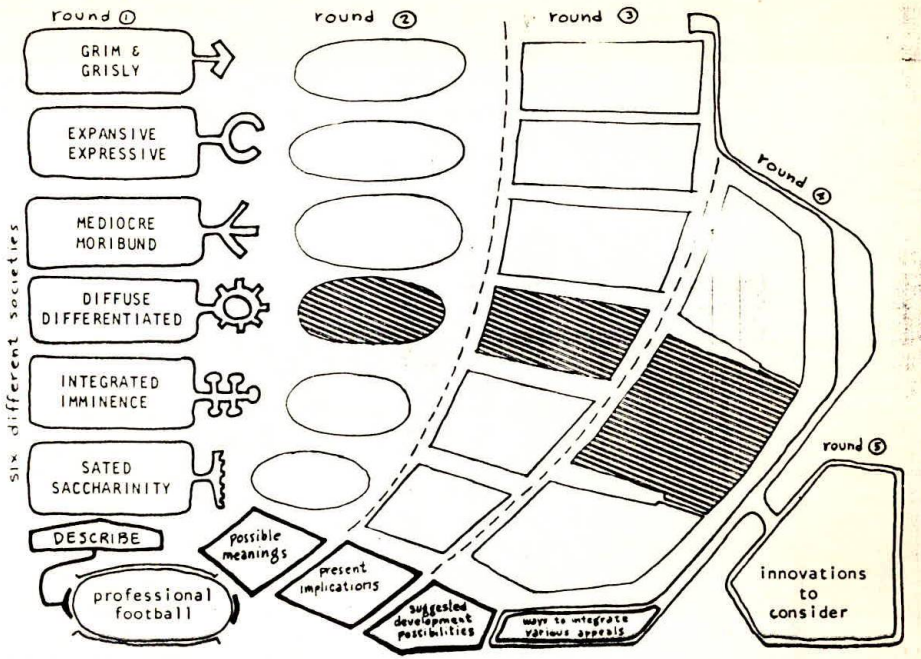


Fig. 4. Professional spectator sports—prevailing views.

Also, you might want to probe the participants' insights regarding items that might be significant in the construction of realities in the future. Here it is useful to suggest items that you believe are important as possible examples. Even include candidate-precipitating events—although their nature and timing are by definition unanticipatable. Knowing that the meaning of events is, so to say, up for grabs can sensitize managers to the use of public-attitude-shaping tools. Introducing these perturbations will begin the discussion and evoke additions and comments from the panelists. An important corollary point is: the meaning of the future in the reality of those taking actions (of necessity in the present) may change. It has dramatically in the past ten years or this book wouldn't have been considered for publication, possibly not even written. One's vision or concept of the future has not always shaped reality in the same way. It is as important to know in what ways images of the future are given meaning as to have a full complement of alternative futures (see Chapter VI, Section D).

The medieval glory-beyond-the-travail-down-here view of the future would prompt a different kind of action from the expectation of a better tomorrow through hard work and technology which has characterized the industrial revolution.

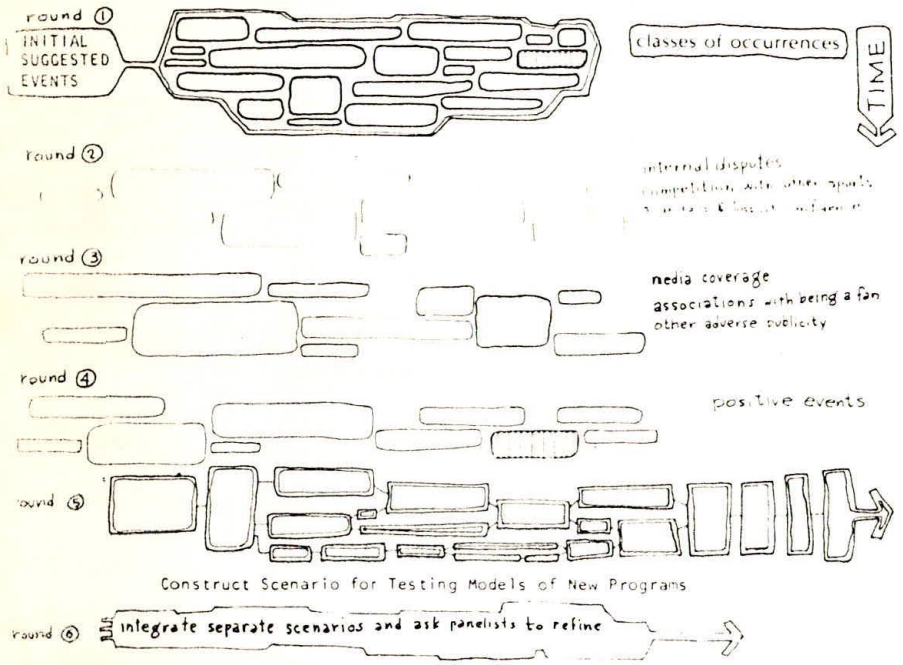


Fig. 5. Professional spectator sports—reality shapers.

Mental Climates and Styles of Thinking

The dominant reality of the "modern" world has been called the positivist-functionalist view. In this view reality is "out there" (we are in it, but it is not in us), discoverable by improved access and greater attention and presumed to be ultimately knowable—given sufficient time and diligence. Lukacs[10] and Panofsky[11] call these global realities the *habitus mentalis* of an era. A *habitus*

[10]Georg Lukacs, *History and Class Conciousness* (trans. Rodney Livingstone), M.I.T. Press Cambridge, Mass., 1971.

[11]Erwin Panofsky, *Meaning in the Visual Arts*, Doubleday, Garden City, N.Y., 1955.

mentalis has at least these three characteristics:

- an epistemology—shared assumptions about what is valid knowledge or "truth," how it can be discovered, and what constitutes acceptance or "proof."
- a social theory—concepts that a society uses to explain its own workings to itself, including the identifiers for constituting aggregates.
- a guidance process—methods and ground rules worked out and accepted by the members of a society—what Harold Lasswell describes as "who gets what, when, where, and how." At different times this is also called "politics" or "management."

At a macro scale the dominant *habitus mentalis* of society is being negotiated. Ours is now. I would like to summarize some of the changes we expect in the postindustrial society. To begin, I'd like to share a name we have coined for the postindustrial society that is descriptive of it, instead of merely indicating *after* industrial. This new label is "the idiomergent era." This is a neologism. *Idio* is a Greek root indicating uniqueness, separateness, distinctness, as in the word idiosyncrasy, and in its most peculiar sense, idiot. It has been combined with the Latin root *mergere*, to cause to be swallowed up, to immerse, to combine or coalesce, to lose identity. The result describes a more separate and a unique kind of interdependent aggregation. It seems that the idiomergent era will be characterized by both extreme variety and increased uniqueness in individual behavior and organizational purposes and at the same time be dependent on the greater integration of the individual and organization into the structures of a variety of specialized communities of interest, producing networks of cohesiveness for the society as a whole. Thus the term "idio-mergent" denotes greater individuality and autonomy combined with a deeper involvement of the individual in fluid social arrangements, innovative organizational structures, and experimental personal relationships. This describes change in the vanguard of society: a new *habitus mentalis* that will spawn new realities.

The *habitus mentalis* for the idiomergent era will no doubt affect the design of Delphi inquiries. Exactly how depends on the response of practitioners and monitors to their perceptions of what is relevant and their selection and advocacy of ideas that come to be in good currency. I will try to share my construct for this new mental climate. You realize that I have no way of knowing the future, nor am I merely asserting an explanation to serve as a basis for judgment. These and many other possible descriptions of what this image of the future is can be offered. My view is that I am engaged with you in negotiating what to expect from Delphis. Use it as you will.

Before getting to differences in the epistemology, social theory, and guidance process of the thinking style of the idiomergent era and what these augur for the design of Delphi inquiries, I would like to suggest another feature of this era —an increasing concern with the future. Greater interest in the future has resulted from a series of changes in our dominant social reality. These changes seem relevant to consider in designing systems for assembling useful informa-

tion. First, change became accepted as normal. Reality became dynamic, but this alone could not make the future more important now; things would just work out differently. Second, technological development and social and physical mobility have increased the number of options available to decisionmakers. This made their reality wider, but not necessarily extended deeper into the future. Third, the number of self-evident paths, inevitable selections, and unquestioned preferences was reduced by an illusion of choice fostered by great increases in the availability of education—both schooling and media-supported awareness. This ballooning of information has expanded reality to include the possible, the imagined, the dreamed, the fantasied, and all other "residents of the future in good standing."

The itch for a fix on the future afflicts both people and organizations. Individuals want to know more about the future so they can choose places to live, things to experience, roles to adopt that will be highly desirable when the future occurs. Businesses analyze future opportunity areas and establish organizations to prepare to exploit them. Institutions assess their roles in the future and begin reconceptualizations. Government investigates the future and identifies many unacceptable contingencies that require more controls. In most cases the future reality is portrayed as an unfolding panorama of large-scale changes in general circumstances and situations whose meanings are anchored in the present. The inference to be made for the sponsor of an inquiry into the future was, "If I know more about what's going to happen, I can get ready for it." It was like spying on or getting the drop on the future. These futures were not the world of will, action, and events; but broad backdrops of scattered possibilities from which a series of succeeding "presents" would coalesce on the moving pointer of time. The message was that to "realistically" take effective actions the future must be considered. This future dimension of reality, now so important, has been chiefly portrayed as a passive set of probable possibilities.

The growing recognition that the "meaning" of things and events is not determined, but in fact refers to social agreements that have changed in the past, and which can be renegotiated, even for each circumstance, marks the beginning of the idiomergent era. The reality of the future as a broad techno-cultural script is too ambiguous to support particular undertakings. Attempts have and will be made to bracket specialized future realities that contain only highly relevant possibilities—but in most instances the number of considerations in even a "bracketed" future will not support finite analysis to select a single course of action. But the process can be reversed. Decisionmakers can attempt to negotiate a specific reality for the future which, if they are successful, can come to be widely enough shared that it is realized through successive choices. The future will be invented. Thus in the idiomergent era you can expect attention to focus on the selection of methods to create and portray new realities that have the potential to be actualized through tacit agreement between the intentions of those affected.

In order not to lend credence to the prevailing presumption that Delphi inquiries are concerned only with the future, let us return to look at ways that

Delphis can be used to contribute to the store of knowledge, enlarge the society's understanding of itself, and improve the style of governance in an idiomergent culture. For each of these three characteristics of an era, the currently prevailing concepts will be contrasted with their idiomergent counterparts. This will be done rather telegraphically to provide a background for an example of an item in a Delphi inquiry. Elucidating these points more clearly would require another book.

Differences in Epistemology concern not only what is sought and accepted as knowledge, but how information is categorized and organized to support actions. As a man thinks, so is he.

Industrial	Idiomergent
Reality is external and knowable.	Reality is constructed and negotiable.
Hypotheses are general and are offered for validation.	Hypotheses are context-specific and serve to redirect the ongoing discovery process.
Categories of description are based on observable differences in measurable dimensions.	Categories for description include intentions and selections and are based on an esthetics of knowledge.
Anticipate future conditions by extrapolation of past performances and behaviors (exploratory images).	Invent future images out of creative integrations of expectations normative gestalts).

To illustrate these epistemological differences, the best example seems to be the development of categorizations of consumers for use in market planning for new products. In the industrial era, consumers were defined demographically, geographically, and by aggregates of features that indicated a consumer's relative propensity to buy either specific types of items or to buy in a particular way. One idiomergent approach would be to develop first a model for the process of product introduction, next those categories of consumers likely to be involved, and then where and how to locate them and their probable numbers.

Differences in Social Theory are alternative views of how the society is organized, its tenants perpetuated, and its actions explained. Society is a name for us to call "what is."

Industrial	Idiomergent
Behavior follows laws that can be derived from situational observation.	Behavior occurs in activity sets that presume and assert meaning.
Actions are explained in terms of expression of internal motivations and conflicts.	Actions are taken in response to assessment of interactional consequences.
Organizations and roles are structures which define possible actions.	Actions and the need for their explication define roles and organizations.
Society is categorized by structural and functional properties to permit measurement and management.	Categorizations of society mark limits for special realities in order to facilitate communication and collaborative actions.

For the Delphi inquirer these differences in social theory suggest greater emphasis on finding out about the appropriateness of societal norms, roles, and institutions, limning how they came into being and, probably most importantly, suggest ways they might be reshaped. The example describes an application to one particular field.

Round 1: *Please elaborate on the diagram below—additional relationships between the retarded and blind person living together to provide mutual support.*

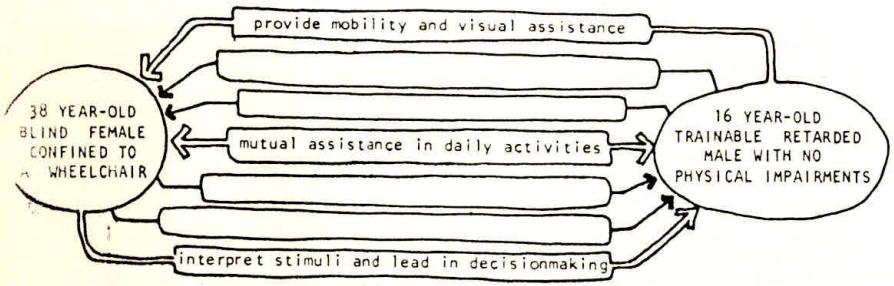


Fig. 6-A

Round 2: *The relationships described by the panel have been aggregated below; suggest where indicated which significant others might contribute to make this relationship more workable, and what they might do. As a neighbor living in the same apartment building, what could you do to contribute to this relationship?*

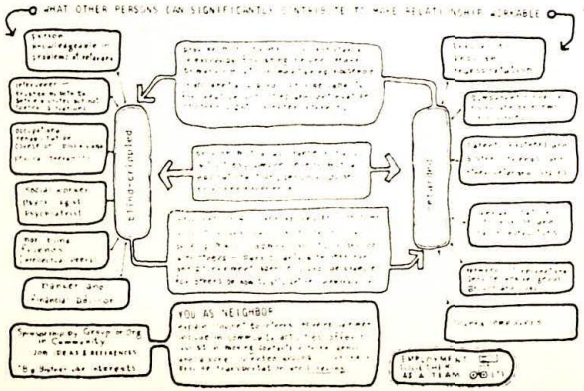


Fig. 6-B

Round 3: *Identify potential difficulties and problems that might be expected to be encountered.*

Round 4: *What might you propose to avoid, ameloriate, or resolve these difficulties?*

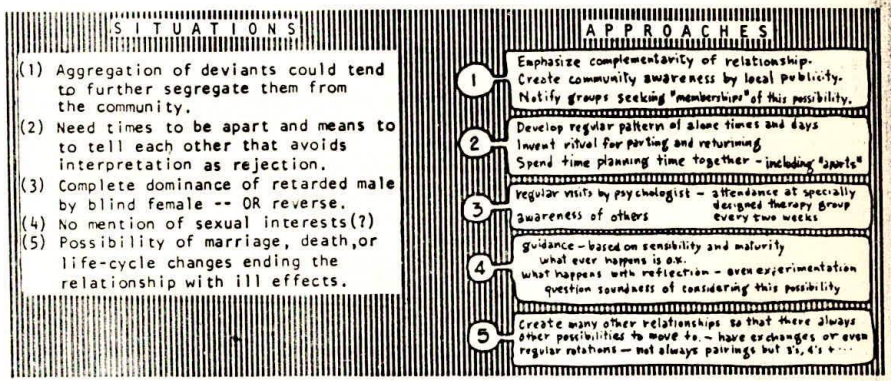


Fig. 6. Symbiotic life-support relationships for the retarded.

Differences in Guidance Processes used in a society reflect implicit, usually unwitting, agreements about the mutual concerns of groups (from pairs to organizations to segments of the society as a whole). These concerns and their mutuality change. Sometimes they change in response to awareness of threats and opportunities. At other times, improvements in techniques such as communications, media, information systems, or analytic schemes change the way fundamental relationships are perceived. Economically and politically the negotiation of power continues. The society's guidance process is not simply decreed—as by a constitution or framework of laws. It includes these, but it is built up rather than laid on. A society's guidance process embodies in conventions all of the expectations of individuals concerning how things ought to be "fairly" and "knowledgeably" decided in order to produce the greater good and what to do if it's not happening as it ought. Societal guidance is all we ourselves are not contending about and in which we assume someone else is watching out for our interests.

Industrial	Idiomergent
Selection of conceptual models based on analytic evaluations including ideological considerations.	Selection of approaches based on evaluation of prototypes and related implications viewed pragmatically.
Competitive determination of potential market for goods and services exchanged in a predominately private marketplace.	Entrepreneurial innovation to create markets for goods and services exchanged in a publicly moderated marketplace.
Representational selection and approval of proposed actions within general jurisdictions based on broad policies.	Collaborative development of proposals for action for particular situations with participative review of specific implications.
Management predicated on adherence to comprehensive plans developed to reach selected goals and objectives defined in limited dimensions to be periodically revised.	Management incorporating participation of those directly affected in the selection of means to respond to short-range opportunities and vulnerabilities identified by multidimensional plans produced by continuing review of an unfolding context.

The changes in style of societial guidance that I expect to characterize the emergence of the idiomergent era will contribute directly to the need for more Delphi and Delphi-like processes for collective consideration of topics. However, evaluation of the usefulness of Delphi inquiries will be made more on their suitability as a process for discussion and choice than on the concerns about the accuracy, insightfulness, or agreeableness of the results. The next example was chosen to illustrate how a Delphi can be used to focus consideration of a topic that has no established institutional advocates.

Round 1: *What actions are possible to reduce peak impact demand for transportation to reduce "overhead" costs of quality urban lifestyle?*

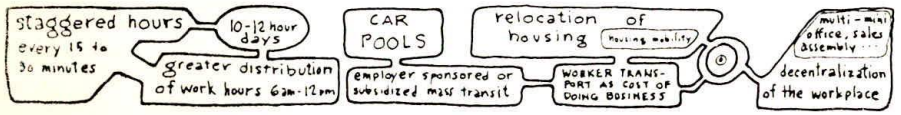


Fig. 7-A

Round 2: *Select most promising approach(es)*

Round 3: *What interests and organizations would be involved in implementation and what interests do they have which could be appealed to?*

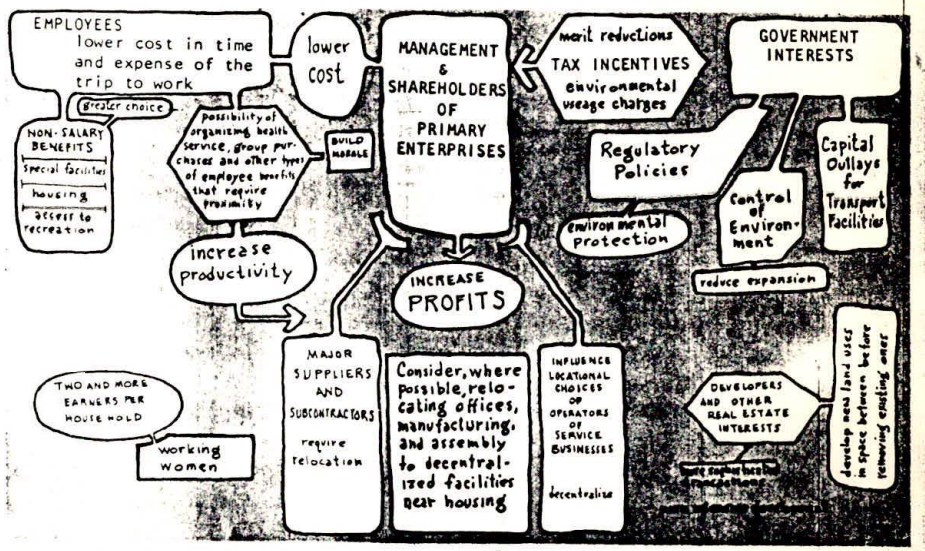


Fig. 7-B

Round 4: *Suggest administrative mechanism to manage program to achieve objectives.*

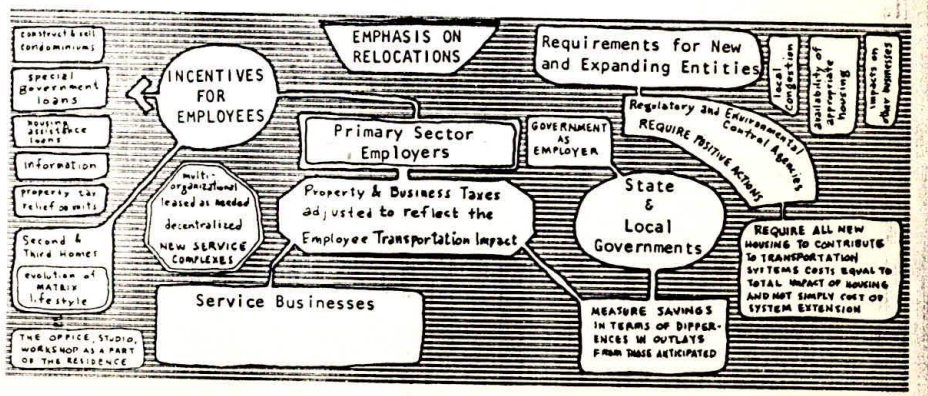


Fig. 7-C

Round 5: Consider implementation: What kind of demonstration program or other approach might be used to rally and mobilize support?

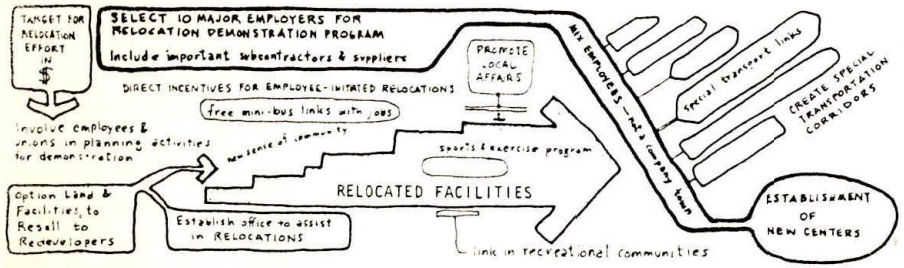


Fig. 7. Relocation incentives in urban transportation policymaking.

Group Influence on Reality Construction

We all have highly idiosyncratic experiences, ideas, and fantasies. While they actually are our own, their meaning is created in the crucible of their interaction with what is going on in our various contexts. For example, a dream shared with people you have recently met has a different impact on the reality being built than the same dream would have if described to an informal caucus during a psychiatry conference. In the first circumstance, your reality is being created from such elements as your image as a person, the meaning of a relationship with you, the place of personal interpretation in your respective motivational structures, and the bearing of dreams on events to come. In the second circumstance, the dream becomes material for negotiation of professional prowess, theoretical differences, clinical inferences, concepts of consciousness—even a trigger for discussions of what this discussion group or the profession is all about. You can imagine that sharing facts, feelings, and proposed actions would also impact differently on the reality being constructed in these contrasting contexts. Try imagining the contrasting reality that could be created in each of the two contexts just described (telling potential friends about yourself *vs.* a discussion about you as a topic in a professional group) for each of the following: the recent death of a relative; your interest in traveling to India; an idea for creating a "kitty" for vacationing by selling the work of local artists to your fellow employees.

Delphi inquiries are conducted with groups where the individual participants might be expected to view the group differently, particularly since they are often anonymous. For most, "being on a Delphi panel" has no particular meaning independent of the topic. Frequently the presentation of the inquiry items and other information is indefinite or ambiguous about the nature of the panel as a group. This directs the panel's attention and energy, in part, to the task of defining the reality of their relationship instead of creating richer realities for the domain of the proported topic. Further, the quality of the reality within which the individual inquiry items are elaborated varies with the ways the panelists view the meaning of the group and its findings, insights, recommendations.

This means that it is not enough that the panelists share a firm idea of the group's identity, but the perception of that identity tends to shape the nature of the individual messages, the quality of the interaction, and the character of what is produced. Table 1 (below) presents a crude taxonomy of the features of groups and some notion of how a group's conception of itself affects the products of Delphi interactions.

The monitor (or modulator) of a Delphi inquiry strongly shapes which conceptualization of the group is assumed by each participant. This usually implicit concept of the group dictates the mode of interaction that each finds appropriate and determines the reality each uses as a frame of reference in attempting to contribute. Several different assumptions can often be made depending on how individuals interpret the implications for reality of the messages in the communications and materials provided them to begin the interaction. Until each panelist is comfortable with his notion of the group's nature and believes it to be confirmed by responses, it is difficult to produce meaningful contributions. In the paragraphs below, I suggest some ways in which a Delphi monitor can shape the group's conceptualization of itself and adopt a mode of interaction that can produce results which will be congruent with their anticipated use.

Transactions will likely become the dominant mode of interaction if there is:

- abstract categorization of the panelists by expertise with inclusion of obviously relevant specialties.
- formal statement of items for consideration, possibilities for estimation, or hypotheses for assessment.
- reiteration of responses categorized by original items with few additions or deletions.
- a product expected of the inquiry as a pooled judgment that will have a "validity" believed to be greater than that of any individual.

In adopting this mode panelists assume that the participants have no other commonalities and expect none. For most this means they will adopt the most familiar model of interaction—probably "answering a questionnaire as

Table 1

Group	Examples	Mode of Interaction	Nature of Realities Produced
Collections of Individuals	Occupants, attendees, shoppers, passengers, gathering, respondents	Transactions	Perfunctory, patterned, provisional, ambiguous—in exceptional cases completely new and perturbing: possibly creating a new pattern.
Casual Groups	Friends, party, clique, players, diners, meetings, trips, class	Experiences	Unique but stylized, comfortable, shaped by random inputs with a strong interest in symmetry and completeness; occasionally evoke powerful new gestalt.
Purposive Group	Colleagues, families, members, associates, collaborators, teams	Episodes	Tightly structured, matter-of-fact, repetitive, presumed heavy with meaning, resist redefinition when alterations (usually catastrophic) occur
Affiliative Groups	Unions, professions, swingers, activists, supporters, delegates	Event	Ritualistic, symbiotic, one-dimensional, subtle shift effected to build power or influence, attempt to be prescriptive, sometimes unsuccessful and produces schism
Defined Group	Students, aged, work force, minorities, dieters, communities	Affairs	Amorphosis, working against stereotype, exploring for unity, proclamatory, infrequently build basis for becoming affiliation
Agents for Society	Citizens, representatives, concerned people, reformers, preservationalists, ecologists	Occurrences	Global, dogmatic, historical, mythological, polemical, sometimes deteriorating into diffuse "glossolalic" diatribe

'accurately' as possible." Exchanges will tend to be formal. They will center on "proof" or "refutation" of ideas identified by the monitor. There will usually be a statistical integration of the group's assessments. Anonymity will likely be used to reinforce the panelists' self-concept as impartial observers in an "analytic" reality. Panelists who are uncomfortable with this reality will devote great attention to selected issues, often digressions from the principal considerations.

Variations in the transactional mode can be effected by (1) beginning the Delphi with open-ended, agendaless items in search of meaning (to explore the agendas and gestalts of the future held by the panelists); (2) selecting the extraneous and derivative responses of the panelists for feedback and later exploring the interpretation of the discourse with the panelists (to identify latent options, issues, and considerations); (3) extracting the explanatory premises from the panelists' responses after two rounds, then focus on estimation, followed by attempts to elaborate a system of relationships (to build theories and models from premises derived from content rather than technique); and (4) starting with a preliminary design of a thing or program, proceed to iterate critique and redesign to produce a better design by successive interaction of multiple viewpoints, or to derive by interpretation of the "designer-panelists'" responses which considerations seemed important or even critical to the design.

Obviously, there are many other manipulations that could be developed, but these are sufficient to illustrate the type of variations that can be created for particular applications. A Delphi should not be undertaken to validate concepts which you already have developed and refined; panelists want to make significant contributions and these will seldom build meaningfully on highly elaborated initial concepts.[12] Delphi inquiries are obviously applicable to much more than obtaining pooled judgements about particular options.

Experiences become the prevailing mode of interaction when:

- the panelists are familiar with each other and identify with the subject or sponsor of the Delphi inquiry.
- the involvement is for a fixed duration, with known consequences that follow a familiar pattern.
- original items serve primarily as jumping-off places for further inquiry.
- the generic form of expected product of the inquiry is clearly indicated.

Inherent in the assumption of this mode of interaction is a commitment to the anticipated process and results. The principal interest of the panelists is the particular qualities or insights each seeks to contribute. Even when anonymity of panelists is maintained, their messages tend to be informal and approval-seeking (other-directed conjectures) instead of raising formal distinctions and

[12]Presenting a less-than-complete concept takes guts. To each panelist there will be "obvious" omissions. You will be severely criticized. But, I believe a stance of calculated naïveté produces the best results, if you live through it.

striving for accuracy and defensibility. Panelists in later rounds will attempt to make contributions that will affirm their brilliance or uniqueness. The result can be omission of important but obvious points and dangling insights (the monitor can point out these). When the panel is largely drawn from a single discipline or field of application, the interaction quickly moves out to the ethereal zone instead of enriching the context for action. Modifications by the experiential mode include: (1) try to organize the panel into "teams" to represent particular interest groups or points of view (to preview negotiations or develop highly detailed but still preliminary agreements for consideration); (2) present initially a clear and detailed description of the expected product of the inquiry; or a unique and special situation that calls out for great imagination and close collaboration to be described first so that the prevailing rules for judging the suitableness of contributions are partially suspended (to spark creativity and develop a relatively complete conceptualization of a viable new approach).

Episodes are the characteristic mode of interaction for groups that:

- are made of individuals who have a significant and continuing relationship, possibly more than one.
- deal with familiar topics or operate in well-known ways.
- are more influenced by consensual implications than external factors in making choices.
- are more concerned with the perceived quality of the interaction than the product—in fact the idea of a product, as such, may not even be considered.

This mode of interaction has the highest emotional content and the most potential for prompting action based on the insights produced during the Delphi inquiry, which is by definition supplemental to the group's normal interactions. In these cases you should consider face-to-face group processes. Other techniques (such as program-planning method—PPM, nominal grouping, multiattribute utility assessment) may be as productive, require less time, and have more spillover benefits than a Delphi. It is also imaginable that a group of strangers could sufficiently internalize instructions and materials that would be dramatic enough to induce an episodic style of interaction for an inquiry. Such an inquiry would have to be virtually a simulation. The substantive insights produced by a small purposive group can be expected to be highly specific to their context. The insights produced by such groups regarding the personal and emotional dimensions of the topics are probably not generally indicative. In fact, it may be necessary to create a simulated purposive group in order to probe the interpersonal and emotional dimensions of circumstances that are themselves the subject of conjectures or competing proposals.

Variations on this mode of Delphi inquiry are limited because alteration in the premises usually induces the group to adopt a new reality resulting in a less

highly bonded mode of interaction. The most interesting results are produced in instances when the group's strongly shared reality is collectively modified in response to new insights and even to incorporation of new members. However, this collective modification of reality is not always completed successfully and it can be expected several realities will prevail. This can result in dissolution of the group, a breaking down of the group process, or the production of highly emotional and idiosyncratic contributions.

Events, often with more form than content, distinguish the interaction of groups which:

- are made up of individuals who see themselves as having a commonality worth recognizing, but not becoming related or involved on other matters.
- center their interests on the specific issues of their commonality and its continuance (survival) and enhancement.
- are guided in their interactions by "the way things are done"—hierarchical considerations and formal divisions of responsibilities within the group.
- make choices to allow greatest individual diver[illegible] while maintaining the essential solidarities that define the group.

The inherent one-dimensionality of the interactions of these groups exhibits itself most clearly in the importance of agenda-setting. Whether highly informal—"wait, should we be discussing this"—or employing a very formally maintained docket, these groups find it essential to screen topics for relevance to the organizing commonality. Issue focus has become more necessary as society becomes more diverse, in order to keep an affiliative group from fractionating into several puposive-type groups, each of which would maintain multidimensional commonalities. To survive and grow, however, such affiliative groups usually have to deal with threats and opportunities that are on the perimeter (if not outside) the group's reason for being. This means that most of the messages exchanged can be assigned to one of two distinct classes: *proforma* statements necessary to group maintenance and expository statements intended to modify the group's prevailing notion of their common interest. There will be very few exchanges about things that affect them as individuals or alter the group's relationship to the society.

Delphi panels who think of themselves as affiliative groups tend to proceed in two-step responses to the items in the initial inquiry. First, they make narrow and direct responses to each item. Second, they have additional considerations. The whole panel is usually asked whether these should be dealt with. If the panel or some substantial subgroup agrees that the additional items should be added, then they are taken up. Consequently, much of the reality that is created has to do with the meaning of considering certain topics and not the meanings derived from actual consideration of those topics. A Delphi should

never be attempted with a "single-purpose" group until they have agreed that there is something that needs their attention which would not take care of itself.

Affairs are the interactions of groups characterized by:

- being asked to consider items that apparently are of interest.
- having an abstract categorical definition of their membership.
- continuing attempts to give direction and create meaning for the interaction impelled by recognizing the insufficiency of the group's label to serve as an adequate definition for their involvement.
- a tendency for speakers to become "spokesmen," to indulge in formalism, and to imply collective interests for the group.

The beginnings of these groups are abstract categorizations which have meaning for the inquirer or the explainer, but seldom for the group itself. So when asked almost anything, these groups scramble in several directions in search of cohesive interests. The direction of a question or inquiry presumes a meaningful response. A variation of this occurs when the inquiry is directed to explore differences within the group. In this instance the inquiry presumes distinct factions. The group will often develop these "factions" in formulating responses, but the factions are transitory. A corollary dynamic in the interactions of these groups is the prefacing of comments with remarks which either attempt to identify the speaker with the whole group or a major faction, or will uniquely distinguish him and his contentions. These interactions tend to be characterized by a series of declarations—"Run it up the flagpole and see who salutes." Occasionally in short interactions, and eventually in almost all, a series of declarations will coalesce a faction (occasionally including the whole group), which will concentrate thereafter on the construction of a position and purpose. Once constituted, such factions interact as large affiliative groups, usually in a way that produces schisms and dissolution.

Because researchers, administrators, and most of the rest of us have been almost conditioned to think of society as divided into distinct kinds of "labeled" groups, the response of abstractly defined groups (e.g., doctors, women, radicals, educated) is frequently sought in inquiries which end in producing results that do not satisfy the inquirer. One way to remedy this is not to expect a coherent position to emerge, but to look for a pattern of diversity in the group's assertions and the ways in which these points are qualified. These groups cannot examine issues that apparently concern them, because, to paraphrase Gertrude Stein's description of Oakland, "there isn't any 'they' there." Instead, such groups can be used to design product or service "lines" or to formulate arrays of options that seem complete—covering most of the considerations. Also, these groups can be used to define and map the factions implied by a number of related specific proposals for action. Panels constituted in this way are used all too frequently. As an alternative, consider making up panels that have two or three distinctive subgroups and manage feedback selectively.

Occurrences are what goes on with groups where:

- the membership is made up of individuals who believe they are acting for the interests of the society in general as they see it.
- topics serve as occasions for participants to expound their general ideologies.
- most messages refer to particular cosmologies about the way things are supposed to be or are appended to statements that describe world-views.
- the envisioned outcome of a set of interactions is a collection of more widely agreed-on precepts that will guide the group, its constituents, and, if possible, all others.

All panels on occasion become forums to prescribe for the ills of society. These groups virtually insist on this mode. In most instances, this is counterproductive. This is because in most circumstances it is necessary to produce a fine-grained, narrowly bracketed reality to support specific actions. The reason for arranging this kind of a panel is to obtain a broad consensus, but not on everything. Since nothing is ever *done* in general, but only in a particular instance, introduction of situation-specific facts and considerations can help to focus the panel on agreeing about society's interest in the instance at hand, instead of on society's interest generally in things of this kind.

Another reason to arrange a broadly based general panel is to develop a contextual mapping that would describe the overlapping large-scale realities which underlie different parts of a society's response to any complex issue. Here it is useful to inquire about "for instances" that are likely to be seen by different elements of society as analogous to the situation being inquired into. "Well, it's just like thus-and-so..." and similar phrases are powerful connectives to ready-made meanings for a proposed idea or approach. These loose, even unrational, associations tend to bring along existing emotional loadings which panelists in this mode tend to deny. In a similar way, apparent metaphors (e.g., like David and Goliath) and apt personifications (e.g., rape of the ecology) should also be explored. This can be done by asking about which ones suggest themselves or provide several examples to be selected from, modified, or added to. Identifying analogies and metaphors is a more fruitful way to define subtle differences in global constructs than asking individuals to examine their core values and beliefs and describe them directly. While there are problems with interpretation of these complex symbols, the interpretations can be iterated for refinement, and also for validation. Since the "picture" is changing, and "observation" using a Delphi inquiry is not a snapshot, there is a gross limit on the fineness of the picture of a whole society that can be produced. But, even faint shadows on pale gray are helpful in many situations.

Do not do a Delphi with representatives or spokesmen for society when you want information for particular practitioners; use experts "about society" who are respected by the practitioners. In fact, you may want a split panel

containing both experts and practitioners to create a joint reality shared by both that can be potentially extended to society.

What I Meant to Say

Most of the foregoing is tentative and much of it highly abstract. It is even embryonic. I share it for whatever use you can make of it and apologize for its density. There are obviously many hypotheses here for research on the process of Delphi interaction that go well beyond the issue of "accuracy" or "reliability." My only suggestion is that you try to supply or create examples from your own experience that seem to explicate these generalities. I attempt to summarize the main points below. In the next—and final—part of this article, I will offer a few pragmatic suggestions for the designer—monitor of Delphi inquiries. Obviously, not all of these strategies and techniques should be attempted simultaneously.

Summary

(1) Each Delphi interaction produces a shared reality which is initially formulated by the panelists from their expectations and the style of presentation used in the initial materials; this particular reality is elaborated and modified by the succeeding interactions.

(2) During the process of interaction the panelists' responses can be expected to deal with personal esteem, group self-concept, and relevant world-views as well as to convey substantive ideas, forecasts, and estimates.

(3) Patterns of styles of interaction can be fostered, retarded, even transformed in order to produce results that will have greater insight, more usefulness, higher impact.

(4) The size and shape of the reality within which things come to be viewed is more important than the specific substantive descriptions produced by the panelists.

(5) The believability and significance to the user of the results of a Delphi inquiry depend as much on the user's perception of the clarity, compellingness, and fit of the reality implied and possibly defined by the results as on the perceived quality of the information.

Some Design Considerations

I shall now briefly focus on some design considerations to help you develop a better Delphi and orchestrate the interaction. These suggestions do not constitute a checklist. I hope they will trigger some ideas and preparations that otherwise would not have occurred to you. Each could separately be the subject of a much longer discussion. Supplying these discussions is left to you, the reader, when you pursue particular applications.

Making a Delphi Context Specific

The conceptual productiveness of the interplay of intentions and circumstances is lost in many Delphi processes because of the lack of concreteness in the context. The mental life of most individuals involves attribution of an intimately known context, even when this context is the world of abstract ideas. However, the subjects of most inquiries, including Delphis, are stated in terms of general propositions; those things that apply only to an instance are defined as being of no significance for action (those things yet to be determined, decided, or done). So the difficulty lies in the extraction of generalizable propositions from particular instances.

I can offer two approaches; there are undoubtedly more. The first approach is to develop a concrete example with detailed features that are typical of a more general case. Another source of specific contexts is to create hypothetical constructs which make a distinct break with most present circumstances and are almost fanciful, e.g., on another planet. Because the hypothetical construct will differ in important respects from the panel's conventional circumstances, it can often provoke new thinking, retard discussions of routine experiences that are rooted in presently accepted and unquestioned conventions, and create an unclaimed and common experience base that can facilitate collaborative (*vs.* competitive or compliant) interaction. Emergencies and other kinds of special circumstances can also be used as hypothetical constructs within which possibilities and insights can be developed for later transfer to the more general everyday circumstances.

Once a case is selected, proceed with the Delphi, focusing on the detailed case initially. Then shift the inquiry by asking the panel to create generalizations—possibly proposing some yourself and iterating the ones they proposed. You can next elaborate these general propositions by asking,"What other circumstances do you know or can you imagine might prevail, and within each would these propositions be valid?" The second approach is to develop several special cases for initial inquiry using subpanels. Later, the results of each subpanel can be shared with the whole. Then, ask for general propositions. What is forthcoming can then be refined and elaborated with examples.

The Domain of Time as a Context

Another shortcoming of "looking at things in general" is that time is assumed as an undifferentiated flow—that one minute or hour is as good as any other. If asked, we all would acknowledge perceptual differences—time dragging or racing. But, unless our attention is specifically directed to it, we tend to

overlook structural distinctions in time. The structural features that limit the allocation of time are of increasing importance as time becomes a scarce resource. When inquiring "What will people do?" "How can resources be used?" "Which group is being designed for?" the time domains become critical concepts. Time domains can be categorized or partitioned in many different ways, as can any continuum. Conventions in the dominant culture have created workdays (and, increasingly, work evenings and nights), regular times off, lunches, weekends, holiday weekends (recently added to by Congressional action), annual vacation times, "the holidays," special events, and so forth. There are also times in the so-called life cycle—youth, teenage, young married. Time domains can be treated as the subject of policy and design, for example, staggered work schedules and new work patterns (4 ten-hour days, 6 twelve-hour days with a week off, etc.) You can also invent new "times" for things, like the enculturation of daily exercise periods. Figure 8, below, deals with several of these considerations and relates them to the selection of market opportunities.

Demand for recreational services is limited more by the time consumers have available for their pursuit than by the expenses involved. Time is experienced in periods that have each come to have their own meaning and expectations. To create new markets for recreational services, the meanings of particular periods or domains can be altered and new periods can be created and given meaning.

Round 1: *Suggest new time domains you think could be developed and indicate how existing domains could be revised or differentiated to create new markets for recreational services.*

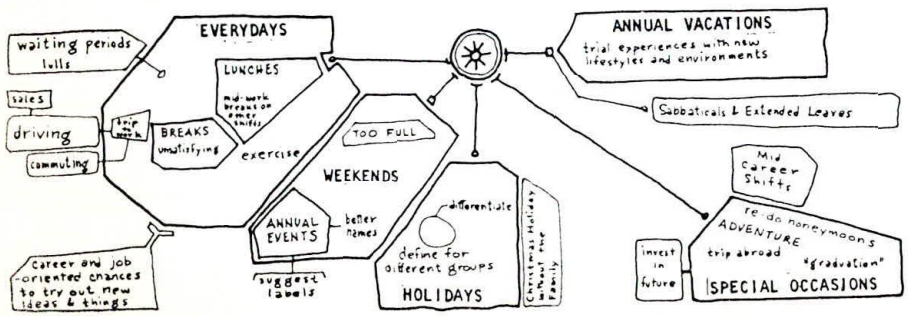
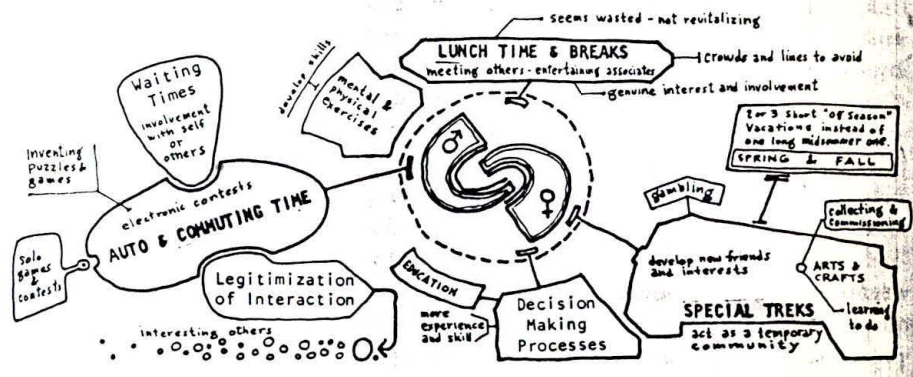


Fig. 8-A

Round 2: *Identify time domains where additional involvements for working adults without children would be most welcome and suggest recreational pursuits.*



Round 3: *Select market opportunities you believe are the most attractive and suggest specific types of products and services and actions to create expectations and aggregate markets.*

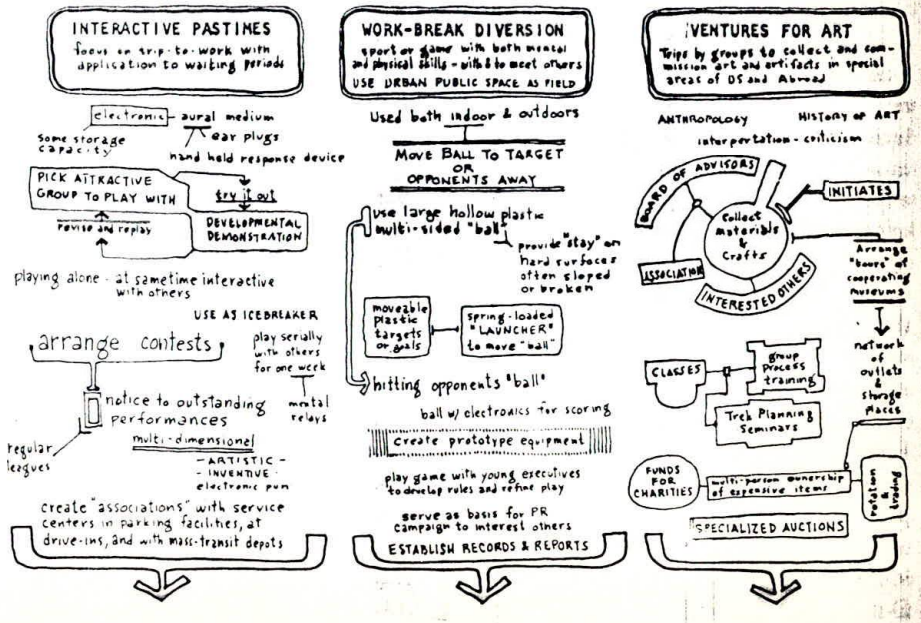


Fig. 8 Time domains and market opportunities.

Getting a Product Out of Results

Every Delphi inquiry can be expected to produce results of some kind. Usually, if they are thought about at all, results are seen as things to be "captured" from their existence "out there somewhere" by the Delphi process. Instead, try to visualize, or think about, the results of a Delphi as being created by the interaction. Just as the important properties of a building are not simply those of the sum of its "stones," the results of a Delphi are not just the individual items produced in the interaction but the reality comprised by the whole. I believe that a Delphi will be more productive if the monitor sees his role as producing results and not as "surveying" things that are already there.

Here are some more design tips:

Creating panels is usually the first task, unless they are a given—as with an organization that wants to study its own future in a participating manner. Even then you may want to augment the group. Three kinds of panelists are ingredients for creating a successful mix: *stakeholders*, those who are or will be directly affected; *experts*, those who have an applicable specialty or relevant experience; and *facilitators*, those who have skills in clarifying, organizing, synthesizing, stimulating... plus, when it seems appropriate, individuals who can supply alternative global views of the culture and society. The proportion of a panel from each category should be tailored for each application. *Note*: There are almost never enough women on panels.

There are no general rules of thumb for creating panels. For example, where options and interests are clear but acceptance of direction and action is fractionated, stakeholders might predominate. If it is clear who has to act, but not clear how, a heavy salting of experts may be best. Where issues, relationships, and values are unclear, a preponderance of facilitators may prove most useful. Pay attention to the minority as well as the majority views expressed by each type of panelist. Also, watch that panelists contribute as you expect—experts, in particular, drift into acting as stakeholders in aggressive moments and facilitators in passive ones.

After abstractly thinking about the panel, it comes down to names. Most of the time you will not know everyone you want. Sometimes you can start with a small group of potential panelists and begin by discussing possible names and searching as a group for interesting and appropriate candidates. This is a better strategy than searching lists of relatively unknown names by categories. Frequently, staff members of professional societies and other "people brokers" can be utilized to suggest panelists. Often the most fruitful part of a Delphi process is assembling a panel, not simply for the particular effort, but because it enhances your contacts. At least, try to develop two alternative tactics for identifying panelists and consider their trade-offs before embarking on the Delphi.

Stimulating response of any kind from quality panelists is not easy. Getting quality inputs is even harder. Few people like "questionnaires." And, from their point of view, engaging in abstract speculation with people you do not know or whom you want to impress about a subject that is not central to your interest (and where your performance is unrelated to your survival) is hardly compelling. Motivation requires more than a good cause, a pep talk, and thanks. If prospective panelists tend to be uninterested, find a "worthy" or prestigious sponsor, or make participation of significant publicity value. Then a token payment or "honorarium" will stimulate interested responses. Payments alone, even of fairly sizable sums, will not assure quality participation. It is difficult to pay panelists based on performance since it is difficult to agree on what is expected; panelists generally believe that "what I do is what's enough." Attractive and potentially stimulating peers are probably the most powerful incentive for participation. Consider gifts and "in kind" rewards for participants, because often a sponsor can provide these at modest cost—particularly travel. Still another way to organize participants is to employ a two-step approach: suggest that key participants use their staff, students, or others they can induce to participate to formulate the specific responses. The key participant then collects, reviews, and submits them for each round. Here publicity and recognition of importance by other individuals are powerful incentives.

Once initial participation is secured, the next important consideration is the quality of the materials, their tone, style, and presentation. Use lots of color. Give the materials style. What you send out reflects the significance of the inquiry and the value that is placed on it. Use emotive language and vernacular expressions to engage panelists and convey importance of results—not another abstract study. Detailed situational descriptions, mini-scenarios, bits of conversation—all help. Other media besides print can be used effectively to make response easier and more scintillating—even speedier. These include tape cassettes and even local "interviewers" in cases where the Delphi sponsor or manager has a large potential manpower pool that could be employed at low cost. Don't forget to pay attention to creating an audience for the results and communicating this to panelists. Examples of good and bad responses are also useful, particularly for discouraging stereotypical remarks and obvious but unhelpful "insights."

Orchestrating interaction requires attention to details in the panelists' responses and a feedback of an overall movement, countervailing forces, or whatever macro-observations are appropriate to describe what seems to be going on between and with the individual items. Highlight divergence and consensus, even when both are true for different sets of responses to a quirk item. Be sure to cheer at least one item from each panelist—everyone needs to be appreciated. Introduce ambiguous factors if discourse focuses on individual items when you would rather explore relationships between items. You can also "stipulate" certain constraints when particular items have evoked massive, but trivial discussions of immaterial distinctions. Where appropriate, supply tenta-

tive theoretical constructs and cosmological frameworks. Indicate the way responses are being categorized for review and whether your interest centers on enriching the decision environment or identifying and refining options for consideration. Also, you can employ cross-impact analysis (see Chapter V), comparisons with analogous situations, and other techniques to provide a basis for evaluating responses in a different light. Again, at the risk of overemphasis in this essay, as interaction progresses, share with the panelists the reality construct(s) you believe they are referring to and shaping by their responses.

Interpretation and summation of responses is never complete, because the panelists do not send in their "heads" but only their responses. A complete record of the interactions is unwieldy and diffuse and difficult to use. It is important to begin "interpreting" responses during the interactions, even at the start. This makes interpretation subject to review by the panelists and can include their refinements, which I have found most insightful. Summation can be made from several points of view—using the one best for each major consideration in the inquiry. It is often useful to point out nil findings, omissions, and ignored items. Try to capture and describe the reality that was negotiated by the panelists and the monitor, because this provides a perspective for understanding and indicates the application foreseen for the results.

Communication of the results may not be felt by some to be a legitimate concern for researchers. Findings, it is supposed, have their own importance. However, the process of acquiring "understanding" carries with it an impetus for action that is not conveyed by presenting the conclusions alone. Most Delphi users, I believe, intend to influence the formulation of policy or the making of decisions. Doing this requires attention to the communication process. Well-organized, lucid, attractively presented reports help—but they are not enough. To augment the report you should attempt to: (1) create involvement—one way being to include the intended users or members of their staffs as panelists in the Delphi, or else to engage users in parallel deliberations that focus on the same issues, infusing from time to time interim considerations developed by the other Delphi panel; (2) generate interest—this usually means fanning respectable controversy and building a climate of expectation for, and awareness of, the Delphi inquiry among those who are important to the users, particularly colleagues and constituents; (3) present the results interactively—this can be done by organizing a simulation exercise where the panelists, the staff of the Delphi monitor, or some of both take roles that are significant in the user's operating environment and then enact a series of situations that can demonstrate which Delphi results are applicable, and when and how they might be used. It may be useful to define a communication process as part of the initial conceptualization of the Delphi design so that the implications for its communication are anticipated in the way the interaction of the panel is orchestrated.

I cannot imagine that all the recommendations in this Chapter will prove unassailable or always useful. I hope they will spark some interest and provoke a few insights. This will require a lot more of you than of me. I have benefited in putting this down, but you will have to work to make it apply to your situation in order to profit. To quote Henry James: "Many things have to be said obscurely before they can be said clearly."

III. A. Introduction

HAROLD A. LINSTONE and MURRAY TUROFF

In this chapter we sample the rich menu of applications. The purposes of the Delphis are as varied as the users. Seven authors focus on specific planning tasks in the areas of government and business. Additional studies are also sketched in this introduction ("Comments on Other Studies").

Government Planning

The four articles covering this field address national, regional, and organizational planning problems. Turoff deals with the basic concept of the Policy Delphi and reviews several efforts of this type. Ludlow's concern is resource management in the Great Lakes region and a major aim is to establish improved communications between technical experts and interested citizens. Jillson focuses on the use of Delphi as an integral instrument in national drug abuse policy formulation, operating from three distinct perspectives: "top-down", "bottom-up", and "issue oriented". Jones explores priorities in System Concept Options for the U.S. Air Force Laboratories, with emphasis on comparing views of four in-house organizations competing for funds.

All four move beyond the use of Delphi as a forecasting tool and stress its value as a communications system for policy questions. A policy question is defined here as one involving vital aspects, such as goal formation, for which there are no overall experts, only advocates and referees. Its resolution must take into consideration the conflicting goals and values espoused by various interest groups as well as the facts and staff analyses. It should be clearly understood that Delphi does not substitute for the staff studies, the committee deliberations, or the decision-making. Rather, it organizes and clarifies views in an anonymous way, thereby facilitating and complementing the committee's work.

Whereas Turoff's panelists constitute a homogeneous group, Ludlow seeks to establish a communication process between the potential users of new knowledge and a team of interdisciplinary researchers. He raises a point which is of concern for Delphi studies generally. The probability used is of the personal or subjective type; it can be interpreted as a "degree of confidence". Scientists and engineers are brought up on a different kind of probability—frequency of

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

currence, i.e., the limit of the ratio of the number of successes to the total mber of trials as the latter approaches infinity. Thus the frequency type of bability assumes repeatability of the experiment (e.g., tossing a coin). But e subjective probability has meaning even if an event can occur only once. A xing match is a one-time event; the odds usually associated with such a tch indicate the "degree of confidence" in the outcome on the part of formed bettors. Both definitions are mathematically valid and have been used develop distinct probability theories. Businessmen intuitively use the "degree onfidence" concept and therefore have no built-in resistance when faced h it in Delphi questionnaires.

Ludlow also presents an evaluation of Delphi by the three participating ups—technicians, behaviorists and decision-makers. Not surprisingly the tter prove to be the strongest proponents of the technique. They are, after all, one group which must regularly seek a consensus and usually has to make isions on complex issues without adequate information.

Miss Jillson's article is a progress report on a study designed to develop drug se policy options, explore the applicability of the Policy Delphi to questions ocial policy generally, and determine the practicability of using it on both "as-needed" and "on-going" basis (i.e., indefinite duration). The partici- ts include researchers, administrators, and policymakers—both in the field in impacted areas (e.g., police chiefs). A unique feature is the use of the ree perspectives. In the "top-down" approach, the objectives for the next five rs are emphasized; the "bottom-up" approach deals with factors which trol transition between various states or levels of drug use and employs a trix format; the "issue-oriented" approach crystallizes statements of policy es in "should/should not" form.

Jones' Delphi reflects the typical consensus or Lockean oriented approach in design to gain consensus among representatives or organizations subject to erent pressures in their competition for limited financial resources. He uses senior managerial and technical personnel (both military and civilian) presenting most departments in the four organizations. Different organiza- al viewpoints are apparent although significant self-interest biases are not tectable. This effort contrasts very nicely with the Kantian nature of Ludlow, Hegelian approach of Turoff, and the mixed Kantian-Hegelian work of on.

iness and Industry

a corporate environment Delphi fills several roles. Bell-Canada's Lawrence lists three: educational device for senior management, environmental nd background material for technological planners in research laboratories, trading material for use with planner-counterparts in other organizations. the corporation is large and diversified it may have the analytical staff to

run the study and the expertise to form the panels "in house". TRW used 140 of the "the most imaginative and creative members of TRW's technical staff of more than 7,000 graduate scientists and engineers". It should be noted that in the hierarchical environment of a business, the rate of response or participation tends to be higher than average. A university professor may feel no compunction about ignoring questionnaires or giving perfunctory or dilatory answers; an employee has stronger incentives to cooperate in a company exercise. Goldstein's experience in this regard is echoed in numerous other cases.

The ability to conduct a Delphi without bringing the respondents together physically is another advantage in the large corporation with units spread over a wide geographical area (e.g., multinational corporations). Overseas personnel can be drawn into a Delphi with relative ease and at minimal cost.

Day covers four Delphis performed by Bell-Canada's Business Planning Group: Educational Technology, Medical Technology, Business Information Processing Technology, and Home Communications Services.

In cases where the corporation does not have the expertise in either the subject of the forecast or the Delphi procedure, it may turn to a professional consulting organization. Enzer's article describes a Delphi on the subject of plastics undertaken for a client by the Institute for the Future. The field of materials for the future is a particularly difficult one for the forecaster. First, the number of possibilities is overwhelming (e.g., tailormade plastics). Second, in a hierarchy (e.g., relevance tree) which has metasystems at the top, followed by systems, subsystems and components, materials are close to the bottom. This means that a given material can branch upward in many ways, i.e., it can find use in a multitude of old and new systems by 1985. Third, each substitution of a new material for an old one may prove sensitive to slight variations in the relative prices, with major deviations resulting in the 1985 forecasts. These factors strongly suggest the desirability of using at least two entirely different methods of forecasting to provide a reasonable degree of confidence.

Nancy Goldstein gives us a hint of this in her article on a Delphi covering steel and ferroalloy materials. She compares the Delphi results with those of a conventional panel study conducted simultaneously. There is considerable agreement in the forecasts. Ideally one would seek a comparison between methods which are more radically different than a conventional panel and a Delphi. However, this particular comparison makes one startling methodological point: the conventional panel study brooks no uncertainty and no areas of disagreement in its forecasts; it presents a definitive view and a set of conclusions. The Delphi, on the other hand, reflects the uncertainties and highlights the differences among its participants. It is more concerned with exploring minds than setting down precise recommendations (cf. Scheele, Chapter II, C).

The Delphis by Day, Enzer, and Goldstein are largely Lockean in nature. However, the use of panels of differing backgrounds by Day has Kantian aspects, and Goldstein suggests Hegelian overtones.

Another use of Delphi has recently evolved in business in connection with risk analysis.[1] It concerns the uncertainties associated with new projects or investments. Normally decisions must be made in the absence of adequate information. The potential market for a new product is uncertain and the development costs may exceed the engineers' estimates. Marketing personnel in an operating unit of the corporation frequently exhibit a glowing optimism which neglects to credit the competition with high intelligence or quick reaction capability. Engineers tend to assume that the cost of a complex product is a linear function (i.e., sum) of the cost of the components which comprise it. They neglect the interactions which result in nonlinear behavior: the total cost is much greater than the sum of the parts. Thus the cost is grossly underestimated. One recent study of a large number of defense-oriented development projects indicated approximately a 50 percent chance of 50 percent cost overrun.[2]

Delphi may be used to advantage to provide input to the risk analysis. The most critical part of such analysis is the subjective probability distribution assumed for the uncertainties. Delphi can serve to probe the views of personnel connected with the project as well as outsiders (e.g., corporate offices or other units), senior executives as well as junior engineers and scientists. The anonymity is particularly valuable in a highly structured environment where individuals may feel constrained in expressing their own views.

Comments on Other Studies

Many applications of Delphi are carried on as integral parts of planning projects or as staff work of a proprietary nature. Therefore, a considerable amount of very good Delphi work has not been published in the open literature or in a form adequately explaining the process used. Before proceeding to the separate application papers it is worthwhile to note a number of Delphi studies of a unique nature which are not yet otherwise documented explicitly for those interested in the technique itself.

The Delphi method has been applied extensively in the medical area. The initial work by Bender *et al*[3] was largely of a straightforward forecasting variety. However, a number of Dr. Bender's Delphis did deal with estimating

[1] See D. B. Hertz, "Risk Analysis in Capital Investment," *Harvard Business Review*, Jan-Feb. 1964, p. 95, and D. H. Woods, "Improving Estimates That Involve Uncertainty," *Harvard Business Review*, Jul-Aug. 1966, p. 91.

[2] S. H. Browne, "Cost Growth in Military Development and Production Programs," unpublished, Dec. 1971.

[3] Bender, Strack, Ebright & Von Haunalter, *Delphic Study Examines Developments in Medicine, Futures*, June 1969. George Teeling-Smith, *Medicine in the 1990's*, Office of Health Economics, England, October 1969.

the necessity and desirability of potential medical research accomplishments. And it did not take very long before the application was broadened to include unique objectives other than future projections.

Dr. John W. Williamson, of John Hopkins University's School of Hygiene and Public Health, has utilized Delphi extensively for estimating historical data. Typical questions deal with determining the incidence of a given disease and the estimated rate of success in utilizing various treatment methods. This is, of course, an area where current reporting practices do not give reliable data owing to differing standards across the country and the effect of multiple complications, e.g., death due to pneumonia while ill with cancer. Usually Dr. Williamson would ask respondents for their best estimate of a number, then a low and high value which would *shock* them. Also requested would be an estimate of their confidence in the estimate and a statement whether their estimate was based upon a particular source, such as an article they had read. In a number of these exercises questions were asked which dealt with the results of unpublished new clinical studies. In this manner one could observe how well the Delphi panelists actually did on part of the exercise and utilize this insight to gain an impression of their capability for providing answers to the rest of the exercise.

An excellent and, perhaps, classic example of this is a study Williamson did at the Philips Electric Corporation Plant in Eindhoven, Holland, in 1970. Approximately 50 doctors who are involved with the company's medical program for the 36,000 employees participated in the Delphi. The first part of the Delphi asked the physicians to estimate the percent of male employees absent from work due to sickness during differing intervals of time. The population was further divided by young, old, blue collar, white collar. This required sixteen estimates from each doctor. When the real data was collected from the computer files three months later it was found that 12 of the 16 estimates were within 10 percent error and the other 4 within 30 percent error. Of course, this represented only a small portion of the exercise as the real objective was to determine what effect various potential changes to the health care program would have on the absenteeism rate. However, one could see that the physicians involved had a good feel for the situation as it existed. Also it was possible to examine how well various subgroups did, e.g., general practitioners *vs.* specialists. Dr. Williamson has conducted four major studies of the above type (involving validity checks) with a total of approximately 200 respondents over the past four years.

Professor Alan Sheldon of the Harvard Medical School, together with Professor Curtis McLaughlin of the University of North Carolina Business School, did a Delphi in 1970 on the Future of Medical Care. A unique feature of this Delphi was the process of combining the events evaluated by the respondents into scenarios in the form of typical newspaper articles. The respondents were then asked to propose additions or modifications to the

scenarios and give their reaction to the scenario as a whole. This concept of utilizing the vote on individual items to group events into scenarios classed by such things as likelihood and/or desirability has become a standard technique.

Also with respect to scenarios it has become fairly common to provide the respondents on a forecasting Delphi with a scenario or alternative scenarios providing a reference point on considerations outside the scope of the Delphi but having impact on the subject of the inquiry. For example, in forecasting the future of a given industry the respondents might be given a "pessimistic," "optimistic," and "most likely" scenario on general economic conditions and asked that their estimates for any question be based on each alternative in turn.

While there have been a number of Delphis on the general future of medical care, a recent Delphi by Dr. Peter Goldschmidt of the Department of Hygiene and Public Health, Johns Hopkins University, dealt with the future of health care for a specific geographical entity, Ocean City, Maryland. The problem in health planning in this case is the tremendous influx of vacation people in the summer months. In order to examine the future growth of the Ocean City area and its resulting medical needs, it was felt the Delphi panel should include individuals who resided in the area and simultaneously worked in endeavors related to the mainstream of the local economy—recreation. Therefore, the Delphi involved long-time residents, hoteliers, bar owners, real estate dealers, and civic officials as well as the usual "experts" such as the regional planning people from local government and industry. This widening, or broadening, of the concept of "experts" to that of "informed" is becoming quite customary in the application of Delphi. In this particular Delphi, Dr. Goldschmidt was able to check the "intuition" of his respondents by comparing their estimates on vacation populations in Ocean City currently with estimates he could analytically infer from the processing load history of the sewage plant that serves the area. As in Williamson's case the results were quite good.

A superb example of the Delphi technique was carried out by Richard Longhurst as a master's thesis at Cornell University.[4] The Delphi attempted to assess the impact of improved nutrition, family income, and prenatal care on pregnancy outcome in terms of birth weight and the resulting I.Q. and intellectual development of young children. The resulting output were of a form useful for incorporation into cost-benefit analyses of government programs to improve the nutrition of pregnant mothers and young children. This is, of course, an excellent example where data exist to indicate that malnutrition in the mother or young child has some degree of impact on the long-term intellectual capabilities of a child. However, this evidence is not of direct quantitative utility to the type of analyses an economist would like to perform in evaluating a government program.

The respondent group was divided into two subpanels. One in the area of

[4] Richard Longhurst, "An Economic Evaluation of Human Resources Program with Respect to Pregnancy Outcome and Intellectual Development," M.S. thesis, Cornell University, Ithaca, N.Y. December 1971.

intellectual development comprised experts in child psycholog
ment; the other, in the area of pregnancy outcomes, was composed
pregnancy, nutrition, and medical care. The group was given a sp
of low-income mothers in a depressed urban area as the population
concerned with. This was a real group on which a good deal of
socioeconomic status were available. In the first round the respondent
asked to sort out the relative importance of environmental components
might be manipulated by the introduction of a government program. T
second round presented a set of feasible intervention programs which related t
the factors brought out in the first round. They were then asked to estimate for each program the resulting incidence of low birth weight and the average I.Q. score of 5-year-old children resulting from the pregnancies under the program. They provide the baseline data on these parameters for the target group as it currently existed. Certain programs were estimated to reduce the incidence of low birth weight from 15% to 10% and to raise the five-year I.Q. scores from 85 to 100 points. The shift in I.Q. can then be used to shift the average education and earning power of the children when they are adults. This then is translatable into dollar benefits that can be used to compare the merits of alternative programs in this area. The Delphi itself involved respondents for three rounds and questionnaires were tightly designed to take about fifteen minutes to fill out.

The area of trying to translate scientific knowledge into an informed judgment on evaluating and analyzing decision options is clearly a potential one for the Delphi method.

Another effort in health care planning is the work of Professor David Gustafson at the University of Wisconsin. This work has been tied into the Governor's Health Planning and Policy Task Force effort. One of the Delphis Professor Gustafson conducted dealt with delineation of the current barriers to the performance of research and development in the health services area—the rather interesting topic of trying to clarify what the real "problem" is. The respondents were asked to delineate barriers of three types: (1) solution development barriers; (2) problem selection barriers; (3) evaluation barriers. For each barrier the group developed comments, implications, and possible reactions or corrective measures. A vote was taken on the significance of each barrier. This was an excellent example of utilizing Delphi to try and isolate the significant part of the problem. Very often, in planning areas, preconceptions by one individual lead to tremendous efforts on the wrong problem. The specification of a particular problem usually predetermines its method of investigation and at times its conclusions.

The use of Delphi for regional planning has probably become popular because of the feeling that there is a necessity to establish better communications among many individuals with diverse backgrounds.

In this area a significant number of Delphis have been conducted by various Canadian government agencies such as Health and Welfare, Department of

Public Works, Department of the Environment, and the Postal Service, to name a few. Most of these are being done by internal staff and very often they tend to be short, focused on a very specific issue, and require a diverse backgroud of respondents. A good example is one done in 1974 by Madhu Agowal on "The Future of Citizen Participation in Planning Federal Health Policy." The Delphi sought to explore and delineate specific options for citizen participation and to determine the consequences of such programs.

There has been very active use of the Delphi in the educational establishment and a survey of that work may be found in an article by Judd.[5] Curiously almost all educational Delphis have been confined to administrative matters and hardly considered as a teaching tool. It is not surprising that educationalists are enthusiastic about the method. There is a high degree of participative planning in higher education. Authoritarianism is eschewed to such an extent that anarchy sometimes results. There is also an entrenched bureaucracy which feeds on well-structured procedures and questionnaires of all kinds.

Delphi is used for several aspects of administrative planning: general goals, curricula, campus design, and development of teacher ratings and cost-benefit criteria. Judd describes many of the problems encountered in the use of Delphi in this environment.

However, to find a clue to what may prove to be the most serious difficulty, we must turn to the conclusion of a (non-Delphi) survey of school administrators conducted recently by R. Elboim-Dror[6] in Israel on the subject of education in the year 2000:

> The lack of creative imagination as revealed in this study, the limited number of new alternatives and innovating ideas expressed by the subjects, and especially the students (of school administration), are a serious sign.

In such an atmosphere Delphi can be as barren as most of the paperwork which traditionally suffocates educational bureaucracies. When the educational field begins to see Delphi in the deeper context discussed in Chapter II, when it starts to consider Delphi as an educational tool as well as a planning tool, then it may be able to escape this trap.

If there is a single message in the philosophy of Singer (see p. 33), it is that the past and the present are often as hard to interpret, or conjecture about, as the future. It is therefore not surprising to see the Delphi method applied to these areas as well as to the future. A recent Delphi devoted to examining the past is an exercise by Professor Russell Fenske of the University of Wisconsin. This involved about twenty-five leading researchers in the field of operations research who utilized the Delphi to review the "state of the art of industrial

[5] R. C. Judd, "The Use of Delphi in Higher Education," *Technological Forecasting and Social Change*, 4, 173 (1973).

[6] R. Elboim-Dror, "Educators' Image of the Future," paper presented at the Third World Future Research Conference, Bucharest, September 1972.

operations research": Each participant could propose significant contributions to the literature in the areas of theory, applications, and economic impact. The group then voted on each item for importance and impact. Also gathered were brief comments on the significance of an item and suggestions for important areas of future research.

Michael Marien has used the Delphi process to elicit from a panel of 14 futurists the most significant books (a "hot list") on the future. Considering the volume of material being produced in most professional areas, these particular applications of Delphi are likely to become quite popular in the future.

A similar concept has been applied in some corporate or organizational uses of Delphi, where the study examines historical performance or factors that have affected the market place for a particular item. The objective is usually to focus on 50 to 100 key items out of hundreds of candidates so that a concise summary of the historical perspective can be prepared for management. This is largely the process of getting a group to filter out the signal of real information from the multitude of communications or noise that may exist on a particular complex topic. This concept is very similar to what Professor James Bright refers to as the monitoring function in technological forecasting, and an excellent example of a Delphi on "Events Leading to the Limitation or Elimination of the Internal Combustion Engine" forms the basis for an exercise in his recent book.[7] The example, based upon a Delphi conducted in 1969 by a chemical company, was, to the best of our knowledge, the first which dealt exclusively with evaluating the past.

A much deeper systemic study of the past is envisioned in the "retrospective futurology" approach which applies dynamic programming to historic societies such as the city-state of Athens. The "hyper-sophisticated polling of experts" mentioned by Wilkinson in conjunction with this concept[8] strongly suggests the Delphi method.

[7] James Bright, *A Brief Introduction to Technology Forecasting: Concepts and Exercises*, Pemaquid Press, Austin, Texas, 1972.

[8] J. Wilkinson, R. Bellman, and R. Garaudy, "The Dynamic Programming of Human Systems," Occasional Paper, Center for the Study of Democratic Institutions, MSS Information Corp., New York, 1973, p. 29.

B. 1. The Policy Delphi

MURRAY TUROFF[1]

A Seer upon perceiving a flood should be the first to climb a tree.
Kahlil Gibran

Introduction

The Policy Delphi was first introduced in 1969 and reported on in 1970.[1] It represented a significant departure from the understanding and application of the Delphi technique as practiced to that point in time. Delphi as it originally was introduced and practiced tended to deal with technical topics and seek a consensus among homogeneous groups of experts. The Policy Delphi, on the other hand, seeks to generate the strongest possible opposing views on the potential resolutions of a major policy issue. In the author's view, a policy issue is one for which there are no experts, only informed advocates and referees. An expert or analyst may contribute a quantifiable or analytical estimation of some effect resulting from a particular resolution of a policy issue, but it is unlikely that a clear-cut (to all concerned) resolution of a policy issue will result from such an analysis; in that case, the issue would cease to be one of policy. In the face of the policy issue, systems analysis, operations research, and other related disciplines can do no more than supply a factual basis for advocacy. The expert becomes an advocate for effectiveness or efficiency and must compete with the advocates for concerned interest groups within the society or organization involved with the issue.

The Policy Delphi also rests on the premise that the decisionmaker is not interested in having a group generate his decision; but rather, have an informed group present all the options and supporting evidence for his consideration. The Policy Delphi is therefore a tool for the analysis of policy issues and not a mechanism for making a decision. Generating a consensus is not the prime objective, and the structure of the communication process as well as the choice of the respondent group may be such as to make consensus on a particular resolution very unlikely. In fact, in some cases the sponsor may even request a design which inhibits consensus formulation.

[1]Murray Turoff, "The Design of a Policy Delphi," *Technological Forecasting and Social Change* 2, No. 2 (1970).

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.) ISBN 0-201-04294-0; 0-201-04293-2

The Committee and the Delphi Process

Traditionally the approach in most organizations to the examination and exploration of policy issues has been the committee process. Certainly it is well documented by a number of writers on the functioning of government organizations, that the committee system is a structure that evolved initially to promote the advocacy process associated with policy analyses.[2,3] The committee approach brings people together across organizational lines in order that all views at similar organizational levels in the whole organization may be represented, and a meaningful view arrived at after the differing interests have been adequately expressed and advocated. It is the contention here, however, that from a pragmatic viewpoint, the committee approach in government and most other organizations no longer functions as effectively in the realm of policy formulation as it once may have.

Many organizations today have become bigger, serve more functions, and span a much wider range of complex interacting functions. Committees that truly represent all interests on an issue are often quite large and unwieldy. By the time one has reached the point of twenty or more people constrained to reach a view in a limited amount of time, a complete and free exchange of views among all concerned is often too time consuming or impossible within the scope of the allocated effort for the job.

With increasing size of organizations, the ratio of the number of people at the top echelons to those in the lower echelons has decreased over the years, particularly in government. This implies that those at the top must spend more time devoted to day-to-day management functions and less time for committee participation on the longer-range issues associated with policy. As a result, the responsibility for committee participation falls more and more on lower-level people. Individuals at the lower levels are less likely to be advocates of anything until they have had ample time to clear it with their supervisors. This often forces the committee into added weeks of delay whenever any new point is made and usually results in the early or premature termination of new considerations that might result from the advocacy process.

If an organization is top heavy a similar problem also develops. Power becomes too diffuse and no one feels he has the authority or jurisdiction to act as an advocate on the broader issues that usually arise at the policy level. There are so many narrowly defined functional responsibilities that everyone is taking care not to tread on their neighbors' territory.

The complexity of issues today usually calls for a great deal of additional staff to supplement the committee process. More often than not, this time or support is not allocated to or available for committee participants. In an atmosphere of budget cuts, belt tightening, and competition for limited funds, it may appear

[2]Charles F. Schultze, "The Politics and Economics of Public Spending," Brookings Institution, Washington, DC., 1968.

[3]Numerous references to Lindblom's writings on committee processes appear in the work cited in reference 2.

dvantageous not to advocate, not to be noticed, and especially not to be held countable for views, promises, or positions which require effort to document substantiate. In addition, in most organizations today, we have individuals ho are not familiar with many of the new decision aids coming out of perations research and systems analyses but who have an intuitive feel for the mplexities of the particular business or function the organization is involved . We also have a good many individuals who have been trained in many of e modern management techniques and who are sometimes a little too nfident that these approaches can be applied to every problem. The lack of fective communication between these two groups has brought about the effectiveness of many committee exercises.

It is the above factors, or any combinations of these factors, which have otivated attempts to seek substitutes for the committee process. Contrary to e above, the earlier writings on Delphi have usually presented a separate but anonical set of problems associated with committees that tend to reflect sychological characteristics of committee processes:

- The domineering personality, or outspoken individual that takes over the committee process
- The unwillingness of individuals to take a position on an issue before all the facts are in or before it is known which way the majority is headed
- The difficulty of publicly contradicting individuals in higher positions
- The unwillingness to abandon a position once it is publicly taken
- The fear of bringing up an uncertain idea that might turn out to be idiotic and result in a loss of face

Given a small committee of around ten individuals with sufficient time to nsider and explore the issues, and some assurance that the privacy of their spective remarks will be respected outside of the committee room, it is oubtful that any of the above issues would greatly inhibit the process. owever, as the size of the committee increases, the time available decreases, nd the organizational considerations listed above present themselves, the sychological problems will also come into play.

Delphi, however, is not a replacement for the committee process. The roposition presented here is that the Policy Delphi can be utilized to revise the ffectiveness of the committee approach.

A Policy Delphi can be given to anywhere from ten to fifty people as a recursor to a committee activity. Its goal in this function is once again not so uch to obtain a consensus as to expose all the differing positions advocated nd the principal pro and con arguments for those positions. In many policy reas, a larger number of respondents, in the area of twenty or more, is mmensurate with the number of differing interests that must often be nsidered in the increasingly complex issues facing organizations.

Once the Delphi has been accomplished, a small workable committee can ilize the results to formulate the required policy. This then is the author's

view of the role of the Policy Delphi—a mechanism for reviving the advocacy process in organizations through improving the effectiveness of lateral policy formulating committees. In this way, Policy Delphis operate as precursors to the committee activity.[4]

The Policy Delphi, therefore, is not in any way a substitute for studies, analyses, staff work, or the committee. It is merely an organized method for correlating views and information pertaining to a specific policy area and for allowing the respondents representing such views and information the opportunity to react to and assess differing viewpoints. Because the respondents are anonymous, fears of potential repercussions and embarrassment are removed and no single individual need commit himself publicly to a particular view until after the alternatives have been put on the table. Even in those cases where the Policy Delphi uses only the committee or sponsoring body as the respondent group, it has the advantage of eliminating the principal bottleneck in the committee procedure by providing a clear delineation of specific differing views, thereby providing an opportunity for the committee members to prepare their respective cases adequately.

A Policy Delphi should be able to serve any one or any combination of the following objectives:

- To ensure that all possible options have been put on the table for consideration
- To estimate the impact and consequences of any particular option
- To examine and estimate the acceptability of any particular option

The ability of the Delphi technique to improve current practices for handling the first objective seems quite clear. Whether or not it can meet or fulfill any portion of the other objectives probably depends on whether the design team can distinguish the motivation of the respondents in making a particular judgment on an option. More specifically, when a difference in judgment does occur on an option, is it based upon uncertainty and/or lack of information with respect to consequences, or is it based upon differences among the self-interests as represented by the respondent group? If the Delphi can be designed to make this distinction it should be able to serve these latter objectives of examining and distinguishing consequences and acceptabilities. Because in some cases people are not fully aware of the motivating factors behind their views, the exposing of these factors could require fairly sophisticated approaches, such as multidimensional scaling.

The Mechanics of a Policy Delphi

A Policy Delphi is a very demanding exercise, both for the design team and for the respondents. There are six phases that can be identified in the communica-

[4] Jerry B. Schneider, "The Policy Delphi: A Regional Planning Application," *Technological Forecasting and Social Change* 3, No. 4 (1972).

tion process that is taking place. These are:

(1) Formulation of the issues. What is the issue that really should be under consideration? How should it be stated?

(2) Exposing the options. Given the issue, what are the policy options available?

(3) Determining initial positions on the issues. Which are the ones everyone already agrees upon and which are the unimportant ones to be discarded? Which are the ones exhibiting disagreement among the respondents?

(4) Exploring and obtaining the reasons for disagreements. What underlying assumptions, views, or facts are being used by the individuals to support their respective positions?

(5) Evaluating the underlying reasons. How does the group view the separate arguments used to defend various positions and how do they compare to one another on a relative basis?

(6) Reevaluating the options. Reevaluation is based upon the views of the underlying "evidence" and the assessment of its relevance to each position taken.

In principle the above process would require five rounds in a paper-and-pencil Delphi procedure. However, in practice most Delphis on policy try to maintain a three- or four-round limit by utilizing the following procedures: (1) the monitor team devoting a considerable amount of time to carefully preformulating the obvious issues; (2) seeding the list with an initial range of options but allowing for the respondents to add to the lists; (3) asking for positions on an item and underlying assumptions in the first round.

With the above simplifications it is possible to limit the process to three rounds. However, new material raised by the respondents will not get the same complete treatment as the initial topics put forth by the monitor team. Still, very successful Delphis have been carried out within a three-round format. Ultimately, however, the best vehicle for a Policy Delphi is a computerized version of the process in which the round structure disappears and each of these phases for a given issue is carried through in a continuous process.[5]

It is also necessary on a Policy Delphi that informed people representative of the many sides of the issues under examination are chosen as participants. These individuals will not be willing to spend time educating the design team, by way of the Delphi, on the subject matter of concern. The respondents must gain the feeling that the monitors of the exercise understand the subject well enough to recognize the implications of their abbreviated comments. Therefore, the initial design must ensure that all the "obvious" questions and subissues have been included and that the respondent is being asked to supply the more subtle aspects of the problem.

In some instances, the respondent group may overconcentrate its efforts on some issues to the detriment of the consideration of others. This may occur because the respondent group finally obtained was not as diversified as the total scope of the exercise required it should be. With proper knowledge of the subject material, the design team can stimulate consideration of the neglected issues by interjecting comments in the summaries for consideration by the group. It is a matter of the integrity of the design team to use this privilege sparingly to stimulate consideration of all sides of an issue and not to sway the respondent group toward one particular resolution of an issue. If, however, the respondent team is as diversified as required by the material, there should be no need to engage in this practice.

[5]Murray Turoff, "Delphi Conferencing," *Technological Forecasting and Social Change*, 3, No. 2 (1972).

A Policy Delphi deals largely with statements, arguments, comments, and discussion. To establish some means of evaluating the ideas expressed by the respondent group, rating scales must be established for such items as the relative importance, desirability, confidence, and feasibility of various policies and issues. Furthermore, these scales must be carefully defined so that there is some reasonable degree of assurance that the individual respondents make compatible distinctions between concepts such as "very important" and "important." This is further complicated by the fact that many of the respondents may not have to think through their answers in order to remain consistent in answering different parts of the questionnaire.

The Delphi technique is not just another polling scheme, and the practices that are standard in polling should not be transferred to Delphi practice without close scrutiny of their applicability. Consider, for example, a poll of different groups in an organization asking for their budget projections over the next five years. This is a comparatively straightforward request which does not ask any one group to place itself in context or to worry about consistency with other groups in the organization. A Delphi on the same subject would ask each group to make projections for every group's budget and, in addition, to project separately a feasible total budget for the organization as a whole.

The normal budget process in an organization is essentially a poll. A few research laboratories have in recent years attempted a budget review process via the Delphi mode, but unfortunately these are never reported in the literature because of the proprietary nature of the subject material. In principle, it would appear that the Delphi offers more opportunity for people to support budget items outside of their current management function and often to obtain a better appreciation of the budget trade-offs that have to be made.

There are many different voting scales that have been utilized on policy type Delphis; however, there are four scales, or voting dimensions, that seem to represent the minimum information that must be obtained if an adequate evaluation is to take place. On the resolutions to a policy issue it is usually necessary to assess both desirability and feasibility. One will usually find a significant number of items which are rated desirable and unfeasible or undesirable and feasible. These types of items will usually induce a good deal of discussion among the respondents and may lead to the generation of new

ptions. The underlying assumptions or supporting arguments are usually valuated with respect to importance and validity or confidence. In this case a erson may think an invalid item is important (because others believe it to be ue) or that a true item is rather unimportant. It is usually unwise to attempt ask for a vote on more than two dimensions of any item. However, if one has stablished a significant subset of items utilizing these scales then further uestions can be introduced focusing on the significant subset. For example, here is the possibility of taking desirable options and asking the probability for ach, given certain actions are taken.

Typical examples of these scales follow. Note that no neutral answer is llowed other than No Judgment (which is always allowed on any question). A eutral position offers very little information in policy debates and it is usually esirable to force the respondent to think the issue out to a point where he can ake a nonneutral stance. In other words, the lack of a neutral point promotes a ebate which is in line with developing pros and cons as one primary objective. his design choice has sometimes upset those who feel consensus is the only alid Delphi objective.

Desirability (Effectiveness or Benefits)

Very Desirable	Will have a positive effect and little or no negative effect extremely beneficial justifiable on its own merit
Desirable	will have a positive effect and little or no negative effect beneficial justifiable as a by-product or in conjunction with other items
Undesirable	will have a negative effect harmful may be justified only as a by-product of a very desirable item, not justified as a by-product of a desirable item
Very Undesirable	will have a major negative effect extremely harmful not justifiable

Feasibility (Practicality)

Definitely Feasible	no hindrance to implementation no R&D required no political roadblocks acceptable to the public

Possibly Feasible	some indication this is implementable some R&D still required further consideration or preparation to be given to political or public reaction
Possible Unfeasible	some indication this is unworkable significant unanswered questions
Definitely Unfeasible	all indications are negative unworkable cannot be implemented

Importance (Priority or Relevance)

Very Important	a most relevant point first-order priority has direct bearing on major issues must be resolved, dealt with, or treated
Important	is relevant to the issue second-order priority significant impact but not until other items are treated does not have to be fully resolved
Slightly Important	insignificantly relevant third-order priority has little importance not a determining factor to major issue
Unimportant	no priority no relevance no measurable effect should be dropped as an item to consider

Confidence (In Validity of Argument or Premise)

Certain	low risk of being wrong decision based upon this will not be wrong because of this "fact" most inferences drawn from this will be true
Reliable	some risk of being wrong willing to make a decision based on this but recognizing some chance of error some incorrect inferences can be drawn

Risky	substantial risk of being wrong not willing to make a decision based on this alone many incorrect inferences can be drawn
Unreliable	great risk of being wrong of no use as a decision basis

The first and foremost problem in conducting a Policy Delphi occurs with the initial steps in the process. If the respondents feel strongly about the issues, and this should be the case, they will generate a large amount of written material. If they are provided a certain number of items to deal with on the first round then each of them will make approximately the same number of written comments or additions in response. These must be abstracted carefully and duplications among the respondents eliminated. On the average, the written material in the questionnaire for the second round will be five to ten times that of the first round.

After the votes are taken on the second round, the material should be rearranged by the average vote on the third round. In other words, referring to the preceding scales, the options should be reordered by Desirability and the supporting arguments reordered for each option by Importance. When the votes are in, the resulting summary for the third round should clearly point out which items exhibited polarized distributions, which ones exhibited a flat distribution across the whole range, skewed distributions, or on which items only a very small sample of the respondents were able to make a judgment. For these items, additional comments should be solicited. If possible, the revote should be put off until a fourth round when everyone can see the additional remarks. In a three-round exercise a revote is taken on the third round.

In many cases it may be desirable to keep track of certain subgroups making up the respondent group as a whole. This provides a mechanism to check whether polarized views reflect the affiliations or the backgrounds of the respondents. Depending on the application, this information can be fed back to the group. Schneider[4] in his article on Policy Delphis proposed a very concise "Measure of Polarization" among the subgroups. Take all two-by-two combinations of subgroups and add the absolute value difference of the average vote on a given item. This sum of first differences is now an index which provides an appropriate ranking of the degree to which differences exist for each item relative to the group of items as a whole. The same measure may be applied to each individual who voted on the item when a subgroup breakdown is not appropriate. Note that in opposition to average and standard deviation this measure is a strong function of the number who voted when applied on an individual basis.

Some additional guidelines on carrying out the Policy Delphi process are as follows:

- The number of professionals acting as the design-monitor team must be at least two so that one can check the other. Ideally, one should be knowledgeable in the problem at hand (but not precommitted) and the other should have editorial talents.
- A month or more is needed to develop the first-round questionnaire. In addition to the questionnaire, a factual summary of background material is usually supplied, and in some cases single or multiple sets of scenarios specifying certain items the respondents are to assume as given are provided for the purpose of evaluating the issues. Typically these scenarios deal with future economic conditions such as the rate of inflation. Sometimes it is more appropriate to introduce a set of alternative assumptions making up scenarios and let the respondents form a group scenario by voting on the validity of each.
- Each questionnaire should be pretested on coworkers who have not been involved in the design. There is a very high probability that this will identify items that are stated in a confusing manner.
- Take care to avoid compound statements to be voted upon. The question "if *A* and *B* are true" should be broken into two separate items. The exception is statements of the form "if *A* then *B*," which are quite useful in some situations.
- The respondents, if new to Delphi, will respond with compound and sometimes lengthy comments. Therefore it is a good idea to show them some examples of the form you would like comments to take, in terms of being short, specific, and singular in nature.
- If there is a trade-off between the ease of summarizing the results and the ease of the respondents in providing the answers and understanding the results, the choice should always favor the respondent.
- The respondents should be allowed to suggest changes in the wording of items which should then be introduced as new items. Experience has shown that the vote on a policy item is very sensitive to wording. Because of this property, the material on Policy Delphi can mushroom in size and represents considerably more effort than the traditional forecasting Delphi oriented to largely quantitative response after the first round.
- When asking for revotes on an item, the individual respondent should be shown his original vote. The respondent should also be provided two copies of the questionnaire so that he may retain one for later reference or to do rough work. He can type his answers on the other copy if he is concerned with security. On Policy Delphis security can be a problem with respect to convincing the respondents that it will be maintained. The design team should set up a procedure where they themselves cannot identify the returns with the individuals involved.
- The respondents must be convinced that they are participating in an exercise which involves a peer group. Therefore it is usually necessary on the letter of

invitation to indicate the types of backgrounds reflected in the participant group. In some cases, a list of the respondents involved can be provided if there is no other effective way to convince the group of the significance of the exercise.

As can be seen, there are many things to be considered in running a Policy Delphi, or any other Delphi for that matter. The Delphi concept seems so simple that many people have thought it an easy thing to do. Consequently there have probably been more poorly done Delphis than ones that have been well done.

One additional aspect of the Policy Delphi which usually argues for four or more rounds arises in the situation where the respondents feel very strongly about their respective views. In such a case they sometimes have an attitude where they cannot imagine that there are rational and intelligent people who hold a contrary view. Even with a vote on the first round on a given issue, the reaction of this type of respondent to the vote presented on the second round is that the individuals holding the opposite view to his just don't understand the problem completely. A few simple comments will clear up their ignorance. It is only until the third round comes back that this type of respondent feels the shock resulting from a realization that the other side also feels it has some valid points to be made. Therefore, it is only at the third round that this type of respondent begins to put a great deal of careful effort into the points he is making and to consider more carefully what the other side is saying. The material generated out of this type of process could have a significant impact on the group views if carried back in a fourth round.

The selection of respondents is one of the most difficult tasks. However, this problem applies to any committee or study effort. The sponsor is likely to have a certain candidates in mind. The design team should try to structure the problem in order to get a comprehensive coverage of the topic. It is also a good idea to mix in a couple of lateral thinkers and devil's-advocate types, just on a matter of general principle—i.e. those individuals who always manage to come up with the unexpected.

It is possible on a Policy Delphi to observe two very different phenomena taking place. One is when the exercise starts with disagreement on a topic and ends with agreement. This can be very useful to those sponsoring the study if it does occur, but, as has been said, is not a necessary result. Another process is to start with agreement on a topic and end with disagreement. In a sense this can be viewed as an educational process taking place among the respondents who suddenly realize, as a result of the process, that the issue is not as clear-cut or simple as they may have thought. Unfortunately, to this point in time there has not been sufficient exploration of the use of the Delphi technique as an educational process. Schneider[1] also discusses this point. As he pointed out, Delphi could be used by a planning agency to interface more effectively with representatives of the community and serve an educational function for both groups.

Another unexplored use of the Policy Delphi is the investigation of the performance of past policy actions. Too many organizations do not have an appropriate mechanism for taking stock of what they have accomplished. Understanding of what has occurred is often lacking and can lead to future mistakes in policy formulation.

Examples of Policy Delphis

One of the first Delphis that bordered on being policy oriented was an exercise undertaken in 1968 by the National Industrial Conference Board. It was titled "An Experimental Public Affairs Forecast." It involved 70 people representing the following areas of expertise:

Economy, Business, and Labor	17
Science, Technology, and Change	9
Government, Law, and Politics	6
Resources	9
Education and Training	5
Communications	8
Culture, Family, and Behavior	12
International Security	4

The vast majority had titles of chief executive or director. All were considered by the Conference Board to be distinguished in their field.

The overall objective of the study was to obtain a rank ordered list of National Priorities or Areas of Major Concern to the Nation, areas which could create major public problems in the seventies and eighties and should receive attention by U. S. leadership. The top ten in that list in order of priority were: (1) division in U. S. society; (2) international affairs; (3) education; (4) urban areas; (5) law and order; (6) science, technology, management of change; (7) economy; (8) resources; (9) values; (10) population.

The Delphi was completed before the presidential campaign and one may note a degree of correspondence between the priorities set by this exercise and the Republican campaign themes. While the Delphi dealt with policy considerations, it was largely oriented to putting the pieces of the problem together by collecting information and views from a diverse set of respondents. Therefore it largely reflected a Kantian-type exercise. The bulk of the material produced was a collection of commentaries on the problem areas with some estimate of when particular problems would arise. Each item was handled in terms of the following categories of information:

- description of the item
- description of public reaction to the item
- beginning date of maximum impact on U. S.
- intensity of impact on U. S.
- opportunity for leadership to change the expected

The Delphi appeared to be quite adequate in meeting the needs of its sponsors; however, the exercise has never been described in the literature so one can only infer this from the final report, which unfortunately did not receive public distribution beyond those immediately involved and some individuals working in the Delphi area at that time. One major fault of the study was the decision made by the staff people not to abstract the comments of the panelists but to retain the full text. In part this decision was probably influenced by the distinguished nature of the respondent group. The result was a very large volume of material which is a little painful to wade through to gather the particular nuggets of wisdom that were produced. One goal of a Delphi design should, therefore, always be to obtain a filtering of the essential from the superfluous.

The next Delphi in the policy area was one conducted by Emory Curtis as a consultant to San Mateo County in California. This one involved around 80 community people representative of the many different constituent groups making up the public body. A great deal of effort went into obtaining a broad-based distribution of respondents. They were provided a large number of policy options dealing with the structure and functions of the county government, and asked to vote on these for relative agreement on a seven-point agreement scale. Additional items were added as a result of the first round. However, the one shortcoming of this exercise was the lack of exploration of the factors underlying disagreement when it did occur. The exercise produced some new options and exhibited consensus where it occurred but provided no mechanism for effectively resolving disagreement. However, it represented one of the first attempts to use the Delphi in policy areas related to community government.

In 1970 a Delphi was conducted by the Office of Emergency Preparadness and the Rand Corporation on the subject of Civil Defense Policy. This Delphi introduced a number of unique features. It exhibited the structure of a Hegelian inquiring system as opposed to the earlier Lockean- and Kantian-type Delphis. First, it recognized in the design that strong disagreement already existed on a number of the issued involved. For a number of items the respondents were asked to choose sides by circling "could/could not," "should/should not," these being choices in the wording of the items. They were also asked to develop the strongest arguments on the various sides of a given issue. The sponsor was not interested in having the group make a decision for him, but in having the group develop, compare, and evaluate the best possible arguments on each side of an issue.

The details of this exercise are well documented in the literature.[1] As typical of these types of Delphis, the respondents generated about eight times the amount of material they were initially given on the first round, which contained some seventy items for evaluation. Basically policy options were evaluated on scales of desirability and feasibility, while supporting points were evaluated on importance or validity. This Delphi was really the first to incorporate a structured debating-type format, which appears to be the useful approach for the exploration of policy issues.

In 1970 Professor J. B. Schneider at the University of Washington adopted the same approach to the exploration of transportation planning as it applied to highway development in the Seattle area. His report of the exercise[4] is an excellent example of applying these techniques to urban planning problems. He also contributed some very useful observations on the methodology for handling disagreement in that particular context.

Following the same line of development, Joel Goodman of the College of Marine Studies, University of Delaware, conducted a policy-type Delphi on the Coastal Zone Land Use Planning Issue. This involved a large number of people representing government business, public groups, and specialists. This exercise was done in 1971 and 1972. It converged the following types of items into different sections of the questionnaires: respondent characteristics; respondent attitudes; arguments pro and con; general policy and budget items; specific policy issues; specific programs; strategic issues.

Some sample questions from that exercise follow:

- As an individual, list in order of priority your 5 principal concerns with respect to the way in which the coastal zone is developing.

  (a) Health hazard
  (b) Unsightly buildings
  (c) Dirty water (visual appearance)
  (d) Too much land going to waste
  (e) Too crowded
  (f) Not enough housing
  (g) Not enough boating facilities
  (h) Not enough camp ground
  (i) Beaches too narrow
  (j) Too many fisherman
  (k) Other

- Why are you as an individual concerned with pollution in the coastal zone and its effects upon the marine environment? Check up to three responses and signify relative importance by numbering principal reason as "1."

  (a) biological danger
  (b) potential loss of recreational opportunity, i.e., swimming, boating, etc.
  (c) potential loss of aesthetic values, i.e., vistas, landscape, etc.
  (d) potential loss of income or revenues
  (e) community involvement
  (f) other (specify)

Indicate by check mark who should assume responsibility for establishing:

	Quality limits for coastal zone environment (1)	Use patterns for coastal zone shore lands (2)	Use patterns for coastal zone submerged lands (3)	Use patterns for coastal waters (4)
a. State				
b. County				
c. Local Community				
d. Other (Specify)				

If standards for the quality of the marine environment are to be maintained, then the authority and responsibility for regulation should be vested in:

SELECT ONE

a. A state agency within the executive branch

b. A county agency

c. Individual municipalities

d. Criteria established by state; regulation by municipalities and counties

e. A new organization responsible to———
with elected/appointed officials (Fill in blank and select one or the other means of acquiring the officials.)

A large number of the current Delphis have started to incorporate policy ues even when that was not the primary concern. Such issues have the sychological advantage of making the exercise of more interest to the respon- its. The policy orientation has been introduced in some different ways. stead of asking individuals to extrapolate data into the future in terms of eir best estimate of what they think will occur, a policy approach would be to what would be a desirable and possible extrapolation as well as an desirable and possible extrapolation. Based upon those estimates one can ask at are the factors that could make the curve go one way or the other.

A Delphi study conducted by the Federal Department of Public Works in Canada illustrates the incorporation of policy options into an essentially non-Policy Delphi. The department's major role is providing accommodation for federal civil servants, and the Delphi was undertaken as part of a model for forecasting government employment with the purpose of determining future accommodation needs. But the department's mandate extends beyond simply providing buildings to house federal employees. It is concerned with the total work environment of the civil service.

Consequently, the Public Works Delphi also explored the existing procedure for space allocation, which at present is based on the average salary of all employees using that space, and asked respondents to comment on that process. In the first round, after reviewing the present process, the respondents were asked to list what they felt were the strengths and weaknesses in the process, and asked to suggest possible options for change. In the second round these options were voted upon according to *Desirability* and *Feasibility*, keeping in mind that a particular option could be desirable and unfeasible at the same time, or vice versa. Some examples of suggestions for change according to *Desirability* were:

- formula approaches, if used, must reflect the quality of space as well as the quantity
- relate space to function not salary
- more emphasis on multipurpose facilities
- DPW should lead the way in educating agencies in new building concepts

The Delphi also looked at possible parameters for measuring building performance that would go beyond the usual cost/benefit measures, such as the ratio of rentable square feet to total square feet. Specific suggestions or concepts for consideration fell into the following categories:

- psychological and motivational impact on employees
- transportation to building
- aesthetic value of building
- community and public service
- energy and environment

The respondents were asked to vote on the *Desirability* and *Feasibility* of specific suggestions and to suggest ways in which some of these concepts could be measured..

Public Works then used the Delphi exercise not only to fulfill its immediate objective of forecasting federal government employment but also to explore policy options relating to its mandate of fulfilling broader social, economic, and environmental objectives.

Another excellent example of a Delphi mixing policy issues with future forecasts was one done by the Canadian Department of Health and Welfare on

ie Future of Genetic Counseling Services in Canada. The exercise involved ome sixty respondents ranging from research geneticists to public health orkers. The design was well balanced between "technical issues" of what was ssible at what point in time and "policy issues" of who could, would, or hould do what. In this latter area, the issue of a genetic registry and its otential abuses as well as uses were explored. The Delphi used the same sort of ales that were mentioned earlier. However, it tended to redefine a scale such s "importance" for each question it was used on. This had the merit of inimizing what the respondent had to remember, since each question was rgely self-contained. It also minimized the chance of confusion by placing the cale within the context of the particular question. Furthermore, it allowed ore variety in the sequencing of questions. Most other designs, by grouping estions of a given type under one explanation, can produce a feeling of notony as the respondent goes through the exercise.

he Problems of a Policy Delphi

'e have already mentioned the danger that a Policy Delphi can be misin-rpreted as a decisionmaking tool as opposed to a decision-analysis tool. eryone at heart is a decisionmaker, or wishes to be, and it is all too easy on e part of the designer to appeal to this unrequited desire. It should be a atter of intellectual honesty for designers to make clear just what the objective the exercise is. If we have a problem in organizations today, especially overnmental ones, it is that the responsibility for a given decision is not clearly used on one individual. A decision should be made by one individual, and role of the Policy Delphi and other tools is to provide the best possible formation and ensure that all the options are on the table. To do this the lphi must explore dissension. Both Dalkey and Helmer in the early writings Delphi expressed the need to establish clearly the existent basis for observed ssension. However, this implies a good deal more work for the design team d has often been neglected in the majority of the early exercises. When a ong minority view exists and is not explored, the dissenters will often drop t, leading to an "artificial" consensus on the final product.

Once a Policy Delphi has been started, there is no way to guarantee a ecific outcome if it is to be an honest exercise. This is something the sponsor ust be well aware of. Occasionally a sponsor, particularly in a policy exercise, desire that the group not reach a consensus on any particular option. hile it is consistent with the objective of a Policy Delphi to choose a spondent group such that a consensus is unlikely to occur, it can never be aranteed that it will not be a result. However, there is a fine line between lphi as an analysis tool and Delphi as an educational or persuasion device. It possible to consider using a Delphi to educate at least a part of a respondent

group on options they may not be aware of. Unfortunately, very little work has been done on the use of Delphi in an educational mode even though most designers would agree that educational processes take place in most exercises.

A Policy Delphi is a forum for ideas. In opening up the options for review, items may arise which can be disconcerting to members of the group. If a sensitive area is under review and an attempt has been made to have diverse representation in the group, then premature leakage of the results can occur. In such a case, individuals outside the exercise may misinterpret what is taking place. This problem of lifting items out of context occurs all the time in the committee process. A workable approach to this problem in the Delphi process is to incorporate members of the press into the respondent group when dealing with major public policy items.

As with any policy process, there are many ways to abuse the use of the Policy Delphi: the manner in which comments are edited, the neglect of items, the organization of the results. However, such a process is a rather dangerous game and not likely to go unnoticed by some segment of the respondents. There are very few greater wraths than that of a respondent who discovers himself to be engaged in a biased exercise. Furthermore, Delphi has reached the point where there is no longer any excuse on a professional basis for making many of the mistakes found in earlier exercises. The person seeking to undertake a Delphi today should be reasonably familiar with what has taken place in the field.

B. 2. Delphi Inquiries and Knowledge Utilization

JOHN LUDLOW

Introduction

The development of methods to obtain, refine, and communicate the informed judgments of knowledgeable people is one of the most crucial problems in planning and decisionmaking. The task is particularly challenging in the Michigan Sea Grant Program, which emphasizes a systems approach by a multidisciplinary group of researchers. Some of these researchers are experts in extremely specialized areas, representing a wide range of technical, economic, social, legal, and political disciplines.

From its inception the general goal of the Michigan Sea Grant Program[1] has been to provide the common management effort necessary to develop and bring to bear university expertise on short- and long-term resources management problems in the Great Lakes. The major approach of the program has been the development of basic information and predictive models for resolution of resource problems, followed by applications and/or demonstrations of such information and models to appropriate agencies and groups. Over 120 research and faculty personnel from practically every major school or college in the university are presently active in the program. Research and planning groups representing federal, state, and local government agencies, industry, and concerned citizen groups are also part of the problem-solving team.

The Grand Traverse Bay watershed region was selected as the focus of pilot efforts to develop research and planning methodologies that will be applicable in dealing with problems and opportunities of all the Great Lakes, and in particular Lake Michigan. In the area of finding mechanisms to improve the coordination of the Sea Grant effort at the University of Michigan it was decided to investigate the potential for utilizing the Delphi technique.

The Michigan Sea Grant Delphi inquiries[2] were designed to obtain and refine an interdisciplinary group of researchers' judgments about issues and developments that should be considered when planning for intelligent management of the water resources of the Great Lakes.

An important objective of the exercises was to convey the judgments of the researchers to the communities which are to benefit from the research. One approach toward this objective was to include on the panels—on the same basis as the researchers—people who were believed to be influential in the political processes through which regional planning is accomplished. Their knowledge of the issues and the region was beneficial to the deliberations, but more importantly, their participation was judged to be an effective way of communicating information to regional planners and decisionmakers.

Two of the three panels were made up of researchers who were designated as technicians and behaviorists. The third group was made up of concerned citizens who were designated as decisionmakers. In addition to forecasting, the method was used in several other roles involving the quantification of subjective judgments. The exercises were designed to be progressive and cumulative, with an emphasis on an orderly development of informed judgments.

The Delphi inquiries were one of several Michigan Sea Grant projects related to the general task of transmitting new knowledge to people and organizations in a way that results in effective use. Respondents in these exercises—a group with exceptional qualifications—served as the primary resource in evaluating the methodology.

The technical panel was composed of thirty-three individuals whose expertise was primarily in the physical sciences and who were divided about equally between Sea Grant researchers and faculty, graduate students, and others in the School of Engineering. A second panel included Sea Grant researchers who were not selected for the technical panel. Generally their academic backgrounds and interests were oriented more to the behavioral sciences, and for this reason they were labeled behaviorists. They represented a wide range of ages, academic disciplines, and university schools and laboratories. Participants for the third panel were randomly selected from groups of Grand Traverse Bay area residents believed to be influential in the following fields: civics, business, planning, politics, natural resources, government, education.

The names associated with the panels, although somewhat arbitrary, are reasonably consistent with the roles each group would be expected to play in planning the management of regional water resources. The technical panel operated independently of the other two panels and its output was fed into the deliberations of two broader-based panels, which operated independently in the earlier rounds and as a combined panel in the final round. The nature of their participation is summarized in Table 1.

In order to provide continuity, a person's judgments on the previous round were used whenever he or she could not respond on a particular round.

Several significant modifications and refinements in the basic Delphi methodology were tested in the Michigan Delphi inquiries. These changes were motivated by the perceived threat of a manipulated consensus, the desire for constrained or conditional judgments, and recognition of desirable aspects of interpersonal methods not obtainable using the Delphi technique exclusively. The concept of informed judgments as contrasted with expert opinion provided

[1]The term "Sea Grant Program" was derived from the National Sea Grant College and Program Act, whose intent was to involve the nation's academic community in the practical problems and opportunities of the marine environment, including the Great Lakes.

[2]The term "Delphi inquiry" was propounded by Turoff and refers to the complete Delphi process. He observed that any particular Delphi design can be characterized in terms of the "inquiring systems" specified in Churchman's writings. See reference 1 at the end of this article.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

Table 1
Participation in Michigan's Sea Grant Delphi Probes

Activity	Technical Panelists	Behaviorists	Decision-makers
Contact established with Delphi administrator	33	16	22
Unavailable after the start of the Delphi probe	6	0	3
Written comments and evaluations made on at least one round	28	11	21
Written comments and evaluations made on 3 or more rounds	14	6	9
Written comments and evaluations on final round	20	9	11
Written evaluation of methodology or evaluation interview	29	12	16

panels; it also provided an opportunity to exploit an inherent characteristic of the method—*to inform* during the process of soliciting judgments.

1. Outline of the Procedures: Social, Political, and Economic Trends

The portion of the Delphi inquiry concerned with social, political, and economic trends was designed to provide respondents on the broader-based panels with some basic reference points in making subsequent judgments regarding future social and technical developments.

The information package for round one presented the trends for eight measures which have commonly been used to indicate the social and economic development of a region. Curves were plotted from 1950 to 1970, taking advantage of the 1970 census and the standardized enumeration procedures of the Bureau of the Census. Panel members were asked to extend the curves through 1990 and to indicate the numerical values for 1980 and 1990 [2].

In the second round, curves representing the medians and interquartile ranges were provided for the panelists, as well as pertinent comments submitted by respondents on the previous round. Panelists were asked to reconsider their estimates, and if any of the new estimates were outside the designated consensus range for the previous round they were asked to support their position briefly.

On this round the graphs of three additional statistical measures were introduced for consideration. A cumulative summary of the group response was provided in the information package for round three to serve as background information for other panel deliberations.

Important Developments and Requisite Technology

The Delphi method has had its greatest application and acceptance as a means of compiling a list of future technical events or developments and collecting subjective judgments regarding them. In the Michigan inquiries social, political, and economic developments were also solicited and evaluated so that panelists would be encouraged to consider all environments in making judgments regarding water quality, waste-water treatment systems, and research priorities.

The initial evaluation matrix for the technical panel did not present a list of potential developments, something which is usually done in order to facilitate participation and generate additional items. It was believed that this unstructured approach would result in a wider range of suggestions; however, the information feedback of the second round did include—in addition to items suggested by respondents—thirteen events that were taken from Delphi exercises conducted at Rand and the Institute for the Future. These events covered areas considered by the researcher to be of interest to the panel and were also good examples of how developments should be specified to avoid ambiguity, particularly with respect to occurrence or nonoccurrence.

The evaluation matrix for the third round provided the respondent with his estimates for the second round and a summary of the group's response. Comments submitted by respondents were also provided, as were the median estimates for technical and economic feasibility if they differed significantly. The evaluation matrix for the third round was designed so that a panel member could easily determine if his reassessed estimates for a specific development were outside the group's consensus range—arbitrarily identified as the group's median 25 percent and 75 percent estimates. If a respondent's latest estimate was outside the consensus range for the previous round he was asked to support this "extreme" position briefly.

The evaluation matrix for the fourth round presented a more comprehensive summary of the previous round than had been provided up to this point in the exercises. Statistical summaries were presented not only for all the respondents but also for those who rated their competence relatively high and for those in the latter group who indicated a familiarity with the Grand Traverse Bay area. In addition, the persons arguing for an earlier or later probability date than that indicated as the consensus were identified by a number which correlated to a list of biographical sketches.

On the final round of the technical-panel exercises, respondents were also asked to make specific conditional probability estimates for pairs of events that

panel members had suggested were closely related. First they were to consider the effects of the occurrence of the conditioning event and then the effects of the nonoccurrence of the conditioning event (see Fig. 1). One of the objectives of this procedure was to encourage panelists to reexamine their estimates for individual events in the light of the influence and probabilities of related events. Analysis of all individual responses reveals that a relatively high percentage of respondents altered their final estimates for those developments included in the set of events which was subjected to conditional probability assessments. Since this was the third iteration of feedback and reassessment for many of these developments, it is not unreasonable to assume that the change in estimates primarily resulted from the evaluation of relationships among events—*relationships which previously had not been fully considered*. This assumption is further supported by the fact that these respondents made almost no changes in their estimates of other developments, which were not subjected to the specific routine of estimating conditional probabilities (but were given the benefit of the feedback of all of the other types of information used in these exercises). In view of the fact that the relationships among events were stressed throughout these exercises, any movement in the final estimates as a result of the consideration of specific conditioning effects is believed to be significant.

Developments and Events that Respondents Have Suggested Are Interrelated	Probability 1971-80	50% Probability Date
D-32 Requirement by the state, calling for tertiary treatment of municipal sewage for Traverse City		
Your Previous Estimates	80	1977
Panel Estimates, Round 3	75 (50-85)*	1977 (1975-80)
Those Who Rated Competence ≥ 3	83 (62-95)	1978 (1975-80)
Your Next Estimates for D-32		
D-31 Construction of a spray irrigation system for waste water disposal in the Grand Traverse Bay region		
Your Previous Estimates	50	1980
Panel Estimates, Round 3	50 (50-50)	1980 (1980-90)
Those Who Rated Competence ≥ 3	50 (15-50)	1980 (1978-80)
Your Next Estimates for D-31		
If you were certain that D-32 would occur before 1980, your estimates for D-31 would be		
If you were certain that D-32 would not occur in 1971-80, your estimates for D-31 would be		

* Interquartile Range

Fig. 1. Example of interrelated developments.

An analysis of the estimates of the technical panel showed that some respondents appeared to have considerable difficulty making probability estimates both for a fixed period (1971–80) and for fixed levels of probability (25, 50, and 75 percent). In some cases inconsistent estimates were made (for example, the probability of occurrence during 1971–80 was estimated to be greater than 50 percent, but the year associated with a 50 percent probability was later than 1980).

Fixed probabilities of 25, 50, and 75 percent were selected for personal probability assessments by the broader-based panelists for several reasons:

(1). There was strong agreement among the three groups involved in the exercises—technical, behavioral, and decisionmakers—on the words and phrases that they associated with the numerical probabilities of 25, 50, and 75 percent [3].

(2) Individual distributions provided the decisionmakers with more information than single probability estimates and were believed to be helpful to the estimator in making assessments that were consistent with his judgment [4].

(3) The 25, 50, and 75 percent levels of probability were ideal for using a betting rationale, that is, systematically dividing the future into equally attractive segments.

(4) It was believed that group medians associated with these fixed probabilities would provide an easily identifiable consensus range.

Since it was likely that many of the decisionmakers would have had little experience with the notion of personal probabilities, a guide for making personal estimates of probability was sent to all members of the broad panels—researchers as well as decisionmakers. The guide presented a systematic method for arriving at the timing estimates for each technical and social development. The assessor was asked to visualize a movable pointer below a sequence of numbers representing years, as in the diagram below. He was asked to move the pointer mentally so as to divide the future into two periods in which the development was equally likely to occur.

```
1                 1                 1
9                 9                 9
7                 8                 9
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 Later
                  A
                  ↑
            50% Probability
```

If the result appeared as it does in the diagram above, 1983 should be entered as the 50 percent probability date. It could also be described as the "1-to-1" odds or "even chance" date. If the pointer came to rest beyond 1990, "Later" would be recorded, and the assessor would go on to consider the next develop-

ment. The assessor was then instructed how to divide up the results to estimate the "3-to-1" and "1-to-3" odds.

Because of the interest in technology transfer and knowledge utilization in Michigan's Delphi inquiries, there was a special interest in the judgment patterns of the technicians and decisionmakers, which were displayed as in Fig. 2. For each round the panel medians (connected by a solid line) and the interquartile ranges (connected by dashed lines) were shown. The rounds were numbered from left to right for the researchers and from right to left for the decisionmakers, to facilitate the comparisons. The average judgments of respondents in each group who rated their competence in the area being considered relatively high were indicated by asterisks. For most items generally, each group's median estimate for the final round was very close to the median estimates of those who considered themselves relatively competent in the subject. Also, the consensus—as measured by the interquartile range—narrowed and the average estimates of the two groups tended to come closer together. Some of the other patterns, while not ideal from the standpoint of movement toward a narrower consensus, provided a decisionmaker with information as to a course of further inquiry.

Sources of Pollution

A crucial consideration in planning for intelligent management of water resources is the identification of the most important sources of pollution. In making their judgments, panelists were asked to assume a future social and political environment consistent with present trends. However, it was expected that concurrent Delphi inquiries regarding important developments and requisite technology would influence their estimates.

On the first round the technical panel was provided with a list of sources of pollution and specific pollutants thought to be important. Panelists were requested to add other items that they felt would affect a body of water comparable to Grand Traverse Bay in the next twenty years. The collated responses identified seventeen additional sources of pollution and eighteen additional pollutants for the panel to consider. Since there were too many alternatives to present in a matrix designed to encourage the careful consideration of several evaluation factors, the primary objective of round two was to narrow the number of alternatives. The evaluation matrix of round three presented the ten most important sources of pollution as determined by a statistical summary of the estimates made in round two. Panelists were asked to distribute 100 points among the sources of pollution, according to each one's relative importance, for two future periods. The information feedback for the following round provided statistical summaries for Group A, all respondents; Group B, those who rated their competence on sources of pollution relatively

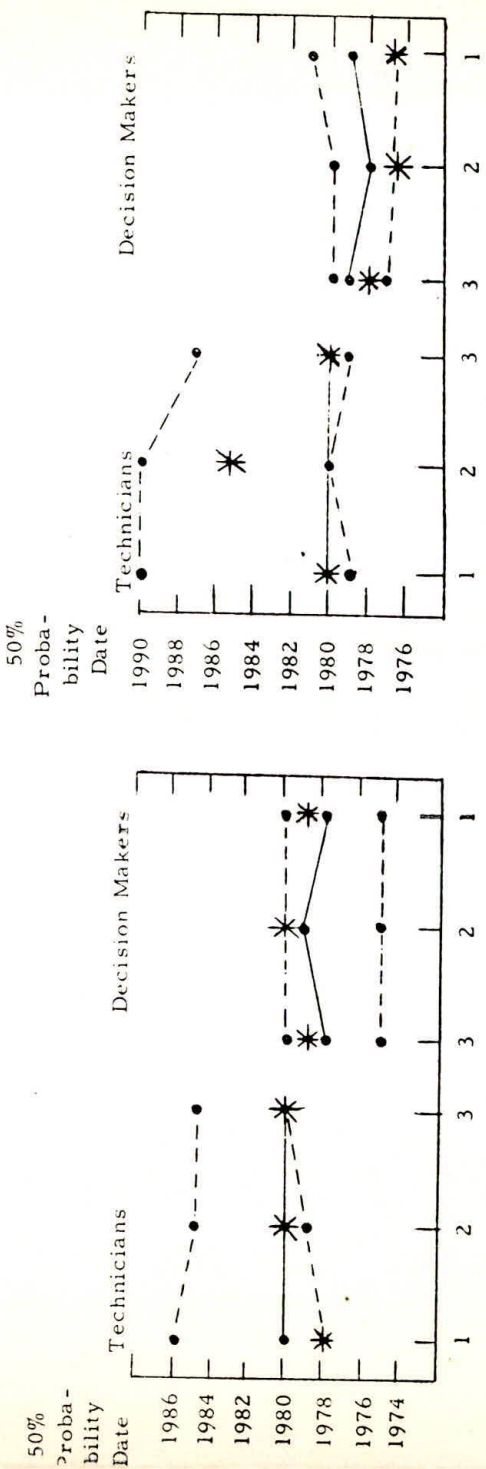


Fig. 2. Statistical summaries—technicians and decisionmakers.

high; and Group C, respondents in Group B who were also relatively familiar with the Grand Traverse Bay watershed area. Although Group B differed considerably in size from Group C, the average estimates of the two were remarkably close. This finding might suggest that technical competence is a more important requisite for panel membership than familiarity with a specific region, an idea that could have important implications for interdisciplinary programs such as Sea Grant, in which research methodologies developed for a subregion are to be applied to a larger socioeconomic system.

In the broad panel exercises the evaluation matrix for round two was similar to the final matrix used in the technical panel. The evaluation matrix for the following round provided statistical summaries of the estimates of both the technical panel and the broader-based panels (Fig. 3).

A significant difference in the final estimates regarding the relative importance of the effluent from the Traverse City sewage system suggests that a series of estimates conditional on specific social, political, or technical developments could be used to determine the assumptions on which the evaluators based their estimates. The reason for the differences in estimates could also be sought through interviews and other means of communication.

Recommended Waste-Water Treatment and Disposal Systems

Many communities in the Great Lakes basin are confronted with decisions on waste-water treatment and disposal systems that will have important consequences for the future socioeconomic development of their region. This is a highly technical and complex issue, and decisionmakers must intuitively assess the judgments of experts in many specialized areas.

A systematic consideration of the available alternatives and the identification of areas of agreement and disagreement within and between the three general groups involved in these exercises may aid planners from this region as well as those from many other communities in the Great Lakes region facing similar problems and decisions.

Included in the technical panel's round-three information package was an evaluation matrix that listed six alternative waste-water treatment and disposal systems. Panel members were asked to suggest other alternatives and to evaluate each of them in terms of two different starting dates for the construction of the necessary facilities. Variances in the estimates were to be attributed to assumptions about the technology that would be available at the two starting dates. Panel members were instructed to give 100 points to their first choice for each time period and a portion of 100 points to the remaining alternatives according to their value relative to the first choice.

The round-four information package provided panel members with a summary of the estimates made in the third round. The evaluation matrix for

Ten Most Important Sources of Pollution	ABATEMENT FEASIBILITY 1971-80				RELATIVE IMPORTANCE							
	ECONOMIC Scale 1-5 1=Extreme subsidy required 2=Substantial subsidy 3=Moderate subsidy 4=Slight subsidy 5=Routine		TECHNICAL Scale 1-5 1=Extremely difficult 2=Very difficult 3=Moderately difficult 4=Slightly difficult 5=Routine		Panel members were asked to distribute 100 points among the most important sources of pollution during the two periods indicated.							
					1971-80			1981-90				
	Technical Panel	Broad Panel	Technical Panel	Broad Panel	Those on technical panel who rated familiarity 3 or more (13)†	Broad Panel (13)†	Final Estimates	Those on technical panel who rated familiarity 3 or more (12)†	Broad Panel (12)†	Final Estimates
Effluent, Traverse City sewage system	3(2-3)*	2(1-2)*	4(3-5)*	3(2-4)*	31	22		27	15	
Septic tanks in the region	2(2-3)*	5(2-5)*	3(3-4)*	4(3-4)*	15	13		10	14	
Storm water run-off, urban	2(2-4)*	3(1-3)*	3(2-5)*	3(2-4)*	12	9		9	9	
Food-processing wastes	5(3-5)*	4(4-5)*	3(3-4)*	3(3-4)*	10	13		7	8	
Industry wastes (less food processing and agriculture)	4(3-5)*	4(4-5)*	3(3-4)*	3(2-3)*	7	8		9	9	
Power production utilities	4(4-5)*	5(4-5)*	3(3-5)*	4(2-5)*	5	4		10	8	
Agriculture	4(2-5)*	4(3-4)*	4(3-4)*	3(2-4)*	9	13		9	12	
Oil spillage, bulk marine vessels	5(2-5)*	5(5-5)*	3(2-3.5)*	4(4-5)*	4	4		4	6	
Direct contact inputs, human use	4(1-5)*	5(4-5)*	3(2-3.5)*	3(2-5)*	4	6		8	9	
Ground water run-off carrying natural pollutants	2.5(1-4)*	3(1-5)*	3(1-4)*	2(1-5)*	3	8		8	10	

* Median--interquartile range in parentheses. Interquartile range contains the middle 50% of the estimates
† Number of respondents who contributed numerical estimates.

Fig. 3. Broad panel's third-round evaluation matrix—sources of pollution.

that round (Fig. 4) requested two evaluations for the six alternative waste-water treatment and disposal systems for two different starting dates. In the first evaluation the respondents were asked to consider all factors, in particular the technology available at the start of construction; in the second evaluation they were to consider only ten-year operating costs.

The broad panels used the same evaluation matrix as the technical panel in their final round of estimates, and they were also given a summary of the results of the technical panel's evaluation of all factors except for cost estimates. The broad panelists were advised that the technical panel probably emphasized technical factors in making their estimates. They were also told that the recommendations applied to a region similar to the Grand Traverse Bay area and could differ significantly if the technical panel had considered a specific situation.

A comparison of the average estimates of those on the technical panel who rated their competence relatively high with the average estimates of the respondents on the broad panels showed a very close agreement for both planning periods. This agreement was evident when panelists considered all factors and also when they considered ten-year operating costs, although the values assigned to each alternative relative to operating costs varied considerably from the values assigned when all factors were considered.

The judgments of the technical experts are believed to embody risk considerations applied to a general situation, whereas the judgments of the broad panels are thought to be more oriented to the benefits of alternative approaches for their specific region. Cost estimates include operating costs only; the consideration of investment costs and financing methods could be equally important to the decision maker.

The waste-water treatment and disposal system issue was undertaken primarily to educate the participants and to explore the problem of gathering a representative group of people and interesting them in the problem. The results could provide important material for gaming techniques and background information for deliberations using a variety of methods of information exchange and analysis.

Regional Opportunities, Problems, and Planning Strategies

A Delphi methodology was used to generate and evaluate suggestions regarding regional opportunities, problems, and planning strategies. The group summaries represent initial individual judgments in terms of a Delphi methodology in that these items were suggested on one round and evaluated on a subsequent round, but not subjected to iterative cycles of reassessments based on statistical feedback. However, many of the assessments were influenced by prior consideration of the following in other phases of the Delphi exercises:

Listed below are several waste-water treatment methods and several disposal techniques which respondents have suggested can be combined to make up six alternative waste-water treatment and disposal systems. You are requested to score each alternative system for its suitability in a region similar to the Grand Traverse Bay area. Two different dates are indicated for the start of construction. For each starting date make two evaluations. First, consider all factors, but particularly the technology available at the start of construction, and second, consider only 10-year operating costs after the system becomes operational. For each evaluation give 100 points to your first choice and a portion of 100 points to each alternative according to its value relative to your first choice.

WASTE-WATER TREATMENT	DISPOSAL	ALTERNATIVE	RECOMMENDED APPROACHES: Start July 1972 — Summaries: Total Panel	Summaries: Your Last Estimate	Next Estimates: All Factors (1)	Next Estimates: Cost Only (2)	Start July 1976 — Summaries: Total Panel	Summaries: Your Last Estimate	Next Estimates: All Factors (1)	Next Estimates: Cost Only (2)
1 Physico-chemical	A Discharge to bay	1A	84				87			
2 Activated sludge or trickling filter biological treatment (sludge disposed of by incineration or land disposal)	A Discharge to scenic river	2A	38				35			
	B Discharge to bay	2B	82				84			
	C Spray irrigation	2C	65				60			
3 Biological treatment followed by "tertiary" treatment	A Discharge to scenic river	3A	55				49			
	B Discharge to bay	3B	100				100			

Please mark the one phrase that comes closest to expressing your familiarity with waste-water treatment and disposal techniques and with the Grand Traverse Bay region.

Waste-Water Treatment/Disposal

(1) ☐
(2) ☐
(3) ☐
(4) ☐
(5) ☐

G. T. B. Region

(1) ☐
(2) ☐
(3) ☐
(4) ☐
(5) ☐

Descriptive Words for Numerical Rating Scale

(1) = Totally unfamiliar
(2) = Casually acquainted
(3) = Well acquainted
(4) = Generally familiar
(5) = Actively studying

1) the trends of statistical measures which have traditionally been used to describe social and economic development; (2) the probabilities and importance associated with potential technical, social, economic, and political developments; (3) the relative importance of future sources of pollution; and (4) alternative waste-water treatment and disposal systems.

On the final round a list of suggestions regarding opportunities, problems, and planning strategies was presented to the broader-based panels. Panel members were asked to indicate whether an individual item should be singled out for special consideration by regional planners using a six-point scale and associated descriptive words as shown in Fig. 5. In interpreting the group means a value of 3.5 was viewed as the neutral point. The boundaries for the descriptive phrases are as shown.

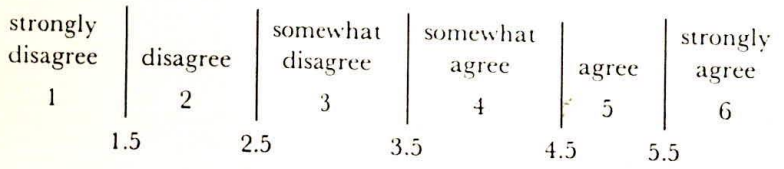


Fig. 5. Singling out individual item for special consideration.

Although the primary interest in these exercises was to identify areas of disagreement and the underlying reasons for them, a Delphi inquiry provides an accounting of the complete set of items that was considered by the respondents—an important concept when an interdisciplinary team of researchers is involved.

Evaluation of the Methodology

There is far from universal agreement on the merits of the Delphi techniques. Rand believes that Delphi marks the beginning of a whole new field of research, which it labels "opinion technology"[5]. However, a paper presented at the joint statistical meetings of the American Statistical Association in August 1971 described the Delphi techniques as the antithesis of scientific forecasting and of questionable practical credibility [6].

According to a recent *Wall Street Journal* article, the Delphi technique is gaining rather widespread use in technological forecasting and corporate planning, although the same article cautions:

> It's easy enough to see the shortcomings of the Delphi procedure; it's much harder to rectify them, as many are struggling to do. Remedial work must be done if the method is to be used in good conscience [7].

The Sea Grant Delphi exercises offered an exceptional opportunity for a critical evaluation of the Delphi techniques in an operational environment. The panelists—the main resource in evaluating the methodology—were interested in the improvement of techniques to integrate the judgments of a multidisciplinary research team and to convey its informed insights to society. Their evaluations were not biased by a strong emotional involvement in the success of the Delphi exercises, as has been true with many of the individual assessments of the method that have been published. From both a program budgeting standpoint and demands on researchers' time, the Delphi exercises competed with a wide variety of other methods for securing and disseminating information.

The primary instrument in evaluating the effectiveness of the method and its potential in other applications was a formal questionnaire. It was developed almost entirely by the respondents themselves using the Delphi technique of feeding back collated individual suggestions to generate additional suggestions. This procedure somewhat reduces the vulnerability of the questionnaire to the biases and shortcomings of the investigator. The six-point scale and associated descriptive words shown in Fig. 5 were used to quantify degrees of agreement and disagreement. To supplement the formal questionnaire, over thirty-five interviews with panelists were conducted.

Summaries were made for the three general groups participating in the Sea Grant Delphi exercises: technicians (Group I), behaviorists (Group II), and decisionmakers (Group III). For some issues the summaries for technical panelists under forty years of age and panelists with previous experience with the Delphi method were shown. Using the sample results, tests of significance were made to test the hypothesis that the distributions of the judgments of the Delphi method are homogeneous across the groups (the test procedure was based on the chi-square test statistic) and to test the null hypothesis that the means of the judgments of the population represented by the groups are identical (based on analysis of variance and the F-test). The results of these tests were used to support the discovery of basic differences in judgments made by different groups which had been formed on the basis of similar backgrounds and experiences. Their evaluations provided evidence that the method is effective not only in its designed role but in two other roles that are important and challenging from a management standpoint: encouraging greater involvement and facilitating communication between researchers and decisionmakers. The evaluations also showed that among the carefully selected samples of people the techniques were more highly regarded among groups which were formed on the basis of broad ranges in training and experience than among technicians—the group most administrators of the techniques have focused on.

The reliability of the method was demonstrated by the fact that the performance of the respondents, as measured by group statistical summaries, was similar for the three groups. Respondents from all three groups were generally willing to suggest future developments, sources of pollution, and research priorities; to utilize scaled descriptors to quantify subjective judgments; to accept a statistical aggregation of weights supplied by a group; and to reassess their judgments on the basis of feedback of information supplied by the group.

Some insight into the nature of the difference between judgments based on panelists' experience in the Sea Grant Delphi inquiries and the panelists' conception of an ideal application of the Delphi techniques can be gained by examining a cumulative summary of the evaluation of the effectiveness of the method in three specific roles shown in Table 2.

Table 2
Comparison of Evaluations Based on Experience in Sea Grant Exercises with Evaluations Based on the Potential of the Delphi Method

Group/Agreement Code =			Distribution of Judgments (proportion of total)								
			1	2	3	4	5	6	n	$\bar{x}$	σ
I	Technicians	Experience	.11	.14	.08	.44	.19	.03	36	3.6	1.340
		Potential	.05	.05	.16	.39	.30	.05	57	4.0	1.172
II	Behaviorists	Experience	.00	.00	.10	.65	.15	.10	20	4.3	.696
		Potential	.00	.00	.04	.15	.42	.32	24	5.1	.859
III	Decisionmakers	Experience	.00	.00	.00	.33	.37	.30	30	5.0	.809
		Potential	.00	.00	.00	.13	.50	.37	30	5.2	.679
II, & III	All Respondents	Experience	.05	.06	.06	.45	.24	.14	83	4.2	1.218
		Potential	.03	.03	.09	.27	.40	.19	111	4.6	1.101
	Respondents under 40	Experience	.07	.02	.07	.54	.22	.07	41	4.0	1.172
		Potential	.05	.00	.09	.21	.45	.20	56	4.6	1.218
	Previous Delphi Experience	Experience	.00	.00	.08	.49	.31	.13	39	4.5	.941
		Potential	.00	.00	.11	.17	.45	.28	47	4.9	.938

strongly disagree.
strongly agree.

The greatest difference in judgments based on experience and potential was found in the behaviorists' group, and the least among the decisionmakers. On the basis of the average judgments for all respondents the Sea Grant Delphi exercises corresponded rather closely to the panelists' conception of an ideal treatment, and there is a similar spread in the judgments of the subgroup which had previous experience with the method.

Table 3
Effectiveness of Delphi in Obtaining, Combining, and Displaying the Opinion of Informed People

Group/Agreement Code =			Distribution of Judgments (proportion of total)								
			1	2	3	4	5	6	n	$\bar{x}$	σ
I	Technicians	Experience	.11	.11	.06	.50	.22	.00	18	3.6	1.290
		Potential	.05	.05	.11	.42	.32	.05	19	4.1	1.177
II	Behaviorists	Experience	.00	.00	.00	.72	.29	.00	7	4.3	.488
		Potential	.00	.00	.00	.43	.43	.14	7	4.7	.756
III	Decisionmakers	Experience	.00	.00	.00	.22	.22	.56	9	5.3	.675
		Potential	.00	.00	.00	.10	.50	.40	10	5.3	.675
I, II, & III	All Respondents	Experience	.06	.06	.03	.47	.23	.15	34	4.2	1.274
		Potential	.03	.03	.06	.33	.39	.17	36	4.5	1.108
	Respondents under 40	Experience	.06	.00	.00	.69	.19	.06	16	4.1	1.025
		Potential	.05	.00	.00	.21	.58	.16	19	4.7	1.098
	Previous Delphi Experience	Experience	.00	.07	.00	.47	.27	.20	15	4.5	1.060
		Potential	.00	.00	.06	.19	.50	.25	16	4.9	.085
Combining Groups II and III:			$P(F > 4.6233) = .0165$								
			$P(\chi^2 > 8.3588) = .2130$								
			$P(\chi^2 > 3.959) = .0464$								

1: strongly disagree.
6: strongly agree.

A summary of the evaluation of the effectiveness of the Delphi method in obtaining, combining, and displaying the opinions of informed people is shown in Table 3. It indicates that the technicians, on the average, agreed somewhat that the method was effective compared to alternative methods. However, there

considerable dispersion in their estimates; some respondents strongly dis-ee that the method was effective in this role. The behaviorists agreed that it is effective in the role, and the decisionmakers displayed strong agreement. persion in the estimates of the two latter groups was much less than it was in technicians' estimates. An analysis of variance on these data gives a value of .233 for the F statistic. The probability of obtaining an F value larger than 233 when the groups are identical is .0165. A chi-square analysis gives a ue for chi-square of 8.3588. The probability of obtaining a larger value, en in fact the distributions come from a homogeneous population, is .2130— descriptive level of significance. A combination of the judgments of the isionmakers and the behaviorists resulted in $P(\chi^2 > 9.165) = .0025$.

valuation of the method's effectiveness in encouraging greater involvement Sea Grant activities provided similar results. The average judgments of ctiveness were $\bar{X}_T = 4.0$, $\bar{X}_B = 4.7$, and $\bar{X}_{DM} = 5.1$, and the dispersion in the mates of the technicians was much greater than for the other two groups. descriptive levels of significance are .0026 based on the F distribution, and 09 based on the chi-square distribution.

erage judgments regarding the method's effectiveness in communicating rmation to regional planners and decisionmakers were $\bar{X}_T = 4.1$, $\bar{X}_B = 5.4$, d $\bar{X}_{DM} = 5.3$. The judgments of the behaviorists were exceptionally high and reflect this group's special concern for the psychological and sociological iers associated with alternate methods. The descriptive level of significance 0003 based on the F distribution, and .0108 based on the chi-square ribution.

or all three roles there appears to be little difference in the average mates aggregated according to the respondent's age and the average esti-es of all respondents. However, respondents with previous Delphi ex-ience showed substantially higher average estimates of the method's effec-eness in obtaining, combining, and displaying subjective judgments and in ouraging involvement.

Average judgments regarding specific applications that are appropriate for a lphi methodology are shown in Table 4. The support of specific applications generally strongest among the decisionmakers and weakest among the hnicians. Only the behaviorists and decisionmakers evaluated the method as aid in decisionmaking, and both groups supported its use in this role.

the evaluation of positive and negative aspects of the method that had n suggested by respondents, the panels agreed that there should be more phasis on the idea that an expert should "not feel obligated to express an ion on every issue." However, the Sea Grant Delphi inquiries stressed the cepts of a systems approach and multidisciplinary teams. Therefore it was ired that each respondent consider all items and attempt estimates on those which he had some familiarity. A self-evaluation index was provided so a panelist could assess his competence regarding each item. The compe-ce index was a control factor in developing the statistical summaries that

were part of the information feedback. This procedure allows an informed person to evaluate such things as relative importance and desirability—evaluations which he can make without being an expert in the area—and gives the administrator additional assurance that panelists considered items outside their specialized areas. In addition, there was some interest in comparing the estimates of experts and nonexperts on specific issues.

Table 4
Suitable Applications for a Delphi Methodology

Application	Technicians Mean	Technicians σ	Behaviorists Mean	Behaviorists σ	Decisionmakers Mean	Decisionmakers σ
Develop Criteria to Recruit Industries	4.3	.996	4.5	1.130	5.3	.675
Establish Program Priorities within a Committee	3.9	1.393	4.6	1.453	5.0	1.080
Long-range Planning by a University	3.8	1.393	4.1	1.453	4.8	1.080
Education of Practicing Politicians	4.1	.793	4.3	.707	4.6	1.424
Budgeting of an Interdisciplinary Program	3.9	1.353	3.7	.866	4.8	.789
Situations Dealing with the Future, Uncertainty, Conflicting Views	4.5	.943	4.3	1.165	4.5	.707
Decisionmaking Aid	–	–	5.1	.835	5.0	1.054

The suggestion that the method "can result in a manipulated and arbitrary consensus" received a neutral judgment from all three groups, perhaps indicating that the respondents felt this danger to be no greater than it would be in alternative techniques for securing group judgments. However, it is this administrator's opinion that the Delphi techniques could be a powerful tool for manipulating opinion and policy [8].

IV. Summary of Findings and Recommendations

The design and administration of a Delphi exercise in which the concept of a multidisciplinary team and a systems approach is desired can best be handled as a project within a professional research organization. The scope of the exercise is generally determined by the respondents, and, as interesting and unexpected issues are suggested, flexibility is needed in designing evaluation matrices and in determining the composition of the panels. Experts knowledge-

able in specialized areas should be available on an adhoc basis to formulate questions and collate responses in order to minimize redundancy and ambiguity. The demand for their services in the course of a Delphi exercise is very uneven, as is the need for designing, editing, typing, and distribution services. There are significant start-up and learning costs associated with the Delphi techniques that can be justified only if the technique will become a routine management tool to be used on a continuing basis. This is particularly true if the benefits of computer processing are to be realized.

The following are some general observations that are consistent with the items suggested and evaluated by respondents in the Sea Grant exercises and with the information gained in personal interviews with panelists.

- Respondents will be more receptive if the techniques are tailored to specific groups on the basis of their training and experience.
- The administrator should consistently emphasize the distinction between the characteristics of a Delphi interrogation and those of conventional questionnaires and polls.
- Panelists—particularly those with technical backgrounds—must be *convinced* that judgments often have to be made about issues *before* all facets of the problems have been researched and analyzed to the extent they would like. (For these situations they must be persuaded that their subjective judgments may be a decisionmaker's most valuable source of information.)

There are several procedural recommendations that may be helpful to designers and administrators of future Delphi exercises.

- Interpersonal techniques, such as interviews and seminars, should be interspersed with the rounds of questionnaires and information feedback.
- The source of a suggested item should be identified (for example, panel member number and basic biographical information), taking care not to compromise the anonymity of specific inputs.
- Standardized scaled measures should be available to a respondent so that he can qualify his response to specific questions. Such measures are relative competence in a technical area, familiarity with a geographical region, or confidence in an estimate.
- If a multidisciplinary approach is desired, respondents should be encouraged to consider all items but to make estimates only on those scaled descriptive phrases with which he feels comfortable. For example, in these exercises it was helpful when respondents indicated their familiarity with a specialized area or the importance of an item even though they did not make probability estimates.
- The panelists should decide through their suggestions and evaluations what items should be considered. The criteria for retaining an item for further evaluation should be made clear at the outset of the exercise.

- Personal comments and arguments submitted by respondents should be part of the information feedback.

The Delphi inquiries have complemented the Michigan Sea Grant gaming-simulation activities by providing the following types of inputs:

- Data which can be helpful in describing social, economic, and political forces affecting the region's development during the next twenty years.
- Regional planning strategies, listed in order of preference for both university researchers and regional planners.
- Problems and issues which provide the link between the simulated regional area and a set of decision roles which are gamed [9].

Integration of a Delphi methodology with the Michigan Sea Grant gaming-simulation exercises will give them a more dynamic aspect and provide greater motivation for the participants. Some particularly interesting applications would be in cost-benefit analyses similar to those used in the Delphi inquiries to evaluate waste-water treatment and disposal systems, selection of research projects through an evaluation of effectiveness in terms of basic objectives and risk factors associated with various levels of funding, and the development of alternate scenarios for a region such as the Grand Traverse Bay watershed area.

According to Michael [10], Delphi inquiries and gaming-simulation exercises are techniques for introducing customers in a nonthreatening way to a more complex way of thinking and a better way of perceiving their needs in terms of the kind of knowledge we have. Knowledge utilization depends upon discovering the nature of the awareness of the problem among potential customers both as a set of variables and as a system of interrelationships; getting new knowledge absorbed by individuals and then by the organizations these individuals are in; and developing a capability and inclination to plan rather then employing an ad lib approach.

Concerned citizens were included as panelists in the Delphi inquiries not only for the purpose of informing them but also to accord the other panel members the benefits of the citizens' knowledge of the area, to take into account political and institutional considerations, and to communicate findings in such a way that the acceptance and implementation of policies and actions on which there appeared to be a reasonable consensus would be encouraged. The behavioral sciences provide support for this type of approach in effective communication [11,12].

The Michigan Delphi inquiries have provided some carefully formulated judgments of a multidisciplinary team of researchers and potential users of research data regarding: the importance and effects of technical, social, economic, and political developments; sources of pollution and recommended waste-water treatment and disposal systems; and regional opportunities, problems, and planning strategies. More important, a critical evaluation of the

method has shown the potential of a Delphi inquiry for improving the dialogue between researchers and regional problem solvers. It has also provided empirical evidence to support further investigation of several innovations which may bring the methodology closer to Jantsch's idealized concept of a forecasting technique, wherein exploratory and normative components are joined in an iterative cycle in which informed citizens can work with researchers in planning for the future [13].

References

1. Murray Turoff, "Delphi and Its Potential Impact on Information Systems," Paper 81, Proceedings of the Fall Joint Computer Conference, Vol. 39. AFIPS Press, Washington, D. C., November 1971.
2. The techniques and procedures used in this series of interrogations and information feedback are similar to those described in "Some Potential Societal Developments—1970–2000" by Raoul De Brigard and Olaf Helmer, IFF Report R-7. Institute for the Future, Middletown, Conn., April 1970.
3. In one phase of the Delphi inquiries panelists were asked to assign numerical probabilities to commonly used words or phrases to indicate the likelihood of an event. The Delphi technique of reassessment based on the feedback of a group response was extremely effective in narrowing the dispersion of the estimates. Verbal phrases associated with numerical probabilities were believed to encourage respondents to think about a probability scale in similar terms and might be more appropriate than numerical probabilities in expressing the likelihood of socioeconomic developments.
4. The Michigan Delphi inquiries provided empirical evidence that the feedback and reassessment techniques which are inherent in the basic Delphi method reduced the number of inconsistencies in personal estimates of probability as the rounds progressed. It also indicated a tendency for a learning "curve" for respondents with respect to the technique itself.
5. "Forecasters Turn to Group Guesswork," *Business Week*, March 14, 1970.
6. Gordon A. Welty, "A Critique of the Delphi Technique" (summary of paper presented at the Joint Statistical Meetings of the American Statistical Association, Colorado State University, Fort Collins, Colorado, Aug. 23–26, 1971).
7. "Futuriasis: Epidemic of the '70s," *Wall Street Journal*, May 11, 1971.
8. For a discussion of the dangers associated with a Delphi devoted to policy issues see Turoff's article, above (Chapter III. B. 1.). For a discussion of deliberate distortion see "A Critique of Some Long-Range Forecasting Developments," by Gordon Welty (paper presented at 38th session of the International Statistical Institute, Washington, D. C., August 1971).
9. The gaming-simulation concept for the Sea Grant Program is presented in "Developing Alternative Management Policies," unpublished report, University of Michigan Sea Grant Office, 1971.
10. Donald Michael is presently program director, Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan. These are a summary of his unpublished remarks to a site visit team of government officials and academicians in Ann Arbor, Michigan, March 4, 1972.
11. Douglas McGregor, "The Professional Manager" (New York: Harper & Row, 1967), p. 153: "...My conception of a two-way communication is that it is a process of mutual influence. If the communicator begins with the conviction that his position is right and must prevail the process is not transactional but coercive."
12. Peter F. Drucker, "Technology, Management and Society" (New York: McGraw-Hill Book Co., 1970), pp. 22–23: "... They must understand it because they have been through it, rather than accept it because it is being explained to them."
13. Erich Jantsch, "Technological Forecasting in Perspective," Organization for Economic Cooperation and Development, Paris, 1967.

Other References

Examples of the key instruments used in the Michigan Sea Grant Delphi inquiries can be found in "Substantive Results in the University of Michigan's Sea Grant Delphi Inquiry," by John D. Ludlow, Sea Grant Technical Report No. 23, University of Michigan, Ann Arbor, 1972, and in "Evaluation of Methodology in the University of Michigan's Sea Grant Delphi Inquiry," by John D. Ludlow, Sea Grant Technical Report No. 22, University of Michigan, Ann Arbor, 1972. Complete sets of the information packages can be found in "A Systems Approach to the Utilization of Experts in Technological and Environmental Forecasting," by John D. Ludlow, Ph.D. dissertation for the University of Michigan, 1971. Available through University Microfilms, Inc., Ann Arbor.

B. 3. The National Drug-Abuse Policy Delphi: Progress Report and Findings to Date

IRENE ANNE JILLSON

Introduction

Rationale

Although the abuse of drugs has been recognized in this country for nearly one hundred years, its popularity as a national problem has resurfaced relatively recently. In 1968, former President Nixon declared drug abuse "public enemy number 1"; from 1969 until 1974, some $2.4 billion in federal funds were obligated to combat the problem, and an industry was created. This expenditure represents funding of programs that were almost exclusively aimed at heroin abuse, rather than the broad spectrum of drug abuse. Since 1968, drug abuse has been the subject of intense public and private debate. Controversy over the government's appropriate response has ranged from debate regarding drug laws (to what extent different drugs should be controlled, and in what manner), to the basic question of whether or not alcohol programs are to be included with "other" drug programs on the federal level.

During the past ten years, the increased concern and expansion of drug-abuse prevention programs has resulted in a swelling of the ranks of professionals who have developed expertise in this field; however, the use of these experts in policy advice and formulation has been sporadic and unsystematic. At the same time, numerous research and evaluation studies of drug-abuse prevention programs themselves have been carried out. The degree to which resultant data from these studies can be, have been, or should be, used in decision-making at the nation level has never been resolved.

In the fall of 1973, it was clear that the problem of drug abuse had diminished in priority, and that substantial reductions in federal funding were imminent. This may be attributed primarily to the apparent abatement of the heroin epidemic, which had served as the stimulus for increased concern and program funding in the late 1960s. Although the crisis associated with the heroin epidemic may have passed, it is by no means clear that the broader problem of drug abuse has been resolved: polydrug use and alcoholism appear to be increasing; and the abuse of prescription drugs, once the "hidden drug problem," is surfacing in many communities. Such a time of decreased public

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

interest and funding for a yet-existing problem calls for careful consideration of the basic issues, and deliberation of the strategies to be followed, in order to maximize effective use of resources available.

Viewed from this perspective, the procedures utilized by Policy Delphi studies seemed most appropriate to explore national drug-abuse policy options for the next five years. The volatility of many of the issues, most of which involve fundamental value and sometimes moral choices; the diverse backgrounds of those who make or influence policy; the apparent differences in the positions held by various experts and groups; and the apparent inability of past policy studies to aid decision-makers led to the conceptualization and implementation of a national drug-abuse Policy Delphi study. The unusually high response rate, the degree of participation achieved to date, and the interest on the part of federal and state decision-makers has borne out this initial hypothesis.

History

The study described in this chapter was originally conceived in 1973, and designed during the fall of that year. Implementation began in December 1973; the first questionnaire was disseminated in March 1974. The first two questionnaires were developed under a contract funded by the National Institute of Drug Abuse.[1] Analysis of the data generated by the second questionnaire, and the further development of the use of the Delphi procedures in the exploration of national drug-abuse policy, as originally conceived, was sponsored by the National Coordinating Council on Drug Education.

The study analyzing the first two rounds was completed and published in December 1974.

Objectives of This Study

There are three primary objectives for this study:

- To develop a range of possible national drug-abuse policy options
- to explore applications of the Policy Delphi methodology to this and other areas of social policy
- to explore the possibilities of applying the technique on an as-needed basis and on an ongoing basis

Since the level of drug abuse in the United States is presumed to be both endemic and epidemic, and since strategies to respond to changes in use patterns need to be both immediate and long range, this study is concerned with ascertaining the feasibility of utilizing the Delphi technique to meet these needs of policy formulation and planning.

- *On an as-needed basis.* This would involve the use of a panel of experts who would respond to queries sent as the need arises. For example, if a decision-maker were to be

[1]NIDA Contract Number B2C-5352/HOIMA-2-5352.

informed that there was a dramatic decline in the number of patients entering treatment programs, a Delphi would be developed to determine the opinion of selected experts with regard to this particular trend.

- *On an ongoing basis.* If one agrees that there is an endemic level of drug abuse, then it would seem appropriate to develop an ongoing Delphi study of indefinite duration. This Delphi would be implemented such that questionnaires would be distributed at regular intervals. Since the panel would be of indefinite duration, membership might be fluid. Current trends in the field would be incorporated into the questionnaires, so that there would be a continuous flow of information for the use of the policymaker, and those operating programs of various disciplines in the field.

Study Design

A number of advisers—experts in the field of drug abuse, policy planning and analysis, and the Delphi technique in particular—assisted in developing the study design.To date, they have continued to provide assistance in all aspects of the study. Included are Dr. Norman Dalkey, Engineering Systems Department, University of California at Los Angeles, who originally was instrumental in developing the technique in the late 1950's and who continues to explore its use as part of his decision-theory research; and Dr. Murray Turoff, who has developed the Policy Delphi and is co-editor of this book. Dr. Peter Goldschmidt has assisted in planning and management of this study, and M. Alexander Stiffman has assisted in developing the analytic approach and designing the computer analysis. Dr. Raymond Knowles, Charlton Price, and Anthony Siciargo have assisted in pretesting the questionnaires.

The final study design was based on the premise that policy may be formulated from a number of different perspectives. In designing this Policy Delphi study we are exploring the formulation of national drug-abuse policy from three perspectives:

- the "top-down" approach—establishing national drug-abuse policy objectives to be achieved over the next five years, based on the respondents' value systems
- a "bottom-up" approach—identifying factors which control the transition between general population and various degrees of drug use; deciding which of these are important and can be affected by national drug-abuse policy; and determining appropriate policies to affect them
- an issue-oriented approach—deriving policies from issues which are the subject of current controversy or debate

The underlying thesis in this design is that different decision-makers may formulate policy predominately using one or another perspective, and that this results in distinct types of policy options and considerations which may appear attractive from one perspective, but turn out to be counterproductive from another. For example, setting an objective using the "top-down" approach may result in its achievement becoming a political issue, while an objective that is formulated as the result of a political issue may never be achieved because it is technically impossible even though its achievement is valued.

This top-down, bottom-up approach speaks to the general concept of the Policy Delphi, in that alternative policy options are drawn from a number of different vantage points (it is not simply a statement of objectives, for example) Additionally, we are not interested in consensus per se, but rather, in exploring alternatives, and *pro* and *contra* arguments for the alternatives.

For these reasons the outputs from the three approaches used in this study will be brought together so that the panel can identify more appropriate alternatives. We anticipate that this will prove more fruitful in terms of exploring national policy options than either approach alone. A description of the approach, round by round, is given in Table 1.

Respondent Population

A list of more than one hundred highly selected potential respondents was developed from among the most notable "experts" in the field and from those who directly impacted on the field (e.g., police chiefs). Invitations to participate in the study were sent to forty-five persons; the remaining names were held in reserve, as a second series of invitations was anticipated in order to secure twenty-five participants. In fact, thirty-eight individuals (84%) responded positively. Since that time, three respondents have withdrawn from the study owing to a change in career orientation from drug abuse. Needless to say, there was no necessity for a second series of invitations.

As the study progressed past the first two rounds, several additional respondents were added. These additional respondents were selected to represent areas of interest which had developed in the study, but for which respondents had not been initially selected. Experts in alcoholism were added to the panel, for example, because a significant proportion of existing panelists expressed the view that a national drug-abuse policy could not be considered separately from alcoholism. The addition of such experts will allow their views to be added to those of the present panelists, and so provide an appropriate additional perspective. The present panel consists of thirty-nine persons.

Our respondent group represents some of the most respected authorities in the field. They include the Deputy Director of the Alcohol, Drug Abuse and Mental Health Administration, a former director of the Bureau of Narcotics and Dangerous Drugs, officials from the Office of the Secretary, and Office of the Assistant Secretary for Health of H.E.W., notable researchers, treatment administrators, law-enforcement officials, and policymakers in the field of drug abuse. It should be emphasized that participation is voluntary, and that no honorarium is paid to respondents.

The Questionnaires

First Questionnaire

A number of study approaches were explored during the developmental phases preparatory to Round One. The final draft of the resultant questionnaire was

Table 1

THE STUDY DESIGN

Perspective	First Questionnaire	Second Questionnaire	Third Questionnaire	Fourth Questionnaire	Fifth Questionnaire	Final Summary
Objectives ("top down")	List objectives	Rate feasibility, and desirability of objectives	Re-rate feasibility and desirability of selected objectives			
	List key indicators	Expand key indicators, perform initial rating				
Transition Factors ("bottom up")	List factors, indicate direction		Final rating of factors Develop policy areas to affect important transition factors which can be influenced by national policy	Identify any contradiction in objectives and policies formulated by the three approaches	Synthesize a consistent and realistic set of national drug abuse policy options	Write a brief summary of the national drug abuse policy options identified by the study, including the normative forecasts for the key indicators Identify future policy research needs
Policy Issue Statements ("political")	Rate selected policy issues; develop for and against arguments Add other important policy issues	Final rating of selected policy issues; rate for and against arguments Rate and give arguments for and against additional policy issue statements				

that the time to complete the questionnaire was decreased considerably by deleting this section.

The first-round questionnaire, disseminated in mid-March, consisted of three primary sections:

1. *Development of Objectives.* The respondents were asked to develop up to five national policy objectives in the field of drug abuse, given a five-year time frame; and to list up to three key indicators for each objective.
2. *Transition Matrix.* Respondents were given a simplified transition model which depicted the flow from one state of drug involvement to another and were asked to list the factors which promoted or inhibited the movement of people from one state to another. (A detailed description of this matrix is provided in a subsequent section.)
3. *Policy Issue Statements.* Twelve issues, in the form of should/should not statements, were posed. These had been culled from a potential list of twenty-five issues felt to be current and controversial.
4. *Additional Items.* These included a self-rated expertise question; two questions relating to expectations and objectives for participating in the study; a request for responses to a list of definitions; and a request for feedback on questionnaire design and/or content.

Twenty-four of the study's thirty-five respondents (69%) actually completed the Round One questionnaire. The breakdown of returns by category of respondent is shown in Table 2. It should be pointed out that mailing posed serious difficulties both in disseminating and in returning the questionnaires. In several cases it took two full weeks for the questionnaires to arrive by air mail to their destinations; several never arrived, and duplicate packages had to be sent out. For this reason, the deadline for completion of the first-round questionnaires was extended by two weeks.

The absolute range of time for completion of the first questionnaire was fifteen minutes to ten hours. The interquartile range was one to three hours; the median time for completion was $2\frac{1}{4}$ hours, which was approximately the median we had anticipated. The respondents included in our predesignated *policymakers* subpanel category spent a median of $3\frac{1}{2}$ hours; the median for all the remaining participants was $2\frac{1}{2}$ hours.

In addition to the substantive-issue questions, respondents were asked to self-rate their expertise in ten drug-related areas (see Table 3). These data were then used to cross-check the categories (subpanels) into which respondents had been placed, and to further analyze responses to particular items. There is one particular point of interest to be noted. As critical as evaluation is held to be in the formulation of national policy, not one of the policymakers rated himself herself as expert in this area.

Respondents were also asked to indicate their expectations for the study, and to list personal objectives with regard to participation. Eighty-two percent of those responding reported that they expected the study to be of direct benefit.

Table 2

PARTICIPATION AND RETURNS BY CATEGORIES OF RESPONDENTS (SUBPANEL)*

Subpanel	Invited	Agreed to Participate	Formally Withdrew	Delphi Panel	Completed Round One	Completed Round Two
Policy Makers	9	9	0	9	5	5
Researchers	12	9	0	9	9	9
Treatment Program Administrators	11	11	2	9	4	6
Criminal Justice Administrators	9	6	1	5	3	3
Other	4	3	0	3	3	2
Total	45	38	3	35	24	25

* As of 1 August, 1974. Does not include panelists added after that date.

Table 3

A SUMMARY OF SELF-RATED EXPERTISE BY AREA AND SUBPANEL*

Area of Expertise	Subpanel: Percent Who Rated Themselves "Expert"(1) Policy Makers	Researchers	C.J.S.(2) Administrators	T.P.(2) Administrators	Other	All Respondents
National Drug Abuse Policy Planning	100	33	33	0	100	50
Prevention	20	22	33	50	0	26
Intervention	0	33	33	75	0	30
Treatment	40	78	0	100	0	57
Law Enforcement	20	11	100	0	33	25
Research	20	100	0	50	0	50
Evaluation	0	78	0	50	33	44
Training	0	33	33	50	0	26
Education	20	56	33	50	0	39
Pharmacology	0	11	0	0	0	4
Number of Respondents	5	9	3	4	3	24

* As of 1 August, 1974. Does not include panelists added after that date.
(1) Percentage of those who responded.
(2) C.J.S. = Criminal Justice System; T.P. = Treatment Program

There were twenty-eight different narrative responses to this item. Twenty-two saw positive utilization of the results for policy planning, idea exchange, consensus development, etc.; three were uncertain of usefulness, and two were skeptical of its utility. There were fifty-four statements of personal-study objectives; these were distilled to twelve clusters, and are shown as Table 4.

Respondents were asked to comment on a list of standard definitions which had been prepared for use in the study. Twelve respondents commented on these definitions. As anticipated, there was little agreement with these definitions on the part of those who commented. Apart from giving each respondent a common base line, one of the reasons for including the list of definitions was to determine what interest would be aroused. In most of the social-service areas, and particularly in the field of drug abuse, there is much dissension, even among policymakers, regarding critical definitions (e.g., drug abuse, modality types). Clearly it is not sufficient to gloss over this issue continuously in the hopes that at some time in the future standard definitions will somehow be devised and agreed upon by a reasonable majority of those in the drug-abuse field. A Delphi study specifically related to the formulation of standard definitions is presently being planned.

Second Questionnaire

Preparation of the second questionnaire began shortly after the first completed Round One questionnaires had been received. It was decided that, because of the complexity and time required for completion of the transition matrix, this section would be deleted from the second questionnaire and included as part of a subsequent round. The second questionnaire and a summary of the first round results were disseminated in mid-May.

The second questionnaire included two sections:

1. *National Drug-Abuse Policy Objectives*. Respondents were asked to rate fifty-five objectives on the basis of feasibility and desirability, and to rate the importance of the key indicators associated with them.
2. *Policy Issue Statements*. Respondents were asked to re-rate the original twelve issue statements; rate the narrative comments associated with them; and rate the fifteen new issues suggested by respondents.

There is usually a decrease in response rates for the second round of a Delphi study, particularly those involving voluntary participation. For this reason, and because of the length of the Round Two questionnaire, we anticipated a response rate of approximately 45 percent to 50 percent. In fact, twenty-five respondents, 71%) completed the questionnaire; a most unusual and gratifying response rate, and one higher than that for Round One. (See Table 2.)

The absolute range of time to complete the second questionnaire was $1\frac{1}{2}$ to $8\frac{1}{2}$ hours; the interquartile range was $2\frac{1}{2}$ to five hours; the median time to complete was three hours. For the *policymakers*' subpanel, the median was $5\frac{1}{4}$ hours; for all others the median was three hours.

Table 4

RESPONDENTS' OBJECTIVES FOR PARTICIPATING IN THE STUDY

To test the Delphi technique; determine the value of the process in sharpening views in social policy fields or in bringing forth practical ideas or new insights.

To explore the limits of possible public policy formulation; to learn more about policy formulation.

To be involved in the formulation and development of drug abuse policy; influence policy through identification of critical policy distructions; to share in a process that may lead to wiser process than we now have in the drug abuse area; help develop policy view in a nonpolitical forum.

To see if any consensus is possible in drug abuse policy; see if the group can solve the problem or give direction.

To distill and synthesize the collective thinking of some of the best minds in drug abuse; obtain the benefit of the ideas of the others as a stimulus to my own thinking; to learn what the group knows about, and applies in responding to drug abuse.

To assess the extent to which my views in drug abuse coincide with or differ from those of my colleagues; check my own opinions against those of the group.

To clarify my own thinking pertaining to drug abuse policy; get a broader perspective.

To test ideas in a multidisciplinary quasianonymous environment.

To develop priorities for the organization or agency in which I work.

To make a useful contribution to knowledge; help solve a problem; provide ideas that may arise from my special knowledge and experience.

To gain personal enjoyment by responsive discussions.

To satisfy my sense of duty to participate in such studies as an active researcher and/or poli[illegible]

Third Questionnaire[2]

For this questionnaire, panelists were asked to respond to two series of questions:

1. *National Policy Objectives*. There were twenty-five objectives included in this section; these had to be re-rated because there was a broad distribution of voting responses, differences in voting between policy experts and policy nonexperts, or because original objectives had been combined or divided.
2. *Transition Matrix*. In this section, the transition factors suggested by respondents in the first questionnaire were further developed.

The policy issues were not included for consideration in this round in order to maintain expected time for completion to a reasonable level. The data from this third questionnaire, and information gathered during previous studies of this type, will be used as a basis for developing policy and program options in future rounds. The respondents were again sent two copies of the questionnaire, and an introduction and summary volume which included results of the previous questionnaire.

Results

Objectives

First Questionaire. In the first questionaire respondents were asked to list up to five national drug-abuse policy objectives (for a five-year time frame), and up to three key indicators for each of these. They listed a total of seventy-eight such objectives; from this list, fifty-five different objectives could be discerned. The 187 key indicators listed by the respondents were culled to 153.

Even after we distilled the original seventy eight objectives to fifty five, there still were similarities between them. Rather than risk a possible misinterpretation of the objectives as stated by the respondents, we decided not to distill the objectives any further. It is difficult to be precise in a Policy Delphi, particularly one in a field as complex as drug abuse; we have therefore emphasized development and creativity.

Second Questionaire. The respondents were asked to rate the fifty-five objectives derived from their responses to Round One Objectives Section. The two rating scales used were *feasibility* and *desirability*. They were also asked to rate the 153 indicators relating to each objective; in this case, the scale used was *importance*. These scales are shown in Table 5. The objectives were presented in arbitrarily defined categories (prevention, treatment, law enforcement, organization, e.g.) simply as a means of ease in completion; these categories had no other significance.

[2]First Round—National Drug-Abuse Policy Delphi.

Table 5

FEASIBILITY/PRACTICALITY SCALE

Scale Reference	Definitions
1. Definitely Feasible	Can be implemented No research and development work required (necessary technology is presently available) Definitely within available resources No major political roadblocks Will be acceptable to general public
2. Probably Feasible	Some indication this can be implemented Some research and development work required (existing technology needs to be expanded and/or adopted) Available resources would have to be supplemented Some political roadblocks Some indication this may be acceptable to the general public
3. May or May Not be Feasible	Contradictory evidence this can be implemented Indeterminable research and development effort needed (existing technology may be inadequate) Increase in available resources would be needed Political roadblocks Some indication this may not be acceptable to the general public
4. Probably Infeasible	Some indication this cannot be implemented Major research and development effort needed (existing technology is inadequate) Large scale increase in available resources would be needed Major political roadblocks Not acceptable to a large proportion of the general public
5. Definitely Infeasible	Cannot be implemented (unworkable) Basic research needed (no relevant technology exists, basic scientific knowledge lacking) Unprecedented allocation of resources would be needed Politically unacceptable Completely unacceptable to the general public

Table 5 (Continued)

DESIRABILITY/BENEFITS SCALE

Scale Reference	Definitions
1. Highly Desirable	Will have a positive effect and little or no negative effect Social benefits will far outweigh social costs Justifiable on its own merit Valued in and of itself
2. Desirable	Will have a positive effect with minimum negative effects Social benefits greater than social costs Justifiable in conjunction with other items Little value in and of itself
3. Neither Desirable nor Undesirable	Will have equal positive and negative effects Social benefits equals social costs May be justified in conjunction with other desirable or highly desirable items No value in and of itself
4. Undesirable	Will have a negative effect with little or no positive effect Social costs greater than social benefits May only be justified in conjunction with a highly desirable item Harmful in and of itself
5. Highly Undesirable	Will have major negative effect Social costs far outweigh any social benefit Not justifiable Extremely harmful in and of itself

Table 5 (Continued)

IMPORTANCE SCALE

Scale Reference	Definitions
1. Very Important	A most relevant point First order priority Has direct bearing on major issues Must be resolved, dealt with or treated
2. Important	Is relevant to the issue Second order priority Significant impact but not until other items are treated Does not have to be fully resolved
3. Moderately Important	May be relevant to the issue Third order priority May have impact May be a determining factor to major issue
4. Unimportant	Insignificantly relevant Low priority Has little impact Not a determining factor to major issue
5. Most Unimportant	No priority No relevance No measurable effect Should be dropped as an item to consider

The feasibility and desirability ratings of the objectives were analyzed so as to develop a summary list of objectives, and determine to what extent there was consensus or polarization.

Respondents' ratings on the feasibility and desirability of objectives were translated into group scores by summing the scale values and dividing the total by the number of ratings. This procedure, it should be noted, treats nominal scales as interval data. The feasibility and desirability scores were used to categorize objectives as follows:

Group Score	Feasibility	Desirability
Less than 1.80	Highly feasible	Highly desirable
Equal to or greater than 1.80 but less than 2.60	Feasible	Desirable
Equal to or greater than 2.60 but less than 3.40	May or may not be feasible	Neither desirable nor undesirable
Greater than 3.40 but less than or equal to 4.20	Probably infeasible	Undesirable
Greater than 4.20	Definitely infeasible	Highly undesirable

Objectives were first grouped on the basis of their feasibility and then sorted on the basis of their desirability. This produced the rating of objectives depicted in Table 6.

The twenty-five objectives which scored Highly Feasible and Highly Desirable, or Feasible and Highly Desirable, are shown as Tables 7 and 8, respectively.

No objectives were rated "Definitely Infeasible" and none was rated as either "Undesirable" or "Highly Undesirable." These results indicate that the majority (55%) of the objectives listed were rated at least "Feasible" and "Desirable." The following items or sets of items deserve special attention because of the distinctions in rating patterns. The feasibility of twenty-one objectives was indeterminable, either because there was polarization (with some respondents rating an objective feasible while others rated it infeasible); a broad distribution

Table 6

NATIONAL DRUG ABUSE POLICY OBJECTIVES: Summary of Feasibility and Desirability Ratings

FEASIBILITY \ DESIRABILITY	Highly Desirable	Desirable	Indeterminate Desirability	Undesirable	Highly Undesirable
Highly Feasible	5	0	0	0	0
Feasible	20	5	1	0	0
Indeterminate Feasibility	9	6	6	0	0
Infeasible	2	1	0	0	0
Highly Infeasible	0	0	0	0	0

Table 7

OBJECTIVES VOTED "HIGHLY FEASIBLE" AND "HIGHLY DESIRABLE"
(in decreasing order of desirability)

Objective	Feasibility Score	Desirability Score
To conduct research on treatment modalities and effectiveness.	1.50	1.14
To train in-line treatment personnel to enhance their skill in helping the drug dependent person.	1.77	1.27
To have available more adequate epidemiological estimates of a) prevalence, by drug type; b) incidence, by drug type and individual characteristics; c) discontinuance, by drug type.	1.73	1.41
To develop, validate and disseminate information on efficacious programs in vocational rehabilitation, and early intervention.	1.79	1.46
To increase research into brain chemistry and psycho-social correlates and clinical-social treatment.	1.70	1.70

Table 8

OBJECTIVES VOTED "FEASIBLE" AND "HIGHLY DESIRABLE"
(in decreasing order of desirability)

Objectives	Feasibility Score	Desirability Score
To increase the efficiency and quality of treatment for drugs and alcohol overuse, based on increased evaluative clinical research.	2.04	1.13
To measure effectiveness of different kinds of treatment treatment programs for different types of users.	2.13	1.13
To reduce overdose deaths and damage.	2.26	1.17
To establish a flexible by comprehensive response system.	2.24	1.19
To minimize the adverse consequences of drug abuse.	2.38	1.23
To measure the extent of the problem by scientific sampling with rigorous respect for reliability of data.	2.04	1.25
To provide enough effective treatment centers for all forms of drug abuse so that no abusers need continue because he has been denied treatment according to his needs.	1.91	1.27
To evaluate training programs.	1.20	1.27
To determine priorities in terms of short- and long-term strategies for achieving our goals.	2.41	1.32
To improve basic progress in social science research pertaining to drug abuse.	2.00	1.35
To assign funding on the basis of priorities.	2.41	1.38
To create an explicit strategy which defines the problem and which includes strategic objectives in order of priority.	2.32	1.41
To delineate the gaps remaining to be filled in -- in knowledge and understanding, through further experience and research.	2.[illegible]	1.46
To reallocate prevention and treatment endeavors independently of particular drug fashions, i.e., emphasis on heroin and not on alcohol so that efforts are aimed at individuals suffering from adverse habits of use of any psychoactive drugs or combination of these.	2.29	1.50*
To make available more trained health professionals.	1.86	1.52
To develop adequate alternative models in prevention and phased intervention.	2.38	1.67
To establish means of transmitting scientific findings to the public.	2.22	1.70
To increase law enforcement pressure on illicit drugs sold in <u>large</u> quantities.	2.00	1.71
To incorporate drug treatment into standard health delivery systems.	2.14	1.77
To increase public awareness of the difference between drug use (responsible) and abuse (irresponsible).	2.46	1.79

* indicates that over one-fifth of respondents did not respond to this item.

Table 9

OBJECTIVES WHICH EXHIBITED POLARIZATION OR A BROAD DISTRIBUTION OF RESPONSES (Feasibility and/or Desirability)

Objective	Scale Value; Percent voting				
	1	2	3	4	5
Polarized or broad distribution on Feasibility					
To achieve improvements in public education which will enhance the quality of the individual's life, strengthen his preference for leisure-time activity not involving the use of drugs, and enhance his capacity for meeting personal problems without resorting to the non-medical use of dangerous drugs.	0	38	25	25	12
To develop a national public discourse and standards on use and abuse of drugs.	14	27	27	23	9
To reduce non-prescribed use of psychoactive drugs.	8	34	21	29	8
To reduce the number of persons engaged in the non-medical or recreational use of dangerous drugs.	0	33	22	33	12
To utilize outreach and coercive techniques to engage those not entering treatment on their own.					
To consider social action approaches, as alternatives to treatment, including concern with housing, employment, education, and counselling, in addition to other novel approaches.	23	23	23	26	5
To reduce/eliminate criminal penalties for personal use and possession of drugs currently defined as "illegal".	4	13	35	35	13
To establish the principle that whether or not an individual uses a mind altering substance is a matter of personal choice with minimum governmental interference.	9	26	13	30	22
To organize an effective, coordinated approach to research, prevention, treatment and rehabilitation among all appropriate Federal, State, local and private agencies in the health and social services areas, thus developing the mechanism for timely government responses.	23	18	27	27	5
To recognize that all substance abuse planning, programming, etc., should be administered together.	5	24	19	47	5
To minimize damage caused by government reaction to drug use.	0*	39*	22*	33*	6*

* Indicates that over one-fifth of respondents did not respond to this item.

Table 9 (Continued)

OBJECTIVES WHICH EXHIBITED POLARIZATION OR A BROAD DISTRIBUTION OF RESPONSES (Feasibility and/or Desirability)

Objective	Scale Value; Percent voting				
	1	2	3	4	5
Polarized or broad distribution on Desirability					
To utilize outreach and coercive techniques to engage those not entering treatment on their own.	32	18	9	23	18
To provide voluntary drug treatment available to anyone who wishes it, without government regulations.	35	5	23	14	23
To reduce legal-political-governmental control of treatment, including urines; dosage; age; identification requirements.	18	27	14	27	14
To reduce/eliminate criminal penalties for personal use and possession of drugs currently defined as "illegal".	35	13	17	13	22
To establish the principle that whether or not an individual uses a mind altering substance is a matter of personal choice with minimum governmental interference.	30	9	26	13	22
To increase med-social research -- hallucinogen for eco-systems and positive social values through use (Ritualistic, etc.)	36*	0*	36*	21*	7*
To develop a group with a primary interest in the problems people have with drugs.	32	14	27	9	18

* Indicates that over one-fifth of respondents did not respond to this item.

(with respondents voting approximately equally for four or more of the five scale values); or truly indeterminable (with the modal response being "May or may not be feasible"). In only eight of the twenty-one objectives which scored "May or may not be feasible" was this the modal response; in the case of eleven objectives the reason was that the voting was either polarized or broadly distributed, as Table 9 shows. All seven of the objectives which respondents scored "Neither desirable nor undesirable" were either polarized or of a broad distribution.

By reviewing the frequency distribution and the scale scores we were able to identify objectives in which there was a significant voting difference between those who rated themselves experts in national drug-abuse policy and those who did not. Table 10 lists these items. Some of the differences were in items that could be of major importance in the formulation of a national drug-abuse policy; most of the differences had to do with the feasibility of attaining objectives.

Table 10

OBJECTIVES WHICH EXHIBITED VOTING DIFFERENCES BETWEEN SELF-RATED POLICY EXPERTS AND NONEXPERTS (Feasibility and Desirability)

Objective	Policy expertise	Scale value; Percent voting 1	2	3	4	5	Feasibility Score	Desirability Score
Differences on feasibility								
To develop adequate alternative models in prevention and phased intervention.	Expert	9	27	55	9	0	2.64	2.00
	Non-Expert	15	69	8	0	8	2.15	1.39
To reduce the supply of illicit drugs available for abuse.	Expert	9	64	18	0	9	2.36	1.36
	Non-Expert	8	25	25	17	25	3.25	1.85
To reduce non-prescribed use of psychoactive drugs.	Expert	0	27	46	27	0	3.00	1.73
	Non-Expert	15	39	0	31	15	2.96	1.69
To reduce prescribed use of psychoactive drugs/diminish misuse by physicians.	Expert	27	9	55	9	0	2.46	2.18
	Non-Expert	32	23	15	15	15	2.62	1.69
To incorporate drug treatment into standard health delivery systems.	Expert	36	19	36	9	0	2.18	1.73
	Non-Expert	27	55	9	0	9	2.09	1.82
To establish an effective social rehabilitation system for drug abusers who have become desocialized.	Expert	0	30	40	20	10	3.10	1.20
	Non-Expert	46	9	36	0	9	2.18	1.09
To consider social action approaches, as alternatives to treatment, including concern with housing, employment, education, and counselling, in addition to other novel approaches.	Expert	0	20	0	80	0	3.60	1.80
	Non-Expert	25	25	34	8	8	2.50	1.75
To develop a group with a primary interest in the problems people have with drugs.	Expert	30	10	40	0	20	2.70	3.50
	Non-Expert	42	25	8	17	8	2.25	2.00
Differences on desirability								
To train in-line treatment personnel to enhance their skill in helping the drug dependent person.	Expert	40	60	0	0	0	2.00	1.60
	Non-Expert	100	0	0	0	0	1.58	1.00
To develop a group with a primary interest in the problems people have with drugs.	Expert	10	10	30	20	30	2.70	3.50
	Non-Expert	50	17	25	0	8	2.25	2.00

In five cases, the modal response of the policy experts was "May or may not be feasible," while nonexperts voted the same objective feasible. These related to major strategies such as "to develop adequate alternative models in prevention and phased intervention," to reduce prescribed use of psychoactive drugs/diminish misuse by physicians," and even "to incorporate drug treatment into standard health delivery systems." Since these objectives were all held to be at least desirable, and since one is unlikely to propose objectives one is unsure are achievable, bringing to light this additional information may broaden the policy options available to decision-makers. Alternately, it could be that the view of nonexperts in these cases is overly optimistic. In one case ("to consider social-action approaches as alternatives to treatment...") the modal response of the nonexpert was "May or may not be feasible"; the policy experts were sure it was "Probably feasible."

Policy experts and nonexperts differed on two objectives which represent a major effort in the present national drug-abuse prevention strategy. Policy experts scored "to reduce the supply of drugs available for abuse" as "Feasible," while nonexperts were less certain, scoring it "May or may not be feasible." The reverse was true of the objective "to establish an effective social-rehabilitation system for drug abusers who have become desocialized." Policy experts were not sure if this objective was attainable and scored it "May or may not be feasible"; nonexperts, on the other hand, scored it "Feasible."

Only one objective exhibited a voting difference between policy experts and nonexperts on both feasibility and desirability. This objective ("to develop a group with a primary interest in the problems people have with drugs") was scored "Undesirable" and "May or may not be feasible" by policy experts, but "Desirable" and "Feasible" by nonexperts.

Although there was a big difference in the desirability score between policy experts and nonexperts on the objective "to train in-line treatment personnel to enhance their skill in helping the drug-dependent person," this was mostly in agreement with 40 percent of experts and 100 percent of the nonexperts voting this objective as "Highly desirable."

Objectives that score "Probably infeasible" or "May or may not be feasible" (and this scale value was the modal response), were dropped from consideration, unless there was a significant difference in voting between policy expert and nonexpert.

Third Questionaire. In this questionaire, twenty-five objectives were listed, which required revoting (desirability and feasibility). Objectives are presented for revoting because of polarization on the part of the panel; because there was a broad distribution of voting responses; or because there were differences in voting between policy experts and policy nonexperts. In some cases, original objectives were combined or divided after respondents' comments had been reviewed; in this instance, voting was required on the newly developed objective. The remaining objectives will be held over until a later round. Consideration of the key indicators associated with objectives rated at least feasible and desirable will also be held over to a subsequent round.

Transition Model and Matrix

First Questionnaire. Social policy is the result of multiple interacting forces. Policy is often seen as being based on advocacy rather than derived from a careful analysis of empiric findings. Although policy may have to be developed even in the absence of information, a rational examination of the bases on which policies have been built is a fruitful way of providing the policymaker with insight to develop more appropriate policies.

In the case of drug abuse, the factors which cause people to pass through various states of drug dependence can be systematically examined. Such examination allows the policymaker to estimate the importance of specific variables, and the extent to which they are subject to his influence. A systematic examination of factors also allows any counterintuitive effect of the variables to be brought to light.

In this part of the study we hoped to elicit from respondents factors which control the rates of flow from general population through the various states of drug abuse. The simplified model shown in Fig. 1 developed for this purpose was intentionally simplified to allow for examination of the five transition states included in the matrix. More complex (and probably more realistic) models would have diverted attention from the question at hand.The five transitions in which we are particularly interested are:

- General Population To Potential User
- Potential User To Experimental User
- Experimental User To Occasional Abuser
- Occasional Abuser To Drug-Dependent Person
- Drug-Dependent Person To Formerly Drug-Dependent Person

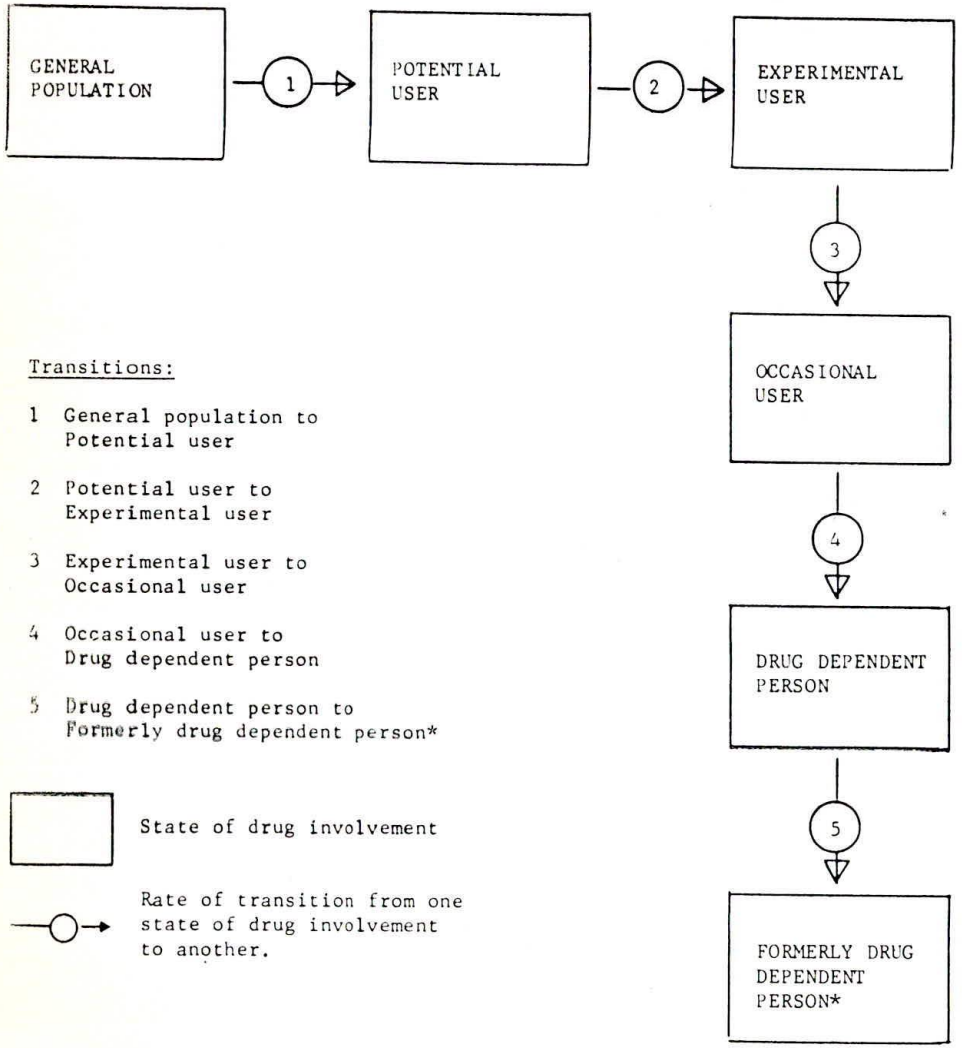
Respondents were asked to list the factors which affected each of the transitions and state whether a specific factor increased (promoted) or decreased (inhibited) the rate of flow of individuals from one state to another.

The number of factors listed by respondents ranged from a low of two to a high of over forty; we identified a total of 128 distinct factors.

Using the criterion that a factor must register three votes for a single transition, twenty-five significant factors were identified from our total list of 128 factors. The twenty-five are shown in Table 11, the vote is shown by transition state. The table shows the number of respondents who thought that the factor increases the transition rate from one state to another, the number who said it decreases the rate, and a residue who did not indicate direction. Because the number of votes is small, and also because the interpretation of respondents' indication of direction was sometimes difficult, reference will be

Figure 1

DRUG INVOLVEMENT TRANSITION MODEL



Transitions:

1 General population to Potential user

2 Potential user to Experimental user

3 Experimental user to Occasional user

4 Occasional user to Drug dependent person

5 Drug dependent person to Formerly drug dependent person*

State of drug involvement

Rate of transition from one state of drug involvement to another.

* one should not presume that this is a permanent state; it indicates that point at which, either through treatment or self-denial, one is no longer drug dependent.

Table 11

THE TWENTY-FIVE TRANSITION FACTORS RECEIVING AT LEAST THREE VOTES ON A SINGLE TRANSITION
(presented by factor category)

Factor	Response	Transition #1(General Population to Potential User)	#2 (Potential User to Experimental User)	#3(Experimental User to Occasional User)	#4(Occasional User to Drug Dependent Person)	#5 (Drug Dependent Person to Former Dependent Person)	Total Impact Score and (Rank)
age	Total	4	5	4	3	2	18
	Inc	4	4	4	3	2	(8)
	Dec	0	1	0	0	0	
	EW	0	0	0	0	0	
family breakdown	Total	3	2	3	3	3	14
	Inc	2	2	2	2	2	(12=)
	Dec	1	0	1	1	1	
	EW	0	0	0	0	0	
peer pressure	Total	6	8	8	7	6	35
	Inc	5	7	7	5	2	(2)
	Dec	0	0	0	1	3	
	EW	1	1	1	1	1	
proportion of drug users in peer groups	Total	5	4	4	4	4	21
	Inc	5	4	4	4	0	(5)
	Dec	0	0	0	0	4	
	EW	0	0	0	0	0	
religious faith/ training	Total	2	3	2	1	2	10
	Inc	0	0	0	0	2	(20=)
	Dec	2	3	2	1	0	
	EW	0	0	0	0	0	
availability of effective stress-relieving alternatives to drug use/lack of boredom	Total	5	5	5	3	3	21
	Inc	3	3	2	1	2	(5)
	Dec	2	2	3	2	1	
	EW	0	0	0	0	0	
opportunity for meaningful social definition/ availability of adult roles for adolescents	Total	1	1	1	1	3	7
	Inc	1	0	0	0	2	(25)
	Dec	0	1	1	1	1	
	EW	0	0	0	0	0	
feeling of social inadequacy or alienation/ alienation from standards of adult-society	Total	3	3	3	3	1	13
	Inc	3	3	3	2	0	(16=)
	Dec	0	0	0	1	1	
	EW	0	0	0	0	0	
inability to resolve frustrating personal problems	Total	3	3	3	3	2	14
	Inc	2	2	2	2	0	(12=)
	Dec	0	0	0	0	1	
	EW	1	1	1	1	1	
personal stress/ situational crises/ increase in intensity of day-to-day pressures	Total	3	3	4	2	2	14
	Inc	2	2	3	2	2	(12=)
	Dec	0	0	0	0	0	
	EW	1	1	1	0	0	
protracted adolescence/ youth	Total	3	2	1	1	1	8
	Inc	3	2	1	1	0	(23=)
	Dec	0	0	0	0	1	
	EW	0	0	0	0	0	
poor school achievement/dropping out of school	Total	3	2	3	3	3	14
	Inc	2	1	2	2	0	(12=)
	Dec	0	0	0	0	2	
	EW	1	1	1	1	1	
meaningful and satisfying life role/ economic opportunity	Total	2	2	2	2	3	11
	Inc	1	1	2	2	1	(18=)
	Dec	1	1	0	0	2	
	EW	0	0	0	0	0	

Inc = Increase
Dec = Decrease
EW = Either Way

Table 11 (Continued)

THE TWENTY-FIVE TRANSITION FACTORS RECEIVING AT LEAST THREE VOTES ON A SINGLE TRANSITION
(presented by factor category)

Factor	Response	Transition #1(General Population to Potential User)	#2 (Potential User to Experimental User)	#3(Experimental User to Occasional User)	#4 (Occasional User to Drug Dependent Person)	#5 (Drug Dependent Person to Former Dependent Person)	Total Impact Score and (Rank)
specific drug effects/ pharmacological properties of drugs of abuse	Total	2	2	4	4	3	15
	Inc	1	1	2	2	0	
	Dec	0	0	2	2	1	(11)
	EW	1	1	0	0	2	
availability of (illicit/euphoria producing/ anxiety reducing) drugs	Total	10	13	13	10	8	54
	Inc	9	12	12	10	0	
	Dec	0	0	0	0	8	(1)
	EW	1	1	1	0	0	
supply of drugs/ heroin	Total	2	4	3	2	9	20
	Inc	2	4	3	2	7	
	Dec	0	0	0	0	2	(7)
	EW	0	0	0	0	0	
advertising of drugs in the mass media especially on TV	Total	3	1	1	1	2	8
	Inc	2	0	0	0	1	
	Dec	0	0	0	0	0	(23=)
	EW	1	1	1	1	1	
dissemination of research results/reported research data (eg. effects of drug use)	Total	3	3	2	2	1	11
	Inc	1	1	0	0	0	
	Dec	1	2	2	2	1	(18=)
	EW	1	0	0	0	0	
drug abuse education programs	Total	6	7	5	2	4	24
	Inc	3	3	2	1	4	
	Dec	3	4	3	1	0	(4)
	EW	0	0	0	0	0	
availability of community based treatment/access to treatment centers	Total	0	0	2	5	9	16
	Inc	0	0	0	1	7	
	Dec	0	0	1	3	1	(10)
	EW	0	0	1	1	1	
severity of legal sanctions/laws against use/legal prohibition of drug use	Total	4	5	3	1	4	17
	Inc	1	1	1	1	4	
	Dec	3	4	2	0	0	(9)
	EW	0	0	0	0	0	
enforcement activities/ law enforcement pressure/application of legal sanctions	Total	4	5	5	5	6	25
	Inc	0	1	2	2	4	
	Dec	4	4	3	3	2	(3)
	EW	0	0	0	0	0	
public acceptance of legal drug taking behavior	Total	3	3	2	1	0	9
	Inc	2	2	1	1	0	
	Dec	0	1	0	0	0	(22=)
	EW	1	0	1	0	0	
economic social deprivation	Total	2	2	2	3	1	10
	Inc	2	2	2	2	0	
	Dec	0	0	0	1	1	(20)
	EW	0	0	0	0	0	
number of persons using the drug	Total	3	3	3	3	1	13
	Inc	3	3	3	3	0	
	Dec	0	0	0	0	1	(16=)
	EW	0	0	0	0	0	

Inc = Increase
Dec = Decrease
EW = Either Way

made only to the total number of votes indicating that a particular factor affected a given transition. It should be noted in passing, however, that for some factors respondents appeared to disagree on the direction in which a factor affected a transition rate. A wide range of factors made the listing; some of them were interwoven into the fabric of society, some were clearly interdependent, and in other cases any relationship between factors was less clearcut. The dominant factor was "the availability of drugs," which scored half again as many votes as the second factor, "peer pressure," which in turn scored almost half as many votes again as the third factor, "enforcement activities/law-enforcement pressure."

The top ten factors, arranged in rank order of voting across all transitions, together with the number of votes cast for the factor *not* affecting a particular transition are shown in Table 12.

It is important to note that the votes of "no effect" were obtained from respondents who completed the matrix *and who specifically recorded the fact that a particular transition was not affected by the factor they had listed.* This means other respondents may also not think a particular factor affects a transition, and so in order to balance the picture the number of nonvotes is also shown. Nevertheless, it is interesting to note that even for the most influential factor ("availability of drugs") some respondents specifically said it did not affect transition #1 (general population to potential user) and transitions #4 and #5 (occasional user to drug-dependent person, and drug-dependent person to formerly dependent person). The factor with most disagreement was "drug-abuse education" (all transitions). We presume that this item refers to drug education as it is practiced now, rather than what its potential impact might be if practiced "correctly." It is particularly noteworthy that no less than eight respondents specifically said that "availability of community-based treatment" did not affect the transition from general population to potential user, and from potential user to experimental user. This number of "no effect on " votes was twice as high as for any other factor, and particularly interesting in that not a single respondent stated that the factor actually affected these two transitions.

The wide range of impacting factors and the fact that some of them are closely related to the kind of society in which we live illustrates the problem of creating specific drug-abuse programs.

Second Questionnaire. The transition matrix was not considered in this round.

Third Questionnaire. In this questionnaire we examined the variables that influence the transition from one state of drug involvement to another. Respondents were asked to rate the twenty-five variables that were most frequently cited in response to the first questionaire; ratings included estimating the importance of each factor in controlling a particular rate of transition, and the extent to which these factors can be influenced by the national drug-abuse policymaker (manageability). These data can then be used by

Table 12

TOP TEN FACTORS ACROSS ALL TRANSITIONS

Factor	Response	Transition #1(General Population to Potential User)	#2 (Potential User to Experimental User)	#3(Experimental User to Occasional User)	#4 (Occasional User to Drug Dependent Person)	#5 (Drug Dependent Person to Former Dependent Person)
Availability of illicit/ euphoria producing/anxiety reducing)drugs	Affect	10	13	13	10	8
	No Affect	3	0	0	2	4
	No Response	4	4	4	5	5
Peer pressure	Affect	6	8	8	7	6
	No Affect	1	0	1	2	3
	No Response	10	9	8	8	8
Enforcement activities/law enforcement pressure/ application of legal sanctions	Affect	4	5	5	5	6
	No Affect	2	1	1	1	0
	No Response	11	11	11	11	11
Drug abuse education programs	Affect	6	7	5	2	4
	No Affect	2	1	3	6	4
	No Response	9	9	9	9	9
Proportion of drug users in peer groups	Affect	5	4	4	4	4
	No Affect	0	0	1	1	1
	No Response	12	13	12	12	12
Availability of effective stress-relieving alternatives to drug use/lack of boredom	Affect	5	5	5	3	3
	No Affect	1	1	1	2	3
	No Response	11	11	11	12	11
Supply of drugs/heroin	Affect	2	4	3	2	9
	No Affect	1	0	0	1	0
	No Response	14	13	14	14	8
Age	Affect	4	5	4	3	2
	No Affect	0	1	1	2	3
	No Response	13	11	12	12	12
Severity of legal sanctions/ laws against use/legal prohibition of drug use	Affect	4	5	3	1	4
	No Affect	1	0	2	4	1
	No Response	12	12	12	12	12
Availability of community based treatment/access to treatment centers	Affect	0	0	2	5	9
	No Affect	8	8	5	3	0
	No Response	9	9	10	9	8

espondents to reformulate priorities for national policy, and develop programs from the objectives and policy-issue-statements sections of this study.

Respondents were also asked to rate the criticality of the five separate ransitions from one state of drug involvement to another, in terms of the priority each transition should receive as part of the national drug-abuse policy. Respondents did this by allocating one hundred points over the five transitions o as to reflect their priorities. One of the intents of this question is to develop alternative methods of determining program-funding priorities.

Policy Issue Statements

First Questionnaire. For the first questionnaire, panelists were asked to respond to twelve should/should not statements, constructed on current policy issues in ne drug-abuse field, and to indicate the importance of these statements. Respondents were also asked to list issues they thought were sufficiently important to be included in the study.

Pro and *contra* arguments for these issue statements were to have been elicited s part of the second questionnaire; in fact, the respondents used their first questionnaire comments regarding each of the statements to do just that. A otal of 184 separate comments were made, and were fed back as part of the econd questionnaire; respondents were asked to rate their importance.

A total of forty-six additional policy issues was suggested by respondents in e first questionnaire; once again, these seemed to exemplify the diverse and omplex nature of this field. From this list we were able to formulate an additional sixteen policy issues in the form of should/should not statements.

Second Questionnaire. The issue statements section of the second questionnaire cluded a statistical summary of responses from the first questionnaire, and a of the narrative comments made by respondents from each issue statement. he respondents were asked to revote the issue statements, and rate the arrative comments with regard to *importance*. In addition, thirteen of the issues osed by respondents themselves in the first questionnaire were fed back in the rm of should/should not statements for initial voting. It should be noted that ree of the original issue statements were rewritten for clarity, following the edback from the respondent panel.

In preparation for this questionnaire, the analysis or the policy issue state- nts was confined to comparing the group's response in Round One and ound Two on those items which were common to both rounds (no major shifts re observed in the group's position); and to comparing the Round Two sponses of self-rated policy experts to those of nonexperts. Both rating sponses and importance ratings were compared.

Respondents were almost unanimous in thinking that "Treatment programs ould offer treatment for more than one type of dependence within the same cility" and that "The federal government *should* allocate funds to community alth centers so that these centers can offer treatment for drug abuse." In this

round, substantially more respondents felt that the private insurance plans should be encouraged by the federal government to provide coverage for the treatment of drug dependence than in Round One. In Round One, 42 percent of respondents felt that marijuana should be legalized. This support dropped to 32 percent in Round Two, but 81 percent of respondents felt that "The personal use of marijuana *should* be decriminalized" (a new item).

Issues were classified according to their importance, determined by means of the group importance score. Table 13 lists all thirty-one issues in order of importance; the most important being "There should/should not be a national registry of drug-dependent persons...." Also shown in Table 13 is a comparison of the way self-rated policy experts and nonexperts voted on the Round Two issue statements. Really major differences were observed on only three items. The greatest difference was in relation to the issue "Patients who enter treatment should/should not be required to remain in treatment for a minimum amount of time"; 70 percent of policy experts voted *should* while 77 percent of nonexperts voted *should not*. Only 56 percent of policy experts voted *should* compared to 96 percent of nonexperts on the issue "When an individual is referred for treatment in lieu of incarceration, he should/should not have the right to choose the treatment he prefers." On the other hand, 100 percent of policy experts voted that "Regulations *should* be passed by Congress prohibiting any and all alcohol and tobacco advertising via any media," compared to only 58 percent of nonexperts.

There were four issues that exhibited marked differences between policy experts and nonexperts in the importance of issues (see Table 13). Policy experts scored the importance of the issue "Treatment programs should/should not offer treatment for more than one type of dependence within the same facility" 2.60, while nonexperts thought it considerably more important, scoring it 1.67. A similar situation pertained to the issue "The manufacture and sale of cigarettes should/should not be outlawed." Although both policy experts and nonexperts considered the issue not to be particularly important, experts scored 3.40, while nonexperts scored 2.42.

Marked differences were also observed between self-rated policy experts and nonexperts in the importance scores of these issues:

- "All drug use or possession for personal use should/should not be decriminalized." (policy experts = 2.30, nonexperts = 1.50)
- "A greater proportion of the funds available for drug abuse should/should not be allocated to basis research." (Policy experts = 2.30, nonexperts = 1.55)
- "Persons who are identified as drug-dependent should/should not be required to receive treatment." (Policy experts = 2.40, nonexperts = 1.64)
- "Minimum mandatory prison sentences should/should not be imposed for certain drug-trafficking violations." (Policy experts = 1.90, nonexperts = 2.64)
- "The personal use of marijuana should/should not be legalized." (Policy experts = 2.60; nonexperts = 1.83)

Table 13

POLICY ISSUE STATEMENTS: RESPONSES AND IMPORTANCE RATINGS BY SELF-RATED POLICY EXPERTS AND NONEXPERTS (in decreasing order of group importance)

Policy Issue Statement	Percent voting "should" (1) Policy Experts	Percent voting "should" (1) Policy Nonexperts	Importance Score (1) Policy Experts	Importance Score (1) Policy Nonexperts
Very Important				
There SHOULD/SHOULD NOT be a national registry of drug-dependent persons (including identifying information such as the social security number of the individuals).	[illegible]	17	1.5*	1.50*
There SHOULD/SHOULD NOT be an explicit national policy regarding drug abuse.	[illegible]	[illegible]	1.70	1.69*
There SHOULD/SHOULD NOT be a maximum amount of time for an individual to be maintained on methadone.	20	31	1.90	1.33
Private health insurance plans such as Blue Cross/Blue Shield SHOULD/SHOULD NOT be encouraged by the federal government to cover treatment of drug dependency in their insurance plans.	[illegible]	65	1.8*	1.66
Personal drug-dependence SHOULD/SHOULD NOT be a criminal offense.	[illegible]	[illegible]	1.70	1.75
When an individual is referred for treatment in lieu of incarceration, he SHOULD/SHOULD NOT have the right to choose the treatment he prefers.	[illegible]	92	1.90	1.50
Important				
The national policy SHOULD/SHOULD NOT be monitored by a non-political group.	[illegible]	91	1.89	1.80*
Personal use of marijuana SHOULD/SHOULD NOT be decriminalized.	[illegible]	[illegible]	1.78	1.92
All drug use or possession for personal use SHOULD/SHOULD NOT be decriminalized.	[illegible]	[illegible]	2.00	1.50
The Federal government SHOULD/SHOULD NOT allocate funds to community health centers so that these centers can offer treatment for drug abuse.	[illegible]	[illegible]	1.90	1.91*
A greater proportion of the funds available for drug abuse SHOULD/SHOULD NOT be allocated to basic research.	[illegible]	[illegible]	2.0	1.53*
Treatment programs SHOULD/SHOULD NOT offer treatment for more than one type of dependence within the same facility.	[illegible]	92	2.00	1.97
THE FEDERAL SPONSORING AGENCY/SINGLE STATE AGENCY should be responsible for evaluating treatment programs.	[illegible]	[illegible]	2.11	1.72
Persons who are identified as drug-dependent SHOULD/SHOULD NOT be required to receive treatment.	[illegible]	81	2.00	1.[illegible]*
Pre-trial detention SHOULD/SHOULD NOT be used in certain drug trafficking cases.	[illegible]	[illegible]	1.70	2.0*
Minimum mandatory prison sentences SHOULD/SHOULD NOT be imposed for certain drug trafficking violations.	[illegible]	[illegible]	1.90	2.0*
The present drug scheduling criteria ARE/ARE NOT adequate.	[illegible]	[illegible]	2.11	2.11*
The National Institute for Drug Abuse and the National Institute for Alcohol Abuse SHOULD/SHOULD NOT be combined to form a single agency.	[illegible]	[illegible]	1.5	1.7*
The United States SHOULD/SHOULD NOT actively assist or influence other governments to cease production of crops such as opium, coca, or cannabis.	[illegible]	[illegible]	[illegible]	[illegible]*

(1) Percent of those responding
(2) Refers to the Federal Sponsors Agency [illegible]
(3) Refers to the "are" response
* Indicates that over one-fifth of respondents did not complete this item

Table 13 (Continued)

POLICY ISSUE STATEMENTS: RESPONSES AND IMPORTANCE RATINGS BY SELF-RATED POLICY EXPERTS AND NONEXPERTS (in decreasing order of group importance)

Policy Issue Statement	Percent voting "should" (1) Policy Experts	Percent voting "should" (1) Policy Nonexperts	Importance Score (1) Policy Experts	Importance Score (1) Policy Nonexperts
Patients who enter treatment SHOULD/SHOULD NOT be required to remain in treatment for a minimum amount of time.	70	23	[illegible].20	2.08
The personal use of marijuana SHOULD/SHOULD NOT be legalized.	33	31	2.60	1.83
Non-medical personnel who are employed as direct servicers in drug treatment programs SHOULD/SHOULD NOT be certified to work in that capacity in those programs.	80	91	2.10	2.33
A drug SHOULD/SHOULD NOT be subject to tight controls (production quotas, export-import regulations, prescription controls, etc.) if it has potential for abuse, irrespective of its historical abuse patterns.	67	82	2.22	2.27*
The Drug Enforcement Administration SHOULD/SHOULD NOT be dissolved (as part of a commitment to decriminalization of drug use and in an effort to decrease poor police practice and unnecessary government expense).	20	9	2.40	2.18*
Regulations SHOULD/SHOULD NOT be passed by Congress prohibiting any and all alcohol and tobacco advertising via any media.	100	58	2.30	2.17
The international treaty actions and commitments of the United States (Single Convention and Psychotropic Convention) SHOULD/SHOULD NOT be removed and our international stance vis-a-vis drug be radically altered.	44	17*	2.11	2.63*
Practicing physicians SHOULD/SHOULD NOT be required to attend continuing education in psychopharmacology or be required to pass periodic tests regarding psychopharmacology.	70	62	2.50	[illegible]*
A citizen-government-pharmaceutical council SHOULD/SHOULD NOT be formed to recommend guidelines to the pharmaceutical industry for its goals in psychoactive drug development and for its practices in sales to physicians.	80	80	2.50	2.55*
Licensed medical personnel who are employed in drug treatment programs SHOULD/SHOULD NOT be certified to work in those programs.	90	80	3.00	2.54
Moderately Important				
Special teams of Federal, State and local agents SHOULD/SHOULD NOT be formed to operate across municipal and state lines for the purpose of suppressing local drug peddling.	78	55	2.67	2.22*
Manufacture and sale of cigarettes SHOULD/SHOULD NOT be outlawed.	0	8	3.40	2.42

(1) Percent of those responding
* Indicates that over one-fifth of respondents did not complete this item

Preliminary Epilogue

Resource Requirements

One of the advantages of the Delphi technique as a tool in policy analysis is its minimal cost for maximum output. The costs for completion of a Delphi study such as this one can range from $15,000 to $40,000 for a nine- to twelve-month effort, depending upon staff and direct-cost expenditures required. For example, if the effort is included as part of ongoing staff assignments, then staff and space costs may not be directly chargeable; if computer services are available, then a sizable cost category is deleted. The amount of data which may be derived, and the opportunity afforded to facilitate a "discussion" of the issues by divergent experts in the field, render the technique unusually cost-effective.

Considerable time should be spent in coceptualization of the study design and development and pretesting of the questionnaire. In any area as complex and diffuse as drug abuse, the study-design team needs to allocate substantial effort to this phase of study development.

Applications

The relative success of this National Drug-Abuse Policy Delphi has resulted in considerable interest in utilization of the technique not only in the drug-abuse field, but in other social policy areas as well. The opportunities it affords for idea exchange among diverse professionals and interest groups; and the continuous flow of significant data for policy review are but two of the positive attributes of the method. The potential for its application is extensive; as this is the first study of its type, all of us who are interested in its future application can profit from the lessons learned from this effort.

The process of conceptualizing and analyzing policy options is supremely complex; it may be that the Delphi policy method will be a significant advance in the field of applied decision theory and policy analysis, as it relates to the social policy area in particular.

Prospectus

This phase of the study was completed in December 1974. During the succeeding rounds, the objectives will be further summarized; the policy issues on which there is significant divergent opinion will be explored; and policy options will be developed from the objectives, policy issues, and transition factors. An interactive conference will be held at the conclusion of the study; part of this conference will involve introduction of the computer conferencing technique to respondents.

We shall evaluate the present effort to determine to what extent the study

objectives were met, and to what extent the respondents' objectives for participating in the study were met. The preliminary steps in this evaluation have already been taken: in the first questionnaire, respondents were asked to list their objectives for participating in the study, for example. At the conclusion of the study, the respondents will be asked to measure the degree to which their objectives have been reached and whether they might have developed other objectives during the course of the study. In addition, they will be asked to evaluate the study on the basis of questionnaire design, content, and other relevant areas. The results of this evaluation will be utilized in developing an ongoing interactive policy planning system which the author is presently designing, as well as other specific studies which are expected to stem from the present effort.

The evaluation of the impact of a study such as this is a much more complex problem, but one which we believe is ultimately of more significance. We have just begun to develop plans for a long-term evaluation of the study. This will include, for example, as assessment of the degree to which study results were reviewed and considered in the formulation of national drug-abuse policy.

Acknowledgments

I should like to acknowledge the extraordinary gratis assistance of Dr. Norman Dalkey and Dr. Murray Turoff; and that of Dr. Peter Goldschmidt and M. Alexander Stiffman of the Johns Hopkins University; without the support and dedicated efforts of these individuals, and the 39 respondents, the study would not be possible.

I should also like to thank the National Coordinating Council on Drug Education for sponsoring the continued development of the study design as originally conceived and implemented.

B. 4. A Delphi Evaluation of Agreement between Organizations

CHESTER G. JONES

Introduction

Delphi [1] is often used to combine and refine the opinions of a heterogeneous group of experts in order to establish a judgment based on a merging of the information collectively available to the experts. However, in this process it is possible to submerge differences of opinion and thus suppress the existence of uncertainty. In many situations it might be advisable to run separate Delphis using more homogeneous groups of experts in order to highlight areas of disagreement. This paper will report on an activity that did just this and point out several areas in which the types of responses obtained were fundamentally very different. In some cases these differences were quite unpredictable, and so, a highlighting of the variations greatly increased the information obtained. Running one Delphi using a subset of the experts from each group would probably not have illuminated some of the differences in opinion. The mere weight of pressure to move toward the median response [2] would have caused a joint Delphi to converge toward a middle position. In addition, the presence of disagreement is much more significant when large groups share similar positions. The traditional approach to Delphi generally results in the using of a small number of experts from any one area.

One concern that is often raised about the credibility of Delphi results is that individual experts may bias their responses so that they are overly favorable toward areas of personal interest. This is of particular concern when experts are asked to evaluate areas in which they are presently working and when the final Delphi results could impact the importance attached to these areas. In this paper results will be presented that indicate that no such bias occurred in the Delphis reported on. It appears that the particular groups of experts used were able to rise above the desire to protect personal interests.

Background

The United States Air Force presently maintains an official list of System Concept Options (SCOs) in order to indicate to the Air Force Laboratories potential future technology needs. This activity is primarily a means of com-

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

municating to the laboratory planners the thinking of Air Force System planners. However the number of potentially worthwhile systems possibilities, and thus the number of technology needs, exceed the resources available to fulfill all the possibilities and needs. Clearly the Air Force Laboratories needed a means of establishing priorities for the System Concept Options. Thus it was decided to undertake a program of Delphi evaluation. This program was run by the Deputy for Development Planning, Aeronautical Systems Division, and was limited to considerations of those SCOs that fell under the Deputy's jurisdiction. Thirty SCOs were evaluated. They covered a rather large spread in need for technological support as well as proposed mission use. Some concepts represented a rather straightforward extrapolation of present technology, while others would require substantial technology development programs. The missions represented included most of the areas of interest to the Air Force including many strategic and tactical possibilities as well as systems intended to meet support and training requirements.

It was decided to conduct separate Delphis utilizing personnel from various Air Force organizations, in order to determine how closely the organizational opinions agreed. In this way it was believed that not only would a basis for prioritizing the systems be obtained, but in addition, the results would help to indicate areas of communication problems between organizations. If organization viewpoints in a particular area differed greatly, there would appear to be a need for increased communication about the area.

Delphis were conducted within the following four USAF organizations: Deputy for Development Planning, Aeronautical Systems Division (ASD/XR); Air Force Avionics Laboratory (AFAL); Air Force Aero Propulsion Laboratory (AFAPL); Air Force Flight Dynamic Laboratory (AFFDL). The experts chosen were senior managerial and technical personnel (both civilian and military), and were selected so that representation of most if not all of the major departments within the organizations was present. A total of sixty-one experts took part in the evaluations which involved three rounds of questioning.

The above organizations are of two different types. The Deputy for Development Planning is a systems planning organization having responsibility for identifying promising aerodynamic system concepts and defining them to the point where development decisions can be made. It has no direct responsibility for research activities. The three laboratories are responsible for developing technologies in their assigned areas which will improve system capabilities. The Avionics Laboratory is concerned with electronic systems, the Aero Propulsion Laboratory with atmospheric engines, fuel, etc., and the Flight Dynamics Laboratory with aircraft structures, controls, aerodynamics, etc. Thus the four groups that were asked to evaluate the list of SCOs are quite different in their areas of expertise. In particular it should be emphasized that the laboratory groups were being asked to compare SCOs some of which required considerable support from their particular laboratory, others of which required little or no

support. All of the participants were, however, senior Air Force personnel and were thus knowledgeable of activities at other Air Force Laboratories.

Results obtained for three of the questions used will be discussed in this paper.

Question 1. Please rank-order the SCO list of systems on the basis of where current Air Force Laboratory Programs will make the greatest contribution toward success of the system.

Question 4. Given that each system becomes a technological success, rank order the SCO list in terms of importance of each system to National Defense.

Question 5. Considering technology, timing, and system importance, rank-order the SCO list according to where you think the Air Force Laboratories can make the greatest contribution to National Defense.

Each of these questions involved a complete ranking of thirty items, which proved to be a trying but not impossible task. It should be noted that succeeding round changes in answers often required a large restructuring of the list. That is, a change in the answer or rank of one system generally changes the rank of other systems (however, the participants were allowed to use a limited number of ties if necessary and thus a few participants avoided this problem). This interrelation of answers tends to make convergence difficult, since disagreement in one area impacts other areas.

Convergence

One indication of the effect of a Delphi experiment is the amount of convergence caused by the iteration process, where convergence is a measure of how much more agreement is achieved on succeeding rounds as opposed to the first-round response. In this effort, one measure of convergence was the change in the spread between the lower and upper quartile values for a given question and a given SCO. In all of the Delphi experiments the spread between the lower and upper quartile values generally showed considerable reduction during the course of the efforts. However, as indicated in Fig. 1, the average amount of convergence varied considerably from group to group for some questions.

All of the groups achieved basically the same degree of convergence for Question 1. However, the convergence on Questions 4 and 5 follow significantly different patterns. In particular, the ASD/XR group achieved less convergence on Questions 4 and 5 than that achieved by the laboratory groups. The ASD/XR group was the only group primarily composed of system planners and so this group's failure to converge as well as the other groups on Question 4 would seem to be quite important. The ASD/XR group should be the best suited to serve as experts concerning Question 4. Thus, for Question 4 the greatest uncertainty is associated with the most expert group. If one Delphi had

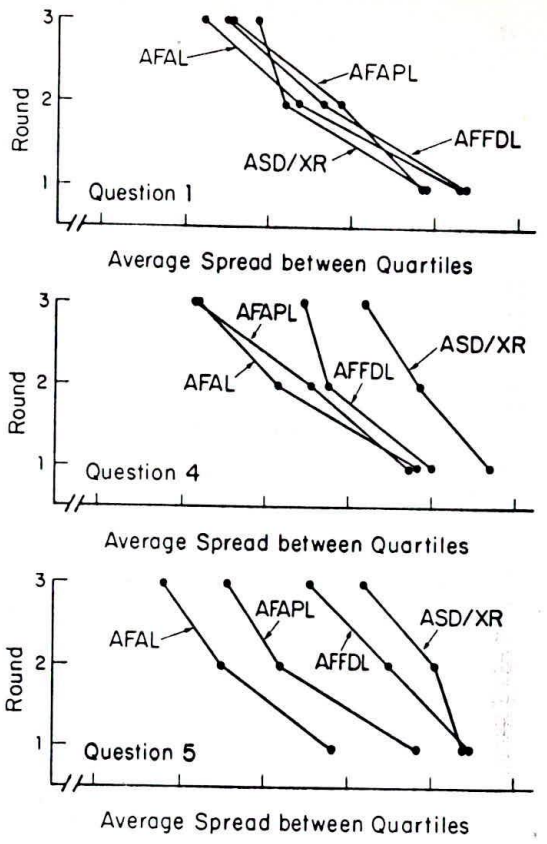


Fig. 1. Average interquartile spreads for questions 1, 4, 5.

been run combining experts from each of those groups, it appears possible that the greater convergence between the laboratory experts might have caused a considerably better overall convergence than that shown in the ASR/XR result. Thus the relatively less expert participants might have caused the creation of a false sense of expert agreement.

Correlation between Questions

In reviewing the results, it was obvious that some groups tended to give SCOs similar rankings for different questions, while other groups changed many of the SCOs rankings drastically from question to question. Table 1 shows the

Spearman rank correlation coefficient for each Delphi for each combination of questions.

Table 1

Spearman Rank Correlation Coefficient for Each Question Combination

QUESTIONS	ASD/XR	AFAL	AFAPL	AFFDL
Q1-Q4	+.295	+.788	+.571	+.448
Q1-Q5	+.315	+.863	+.904	+.698
Q4-Q5	+.746	+.925	+.579	+.844

Clearly the ASD/XR answers suggest a greater change in laboratory emphasis (as shown by the low correlation between Questions 1 and 4, and between Questions 1 and 5) than that indicated by the other three groups. The system planners thus indicated a greater need for laboratory redirection than the laboratory personnel. Again we have an area of disagreement that might be camouflaged had one combined Delphi been utilized.

It is interesting that the AFAPL results indicate the least correlation between Questions 4 and 5. Although it might seem that the answers to these questions should correlate closely, there are several possible reasons to explain lack of correlation:

(1) A system may be important but not need substantial laboratory support.

(2) The necessary laboratory support might best be supplied by non-Air Force Laboratories.

(3) A system might be important if technologically feasible, but the necessary technological developments might not be considered likely in the near future.

Thus there might be a logical explanation for this lack of correlation. However, the data are surprising enough to indicate the desirability of a more detailed review of the AFAPL results. A subsequent review of the AFAPL answers indicated that many of the comments used to justify the apparently inconsistent results did involve considerations such as those listed above. However this example shows the value of looking for correlation between answers, and then, highlighting comments that justify departures from expected correlation.

Bias by Time Period

Figure 2 shows the average evaluations for Question 5 when the SCOs are grouped according to date of estimated technological feasibility. Obviously the

system planners (ASD/XR) with their more futuristic interests attach greater importance to the far-term, more advanced systems. This might be a result of the planners' greater awareness of the possible benefits these futuristic systems offer. However, a possible reason for the laboratory viewpoint might be a greater appreciation of the difficulty associated with solving the technological problems.

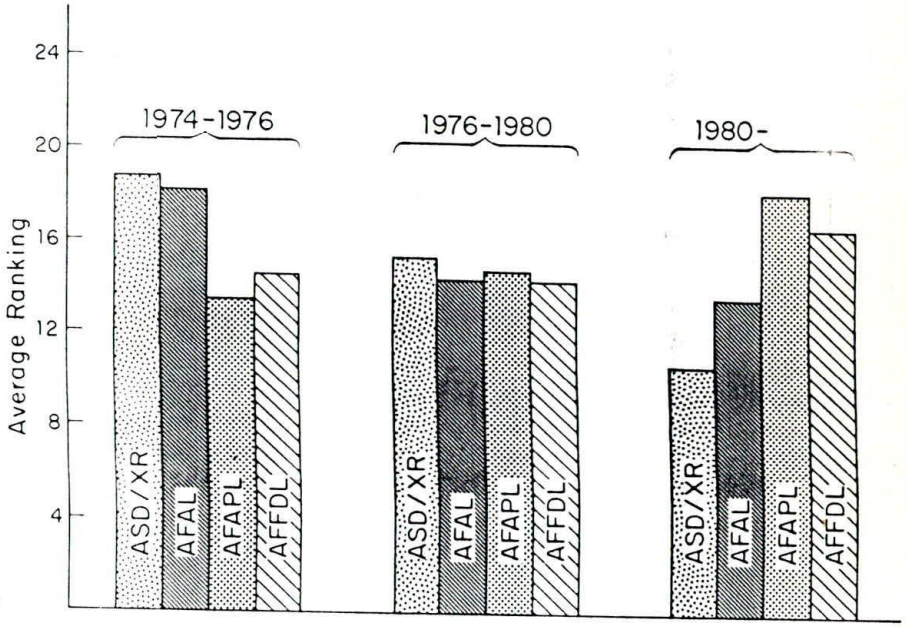


Fig. 2. Average responses for Question 5 when SCOs are grouped by year in which engineering development could start.

Again the results suggest the possibility of a communications gap. Both groups should benefit from an exposure to the reasoning that led to such diverse results. This type of exposure might best go beyond a Delphi-type exchange (which is generally limited in the amount of information transferred). Such a transfer of information is essential if the potential value of the SCO list is to be achieved. It is often not enough to establish priorities, unless all parties concerned accept and understand the logic that led to the priorities.

Laboratory Bias

There was some concern before the laboratory efforts were started that the

results might tend to be biased. Although the laboratory participants were instructed to rank the laboratory efforts by the total efforts from all the Air Force Laboratories, it was hypothesized that the participants' greater knowledge about their own laboratory programs and the natural tendency to promote one's personal interests would lead to a bias in favor of their laboratory's efforts. In order to test this hypothesis, the rankings obtained on Question 5 for the SCOs that received crucial support from each of the laboratories were compared.

In mid-1972, each of the laboratories published reports that reviewed their Technology Planning Objectives (TPOs) and the relevance of each TPO to each SCO. The top relevancy category indicated a TPO that the laboratory felt was essential to a given SCO. Table 2 shows the average ranking given for Question 5 to the groups of SCOs having a top relevancy match with the various laboratory TPOs, respectively. The lowest number in each column indicates the organization placing the greatest emphasis on that laboratory's progress. Therefore, bias would be indicated if the lowest number in a given laboratory's column was on the row corresponding to that laboratory's Delphi. The Delphi conducted in Laboratory B gave poorer (larger numerically) rankings to SCOs that were felt to be essentially related to one or more of their TPOs than any of the other groups, while the Delphi conducted in Laboratory A gave neither the poorest nor the best rankings to SCOs that were felt to be essentially related to one or more of their TPOs. Although the Delphi conducted in Laboratory C did give the best (numerically lowest) ranking to SCOs considered to be essentially related to their TPOs, the average ranking is not too different from those obtained in the other Delphis.

Table 2

Average Answer to Question 5 for SCOs That Are Related to Programs of Particular Laboratories

DELPHI	Relevancy Match with		
	Laboratory A	Laboratory B	Laboratory C
ASD/XR	12.7	15.1	15.6
Laboratory A	12.0	17.6	15.9
Laboratory B	15.1	22.3	15.8
Laboratory C	11.6	20.0	14.9

Thus, the hypothesis that a given laboratory Delphi tends to indicate biased rankings for SCOs that receive crucial support from that laboratory's effort does not appear to be valid. The answers obtained from Question 5 do not indicate the presence of laboratory bias.

Summary

The results discussed in this paper indicate information that was obtained by comparing several Delphi experiments utilizing experts from different organizations that probably would not have been obtained had one Delphi been run utilizing a subgroup from each of the groups of experts. Clearly differing organization viewpoints were identified, despite the fact that all of the groups involved very senior Air Force personnel who shared access to a considerable common information base. That is, all of the organizations had detailed knowledge of many of the same programs.

A noticeable difference in the amount of convergence was observed where in one case the apparently more expert group showed the poorest convergence. Disagreement was also apparent concerning the question of whether or not the laboratory programs should be redirected (as well as the related question of whether laboratory efforts should be directed toward near-term or more futuristic technology needs).

Comparisons of results were also made to determine if the laboratory group gave answers that were biased to support their own program. This investigation failed to show the presence of any real bias. This finding is very encouraging, for it suggests that at least these groups of technical experts were able to place their professional ethics above the common desire to promote personal gain. Had this not been true, the worth of this activity would be greatly reduced.

References

1. Olaf Helmer, and Nicholas Rescher, "*On the Epistemology of the Inexact Sciences*," *Management Science*, 6, No. 1 (October 1959).
2. Norman C. Dalkey, *The Delphi Method: An Experimental Study of Group Opinion*, The Rand Corporation, RM-5888-PR; 1969.

II. C. 1. Delphi Research in the Corporate Environment

LAWRENCE H. DAY

Introduction

The Delphi technique has become widely accepted in the past decade by a broad range of institutions, government departments, and policy research organizations ("think tanks"). These applications are described elsewhere in this book. The use of the Delphi approach in the corporate environment will be discussed in this section. Corporate utilization of Delphi is perhaps one of the least-known aspects of the technique's application. This is a result of corporations regarding the products of their Delphi exercises as proprietary and, hence, restricting their distribution or description in professional literature. A review of the long-term planning and futurist literature has revealed that few of the corporate efforts in this field have been documented in any detail.

The first part of this analysis will examine some uses of the Delphi technique in industry. This general review will be supplemented by an analysis of the application of the methodology in six Delphi studies conducted by the Business Planning group of Bell Canada. The Bell Canada experience will be followed by a description of some of the problems and issues that arise when using Delphi in the corporate environment. This review will conclude with some comments on the potential future of the Delphi technique in the business environment.

Examples of Corporate Delphi Research

Industrial Grouping or Professional Association Sponsorship

Delphi studies sponsored by corporations can be classified into three categories. The first category includes those studies sponsored by an industry association or a professional association.

These studies are usually of a broad nature and are concerned with projecting the future of an industry or perhaps even some broader societal field. The logistics of this application usually indicate that the study has to be contracted out to an independent consultant or research organization. While this type of application does not result in the day-to-day use of the methodology in business, the results of these Delphi studies are often exposed to a broad range

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

of high-level managers and executives in business. Thus, in this situation, corporations are the consumers of Delphi research rather than users of the technique itself.

Parsons & Williams Inc., an international consulting firm, conducted a Delphi study entitled "Forecast 1968—2000 of Computer Developments and Applications" in 1968.[1]. This study was undertaken for a conference on computer file organization held in Denmark sponsored by the International Federation of Information Processing Societies. This study examined future computer applications in business, the home, government, and institutions and projected the future of specific computing and technological developments [2]. Another recent Danish conference has also used a Delphi study as an input to panel discussion. This study, for a conference on long-term trends in personnel management called "Delphi 71–80", examined thirty-seven areas and predicted "how far society would be moved in certain directions by 1980" [3].

Sponsorship of Delphi studies by groups of firms are generally examinations of the future of an industry or an industrial segment. Current examples in this area include reviews of the cosmetics, recreation, and insurance markets. A "Delphi Panel on the Future of Leisure and Recreation" has been conducted by Social Engineering Technology Inc. (SET Inc.) [4]. This multiclient study was conducted by SET for a group of companies interested in future market opportunities in the recreational area that could develop through the impact of cultural change.

A Delphi study on the life insurance and other personal financial services markets is being conducted for two life insurance companies by the Canadian consultants Ducharme, Deom & Associes Ltee. The objective of this study is to "design a picture of the Life Insurance market and other personal financial services in the 1980's in terms of external environmental variables..." [5].

Individual Corporate Sponsorship

This second category of Delphi research is similar to the first. The grouping includes individual corporations who sponsor Delphi studies at research organizations on subjects of general or specific interest. The Institute for the Future (IFF) has conducted the largest number of these studies on this basis. In the case of IFF, the study results are in the public domain [6]. Several of these studies have been concerned with the impact of the computer/communications revolution:

(1) The Future of the Telephone Industry; sponsored by the American Telephone and Telegraph Co. (New York, N. Y.) [7].

(2) The Future of Newsprint; sponsored by MacMillan Bloedel Ltd. (Vancouver, B. C.) [8].

(3) On the Nature of Economic Losses Arising from Computer Based Systems in the Next Fifteen Years; sponsored by Skandia Insurance Co. (Stockholm, Sweden) [9].

Another IFF study sponsored by Owens-Corning Fiberglas Corp., examines "Some Prospects for Residential Housing by 1985" [10]. The research program by Owens-Corning also produced an IFF study on "Some Developments in Plastics and Competing Materials by 1985" [11].

An interesting IFF study sponsored by General Telephone and Electronics Corp., examines "Some Prospects for Social Change by 1985 and Their Impact on Time/Money Budgets" [12]. Hence the various corporate-sponsored studies at IFF fall into general research categories (e.g., GTE, Skandia, and Owens-Corning) and specific industry research studies (e.g., AT&T, and MacMillan Bloedel).

Delphi research has also been sponsored by corporations in other research organizations. The Danish study referenced above on personnel management was extended by the consultants (Management Training Division of the Danish Institute of Graduate Engineers) to three groups of employees from the printing firm CON-FORM [13]. The Pace Computing Corporation sponsored a study by marketing research consultants to determine the potential demand for its services [14].

The Delphi technique has recently come to the attention of other marketing research consultants. This should lead to an expansion of corporate sponsorship of these studies, as the market researchers will promote the technique with customers who might not normally become exposed to Delphi or other longer-term planning techniques. One Canadian market research organization has mentioned the technique in its periodic newsletter to clients [15]. As noted in the first category, this sponsorship leads to senior management exposure to the technique even though the corporations do not conduct the studies themselves.

Corporate In-House Delphi Research

This final category includes Delphi studies conducted by research or planning groups within the corporation itself. In this case, members of the corporation staff become very involved with Delphi as they must master the technique as well as use the study results. This category includes most proprietary uses of the studies and their results, and they usually have not been published or distributed widely.

The best-known example of corporate Delphi experience is that of TRW. While the TRW Delphi studies are unpublished and proprietary, a number of papers have been published by North and Pyke on the technique's use and selected study results [16]. TRW's modification of Delphi has been named PROBE. The initial study was started in 1965 and resulted in a fifty-page document containing a set of 401 forecasts published in June 1966 [17]. This has been refined in a second study called "Probe II." The reader is referred to the papers noted above for an elaboration of the TRW experience with Delphi.

A Delphi study was conducted in the U. K. by the Hercules Powder Co. Ltd. on the future of the British Chemical Industry in the 1980s. This has been discussed in several articles by Parker of Hercules [18] and used as a case example in the book *Technological Forecasting* by Wills [19]. Other U. K. experiences with technological forecasting and Delphi were referenced in a recent article "Technological Forecasting in Six Major U. K. Companies" by Curill [20]. While he does not name specific companies, he notes that Delphi has been used by: a "Glass" Company, a "Consumer Goods" Company, two "Chemical Companies," and an "Electrical Engineering" Company and that this is one of the most popular techniques of those companies utilizing technological forecasting methodologies.

The medical field has been explored in a U. S. study by Smith, Kline and French, a major pharmaceutical manufacturer [21]. Three other large U. S. pharmaceutical companies are reported to have conducted studies as well [22].

Industries undergoing rapid change have been frequent targets of Delphi research. The merging computer and communications fields are an example of this phenomenon and a significant number of industrial studies have been conducted. IBM has conducted an internal study on future computer applications. ICL in England have also sponsored a Delphi study. In addition to sponsoring the IFF research, AT&T conducted a study, "Communication Needs of the Seventies and Eighties" (internal document) [23]. Bell Canada has undertaken six studies projecting technological and social trends in four main areas: education, medicine, business information systems, and "wired city" services (all proprietary) [24]. The Trans-Canada Telephone System conducted an internal Delphi study on future data service needs. British Columbia Telephone is conducting a Policy Delphi with senior managers.

Summary: Corporate Examples

The discussion of various forms and examples of corporate sponsorship of Delphi studies is not intended to be all-inclusive. It merely attempts to outline, on an international basis, a few of the known examples of the scope of Delphi usage in industry. The next discussion will center on our experiences in Bell Canada as one example of how corporate Delphi studies have been conducted.

Delphi and Bell Canada

Background

Bell Canada is an operating telecommunications company serving the provinces of Ontario and Quebec in Canada. In addition to offering voice, data, and visual telecommunications services, Bell Canada owns a large manufacturing subsidiary (Northern Electric) and an R & D subsidiary (Bell-Northern Research). There are also several other subsidiaries in the telephone, directory,

and electronic components manufacturing fields. The Business Planning Group in Bell has the responsibility for identification of corporate opportunities (or threats) that will arise through changes in society and/or technology in the next decade or two.

The communications field is in the midst of rapid change which will have a significant impact on its intermediate and long term future. Highlights of these changes include:

- merging computer and communications technologies
- regulatory changes introducing new competitive elements
- emerging visual telecommunications markets
- perceived and projected social changes
- increasing costs of investment options

The Business Planning Group surveyed these various pressures in the late 1960s as it was developing a study plan to evaluate future trends in the visual and computer communications fields. There was a distinct lack of qualitative data on potential futures for these fields, especially in the Canadian environment. An examination of various potential technological forecasting techniques indicated that the Delphi technique would fill the perceived information gap.

Bell Canada Delphi Study Development

The individuals involved in designing, conducting and managing the Business Planning Delphi efforts have generally had a marketing background. This background includes both academic training and professional experience. These background factors were important determinants of the approach followed.

Initial steps relied upon the basic marketing approach of defining the "market segments" that will have the most important impact on future applications of visual and data communications. These segments were chosen after preliminary studies of potential segments and taking account of the time and resources available. The final choices were future applications in the educational, medical, information systems, and residential markets.

The basic philosophy in the studies was to examine the future of applications in these segments from a user point of view, not from the direction of technological imperatives. The initial questionnaires were prepared after extensive literature reviews of potential developments in each of the chosen areas. The approach in questionnaire design was to guide the discussions in some basic areas of interest in a segment rather than start with blank paper and ask the experts to suggest the most important areas of interest. Since the panelists were actively encouraged to suggest new questions or modifications to existing ones, the potential for significant study bias by the designers was low. This approach also helped reduce the number of rounds required for the studies and hence saved time for the participants and study managers alike.

Table 1

VALUE CHANGES IN NORTH AMERICAN SOCIETY
1970 – 2000

	SIGNIFICANT INCREASE	SLIGHT INCREASE	NO CHANGE	SLIGHT DECREASE	SIGNIFICANT DECREASE
TRADITIONALISM				▒	▒
HARD WORK AS A VIRTUE				▒	
AUTHORITARIANISM				▒	
MATERIALISM			▒	▒	
REWARDING WORK AS A VIRTUE		▒	▒		
INDIVIDUALISM					
INVOLVEMENT IN SOCIETY		▒			
PARTICIPATION IN DECISION MAKING		▒			
SELF EXPRESSION	▒	▒			
ACCEPTANCE OF CHANGE	▒	▒			

NOTE: The shaded areas above represent the median responses from the five Bell Canada Delphi studies noted in the footnotes. Shading over two areas indicates differences in opinions between the various panels.

Next, initial questionnaires were pretested with groups of readily available experts. This proved to be a very valuable step, as poorly worded questions or confusing questionnaire design were largely eliminated before the errors could be inflicted upon the Delphi panel. This step adds time to the study and may be somewhat ego deflating at times for the study managers; however, it pays good dividends in higher quality, less ambiguous results, and happier panelists.

Delphi Study Results—Education, Medicine, and Business

The initial studies in education, medicine, and business followed a similar format. The first part of the questionnaire asked the panelists to project their views on the long-term (thirty years) future of some basic North American values. The purpose in asking these questions was more to help the panelists get in a societal frame of mind when answering the rest of the questionnaire than to obtain the societal trend data itself. When the social trend views of the various groups of experts, as shown in Table 1, were compared after all of these studies were completed, it was interesting to note how similar the results were, considering the diverse background of the 165 individuals in the various panels (there was no interpanel communication during the studies).

Other areas of each study also explored nontechnological developments as well as the adoption of systems to serve various applications. Table 2 illustrates some of these nontechnical factors considered in the three studies [25].

Table 2

Nontechnological Factors Considered in the Bell Canada Delphi Studies

EDUCATION	MEDICINE	BUSINESS
1. Value Trends	1. Value Trends	1. Value Trends
2. Evolution in School Design	2. Trends in the Medical Profession	2. Changes in Business Procedures
3. Changing Role of the Teacher	3. Changes in the Medical Environment	3. Trends in Business Physical Environment

The Education Delphi examined potential adoption of three basic types of educational technologies: Computerized Library Systems (CLS), Computer Aided Instruction Systems (CAI), and Visual Display Systems (including IRTV—Instant Retrieval Television). The summary forecasts of the panel are shown below in Table 3 [26]. The projections on the use of terminals for input and output purposes for CAI and CLS are shown in Table 4 [27]. In both

tables, threshold market penetration values of 20 percent and/or 55 percent were used for the technologies. This gives the panelists and readers some feeling for the scope of service acceptance in the markets under consideration.

The Medical Delphi explored acceptance of a number of developing medical technologies. These included: Multiphasic Screening, Computer-Assisted Diagnosis, Remote Physiological Monitoring, Computerized Medical Library Systems, and Terminal Usage. Table 5 illustrates some of the summary results from the medical study [28]. The format and adoption thresholds were similar to those in the Education Delphi.

The Business Information Processing Technology study examined trends in Management Information Systems, Mini and Small Computers, Terminals and Data Processing. Table 6 summarizes the median conclusions of the panel on 20 percent acceptance of various technologies both in business and in the home [29]. It should be noted that this Table, like the earlier ones, summarizes only the median statistical conclusions of the panels. Panelists were sometimes split into "schools of thought" on various issues. These opinion splits were often not reflected in graphic presentation of the results. In these cases the panelists were encouraged to debate their differences in writing through the rounds of the various studies. These differences are reflected in the reports along with supporting assumptions and comments of the panelists. It was found that the panelists' comments and their analyses were often very important modifications of the statistical projections shown in the above tables.

Design of the Future of Home Services Delphi

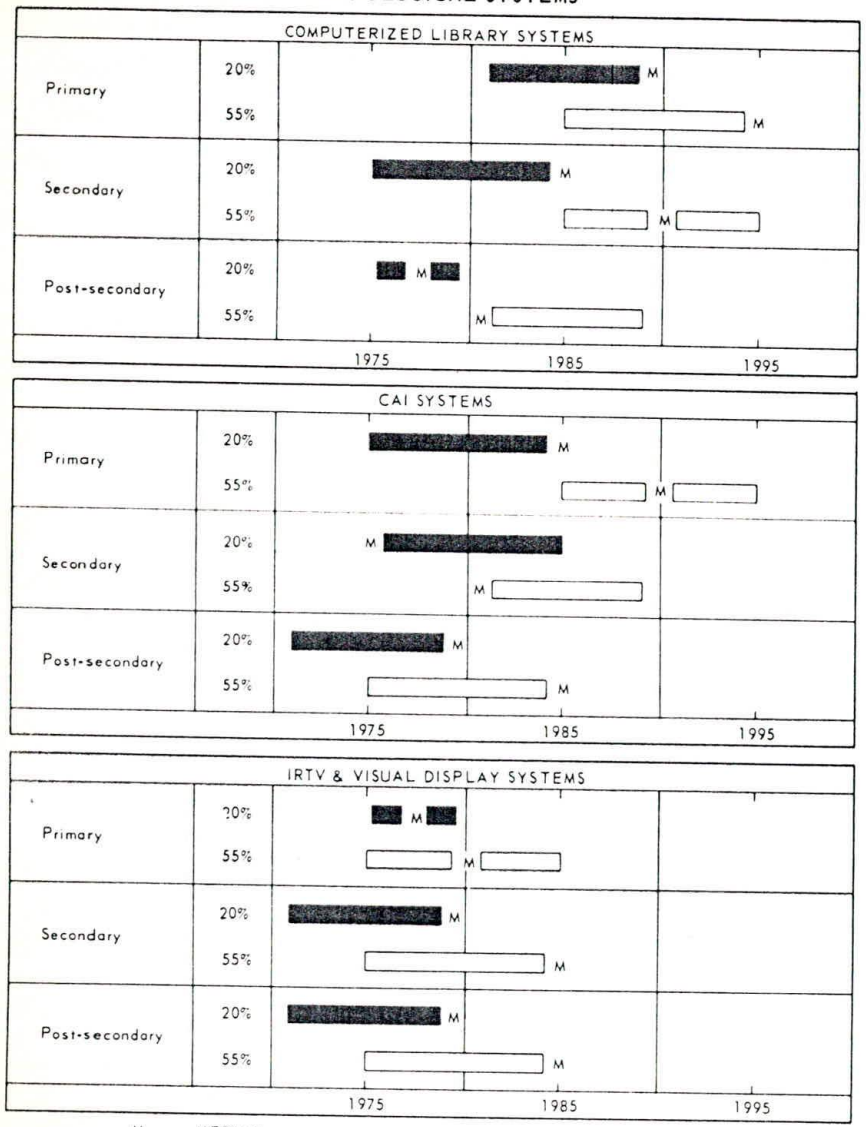
The three studies outlined above provided an important new source of input to the Business Planning group. However, one market area was still largely unresolved: the future of communications services in the residence market. Each of the above studies asked a few questions on home services but their combined answers still left a large information gap. Determining how to obtain the additional information reopened some important internal differences of opinion on the value of Delphi for this purpose.

The main issue revolved around the definition of what is an "expert" in the residential field. This question had developed in creating the earlier expert panels as well. In two cases (Education and Business) the question was whether selected industry specialists within the telecommunications industry were as knowledgeable as experts in the above fields when it came to projecting the future. This question was resolved by conducting two studies in the Education and Business fields. In both cases, independent panels of "internal" and "outside" experts were used.

The question in the residential study was whether housewives or researchers and planners were the best experts on the future adoption of communications services in the home. In this case the study was totally service- and *not*

Table 3

LIKELY TIMING OF ADOPTION OF TECHNOLOGICAL SYSTEMS



SOURCE: DOYLE AND GOODWILL, EDUCATIONAL TECHNOLOGY, P. 64

NOTE: THE MEDIAN SYMBOL 'M' ON THIS CHART AND THE FOLLOWING ONES ONLY INDICATES THE 5 YEAR TIME FRAME THAT THE MEDIAN LIES IN, NOT ITS EXACT LOCATION. THUS, THE MEDIAN AND THE FIRST OR THIRD QUARTILE MAY BE IN THE SAME 5 YEAR PERIOD IN SOME CASES.

Table 4

TERMINAL TO BE USED FOR CAI AND/OR CLS

INPUT CAPABILITY	WILL BE IN 20% OF ALL	70–75	76–80	81–85	86–90	91–99	LATER	NEVER
Touch-Tone Telephone Type Keyboards	Homes			M				
	P. Schools			M				
	S. Schools			M				
	P.S. Schools		M					
Typewriter Type Keyboards	Homes			M				
	P. Schools			M				
	S. Schools		M					
	P.S. Schools		M					
Electronic or "Light" Pens	Homes				M			
	P. Schools				M			
	S. Schools			M				
	P.S. Schools			M				
Human Voice	Homes					M		
	P. Schools					M		
	S. Schools				M			
	P.S. Schools				M			
Specially Formed Printed Characters	P. Schools				M			
	S. Schools				M			
	P.S. Schools			M				

Table 4 (continued)

OUTPUT CAPABILITY	WILL BE IN 20% OF ALL	TERMINAL TO BE USED FOR CAI AND/OR CLS: 70–75	76–80	81–85	86–90	91–99	LATER	NEVER
Computer Voice Reply	Homes				M			
	P. Schools				M			
	S. Schools				M			
	P.S. Schools				M			
Printed Page	Homes				M			
	P. Schools			M				
	S. Schools			M				
	P.S. Schools		M					
Traditional Television Screen Display	Homes			M				
	P. Schools			M				
	S. Schools			M				
	P.S. Schools		M					
Large Flat Television Screen Display	Homes					M		
	P. Schools				M			
	S. Schools			M				
	P.S. Schools				M			
Audio Video Recorder Recording	Homes				M			
	P. Schools			M				
	S. Schools			M				
	P.S. Schools		M					
Picturephone Screen Display	Homes				M			

P. SCHOOLS = PRIMARY SCHOOLS
S. SCHOOLS = SECONDARY SCHOOLS
P.S. SCHOOLS = POST SECONDARY SCHOOLS

Source: Doyle and Goodwill, Educational Technology, P. 41–2

Table 5

PERIOD OF SYSTEM REFINEMENT AND EARLY ADOPTION OF TECHNOLOGICAL SYSTEMS (20% UTILIZATION)

INSTITUTION	TECHNOLOGICAL SYSTEM	70–75	76–80	81–85	86–90	91–99	LATER	NEVER
Physicians in Solo Practice	MS			M				
	CAD				M			
	CLS				M			
Physicians in Group Practice Who Share Facilities Only	MS		M					
	CAD			M				
	CLS			M				
Physicians in Group Practice Who Share Records & Incomes etc.	MS		M					
	CAD			M				
	CLS			M				
Acute General Hospitals	MS		M					
	CAD			M				
	CLS			M				
Occupational Medicine	MS		M					
	CAD				M			
	CLS				M			

MS– ▬ – MULTIPHASIC SCREENING
CAD– ▨ – COMPUTER ASSISTED DIAGNOSIS
CLS– ▩ – COMPUTERIZED LIBRARY SYSTEMS

Source: Doyle and Goodwill, Medical Technology, P. 60

Table 6
An Exploration of the Future in Business Information Processing Technology

DEVELOPMENTS*	1970–75	1976–80	1981–85
THE WORK LOCATION (WHERE WHITE-COLLAR AND CLERICAL EMPLOYEES WILL PERFORM THEIR JOB DUTIES)		OFFICE CENTER	
MANAGEMENT INFORMATION SYSTEMS		MIS FOR LARGE FIRMS	MIS FOR MEDIUM FIRMS INTEGRATED MIS
MINI AND SMALL COMPUTERS		FREE-STANDING COMPUTERS COMMUNICATION COMPUTERS DUAL-PURPOSE MACHINES	
TERMINALS		T-T TELEPHONE (BUSINESS TYPEWRITER KEYBOARDS (BUSINESS COMPUTER VOICE REPLY (BUSINESS PRINTED PAGE (BUSINESS TRADITIONAL TV SCREEN (BUSINESS ACOUSTICALLY COUPLED TERMINALS (BUSINESS PLUG-IN PORTABLE TERMINALS (BUSINESS	T-T TELEPHONE (HOMES) PICTUREPHONE FOR FACE-TO-FACE COMMUNICATION (BUSINESS) TOUCH-SENSITIVE INPUT (BUSINESS) OCR (BUSINESS) 2-WAY LARGE SCREEN COLOR TV (BUSINESS) AUDIO-VIDEO RECORDERS (BUSINESS) PICTUREPHONE DISPLAY (BUSINESS)
DATA PROCESSING	MEDIUM-SIZE FIRMS IN MANUFACTURING AND SERVICE INDUSTRIES WILL BE UTILIZING D.P. FACILITIES	SMALL-SIZE FIRMS IN MANUFACTURING INDUSTRIES WILL BE UTILIZING D.P. FACILITIES	SMALL-SIZE FIRMS IN SERVICE INDUSTRIES WILL BE UTILIZING D.P. FACILITIES
TECHNOLOGY IN THE HOME			ONE-WAY (INCOMING) AUDIO-VISUAL COMMUNICATION

DEVELOPMENTS*	1986–90	1991–95	LATER
THE WORK LOCATION (WHERE WHITE-COLLAR AND CLERICAL EMPLOYEES WILL PERFORM THEIR JOB DUTIES)	NEIGHBORHOOD R.W.C. THE MOBILE WORKER	HOME REMOTE WORK CENTER	
MANAGEMENT INFORMATION SYSTEMS	MIS FOR SMALL FIRMS DISTRIBUTED MIS		
MINI AND SMALL COMPUTERS			
TERMINALS	LIGHT PENS (BUSINESS) HUMAN VOICE INPUT (BUSINESS) PICTUREPHONE FOR FACE-TO-FACE COMMUNICATION (HOMES) PICTUREPHONE FOR DATA INPUT (BUSINESS) HANDWRITTEN INPUT (BUSINESS) COMPUTER VOICE REPLY (HOMES) TRADITIONAL TV SCREEN (HOMES) AUDIO-VIDEO RECORDERS (HOMES) FULLY PORTABLE WIRELESS TERMINALS (BUSINESS)	PICTUREPHONE FOR DATA INPUT (HOMES) SCRIBBLEPHONE (HOMES) SCRIBBLEPHONE (BUSINESS) PRINTED PAGE OUTPUT (HOMES) LARGE FLAT COLOR TV DISPLAY (HOMES) LARGE FLAT COLOR TV DISPLAY (BUSINESS) PICTUREPHONE DISPLAY (HOMES)	TYPEWRITER KEYBOARDS (HOMES) LIGHT PENS (HOMES) HUMAN VOICE INPUT (HOMES) HANDWRITTEN INPUT (HOMES) TOUCH-SENSITIVE INPUT (HOMES) OCR (HOMES) 2-WAY LARGE SCREEN COLOR TV (HOMES)
DATA PROCESSING			
TECHNOLOGY IN THE HOME	TWO-WAY AUDIO-VISUAL COMMUNICATION		

*THE RESULTS INDICATE WHEN THE DEVELOPMENT WILL REACH A 20 PERCENT LEVEL OF ACCEPTANCE BY THE APPROPRIATE UNIVERSE. THE MEDIAN "EXPERT" RESULTS ARE PRESENTED.

technology-oriented. This issue was resolved by establishing two competing panels to forecast the future in this area. One panel consisted of housewives (experts through experience) and the other of experts through research or planning for "wired city" services. The study design and steps followed are shown in Table 7.

Table 7

Study Design: Future of Communications Services into the Home

1. Literature Search
2. Assemble Panels of "Experts" and Housewives
3. Design Draft Questionnaire
4. Pretest Questionnaire
5. Print and Distribute Revised Questionnaire (Identical to Both Groups)
6. Prepare Statistical Analysis of 1st-Round Answers
7. Prepare Analysis of Supporting Comments from Each Group
8. Design, Pretest, Print and Distribute 2nd-Round Questionnaire showing:
   a) 1st-Round Statistical Results from Each Group on One Page
   b) 1st-Round Supporting Comments from Each Group on Opposite Page
   c) Ask for Resolution of Answers within Each Panel
   d) Highlight Differences between Panels and Ask for Resolution
9. Prepare Final Analysis

The important steps that are different from normal Delphi studies are 7 and 8 in Table 7. The results should stimulate debates between the panels if this approach is going to derive the maximum benefits from both panels. Table 8 shows a typical two-page feedback and question set from the Home Communications Delphi [30]. The importance of obtaining feedback comments from the panelists is illustrated in the table.

This study examined future acceptance of electronic shopping from the home, remote banking, electronic home security services, and electronic programmed education in the home. The study also explored the future of ten types of information retrieval services that may be offered to homes. Table 9 illustrates some summary results of the study [31].

Summary: Bell Canada Delphi Studies

Business Planning efforts in the six studies outlined above have resulted in an important increase in the availability of qualitative data for planning purposes. Experience with the technique resulted in significant modifications from the original RAND approach, especially with the emphasis on analyzing the panelists' comments and establishing threshold levels of acceptance. The use of

Table 8

SHOP–FROM–HOME SERVICE

In predicting which types of products will be purchased through a Shop–from–Home system, the housewives and experts disagreed on a number of items. The summarized answers are presented below with the answers for the expected costs of such a service. Some typical comments are presented on the facing page. There appear to be significant differences in such items as produce, small and large appliances.

TYPE OF PRODUCT:

E / H	% YES	% NO	over 20% more	5–20% more	0–5% more	same	0–5% less	5–20% less	over 20% less
meat	48 / 58	52 / 42			E / H				
produce	**52** / **75**	**48** / **25**			E / H				
other perishables (dairy, bread, etc.)	100 / 83	0 / 17			E / H				
groc. dry goods	100 / 96	0 / 4			E / H				
clothing	58 / 48	42 / 52			E / H				
small appliances	**74** / **96**	**26** / **4**			E / H				
drugs and cosmetics	90 / 75	10 / 25			E / H				
large appliances	**47** / **79**	**53** / **21**			E / H				

NOTE: Shaded area represents significant differences between panels
E = Expert Panel: median or percentage response.
H = Housewife Panel: median or percentage response.

Do you wish to change any response? Could you comment on the differences between the two groups? Do you have any concluding comments

ROUND II COMMENTS

(Cont'd.)

Table 8 (continued)

SHOP-FROM-HOME (CONTINUED)

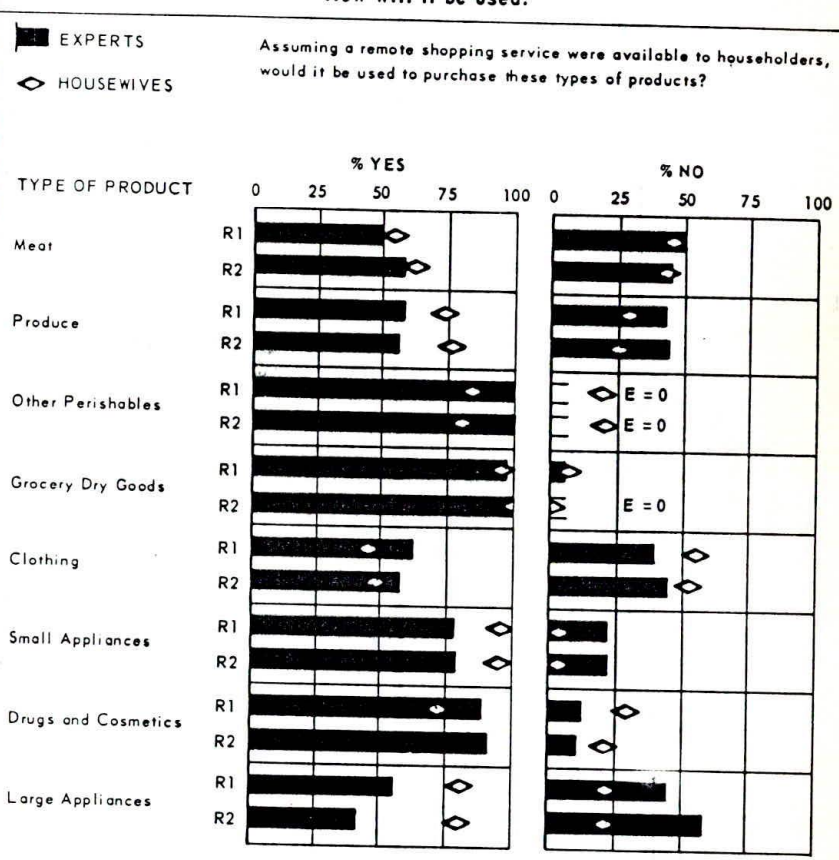
EXPERTS	HOUSEWIVES
The panelists had the following generally favorable comments on the service and its expected costs.	
1. "Will succeeds only if it offers an economic advantage (savings on sales people, expensive store space, electricity and display, etc.)"	1. "A useful service for those living in suburbs or remote rural areas whose access to cities and stores is difficult, also for elderly people."
2. "Assumption – meat and produce supplied from a familiar supplier and good previous experience."	2. "I don't think it should cost any more because it would not need a huge store or many salesclerks, and certainly shoplifting would be a thing of the past for anyone participating in the service."
3. "Better price and product comparisons can allow the purchaser to save money as well as time."	3. "Small or large appliances (if brand name is known) present no problems for Shop-from-Home."
These comments reflect some of the typical reservations about the service:	
1. "Meat, produce – these items are bought by feeling, smelling, and seeing the specific items in question and comparing them to the others available in the display. This cannot be done remotely, hence the continued existance of meat markets, fruit stands, etc."	1. "For most housewives, shopping is a diversion and a break in a routine."
2. "Insuring quality will raise the price of perishables above attractive cost."	2. "Meat and produce – Individual specific quantities required; bult packaging not always desirable."
3. "People will expect large cost reductions for appliances in return for not being able to see the merchandise. . .Several visits to warehouses will be typical even with home shopping."	3. "The cost of these items in our daily living budget is expensive enough as it is; if it would cost more to buy these things at home, not many people would take advantage of it unless they were unable to get out of the house. . . Clothing and large appliances would be difficult with this service; with clothing you like to examine the fabric to see how well it is made; with large appliances you would want to discuss with the salesman the pro's-con's of the appliance."

Source: Bedford, Questionnaire, pp. 2-3

Table 9

SHOPPING FROM THE HOME

How will it be used?



While there seemed to be general agreement between panels that grocery dry goods and certain perishables would be purchased through a remote shopping service, the housewives believed that a number of other products would also be purchased with the aid of this service. The comments reinforce this pattern ("that housewives would 'trust' the store . . . . but experts have a smaller trust"). At the same time however, the comments reflect the feeling that this service would eliminate a significant element in housewives' social activities. It seems that if remote shopping is to become widely accepted, housewives will need to have recourse to alternate modes of socializing.

There did not appear to be any major change in panelists' attitudes between rounds. The largest shift was the decrease in the number of experts expecting large appliances to be purchased from the home.

(Cont'd.)

Table 9 (continued)

COMMENTS

EXPERTS	HOUSEWIVES
ROUND ONE	
"Meat, produce – these items are bought by feeling, smelling, and seeing the specific items in question and comparing them to the others available in the display. This cannot be done remotely, hence the continued existence of meat markets, fruit stands, etc."	"A useful service for those living in suburbs or remote rural areas whose access to cities and stores is difficult, also for elderly people. Clothing would be difficult in my opinion as there would be no way of trying on and fitting. Endless returns and exchanges would defeat the original purpose of the shop-from-home service. Otherwise, most things could conveniently be bought from the home."
"In regard to meats, the consumer will always want hands-on experience. The butcher has a very poor public image. Consumers will want to see how much fat the steak has."	"Small or large appliances (if brand name is known) present no problems for Shop-from-Home."
"It's not a function of the product but rather how a given purchaser perceives the product. The gourmet thinks of food as an art form, and wants to be there. The cafeteria operator couldn't care less. Since many stores are our contemporary art museums, the process of shopping is quite complex."	"For most housewives, shopping is a diversion and a break in a routine."
	"Clothing and large appliances would be difficult with this service; with clothing you like to examine the fabric to see how well it is made; with large appliances you would want to discuss with the salesman the pro's-con's of the appliance."
ROUND TWO	
"I still don't accept the idea that housewives will buy meat, produce, and large appliances via shop-from-hone service. The first time the housewife gets burned using S-F-H (ie. a poor steak, unfresh meat, etc.), she will go back to using the store."	"Change my response for meat, produce and clothing to NO because one of the main reactions to the product is through a confrontation with it which results in the buying or rejecting of it."
"My 'NO' responses were based on the belief that an intelligent housewife would wish to carefully choose meat and produce herself unless her butcher and grocer were intelligent people who knew her states and preferences well."	"It is rather amusing that housewives would 'trust' the store for meat and produce but the experts have smaller trust. Go with the housewives – we do most of the shopping."
"Experts tend towards NO, probably because they buy these products as men with a technical bent."	"I still don't like the concept of shopping on a larger scale from home. It strikes me that this system is geared to the larger chain stores which already offer catalogue service as a type of shop-from-home feature (with regard to clothing, appliances, etc.) But I wonder how such a service would affect smaller businesses and specialty stores which definitely couldn't operate such a service themselves."
"Experts seem to ascribe greater weight to first hand observation of product in buying decision than do housewives. Husbands may be more similar to experts in this propensity."	

Source: Bedford, Future of Communications Services in the Home, pp. 28-29.

Delphi to evaluate the marketability of services by users rather than predicting the median dates of potential technological development was also helpful. An analysis of completed studies has also revealed comparison information on the use of internal panels vs. external panels. Thus, Business Planning has learned much about the technique while obtaining useful information. This three-year intensive involvement with the technique has also given Business Planners a realistic view of some of the issues that arise when operating with Delphi in the corporate environment.

Delphi in the Corporate Environment

The issues discussed below are based upon Bell Canada experience and on discussions with individuals who have conducted similar studies in other corporations. These issues will probably be faced by any group in industry that launches a serious attempt to conduct professional quality research in this area.

Should Corporations Pay for Basic Delphi Research?

This, of course, is the fundamental question that must be answered. The emphasis here is on in depth research, since this will often result in a significant allocation of time and money resources in an area where immediate payoff is not clear to senior management. Other forms of business research (market research, operations research, economic research, etc.), have more precise goals and utilize more understandable techniques. The benefits of Delphi research will not be reiterated here, but the corporate planner has to recognize that this is one area not easily understood by busy executives.

This is part of the more basic question on the value of long-term planning in business. Generally, long-term planning has become an accepted part of business today. Delphi research is most needed in the long-term planning function where the conditions of uncertainty are the most evident.

Basic research of this nature is beginning to fall into the general area of corporate social responsibility. Many corporate decisions made today will have important secondary effects for decades to come. The rapid rise of interest in government, academia, and business in what is termed "Technology Assessment" [32] is one reason for considering this as a part of corporate social responsibility. Delphi study results can be used as corporate inputs to the development of technology assessment equations [33].

The North American telecommunications industry is generally privately owned but regulated by government agencies. Recent studies in the U. S. [34] and Canada [35] have noted and projected an accelerating trend on the sharing of planning data between the regulators and the corporations. While many industries are not formally regulated, none can escape the growing governmental and public scrutiny of the consequences of their actions and

plans. Sharing of basic planning information such as Delphi study results can help develop a common assessment data base on both a corporate and a public basis.

The social/political reasons outlined above are especially important when evaluating the cost/benefit analysis of undertaking corporate Delphi research. However, this is a more obvious reason for doing this type of research: the results can be used to help make business decisions.

Using Delphi Results in Business

Many corporate Delphi studies conclude with the publication of a report to the panelists and management outlining the study findings. The problem of recognizing the value of this research develops if that is where the Delphi studies end. These basic Delphi reports are important as:

(1) Educational tools to inform senior managers of the panelists' views of potential futures or various areas of interest to the business.
(2) Trading documents with other planners and researchers.
(3) Environmental trend documents that can help technological planners in research labs.

The use of the Delphi results must then become more directed. One useful way of using the specific results is to regard them as a data base to be drawn upon when preparing corporate recommendations in specific topic areas. The Delphi forecasts should be combined with other relevant material (trend extrapolations, multiclient study results, market research data, etc.) in order to present a comprehensive estimate of the impact of a forthcoming decision [36]. These combinations may be in the form of cross-impact matrices, scenarios, market analyses, etc. The use of the Delphi data with other material helps create confidence in the overall package. It is rare that the Delphi results alone can help resolve an issue when preparing a recommendation. Of course, this approach is useful in the nonbusiness environment as well.

The Bell Canada Delphi study results are regarded as part of a data base. Each of the Delphi forecasts has been abstracted, key-word indexed, and stored in an on-line computerized information retrieval system. Other items in the data base are also stored in the same manner. These items may be forecasts from trend studies, material from other internal research, appropriate forecasts from studies available from government institutions, policy research institutes, corporations, etc. The data base is used in the creation of several types of Business Planning outputs (Note: the Delphi material is an *input*, not an output). These outputs include:

(1) Specific Service and Business Proposals designed to exploit identified opportunities.

(2) Environmental Outlook Reports that identify trends and potential future events which may impact on the company or a specific corporate function (i.e., marketing).
(3) Targeted Outputs designed to present selected material to various government commissions, task forces, as well as other research organizations.
(4) Subject Sourcebooks and Information Packages which combine all of the available information on a specific field of interest into an annotated document for use by other planners in the Bell Canada Group.
(5) Methodology Analyses that document what we have learned using a particular technological forecasting technique.

In all of the above cases, the Delphi research material has been combined, massaged, analyzed and placed in perspective vis-à-vis other future information.

Delphi research can also be used in obtaining certain types of information not usually available from normal marketing research activities. Statistical polling of consumers can only produce a limited base of attitudinal data. Feedback and interaction are not possible here. On the other hand, group depth interviews can run into many of the problems that the lack of anonymity produces. The modified version of Delphi used by Bedford enables the researcher to generate opinions and conflicts between potential consumer groups for new products or services. This controlled conflict with feedback produces valuable behavioral information that would not emerge using other techniques. This data can be used for product or service modification or redefinition of market opportunities.

Delphi research in business must be regarded as a means toward ends rather than as an interesting intellectual end in itself. Use of the technique as indicated above can result in an affirmative answer to the question: "Should corporations pay for basic Delphi research?"

Misusing Delphi Results in Business

Delphi study results can be used to advantage in the corporate environment. The reverse situation is also possible. One of the most common situations is for the results of the study to be viewed as representing a corporate position, policy, or forecast.

This is not the case for the results of the vast majority of Delphi studies which represent the combined and refined wisdom of the particular panel of experts on the study [37]. One of the recurring problems with the Bell Canada Delphi's has been the suggestion that the studies represent a corporate position even though this suggestion is explicitly refuted in the reports.

A related problem is the temptation of corporate public relations groups to distribute the studies as another P. R. tool. This can be especially problematic,

nce Delphi panelists are assured that their contributions are provided in onfidence on a professional basis. The use of the study results in this manner ould backlash on the study director, especially if he hoped to conduct future udies using panelists drawn from the same population. Of course, the value of he documents as trading vehicles would diminish as well if they were handled this manner.

A further issue is related to the perceived precision of the study results. Many Delphi studies process the interim and final results using computers. This ermits the presentation of statistical results that "appear" very precise to the sual observer or individuals accustomed to dealing with the results of conomic and statistical research. The findings of Delphi studies are subject to ore interpretation than are most research results. The planning group should y to ensure that others using the results as a data base are aware of the arious strengths and weaknesses of the information.

n-House vs. Consultant-Conducted Studies

Another question that must be resolved is whether or not to conduct the study sing in-house or consultative resources. This decision can be analyzed by onsidering the following factors.

(1) Single vs. Multiple Studies. There is a definite learning curve involved hen conducting Delphi studies. Serious attempts utilizing the technique equire an initial time and resource investment to learn how Delphi studies are ffectively conducted. This investment will pay continuing dividends if a umber of studies are planned. These rewards include the development of a ore knowledgeable planning staff that fully understands the strengths and eaknesses of the data obtained from the studies. On the other hand, conduct- ng a single study with a planning group unfamiliar with the use of the chnique may be a costly venture that produces mediocre results.

(2) Study Sophistication. The corporation may be dealing with a subject matter that is changing rapidly and is very complex (e.g., computer tech- ology). The firm may also want a large number of factors considered in the udy. In this case, the use of outside consultants who have considerable xperience in conducting large complex Delphi's (i.e., I.F.F.) may be more roductive. The use of these consultants will also ensure that the best modifica- on of the technique is applied to the company's problem. Experienced onsultants are constantly learning more about the technique and are modify- ng it as a result of that experience. The time delays that occur before this xperience is reflected in the professional literature may mean that the orporate researcher is using a somewhat less than optimum version of the chnique.

(3) Proprietary Research. The problems involved with proprietory research that are discussed in the next section are also factors to consider when choosing etween in-house and consultant conducted studies.

Proprietary Nature of Delphi in the Corporate Environment

One of the usual descriptors of corporate market research is that the results are considered proprietary. Many studies are conducted to further a competitive advantage. Corporate Delphi research is often conducted in this environment with similar objectives. In these instances the results of the studies are not designed for outside consumption. This creates problems if external expert panelists are used in the study. The usual contract with the panelist is a full or partial payment with a copy of the study results. This usually attracts high-caliber panelists who are interested in adding the study results to their own store of information. The presence of the report in turn results in dissemination of its contents to the panelist's professional colleagues, either by photocopying or by requests to the study director for additional copies. This process of information dissemination through "invisible colleges" usually means that proprietary studies are not too practical with external panelists.

One solution to this situation is to utilize in-house experts. Of course, this is practical only if there are a significant number of internal experts in the subject matter of interest. The penalty of using this approach is the loss of the independent outside viewpoint.

The use of mixed panels of in-house and external experts creates another potential problem. The in-house panelist may have access to confidential corporate market or technological research and use this in making and justifying his projections. The study director may have to edit this data out of the panelist feedback material unless the company is prepared to let the corporate information out to the external panelists. This situation can create some intellectual dissonance for the study director, since the secret data could help resolve specific questions under consideration by the panel. The best solution in this case may be to try in advance to avoid subject matter in the study where confidential company research is underway.

Conclusions

The preceding list of issues that must be considered when conducting corporate Delphi research is not exhaustive. The main purpose of this section was to examine some of the common or most important issues that the business planner must face when deciding whether or not, or how, to use Delphi research. As with many situations, a heavy application of common sense when planning Delphi research will avoid some of the potential problems outlined above.

Future Use of Delphi in Corporations

The near future should see continued rapid expansion of the Delphi technique in business. The methodology appears to be currently reaching the "faddish" stage. Many low-quality studies (which may be mislabeled "Delphi") will be

conducted. This could result in a credibility gap with those trying to use the technique to its best advantage. If this credibility gap does occur, there may be a numerical decline in the number of studies conducted, but a general improvement in the overall quality of corporate Delphi research.

Widespread use of the methodology will result in continued rapid modification of the original RAND design. Mini-Delphi's will be used to develop specific forecasts or evaluate potential policy changes. The latter area will receive further attention with the continued development of interest in technology assessment. The use of on-line Delphi techniques will spread, especially as corporate management information systems and remote access terminals become widespread [38]. The availability of standard packages that permit any researcher with access to an on-line Delphi system to act as a study director will also encourage further use of the technique [39].

Delphi will become popular for certain types of market research studies. This will probably occur more as a result of the promotional activities of market research firms than from the conscious decision of corporate researchers or marketing academics. This opinion is held since there is little overlap between the current professional literature of the marketers and the long-term planners [40], whereas consultants are presently indicating interest in the technique.

In conclusion, Delphi has a healthy future in the corporate environment. This is a future for a whole *family* of Delphi-inspired techniques in a broad range of applications. Use of the term "Delphi" to describe a monolithic technique has rapidly become obsolete in this environment. This expanding family of techniques will be the property of the market researcher, market planner, policy planner, systems researcher, etc., as well as the long-term business planner.

References

1. *Forecast 1968-2000 of Computer Developments and Applications*, Parsons and Williams, Copenhagen, 1968.
2. Sample results are shown in "Computers in the Crystal Ball," *Science*, August 1969, p. 15; see also Joseph Martino, "What Computers May Do Tomorrow," *The Futurist*, Oct. 1969, pp. 134-135.
3. Ole Lachmann, "Personnel Administration in 1980: A Delphi Study," *Long Range Planning*, 5, No. 2 (June 1972) p. 21.
4. *Delphi Panel on the Future of Leisure and Recreation*, SET Inc., Los Angeles, 1972 (multiclient proprietary study).
5. Letter from Ducharme, Déom, and Associés Inc., August 18, 1972.
6. See the Articles of Incorporation; the Institute aim is "to make available without discrimination the results of such research and scientific advances to the public"; also quoted on inside front covers of many IFF reports.
7. Paul Baran and A. J. Lipinski, *The Future of the Telephone Industry*, R-20, Institute for the Future, Menlo Park, Sept. 1971.
8. Paul Baran, *The Future of Newsprint*, R-16, Institute for the Future, Menlo Park, December 1971.
9. J. R. Salancik, Theodore J. Gordon, and Neale Adams, *On the Nature of Economic Losses Arising from Computer-Based Services in the Next Fifteen Years*, R-23, Institute for the Future, Menlo Park, March 1972.

10. Selwyn Enzer, *Some Prospects for Residential Housing by 1985*, R-13, Institute for the Future, Menlo Park, January 1971.
11. Selwyn Enzer, following article (III C 2).
12. Selwyn Enzer, Dennis L. Little, and Frederick D. Luzer, *Some Prospects for Social Change by 1985 and Their Impact on Time/Money Budgets*, R-25, Institute for the Future, Menlo Park, March 1972.
13. Ole Lachmann, "Personnel Administration," p. 23.
14. Marvin A. Jolson and Gerald L. Rossow, "The Delphi Process in Marketing Decision Making," *Journal of Marketing Research* 8, (November 1971). 444.
15. Contemporary Research Center, "Thinking about Tomorrow," *Printout*, April 1972.
16. See, for example: Harper Q. North, *Technological Forecasting in Industry*, NATO Defense Research Group Conference, National *Physical Laboratory*, England, Nov. 1968.
    Harper Q. North and Donald L. Pyke, "'Probes' of the Technological Future," *Harvard Business Review*, May-June 1969.
    Harper Q. North and Donald L. Pyke, "Technological Forecasting to Aid Research and Development Planning," *Research Management*, 12 (1969).
    Harper Q. North and Donald L. Pyke, "Technological Forecasting in Planning for Company Growth," *IEEE Spectrum*, Jan. 1969.
    Donald L. Pyke, "A Practical Approach to Delphi," *Technological Forecasting and Long Range Planning*, American Institute of Chemical Engineers, Nov. 1969.
    Donald L. Pyke, *The Role of Probe in TRW's Planning, Technological Forecasting—An Academic Inquiry*, Graduate School of Business, University of Texas, April 1969.
    onald L. Pyke, "Long Range Technological Forecasting: An Appl. ation of the Delphi Process," Fifth Summer Symposium ofEnergy Economic Division of ASEE, UCLA, June 21, 1968.
17. North and Pyke, "Technological Forecasting in Planning for Company Growth," *IEEE Spectrum*, Jan. 1969, p. 32.
18. See, for example: E. F. Parker, "Some Cexperience with the Application of the Delphi Method," *Chemistry and Industry*, Sept. 1969. p. 1317.
    E. F. Parker, "British Chemical Industry in the 1980's—A Delphi Method profile," *Chemistry and Industry*, Jan. 1970, p. 138.
19. Gordon Wills *et al. Technological Forecasting*, pp. 194-203, 208-9 Penguin Books, London, 1972. Also discussed in Gordon Wills, "The Preparation and Deployment of Technological Forecasts," *Long Range Planning*, Mar. 1970, pp. 51, 53, 54.
20. D. I. Currill, "Technological Forecasting in Six Major U. K. Companies," *Long Range Planning*, March 1972, p. 72.
21. A. D. Bender, A. E. Strack, G. W. Ebright, and G. von Haunalter, "Delphi Study Examines Developments in Medicine," *Futures*, June 1969, p. 289.
22. George Teeling-Smith, "Medicine in the 1990's: Experience with a Delphi Forecast," *Long Range Planning*, June 1971, p. 69.
23. *Communications Needs of the Seventies and Eighties*, Bell System, April 1972 (internal document).
24. Frank J. Doyle and Daniel Z. Goodwill, *An Exploration of the Future in Educational Technology*, Business Planning, Bell Canada, Montreal, January 1971 (proprietary).
    Frank J. Doyle and Daniel Z. Goodwill, *An Exploration of the Future in Medical Technology*, Business Planning, Bell Canada, Montreal, Oct. 1971 (proprietary).
    Daniel Z. Goodwill, *An Exploration of the Future in Business Information Processing Technology*, Business Planning, Bell Canada, Montreal, Oct. 1971 (proprietary).
    Daniel Z. Goodwill, *Perspectives D'Avenir des Techniques de l'Informatique*, Planification du Service Téléphonique, Business Planning, Bell Canada, Montreal, August 1971 (proprietary).
    Michael T. Bedford, *The Future of Communication Services in the Home*, Business Planning, Bell Canada, Montreal, November 1972 (proprietary).
25. Sources: Doyle and Goodwill, *Educational Technology* and *Medical Technology*; Goodwill, *Business Technology*.
26. Doyle and Goodwill, *Educational Technology*, p. 64.
27. *Ibid* pp. 40-41.
28. Doyle and Goodwill, *Medical Technology*, p. 60.
29. Goodwill, *Business Technology*, p. 99.
30. Michael T. Bedford, *The Future of Communication Services in the Home*, Round II Questionnaire, Business Planning, Bell Canada, Montreal, Jan. 1972, pp. 2-3.

31. Bedford, *Communications in the Home*, Nov. 1972.
32. One recent indication of this interest was the formation of the "International Society for Technological Assessment."
33. Bell Canada has recently attempted a modest pilot assessment: Philip Feldman. *A Technology Assessment of Computer Assisted Instruction*, Business Planning, Bell Canada, Montreal, Sept. 1972.
34. Baran, *Future of the Telephone Industry*.
35. Canadian Computer/Communications Task Force, *Branching Out*, Vol. 1, Information Canada, Ottawa, May 1972.
36. The Bell Canada Business Planning approach has been described in J. Martino, "Technological Forecasting is Alive and Well in Industry." *The Futurist*, IV, No. 4 (Aug. 1972), pp. 167-68.
37. An exception to this is the Policy Delphi conducted within a corporation.
38. See Chapter VII. For information on on-line Delphi see: Murray Turoff, "Delphi Conferencing," *Technological Forecasting and Social Change*, 3 (1972) pp. 159-204.
Murray Turoff, "Delphi and Its Potential Impact on Information Systems," *Fall Joint Computer Conference Proceedings 1971*, Vol. 39, Afips Press, 1971.
A. J. Lipinski, H. M. Lipinski, R. N. Randolph, *Computer-Assisted Expert Interrogation—A Report on Current Methods Development*, *Technological Forecasting and Social Change* 5 (1973) pp. 3-18.
39. A. J. Lipinski *et al.* pp. 13ff.
40. One exception is the book by Wills, which takes a marketing approach when reviewing forecasting and Delphi studies.

Acknowledgments

Many of the current and past members of the Bell Canada Business Planning Group have been involved with the Bell Delphi studies outlined in this paper. Mike Bedford, Frank Doyle, and Dan Goodwill spent many months on the research, design, conduct, and management of those studies. Don Atkinson, Ogy Carss, Ken Hoyle, and Sinc Pritchard provided the necessary management support and maintained a belief that these efforts would produce useful results.

I would also like to thank Phil Feldman for his efforts in tracking down many of the references listed. In addition to some of the individuals mentioned above, Tony Ryan and Phil Weintraub provided many useful comments and suggestions on earlier drafts of this article. L.H.D

III. C. 2. Plastics and Competing Materials by 1985: A Delphi Forecasting Study

SELWYN ENZER

The application of Delphi to the identification and assessment of possible developments in plastics and competing materials[1] posed a severe challenge to the technique. Before launching into a discussion of this project it is worth considering the advantages offered by the technique for this application. Since the study was conducted with questionnaires transmitted through the mails, it permitted many widely separated people to participate without the difficulty of having them travel to be co-located at any specific time. It permitted the group to focus on what they regarded as major developments very quickly and discuss only those prospects in detail. Furthermore, because anonymity was employed, each participant was forced to judge the potential of each possibility on the basis of his knowledge and the supporting arguments presented. In other words, the tendency to judge those developments suggested by the most notable panelists were eliminated by virtue of anonymity.

This study was originally scheduled to be completed in three rounds of interrogation. However, as it evolved, only two rounds appeared necessary. This occurred by virtue of the high degree of specialization which appeared in the first-round responses and became even more evident in the second round.

The ability to tailor-make plastics for various applications, enhanced by growth in understanding of organic chemistry, alloying, reinforcing, etc., plus the responsiveness of the material itself, have led many researchers to believe that the types of plastics produced in the future will be determined more by what is desired (and pursued) than by what is possible. Thus in many ways this study was more an investigation of material needs and resource allocations than of technological possibilities.

The study focused upon possible combinations of material property[2] changes that are likely to affect widespread material usage. A prime difficulty encountered in this study arose from discussing yet unknown (and hence unnamed) materials. In general, it is easier to discuss improvements in the properties of steel, aluminum, concrete, boron, niobium, etc., than to discuss the prospects for development of, and properties of, material *X*, *Y*, or *Z*. Yet in many cases

[1]Selwyn Enzer, *Some Developments in Plastics and Competing Materials by 1985*, Report R-17, Institute for the Future (January 1971).

[2]The term "material property" included not only physical properties such as strength, density, toughness, and others, but also processability and cost.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

this study had to do exactly that. As a result, it probably tended to focus more on changes in existing materials than it did on totally new materials.

Since the study focused on material property changes that may be realized in existing materials as well as new materials and their properties, the number of alternatives to be contemplated was vast. To address this challenge a matrix-type categorization of materials and properties was used as the point of departure. For this purpose a breakdown similar to that presented in "The Anatomy of Plastics," *Science and Technology* (F. W. Billmeyer and R. Ford), was used. This matrix of materials and properties was divided into five subcategories:

- Engineering Plastics
- General Purpose and Specialty Plastics
- Glass Fiber Reinforced Plastics
- Foamed Plastics
- Nonplastics

The panel was asked to: (1) review the materials and properties presented, indicating where they thought changes were likely to occur within the next fifteen years which would significantly affect the widespread use of that material; and (2) add and describe the anticipated properties of new materials which they thought were likely to evolve and gain widespread use by 1985. In both of these steps the panel was also asked to describe the new chemical, physical, or other technological developments that they believed would lead to the creation of the new material.

These inputs from the first Delphi round were used to prepare a three-part questionnaire for the final round of interrogation. These parts were: (1) a summary of the assessments of anticipated changes in existing material properties, indicating those selected for more detailed investigation; (2) a listing of both plastic and nonplastic materials with the nature of the anticipated major changes described (those respondents who had anticipated these changes were asked to estimate the new material properties they expected would exist by 1985 and to estimate the 1985 annual consumption by application); and (3) a list of new materials anticipated by 1985 and a description of their properties (those respondents who had anticipated these items were asked to estimate the properties and consumption patterns they expected for these by 1985). All of these parts were open-ended in that any of the respondents could still add additional items or comment on any item.

Anticipated Changes in Properties of Existing Materials

The Delphi panel was presented with descriptions of the major uses, properties, and proprietary qualities of 37 plastics and 16 nonplastics all currently in widespread use. These 37 plastics are presented in Table 1. As indicated earlier, they were asked to: (1) identify likely changes in the properties of these

materials which would significantly affect their widespread use by 1985; and (2) identify new materials (in each of the categories shown) which are likely to be developed and would be in widespread use by 1985.

Table 1

Existing Plastics

Engineering Plastics:	Glass Fiber Reinforced Plastics:
ABS	ABS
Acetal	Epoxy
Fluorocarbons	Nylon
Nylon	Polyester
Phenoxy	Phenolics
Polycarbonate	Polycarbonate
Polyimide	Polystyrene
High Density Polyethylene	Polypropylene
Polypropylene	San
Polysulfone	Polyethylene
Urethane	
Poly (Phenylene Oxide)	
General Purpose & Specialty Plastics:	**Foamed Plastics:**
Acrylics	Polyethylene
Cellulosics	Polystyrene
Cast Epoxy	Polyurethane (low density)
Ionomer	PVC
Melamines & Ureas	Polyurethane (high density)
Phenolics	
Low Density Polyethylene	
Polystyrene	
Vinyls (PVC)	
San	

The format used for this portion of the assessment is shown in Fig. 1. This figure is divided into four columns. Column 1 lists the material and its typical uses. Column 2 describes the properties of that material which are the key to its current widespread use. Column 3 is divided into subcolumns which contain specific material properties and the current performance ratings of each

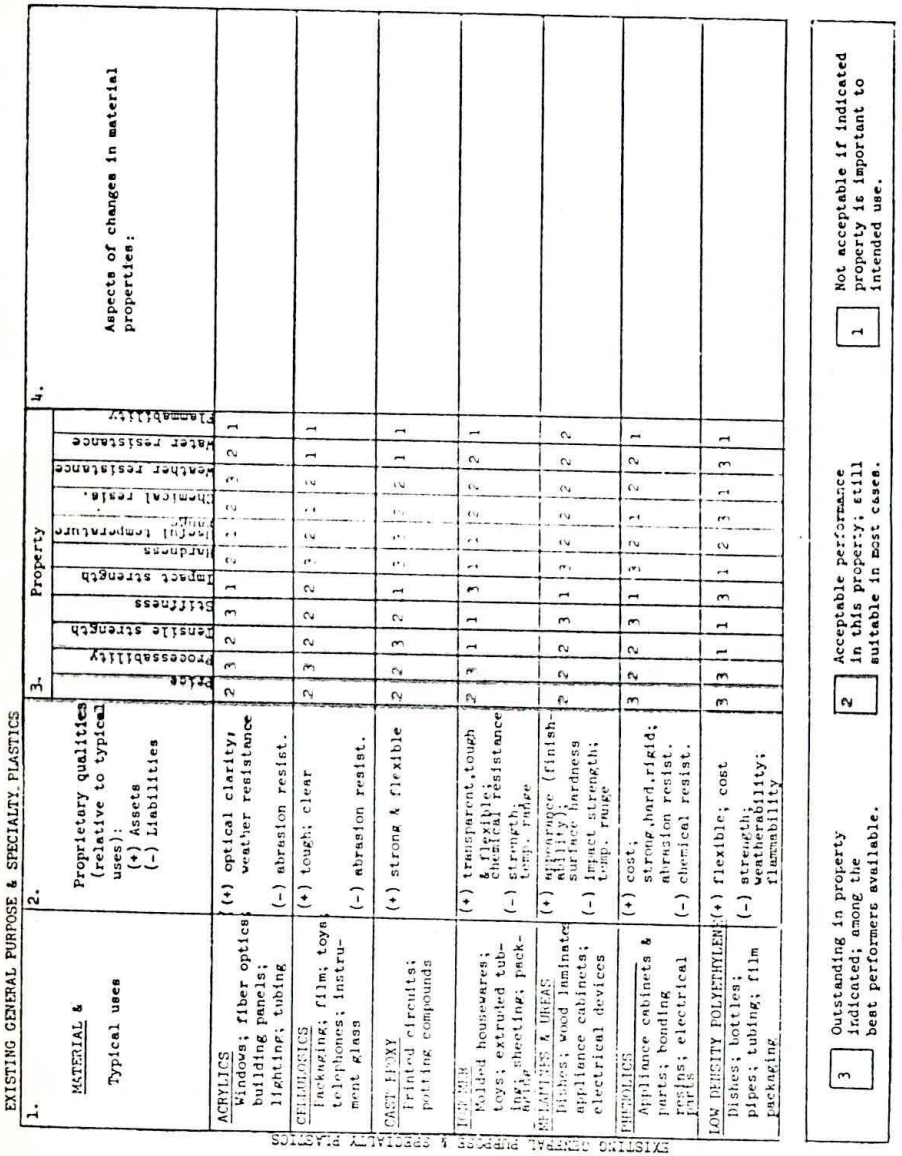


Fig. 1. Typical questionnaire for eliciting changes in existing plastics.

material relative to the others in that category. This rating is indicated with a "1," "2," or "3" in accordance with the code noted at the bottom of the figure.

The results of this assessment were presented to the panel in the format presented in Fig. 2. Shown in the subcolumns of Column 3 are the changes anticipated by the panel. These changes are noted, using the code presented in the upper-right-hand corner of the figure. Column 4 contains the panel's comments. These comments and all other changes suggested by the panel are in italics.

Items noted as being "included in Package No. 2" were reassessed by the panel in the second round in greater detail. The comments received from the panel regarding these materials were presented in greater detail in a subsequent section of the questionnaire.

In the course of this assessment several variations in the original list of existing plastics were made by the panel. As a result, the existing materials that were investigated further in round two are presented in Table 2.

The questionnaire format used for this interrogation and typical results are presented in Fig. 3. As before, the information presented to the panel is in roman type, the additions are in italics.

Table 2

Existing Plastics for More Detailed Considerations

Engineering Plastics:	Glass Fiber Reinforced Plastics:
ABS	ABS
PVF	Epoxy
Nylon	Nylon
Polyimides	Polyester (Molding Compounds)
High Density Polyethylene	
Polypropylene	
Polysulfone	
General Purpose & Specialty Plastics:	**Foam Plastics—Rigid:**
Acrylics	Polystyrene Foams
Epoxy	Low Density Polyurethane Foam
Ionomer	Variable (High Overall) Density Integral Skin Urethane Foam
Phenolics	**Foam Plastics—Flexible:**
	PVC Foam
	Variable Density, Integral Skin Urethane Foam

EXISTING GENERAL PURPOSE & SPECIALTY PLASTICS

1. MATERIAL & Typical uses	2. Proprietary qualities (relative to typical uses): (A) Assets (L) Liabilities	3. Property: Price	Processability	Tensile strength	Stiffness	Impact strength	Hardness	Useful temperature range	Chemical resis.	Weather resistance	Water resistance	Flamability	4. Comments
ACRYLICS Windows; fiber optics; building panels; lighting; tubing	(A) optical clarity; weather resistance (L) abrasion resist.	2 +	3	2	3	1 + +	2 +	1 + +	2	3	2	1 + +	(Included in package No. 2).
CELLULOSICS Packaging; film; toys; telephones; instrument glass	(A) tough; clear (L) abrasion resist.	2	3	2	2	2	3	2 (-)	1	2	1	1 +	●Not competitive with low cost plastics, e.g., vinyls.
CAST EPOXY Printed circuits; potting compounds	(A) strong & flexible	2 + +	2	3	2	1 + +	3	3	2 +	2	1 +	1 + +	(Included in package No. 2).
IONOMER Molded housewares; toys; extruded tubing; sheeting; packaging	(A) transparent, tough & flexible; chemical resistance (L) strength; temp. range	2 + +	3	1 +	1 +	3	1 + +	1 + +	2	2	2	1 + +	(Included in package No. 2).
MELAMINES & UREAS Dishes; wood laminates; appliance cabinets; electrical devices	(A) appearance (finishability); surface hardness (L) impact strength; temp. range	2	2	2	3	1 + +	3	2	2	2	2	2	
PHENOLICS Appliance cabinets & parts; bonding resins; electrical parts	(A) cost; strong, hard, rigid; abrasion resist. (L) chemical resist.	3	2 + +	2 +	3	1 + +	3	2 +	1 + + +	2	2	1 + + +	(Included in package No. 2).
LOW DENSITY POLYETHYLENE Dishes; bottles; pipes; tubing; film packaging	(A) flexible; cost (L) strength; weatherability; flammability	3	3	1 +	1	3	1	2	3	1 + +	3	1 + + +	●Compounding can improve weather resistance.

CODE

Property changes anticipated by the panel, likely to affect the widespread use by 1985, are noted in *italics* as follows:

Direction of expected change	% of the panel	40 or more	20 to 40	less than 20	None
	Improvement	+ +	+ +	+	
	Degradation	- -	- -	-	

Degradations encircled indicate panel disagreements with the original rating not a forecast.

3 Outstanding in property indicated; among the best performers available.

2 Acceptable performance in this property; still suitable in most cases.

1 Not acceptable if indicated property is important to intended use.

Fig. 2. Typical feedback of results of initial estimates of changes in existing materials.

ENGINEERING PLASTICS

MATERIALS FOR WHICH THE PANEL ANTICIPATES A SIGNIFICANT INCREASE IN USE BY 1985

KEY PROPERTIES IN WHICH IMPROVEMENTS ARE CONSIDERED LIKELY BY THE PANEL	CURRENT PROPERTY VALUE	ESTIMATE PROPERTY VALUE LIKELY BY 1985
1. ABS		
Price, ($/lb.)	.28-.44	*.20-.30*
Maximum Service Temperature *(D 648) °F	180-245	*250*
Flammability, in./min.*(D 265)	1.0-2.0	*0.65-2.0*
Impact Strength Notched Izod, ft. lb./in. *(D 256) @ 73°F	2.0-10.0	*2.5-11.*
@ 40°F	.8-3.5	*.9- 5.0*
Deflection temperature, °F *(D 648) @ 264 psi	214-244	*214-280*
@ 66 psi	215-250	*215-280*

*ASTM Test Method

COMMENTS BY THE PANEL

- Reduction in depolymerization will improve flammability.
- Alloying and blending will improve temperature range and flammability
- Improved shaping of disperse particles as well as processing techniques will orient particles.
- Reduction in compound prices; scale and competition lowering prices.
- **Availability of composite forms as sheet and ability to fabricate in inexpensive equipment will make this material competitive with glass-polyester.**
- **Platability makes this material increasingly attractive for automobile parts, this will be especially important if low temperature strength and crack resistance can be improved.**
- *Properties will be very dependent upon filler.*
- *ABS will include chemical and cross-linked materials.*

	AGREE	NO CONSENSUS	DISAGREE
		x	
	x		
	x		
	x		
	x		
	x		

APPLICATIONS

MARKET BREAKDOWN (MAJOR USES)	VOLUME-MILLION LB/YR CURRENT	VOLUME-MILLION LB/YR ESTIMATE 1985
TOTAL VOLUME	508	*1500*
Automotive	80	*250*
Major Appliances	60	*150*
Pipe & Fittings	60	*200*
Bus. Machines, Phones	40	*90*
Recreational Vehicles	45	*150*
Luggage	27	*65*
Other:	196	*350*
Interior panels & sheeting		

Fig. 3. Typical results of final assessment of important changes in existing plastics.

The comments received from the panel are presented in Column 2 of this figure. Because these generally referred to the reasons why the material property changes were anticipated, the panel was asked to indicate whether or not they agreed or disagreed with each statement. The results of this assessment are also shown in Column 2. Those items presented in italics in this column were added in round two and hence were not assessed by the entire panel.

Column 3 presents the current major markets and their annual volume usage. Shown in italics are new markets suggested by the panel and their estimated 1985 usage.

In that portion of the investigation concerned with nonplastics, many new material developments were suggested, but only a few of these were regarded as threats to the growth of plastics. This can be seen in the following general comments received from the panel.

- The main competition between plastics and aluminum will occur in the construction field, particularly in residential housing and light industrial buildings. New developments in aluminum will hurt plastics in the applications which are primarily structural. On balance, however, these developments will affect the use of other metals more than plastics.
- In general, plastics will continue to replace iron and steel in some applications. This will be significant to the plastics industry; however, it will be a relatively small change to the steel industry. Any development which brings steel closer to "one-step" finishing, with improved environmental resistance, will be important in this regard, since it will blunt some of the basic advantages that plastics have over steel, allowing the use of "conventional" technology and existing capital equipment. Such developments will bring steel and plastics closer to a straight-cost competition. However, these developments must be realized before potential markets have switched from steel to plastics to maintain the continuity of technology and equipment.
- Developments in concrete appear more likely to enhance the demand for plastics than to replace or be replaced by them. Developments in wood and plywood are more likely to be in combination with plastics and hence are apt to increase the demand for such materials. However, unlike the concretes, wood will increasingly be replaced by plastics, particularly in furniture and siding.

Other Materials Suggested by the Panel as Likely to Become Important by 1985

In addition to the changes suggested in the existing materials, other materials (some already in existence) were suggested by the panel as prospects for widespread use by 1985. These are presented in Table 3.

These materials were submitted for consideration by the entire panel in the final round of reestimation. Format and typical results are presented in Fig. 4.

Table 3

Other Materials Suggested by the Panel as Likely to Become Important by 1985

Engineering Plastics:	Other Fiber Reinforcements and Reinforced Plastics:
Polybutadiene (High 1, 2 Content)	Boron Fibers
Polyethyleneterephthalate	Graphite Fibers
Polyphenylene Oxide Derivatives	Fiber Strengthened Oxides
New Thermoplastic	Aluminum Oxide Fiber & Whisker Composites
New Tougher Plastics	Boron/Epoxy
	Boron/Polyimide
	Graphite/Epoxy
	Graphite/Polyimide
General Purpose and Specialty Plastics:	**Foamed Plastics:**
PVC—Polypropylene Copolymers	Phenolic Foams
Ethylene—Polar Copolymer	Vinyl Foam
Acrylic—PVC	Polyolefins (Ethylene, Propylene, etc.)
New Polyolefins	Isocyanurate—Urethane
New Thermosetting Resins	Silicone Foams
Completely Nonburning Organic	Special Hi-Temp Foams
Semiorganic & Inorganic	Structural Foams
	Foamed Thermoplastics
	Injection Molded Urethane Foams
Glass Fiber Reinforced Plastics:	**Miscellaneous:**
PVC	Silicate Glasses & Polymers
Polyimides (or Amidemides)	Titanium Alloys
Polysulfones	Cermets
Polyurethane	
Vinyl Ether	
Thermoplastic Polyester	
Thermoplastic Sheet	
New Thermoplastic Resin	

As seen, this is similar to the format presented earlier. One notable difference is in Columns 5 and 6, which contain estiamtes of the likelihood of these materials being in widespread use by 1985 and the annual production estimated by that time. These estimates should be treated with even greater care than those presented earlier, since they often represent the comments of as few as two or three respondents who were familiar with the development.

There were two principal elements to handling the results of round one. First, a determination was made, for each graph, as to the location of upper and lower limits of the extensions which would include 50 percent of the responses to that graph.

The second element involved the gathering and synthesizing of comments presented under Questions 2 and 3 of the graph sheets. The comments of all the respondents were collected, and each comment was then studied and determined to be either a forecasting assumption, an economic and international consideration, a key development, or a comment to be associated with that particular graph page. The comments were grouped accordingly, and the final product was retyped for inclusion on the second round. It was quite apparent that in many instances one respondent's assumption was another's uncertainty. The frequency with which a topic was brought up influenced the judgment on its choice as a key development for round two.

The process of collecting and editing the large number of comments obtained on the first round represented the largest single task in the exercise, in terms of both clerical time and professional judgment. Assumptions from all respondents were initially xeroxed, cut out, and taped on large sheets. Each sheet represented different topics or curves. On these large sheets duplications were crossed out and editing of assumptions to produce shorter wordings took place. This conglomeration was then retyped once and put through a final polishing, editing, and reordering before the final typing for the second-round questionnaire. The process of putting each set of assumptions through a two-stage editing process allowed each professional to check the other's work. It is noticeable to a certain extent that the availability of the xerox machine is a key feature in making large-scale Delphis possible via paper-and-pencil approaches. This is particularly true where one is handling a large volume of textual comments on the part of the respondents.

Section III in round one, "Suggested Additional Variables and Key Developments," produced few responses from the participants. Some of the responses in this section became key developments for Section III of round two, some became assumptions for round two, and the remainder were dropped from the exercise. Three additional curves suggested by the respondents were prepared for inclusion in round two.

The flow chart appended to Section I in the first round also received scant response (i.e., about ten respondents). The responses that were supplied were averaged for each entry in the flow chart and standard deviations were provided through the use of a simple computer program.

Round Two

Design. The questionnaire sent to the respondents in round two was patterned after the results described in the handling of round one.

The new Section I, "Steel," contained thirty-six "Forecasting Assumptions,"

thirty-five "Economic and International Considerations," and all the graphs contained in round one (Section I) with their associated reasons. The respondent was asked to associate a validity score with each forecasting assumption and economic and international consideration presented. The scores were based on the validity codes which are shown in Fig. 2.

For all graphs the original trend line was presented and the 50 percent confidence limits were indicated. The respondent was asked to reestimate his previous extension, after viewing the 50 percent limits, and to identify his estimate as reliable, as good as anyone's, or risky. Each graph also contained associated reasons given by the respondents for increasing or decreasing the graph extension. In Section I, a total of 116 reasons associated with the graphs were presented for evaluation. The respondent was asked to assign a rank of 1 to 6 to each reason given according to the validity scale. He was also invited to show additional reasons if he wished. A sample of a typical round-two question is presented in Fig. 3. This is exactly the same basic form as used in round one.

At the end of Section I, the flow charts were again presented. Two charts, for 1969 and 1980, showed the means and standard deviations for each chart entry and also showed some additional boxes and paths not included in round one. These modifications were suggested by the respondents and incorporated by the senior professional. The respondent was asked to circle the estimate presented if he agreed with it, or to cross it out and provide a new figure on the blank charts provided if he disagreed. The absence of either action was considered to be a No Judgment vote. It was a surprise to the designers that almost all the respondents to the flow chart chose to modify it, since this was not an action suggested in the instructions.

Section II, "Alloys," was similar to Section I in design. The category of forecasting assumptions included sixty-nine assumptions about individual alloys studied, as well as seventeen general assumptions. The respondents were again asked to assign each statement a validity score of from 1 to 6. There was no category of Economic and International Considerations, but all the graphs from Section II of the first round were included with the 50 percent confidence limits and 128 associated forecasting reasons. The respondents were asked to provide the same information requested in Section I.

Section II was entitled "Key Developments and Added Curves." Under Key Developments, thirty-six items were presented for scoring by the respondents. For each item, the respondent was asked to evaluate the likelihood of occurrence by 1975 on a scale of 1 to 6 and to indicate, on a scale of 1 to 4, the impact on the steel industry if the development were to occur. Figure 1 provides a sample of the form utilized for key developments.

In addition to scoring the developments, the respondents were asked to describe the nature of impacts they had characterized as strong or moderate.

Three new curves, supplied by respondents to round one, were also included in this section. The respondents were asked to handle these new curves in the same way they treated the original curves in round one (Sections I and II).

Numeric Scale	
	CERTAIN (Average of 1 to 1.5)
1	• Low risk of being wrong. • Decision based upon this will not be wrong because of this "fact." • Most inferences drawn from this will be true.
	RELIABLE (Average of 1.6 to 2.5)
2	• Some risk of being wrong. • Willingness to make a decision based upon this. • Assuming this to be true but recognizing some chance of error. • Some incorrect inferences can be drawn.
	NOT DETERMINABLE (at this time) (Average of 2.6 to 3.5)
3	• The information or knowledge to evaluate the validity of this assertion is not available to *anyone* — expert or decision-maker.
	RISKY (Average of 3.6 to 4.5)
4	• Substantial risk of being wrong. • Not willing to make a decision based upon this alone. • Many incorrect inferences can be drawn. • The converse, if it exists, is possibly RELIABLE.
	UNRELIABLE (Average of 4.6 to 5)
5	• Great risk of being wrong. • Worthless as a decision basis. • The converse, if it exists, is possibly CERTAIN.
	NOT PERTINENT (Used to eliminate some assumptions from exercise)
6	• Even if the assertion is CERTAIN or UNRELIABLE it has no significance for the basic issue. • It cannot affect the variable under question an observable amount.
	NO JUDGMENT
blank	• No knowledge to judge this item, but the appropriate individual (expert, decisionmaker) should be able to provide an evaluation I would respect.

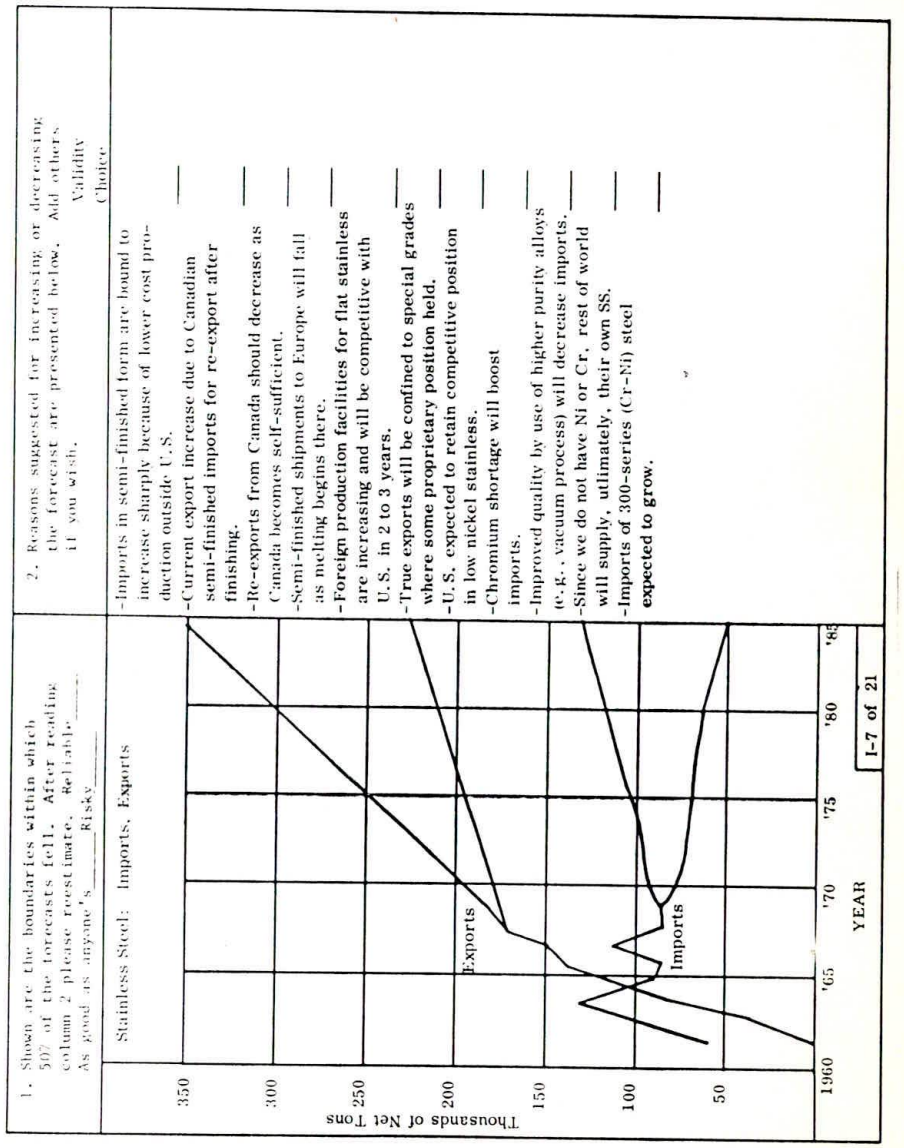


Fig. 3. Form for Trend Extrapolations.

Handling the Results. Round two was sent to fifty-two potential respondents; thirty-four replies were received. Several respondents, representing the polymer industry, were added during this round. They had not been represented in round one and were introduced for the purpose of addressing specific issues on the substitution of plastics which had been generated in round one.

The results of round two required three separate types of handling: (1) new 50 percent confidence limits were supplied for the graphs; (2) verbal comments associated with the assumptions, the graphs, and the Key Development section were collected and considered; and (3) the numerical results, i.e., validity choices, Key Development scores, and flow chart inputs, were collected and tabulated.

There were several steps necessary in handling the large amount of data that was generated by the results of round two. It was first determined that several statistical calculations on the data would be desirable—specifically, the mean and standard deviation of the validity choices for each statement, a distribution showing the percentage of responses falling under each of the scores from 1 to 6 for each of the statements, and a matrix comparing the distribution, by numbers, of the different occupation categories represented by the respondents with the range of scores from 1 to 6. These occupation categories included: primary steel producers, research institutions, steel producers, ferroalloy producers, research institutions, government, and the universities. A computer program was written to carry out these computations. This breakdown allowed us to observe if there were any differences in judgment which may have reflected differences in affiliation of the respondents.

Examples of the statistical presentations are shown in Fig. 4.

The flow chart included in Section I received very little additional information in round two. The scant information received was averaged into previous information on the flow chart and presented in a summary for round three. Due to the differences of opinion among the respondents on how actually to model the flow of steel-making materials, it was felt by the designers that this one question could have constituted the total Delphi exercise among a select smaller group of respondents.

Round Three

Design. Round three again consisted of three major sections: Section I: A Summary of Round Two, Section I, Steel; Section II: A Summary of Round Two, Section II, Alloys; and Section III: Key Developments and Added Curves.

Sections I and II were provided for summary purposes and required no further input by the respondents. For each statement, the mean and standard deviation as calculated from round two results were shown. Several new

Question Number	Mean	Standard Deviation
1	3.4	1.0
2	2.6	0.9
3	3.0	1.2

Question Number	Validity Choice						Number Responding
	1	2	3	4	5	6	
1	.25		.25	.25		.25	25
2		.33	.33		.33		16
3	.20	.20		.20	.20	.20	19

Question 1

Occupation Category	Validity Choice					
	1	2	3	4	5	6
1 Steel	3	2	1	2	0	1
2 Ferroalloy	1	4	0	0	0	0
3 R&D	3	1	0	1	0	0
4 Government	0	0	1	3	2	0

Fig. 4. Examples of statistical presentations.

statements were added in Sections I and II; a few new developments were added to Section III to be assigned a validity score by the respondents. The new 50 percent confidence limits, taken from the results of round two, were also presented in a final form for each graph.

Section III, the only section to be returned by the respondents, contained all the new assumptions and all previously evaluated assumptions which exhibited a large standard deviation (i.e., disagreement). It also called for a reevaluation of all key developments.

The first portion of Section III indicated the percentage distribution of the scores from round two on the likelihood and impact of potential key developments. The estimated average score for each development was indicated and a summary of all verbal comments associated with each development was given. The respondent was asked again to give his preference on the likelihood and impact of each potential development.

The second portion of Section III presented the three curves shown in this section of round two and included a number of reasons given by the respondents of round two for their curve extensions. The respondent was asked to reestimate the curves after reading the associated past reasons and to rate the reliability of his estimation. He was also asked to vote on the reasons given for each curve using the validity scale from 1 to 6 described earlier.

The third portion of Section III contained all assumptions from Sections I and II which exhibited a considerable degree of disagreement. This category generally, although not exclusively, included assumptions with a standard deviation of 1.3 or greater. The respondent was asked to reevaluate his previous validity choice and submit a new score. Several new assumptions were also added and the respondent was requested to provide a validity choice for these new assumptions.

In the final portion of Section III a new chart was introduced showing percentage breakdowns of inputs and outputs for three major steel processes. The figures were supplied for 1969 and a blank sheet was provided for 1980. The respondent was asked to fill in the sheet for 1980 and to change any 1969 figures with which he disagreed. A space was provided for an explanation of any disagreements with 1969 figures. The results of this chart were to be considered a summary response, as the monitors did not plan to feed back the responses for changes.

The monitor also surveyed briefly the attitudes of the respondents toward the Delphi approach by asking the rspondents a number of questions, e.g., was the time spent in participating in Delphi well used; what organizations should sponsor an exercise of this type on a regular basis, etc.

Handling the Results. Round three was sent to thirty-eight respondents on December 10, 1970. Thirty-three respondents actually replied to Round 3.

A computer program provided the means, standard deviations, percentage distributions, and industry category matrices for all key developments, assumptions to be reevaluated, and new assumptions. The percentage distributions were then examined by the senior professional. If 20 percent or more of the vote fell into the "not pertinent" category (a validity score of 6), the items were dropped from the exercise. Eight items were dropped for this reason. The remaining items were regrouped so that every assumption was associated with a curve. The assumptions and reasons for each curve were then reordered according to their mean validity scores.

The final report was then prepared for the National Materials Advisory Board of the National Academies of Science and Engineering.

Comparison of Delphi and Panel Studies

A separate panel appointed by the National Materials Advisory Board approached the same problem considered by the Delphi exercise in a more

"conventional" vein. The conventional study was carried out simultaneously with the Delphi, but the results were not compared until both exercises had been completed.

In the panel approach, individual members of the Panel on Ferroalloys reviewed portions of the problem with which they were most familiar. The Panel Report, NMAB-276, "Trends in the Use of Ferroalloys by the Steel Industry of the United States" consists of chapters, each of which is logical, comprehensive, and definitive with respect to its topic. While appropriate caveats exist, its forecasts are precise. The recommendations and conclusions therein represent the unanimous agreement of the panel and no areas of disagreement are spelled out. The result is typical of a competent panel (or committee) activity. Based upon the expertise of the carefully selected participants, the report is a reliable and comprehensive account of known information and of projections based on this information and on current research and development. In contrast, this Delphi was designed to complement the panel report. The planned approach was to provide an opportunity to indicate uncertainties or disagreements about the subject and to evaluate quantitatively the degree of uncertainty which exists within a large group of experts. The Delphi product attempts to present an awareness of the areas which are subject to differences of view and to highlight the topics which appear to concern the respondent group. The Delphi provides a group evaluation of every statement advanced by the respondents who, presumably, express their beliefs. Although the results of this exercise include a number of statements which were rated uncertain, risky, or unreliable by the whole group, this variation does not imply that one dissenter from the group will be incorrect in retrospect. The group view has a higher probability of being correct than the view of any one individual. However, in the past, developments that significantly affected industries were often unforeseen by most of the involved experts. Therefore, the reader is cautioned not to extrapolate blindly from the group judgments exhibited in a Delphi to assumed facts.

In this case, the Delphi exercise is a literal exploration of the minds of experts in the steel and ferroalloy industries regarding their views on individual items. This exploration allowed a broader coverage of the subject area than was possible in the panel report.The presentation of the Delphi results allows the reader to compare easily his judgments with those of the group. No attempt is made to arrive at conclusions or recommendations, or to present a definitive view as was done in the panel activity.

Where the panel and Delphi activities overlap, there is considerable agreement in their forecasts. Figure 5 compares the consumption in steel of a number of alloys, as predicted by the panel and the Delphi, and some of the qualitative features of the two methodologies.

Other comparisons could be made between information presented in the panel report and the Delphi predictions. For example, NMAB-276 projected that carbon steel shipments would increase by 22 percent in the next ten years;

Quantitative Comparison

	Predicted Ferroalloy Consumption in Steels and Superalloys in 1980 (In Short Tons of Container Element)	
	Panel Report NMAB-276	Delphi
Chromium	319,260	250,000–303,000
Cobalt	2,732	3,000–4,000
Columbium	1,977	1,300–1,850
Manganese	1,011,235	1,100,000–1,250,000
Molybdenum	21,540	17,400–21,000
Nickel	124,200	90,000–115,000
Tungsten	1,850	1,550–2,600
Vanadium	5,796	5,000–6,200

Qualitative Comparison

Type of Activity	Delphi	Committee
Form of Information	Specific Comments	General Discussion
Weighting of Information Provided	Most Rated on Reliability Scale	None
Disagreement among Committee Members or Respondents	Indicated in Reliability Score	Eliminated
Presentation of Back Ground Information	Only as Randomly Generated by Respondents	Thorough and Systematic
Recommendations	Not Specifically Stated	Consensus Recommendations Indicated
Range of Information Provided	Broad, Reflecting Wide Interest of Respondents	Limited to Specific Committee Subject Area

Fig. 5. Comparison of the panel and Delphi approaches.

the increase projected on the Delphi graph for the next decade was 22 to 26 percent over the current figure. Also, the panel report stated that High Strength Low Alloy Steel (HSLA) is the fastest growing segment of the steel industry; the Delphi results were that HSLA is one of the two fastest growing segments of the industry.

As was mentioned earlier, the panel report did not indicate the areas of disagreement. In the Delphi category, "not determinable" with respect to validity reflected either the inability of the entire group to determine the validity of an assumption or the averaging of opposing judgments on the validity of a given assumption. Of the 135 statements that fell into this classification, seventy-three reflected an actual disagreement among the respondents by exhibiting a high standard deviation.

The following assumptions exemplify those that fell into the "not determinable" category of the Delphi as a result of disagreement:

- Cobalt-iron base tool steels will be marketed.
- Continuing nickel shortages will establish permanent substitution.
- New techniques will allow significantly greater flexibility of substitution among alloying elements based upon price changes.
- Critical shortages of nickel will reoccur.
- No important new use for cobalt.
- Present and projected investment in ocean studies is too large to exclude development of economical offshore mining except in short term (i.e., next five years).
- Alloy steels will increase in nickel content.
- Tungsten content of carbides will decrease because of cost.
- Full alloy shipments will parallel automotive production.
- Low-cost method of preparing high-purity iron powder will be developed by 1975.
- More and cheaper scrap will result from urban waste recycling.
- Shortages of natural gas will be a primary limiting factor in the expansion or modernization of the steel industry.
- Electron-beam refining will grow significantly.

An additional fifty-seven assumptions were rejected by the Delphi respondents as either risky or unreliable. One must reflect that while each of the 667 assumptions were suggested by at least one expert in the Delphi respondent group, approximately two hundred of these, or 30 percent, were considered less than reliable by the group as a whole. Furthermore, there is often considerable value to decisionmakers in observing the nature of rejected assumptions. For this reason the final Delphi report listed all the evaluated assumptions pertaining to any curve in the order of decreasing validity so one may observe the complete span of the topics covered. Probably the most significant difference between the Delphi and committee approaches is the itemization of what the Delphi group could not agree on or what they rejected. Usually the psychologi-

cal process in a committee of experts tends to eliminate these categories of information from the final report.

In summary, the two reports did not always cover the same subject matter. However, when they did touch upon the same subject matter, the results were generally compatible.

Advice to Future Delphi Monitors and Designers

The monitor's experience with the Steel and Ferroalloy Delphi gives rise to a number of observations and advice to those planning to monitor Delphis in the future.

(1) When presenting statements for a vote, or synthesizing the respondents' suggestions, be alert for ambivalent wording. Two separate statements may appear as one, leading to confusion as to what should be voted upon. Vague wording or easily misinterpreted wording may also lead to confusion.

(2) When editing respondents' comments for clarity, try to preserve the intent of the originator. When editing from round to round, avoid changing a statement so that it has one meaning in round one and another in round two.

(3) Lay out the expected processing of the data throughout all the rounds of the Delphi before you finalize the design. You may, by circumstance, be forced later to modify the procedure, but the process of planning ahead will usually turn up any gross problems in your initial questionnaire design and its impact on following rounds.

(4) Design the handling of your data so that each response can be processed (or punched for processing) as it comes in. Thus you will not have a frantic rush to analyze all the responses at once when the last tardy return comes in.

(5) Keep track of how different subgroups in your respondent group vote on specific items. This can be very useful in analyzing the results and will occasionally produce situations where you wish to let the respondent group know that polarizations or differences based upon background exist.

(6) If you are covering a number of fields of expertise, make sure that each field is adequately represented in your group.

(7) It should be mandatory that at least two professionals work on monitoring any one Delphi exercise, particularly when the abstracting of comments is a notable portion of the exercise. With two individuals one can always review what the other has done.

(8) Pretest your questionnaire on any willing guinea pigs you can find outside your respondent or monitor group. If you have a sponsor, it is useful to go over the design of each round with some of his people before finalizing it.

IV. Evaluation

IV. A. Introduction

HAROLD A. LINSTONE and MURRAY TUROFF

Skeptics from the allegedly "hard" sciences have at times considered Delphi an unscientific method of inquiry. Of course, the same attitude is often encountered in the use of subjective probability (even in the face of considerable mathematical theory developed to support the concept). The basic reason in each case is the subjective, intuitive nature of the input.

Yet Delphi is by no means unordered and unsystematic. Even in the Gordon-Helmer landmark Rand study of 1964, an analysis of certain aspects of the process itself was included.[1] The authors observed two trends: (1) For most event statements the final-round interquartile range is smaller than the initial-round range. In other words, convergence of responses is more common than divergence over a number of rounds. (2) Uncertainty increases as the median forecast date of the event moves further into the future. Near-term forecasts have a smaller interquartile range than distant forecasts.

It was also observed in all early forecasting Delphis that a point of diminishing returns is reached after a few rounds. Most commonly, three rounds proved sufficient to attain stability in the responses; further rounds tended to show very little change and excessive repetition was unacceptable to participants. (Obviously this tendency should not unduly constrain the design of Policy Delphis or computerized conferencing which have objectives other than forecasting.)

We shall briefly review here some of the systematic evaluations made in recent years.

Dispersion as a Function of Remoteness of Estimate

Martino has analyzed over forty published and unpublished Delphi forecasts.[2] For every event the panel's median forecast dates (measured from the year of the exercise) and the dispersion were determined. A regression analysis was performed and the statistical significance presented in terms of the probability that the regression coefficient would be smaller than the value actually obtained if there were no trend in the data.

The results are quite clear-cut. The remoteness of the forecast date and the

[1] T. J. Gordon and O. Helmer, "Report on a Long Range Forecasting Study," Rand Paper P-2982, Santa Monica, California, Rand Corporation, September 1964.

[2] J. P. Martino, "The Precision of Delphi Estimates," *Technological Forecasting* 1, No. 3 (1970), pp. 293-99.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

degree of dispersion are definitely related. The regression coefficient is in nearly all cases highly significant for a single panel addressing a related set of events. However, there is no consistent relation among different panels or within a panel when addressing unrelated events.

Martino also finds that the dispersion is not sensitive to the procedure used: in cases where only a single best estimate year is requested the result is similar to that where 10 percent, 50 percent, and 90 percent likelihood dates are stipulated.[3]

Distribution of Responses

Dalkey has analyzed the first-round responses by panels asked to respond to almanac-type questions, i.e., those with known numerical answers.[4] All responses to each question are standardized by subtracting the mean value and dividing by the standard deviation for that question. The resulting distribution of "standardized deviates" shows an excellent fit to a lognormal distribution. Martino has applied the same techniques to the TRW Probe II Delphi using 10 percent, 50 percent, and 90 percent likelihood dates for 1500 events.[5] Again there is a very good fit to a lognormal distribution.

Optimism-Pessimism Consistency

Another interesting analysis on the TRW Probe II data was undertaken by Martino to ascertain whether a panelist tends to have a consistently optimistic or pessimistic bias to his responses.[6] With each respondent providing 10 percent, 50 percent, and 90 percent likelihood dates, three standardized deviates can be computed for each individual and a given event. Taking the means over all events of the standardized deviates for a given individual and likelihood, we find an interesting pattern. Most panelists are consistently optimistic or pessimistic with respect to the three likelihoods, i.e., there are relatively few cases where, say, the 10 percent likelihood is optimistic while the 50 percent and 90 percent likelihoods are pessimistic. Considering the totality of events the individual panelist tends to be biased optimistically or pessimistically with moderate consistency. However, the amount of the bias is not very great; an optimistic panelist is pessimistic in some of his responses and vice versa. In other words, each participant exhibits a standard deviation which is comparable to, or greater than, his mean.

[3]In the first case the interquartile range of best estimates was used, in the second case the 10 percent to 90 percent span was taken.

[4]N. C. Dalkey, "An Experimental Study of Group Opinion," Rand RM-5888-PR, Rand Corporation, Santa Monica, California, March 1969.

[5]J. P. Martino, "The Lognormality of Delphi Estimates," *Technological Forecasting* 1, No. 4 (1970), pp. 355-58.

[6]J. P. Martino, "The Optimism/Pessimism Consistency of Delphi Panelists," *Technological Forecasting and Social Change* 2, No. 2 (1970), pp. 221-24.

Accuracy of Forecasts

An apparent indicator of the value of Delphi as a forecasting tool is its accuracy. Since the method was widely publicized only ten years ago, it is difficult to have sufficient hindsight perspective to evaluate its success by this measure. In any event caution is in order. The most accurate forecast is not necessarily the most useful one. Forecasts are at times most effective if they are self-fulfilling or self-defeating. The Forrester-Meadows World Dynamics model has been sponsored by the Club of Rome in the hope that it will act as an early warning system and prove to be a poor forecast. Delphi may be viewed similarly in terms of effectiveness.

We should also observe that long-range forecasts tend to be pessimistic and short-range forecasts optimistic. In the long term no solution is apparent; in the near term the solution is obvious but the difficulties of system synthesis and implementation are underestimated.[7] Thus in 1920 commercial use of nuclear energy seemed far away. By 1949 the achievement appeared reasonable and in 1964 General Electric estimated that fast breeder reactors should be available in 1970.[8] Today the estimate has moved out to the 1980s. The same pattern has been followed by the supersonic transport aircraft. Buschmann has formulated this behavior as a hypothesis and proposed an investigation in greater depth.[9] If this pattern is normal, forecasts should be adjusted accordingly, e.g., forecasts more than, say, ten years in the future brought closer in time and forecasts nearer than ten years moved out. Subsequently Robert Ament made a comparison between a 1969 Delphi study on scientific and technological developments and the 1964 Gordon-Helmer Rand study.[10] Focusing on those items forecast in both studies, he found that all items originally predicted to occur in years before 1980[11] were later shifted further into the future, i.e., the original year seemed optimistic by 1969. On the other hand, two-thirds of the items originally forecast to occur after 1980 were placed in 1969 at a date earlier than that estimated in the 1964 study. Thus we find evidence here, too, of Buschmann's suggested bias.

Grabbe and Pyke have undertaken an analysis of Delphi forecasts of information-processing technology and applications.[12] Forecast events whose occurrence could be verified cover the time period 1968 to 1972. Although six

[7]The large cost overruns on advanced technology aerospace and electronics projects are evidence of this trend (see Chapter III, A).

[8]E. Jantsch, "Technological Forecasting in Perspective," OECD, Paris, 1967, p. 106.

[9]R. Buschmann, "Balanced Grand-Scale Forecasting," *Technological Forecasting* 1 (1969), p. 221.

[10]R. H. Ament, "Comparison of Delphi Forecasting Studies in 1964 and 1969," *FUTURES*, March 1970, p. 43.

[11]T. J. Gordon and H. R. Ament, "Forecasts of Some Technological and Scientific Developments and Their Societal Consequences," *IFF Report R-6*, September 1969.

[12]E. M. Grabbe and D. L. Pyke, "An Evaluation of the Forecasting of Information Processing Technology and Applications," *Technological Forecasting and Social Change* 4, No. 2 (1972), p. 143.

degree of dispersion are definitely related. The regression coefficient is in nearly all cases highly significant for a single panel addressing a related set of events. However, there is no consistent relation among different panels or within a panel when addressing unrelated events.

Martino also finds that the dispersion is not sensitive to the procedure used: in cases where only a single best estimate year is requested the result is similar to that where 10 percent, 50 percent, and 90 percent likelihood dates are stipulated.[3]

Distribution of Responses

Dalkey has analyzed the first-round responses by panels asked to respond to almanac-type questions, i.e., those with known numerical answers.[4] All responses to each question are standardized by subtracting the mean value and dividing by the standard deviation for that question. The resulting distribution of "standardized deviates" shows an excellent fit to a lognormal distribution. Martino has applied the same techniques to the TRW Probe II Delphi using 10 percent, 50 percent, and 90 percent likelihood dates for 1500 events.[5] Again there is a very good fit to a lognormal distribution.

Optimism-Pessimism Consistency

Another interesting analysis on the TRW Probe II data was undertaken by Martino to ascertain whether a panelist tends to have a consistently optimistic or pessimistic bias to his responses.[6] With each respondent providing 10 percent, 50 percent, and 90 percent likelihood dates, three standardized deviates can be computed for each individual and a given event. Taking the means over all events of the standardized deviates for a given individual and likelihood, we find an interesting pattern. Most panelists are consistently optimistic or pessimistic with respect to the three likelihoods, i.e., there are relatively few cases where, say, the 10 percent likelihood is optimistic while the 50 percent and 90 percent likelihoods are pessimistic. Considering the totality of events the individual panelist tends to be biased optimistically or pessimistically with moderate consistency. However, the amount of the bias is not very great; an optimistic panelist is pessimistic in some of his responses and vice versa. In other words, each participant exhibits a standard deviation which is comparable to, or greater than, his mean.

[3]In the first case the interquartile range of best estimates was used, in the second case the 10 percent to 90 percent span was taken.

[4]N. C. Dalkey, "An Experimental Study of Group Opinion," Rand RM-5888-PR, Rand Corporation, Santa Monica, California, March 1969.

[5]J. P. Martino, "The Lognormality of Delphi Estimates," *Technological Forecasting* 1, No. 4 (1970), pp. 355-58.

[6]J. P. Martino, "The Optimism/Pessimism Consistency of Delphi Panelists," *Technological Forecasting and Social Change* 2, No. 2 (1970), pp. 221-24.

Accuracy of Forecasts

An apparent indicator of the value of Delphi as a forecasting tool is its accuracy. Since the method was widely publicized only ten years ago, it is difficult to have sufficient hindsight perspective to evaluate its success by this measure. In any event caution is in order. The most accurate forecast is not necessarily the most useful one. Forecasts are at times most effective if they are self-fulfilling or self-defeating. The Forrester-Meadows World Dynamics model has been sponsored by the Club of Rome in the hope that it will act as an early warning system and prove to be a poor forecast. Delphi may be viewed similarly in terms of effectiveness.

We should also observe that long-range forecasts tend to be pessimistic and short-range forecasts optimistic. In the long term no solution is apparent; in the near term the solution is obvious but the difficulties of system synthesis and implementation are underestimated.[7] Thus in 1920 commercial use of nuclear energy seemed far away. By 1949 the achievement appeared reasonable and in 1964 General Electric estimated that fast breeder reactors should be available in 1970.[8] Today the estimate has moved out to the 1980s. The same pattern has been followed by the supersonic transport aircraft. Buschmann has formulated this behavior as a hypothesis and proposed an investigation in greater depth.[9] If this pattern is normal, forecasts should be adjusted accordingly, e.g., forecasts more than, say, ten years in the future brought closer in time and forecasts nearer than ten years moved out. Subsequently Robert Ament made a comparison between a 1969 Delphi study on scientific and technological developments and the 1964 Gordon-Helmer Rand study.[10] Focusing on those items forecast in both studies, he found that all items originally predicted to occur in years before 1980[11] were later shifted further into the future, i.e., the original year seemed optimistic by 1969. On the other hand, two-thirds of the items originally forecast to occur after 1980 were placed in 1969 at a date earlier than that estimated in the 1964 study. Thus we find evidence here, too, of Buschmann's suggested bias.

Grabbe and Pyke have undertaken an analysis of Delphi forecasts of information-processing technology and applications.[12] Forecast events whose occurrence could be verified cover the time period 1968 to 1972. Although six

[7]The large cost overruns on advanced technology aerospace and electronics projects are evidence of this trend (see Chapter III, A).

[8]E. Jantsch, "Technological Forecasting in Perspective," OECD, Paris, 1967, p. 106.

[9]R. Buschmann, "Balanced Grand-Scale Forecasting," *Technological Forecasting* 1 (1969), p. 221.

[10]R. H. Ament, "Comparison of Delphi Forecasting Studies in 1964 and 1969," *FUTURES*, March 1970, p. 43.

[11]T. J. Gordon and H. R. Ament, "Forecasts of Some Technological and Scientific Developments and Their Societal Consequences," *IFF Report R-6*, September 1969.

[12]E. M. Grabbe and D. L. Pyke, "An Evaluation of the Forecasting of Information Processing Technology and Applications," *Technological Forecasting and Social Change* 4, No. 2 (1972), p. 143.

different Delphi studies were used, eighty-two out of ninety forecasts covering this period were taken from one study: the U.S. Navy Technological Forecast Project. The results appear to contradict the hypothesis that near-term forecasts tend to be optimistic. In this case information-processing advances forecast four to five years in the future occur sooner than expected by the panelists who were drawn largely from government laboratories. There is, of course, the possibility that these laboratories are not as close to the leading edge of technology in this field as industrial and university research and development groups. Alternatively, the meaning of "availability" of a technological application may be interpreted differently by the laboratory forecasters and by the authors of this article.

Delphi Statements

The statements which comprise the elements of a Delphi exercise inevitably reflect the cultural attitudes, subjective bias, and knowledge of those who formulate them. This was recognized by Gordon and Helmer a decade ago and led them to commence the first round with "blank" questionnaires. Every student knows that multiple-choice examinations require insight into the instructor's mode of thought as well as the substance of the questions. Misinterpretations of the given statements can arise in both superior and inferior students. Grabbe and Pyke present examples of good and poor Delphi statements.[13] Statements may be too concise, leading to excessive variations in interpretation, or too lengthy, requiring the assimilation of too many elements. Consequently, we would expect a constraint on the number of words leading to the widest agreement in interpretation. Salancik, Wenger, and Helfer have probed this question more deeply.[14] They use an information theory measure (bits) of the amount of information derivable from a distribution of responses to a Delphi statement to measure consensus and the number of words needed to describe an event as a measure of its complexity. The study uses a computer development and application Delphi study as a test case. The authors find a distinct relation between number of words used and amount of information obtained, i.e., agreement in forecast dates. Low and high numbers of words yield low consensus with medium-statement lengths producing the highest consensus. In the particular case considered, twenty to twenty-five words form the peak in the distribution. This study also finds that the more familiar respondents are with a specific computer application, the fewer words are needed to attain agreement. If many words are used, less information results as

[13] *Ibid.*

[14] J. R. Salancik, W. Wenger, and E. Helfer, "The Construction of Delphi Event Statements," *Technological Forecasting and Social Change* 3, No. 1 (1971), pp. 65-73.

to the occurrence of a familiar event. On the other hand, a longer-word description raises the consensus level for unfamiliar events.

A corresponding pattern is found when expert respondents are compared to nonexperts. The latter develop increasing consensus with longer-event descriptions. The experts, however, come to very high consensus with moderate-statement lengths (higher than the greatest nonexpert consensus) but fall to a very low level of agreement with long statements. Apparently the addition of words brings on an effect somewhat similar to that of disputations by Talmudic scholars about minutiae.

Basis for Respondents' Intuitive Forecast

Salancik has examined the hypothesis that the panelists in a forecasting Delphi assimilate input on *feasibility*, *benefits*, and *potential costs* of an event in an additive fashion to estimate its probable date of occurrence.[15] The subject of the test is again a panel forecast of computer applications. Separate coding of participants' reasons for their chosen dates in the three categories enables the author to make a regression analysis. The second-round median date is made a linear function of the number of positive and negative statements in each of the three categories. He finds that the multiple regression strongly supports the hypothesis. The more feasible, beneficial, or economically viable a concept is judged, the earlier it is forecast to occur. The three categories contribute about equally to the regression.

In a second study independent assessments of feasibility and benefits are rated for twenty computer applications and then combined to form the basis for a rank ordering. This ordering is then compared to the Delphi panelists' responses. Again the correlation supports the suggested model of Delphi input assimilation. This paper adds another beam of support to the idea that Delphi is a systematic and meaningful process of judgment synthesis.

Self-Rating of Experts

Dalkey, Brown, and Cochran tackle another aspect of Delphi: the expertise of the respondents.[16] With a given group we might consider two ways of improving its accuracy: iterating the responses and selecting a more expert subgroup. The latter process implies an ability to identify such a subgroup (e.g., by self-rating) and a potential degradation in accuracy due to the reduced group

[15] J. R. Salancik, "Assimilation of Aggregated Inputs into Delphi Forecasts: A Regression Analysis," *Technological Forecasting and Social Change* 5, No. 3 (1973), pp. 243-48.

[16] N. Dalkey, B. Brown, and S. Cochran, "Use of Self-Ratings to Improve Group Estimates," *Technological Forecasting* 1, No. 3 (1970), pp. 283-91.

size. The authors stipulate a minimum subgroup size to counteract this degradation and they force a clear separation in self-ratings of low- and high-expertise subgroups. The experiments were carried out by the authors using 282 university students and verifiable almanac-type questions. The conclusions: (1) self-rating is a meaningful basis for identification of expertise, and (2) selection of expert subgroups improves the accuracy to a somewhat greater degree than does feedback or iteration.

One must raise the question whether an experiment based on almanac type questions serves as an adequate basis for a conclusion about the validity of self-ratings of expertise for forecasting Delphis. While the lognormality behavior exhibited a similar pattern for factual (almanac-type) and forecasting cases, this similarity might not carry over for self-ratings.

And there are other fascinating unanswered questions. Why do women rate themselves consistently lower than men? Should only the expert subgroup results be fed back to the larger group in the iteration process? How do age, education, and cultural background condition the response of individuals?

The four articles in this chapter provide us with further evaluations of the process. When we use Delphi to draw forth collective expert judgments, we are actually making two substitutions: (1) expert judgment for direct knowledge, and (2) a group for an individual. In the first article, Dalkey strives to develop some mathematically rigorous underpinnings, i.e., a start toward a theory of group estimation. It quickly becomes evident that we still have much to learn about this process. Dalkey emphasizes the concept of "realism," or "track record," to describe the expert's estimation skill and the theory of errors for the group. But the final verdict on their applicability is by no means in.

Scheibe, Skutsch, and Schofer report on several highly instructive findings based on research in the application of Delphi to the derivation of explicit goals and objectives. Analysis of a Delphi goal-formulation experiment for urban systems planning yielded the following important results:

(1) The three-interval scaling methods used—simple ranking, a rating scale, and pair comparisons—give essentially equivalent scales. The rating scale is found to be most comfortable to use by the participants.

(2) Respondents are sensitive to feedback of the scores from the whole group and tend to move (at least temporarily) toward the perceived consensus.

(3) There is only a modest tendency for the degree of confidence of an individual with respect to a single answer to be reflected in movement toward the center of opinion, i.e., less confident members exhibit a somewhat larger movement in the second round.

(4) Stability of the distribution of the group's response along the interval scale over successive rounds is a more significant measure for developing a stopping criterion than degree of convergence. The authors propose a specific stability measure.

Next, Mulgrave and Ducanis discuss an experiment which focuses on the behavior of the dogmatic individual in successive Delphi rounds. Surprisingly, the high-dogmatism group exhibits significantly more changes than the low-dogmatism group. It is the authors' belief that the dogmatic individual looks to authority for support of his view. In the absence of a clearly defined authority, he views the median of the group response as a surrogate.

There clearly exists the possibility of an unnatural overconsensus. Conformists may "capitulate" to group pressures temporarily, on paper. It would be interesting to compare the behavior of such psychological types in a Delphi with that in a conventional committee.

Finally, Brockhoff examines a series of hypotheses on the performance of forecasting groups using the Delphi technique and face-to-face discussions in a Lockean context. He focuses on short-range forecasting and small homogeneous groups. Staff members of local banks trained in economics are queried about data concerning financial questions, banking, stock quotations, and foreign trade. Groups vary in size from eleven to four participants (the latter below the size considered minimal by Dalkey, Brown, and Cochran[17]). The Delphi process uses an interactive computer program for structuring the dialogue as well as computing intermediate and final results. The correlation of self-rating of expertise with individual or group performance, the relation between information exchange and group performance, and the relevance of almanac-type fact-finding questions for short-term forecasting analysis are among the questions examined. One may speculate whether the dogmatism aspect raised by Mulgrave and Ducanis plays a significant role in groups of the type used in Brockhoff's experiments.

For the reader the thrust of this chapter is that, to develop proper guidelines for its use, we can and should subject Delphi to systematic study and evaluation in the same way as has been the case with other techniques of analysis and communication. Much still needs to be learned!

[17] *Ibid.*

V. B. Toward a Theory of Group Estimation *

NORMAN C. DALKEY

Introduction

The term "Delphi" has been extended in recent years to cover a wide variety of types of group interaction. Many of these are exemplified in the present volume. It is difficult to find clear common features for this rather fuzzy set. Some characteristics that appear to be more or less general are: (1) the exericse involves a group; (2) the goal of the exercise is information; i.e., the exercise is an inquiry; (3) the information being sought is uncertain in the minds of the group; (4) some preformulated systematic procedure is followed in obtaining the group output.

This vague characterization at least rules out group therapy sessions (not inquiries), team design of state-of-the-art equipment (subject matter not uncertain), brainstorming (procedure not systematic), and opinion polls (responses are not treated as judgments, but as self-reports). However, the characterization is not sufficiently sharp to permit general conclusions, e.g., concerning the effectiveness of types of aggregation procedures.

Rather than trying to deal with this wide range of activities, the present essay is restricted to a narrow subset. The subject to be examined is *group estimation*—the use of a group of knowledgeable individuals to arrive at an estimate of an uncertain quantity. The quantity will be assumed to be a physical entity—a date, a cost, a probability of an event, a performance level of an untested piece of equipment, and the like.

Another kind of estimation, namely, the identification and assessment of value structures (goals, objectives, etc.) has been studied to some extent, and a relevant exercise is described in Chapter VI. Owing to the difficulty of specifying objective criteria for the performance of a group on this task, it is not considered in the present paper.

To specify the group estimation process a little more sharply, we consider a group $I = \{I_i\}$ of individuals, an event space $E = \{E_j\}$ where E can be either discrete or continuous, and a response space $R = \{R_{ij}\}$ which consists of an estimate for each event by each member of the group. In addition, there is an external process $P = \{P(E_j)\}$, which determines the alternatives in E which will occur. Depending on the problem, P can either be a δ-function on E—i.e., a specification of which event will occur—or a probability distribution P_j on the

*Research reported herein was conducted under Contract Number F30602-72-C-0429 with the Advanced Research Projects Agency, Department of Defense.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.) ISBN 0-201-04294-0; 0-201-04293-2

event space. In general P_j is unknown. For some formulations of the group estimation process, it is necessary to refer to the a priori probability of an event. This is not the same as the external process, but rather, is (in the present context) the probability that is ascribed to an event without knowing the individual or group estimates. This a priori probability will be designated by $U = \{U(E_j)\}$.

In many cases the R_{ij} are simply selections from E. The weatherman says, "It will rain tomorrow"—a selection from the two-event space *rain tomorrow* and *no rain tomorrow*. The long-range technological forecaster says, "Controlled nuclear fusion will be demonstrated as feasible by 1983"—selection of a single date out of a continuum. In these cases the R_{ij} can be considered as 0's and 1's, 1 for the selected event and 0 for the others. Usually, the 0's are left implicit. More complex selections can be dealt with—"It will either rain or snow tomorrow." "Controlled nuclear fusion will be demonstrated in the interval 1980–1985"—by allowing several 1's and interpreting these as an or-combination. Selections can also be considered as special cases of probability distributions over the event space. In the case of probability estimates, the R_{ij} can be probability assignments for discrete alternatives, or continuous distributions for continuous quantities.

A kind of estimate which is sometimes used in applied exercises, but which is not directly expressible in terms of elementary event spaces, is the estimation of the functional relationship between two or more variables (e.g., the extrapolation of a trend). Such an estimate can be included in the present formalism if the relationship is sufficiently well known beforehand so that all that is required is specification of some parameters (e.g., estimating the slope of a linear trend). Although of major practical importance, estimates of complex functional relationships have received little laboratory or theoretical treatment. In particular, there has been no attempt to develop a scoring technique for measuring the excellence of such estimates.

In addition to the group I, event space E, and response space R, a Delphi exercise involves a process $G = G[I,E,R]$ which produces a group response G_j for each event E_j in the event space. Square brackets are used rather than parentheses in the expression for G to emphasize the fact that generally the group estimation process cannot be expressed as a simple functional relationship. The process may involve, for example, discussion among members of the group, other kinds of communication, iteration of judgments with complex selection rules on what is to be iterated, and so on.

One other piece of conceptual apparatus is needed, namely, the notion of *score*, or measure of performance. Development of scoring techniques has been slow in Delphi practice, probably because in most applied studies the requisite data for measuring performance either is unavailable, or would require waiting a decade or so. But in addition, the variety of subject matters, the diversity of motivations for applied studies, and the obscuring effect of the radical uncertainty associated with topics like long-range forecasting of social and tech-

nological events have inhibited the attempt to find precise measures of performance.

In the present paper, emphasis will be put on measures related to the accuracy of estimates. There is a large family of such measures, depending on the form of the estimate, and depending on the interests of the user of the estimate. For this essay, measures will be restricted to what might be called scientific criteria, i.e., criteria which do not include potential economic benefits to the user (or potential costs in terms of experts' fees, etc.) or potential benefits in facilitating group action.

For simple selections out of discrete event spaces a right/wrong measure is usually sufficient, for example, crediting the estimate with a 1 or 0 depending on whether it is correct or incorrect. However, as in the related area of performance testing in psychology, the right/wrong measure is usually augmented by computing a score—total number right, or proportion right, or right-minus-wrong, etc.—over a set of estimates.

For simple selections out of continuous spaces (point estimates), a distance measure is commonly employed, for example, difference between the estimate and the true answer. However, if such measures are to be combined into a score over a set of estimates, some normalizing procedure must be employed to effect comparability among the responses. One normalizing procedure for always positive quantities such as dates, size of objects, probabilities, and the like, is the log error, defined as

$$\text{Error} = \log\left|\frac{R_i}{T}\right|,$$

where T is the true answer and R_i is the individual response. The vertical bars denote the absolute value (neglecting sign). Dividing by T equates proportional errors, and taking the logarithm uniformizes under- and over-estimates. Comparable scoring techniques have not been worked out for quantities with an inherent zero, i.e., quantities admitting both positive and negative answers. Such quantities are rare in applied exercises. Whether this is because that type of quantity is inessential to the subject matter or whether it is due to avoidance by practitioners is hard to say.

For probability estimates, some form of probabilistic scoring system appears to be the best measure available. The theory of probabilistic scoring systems is under rapid development. It is usually pursued within the ambit of subjective probability theories, where the primary property sought is a reward system which motivates the estimator to be honest, i.e., to report his "true" belief.

This requirement can be expressed as the condition that the expected score of the estimator should be a maximum when he reports his true belief. If $q = \{q_j\}$ is the set of probabilities representing the actual beliefs of the estimator on event space $\{E_j\}$, $R = \{R_j\}$ is his set of reported probabilities, and $S_j(R)$ is the reward he receives if event E_j occurs, then the honesty condition can be written

in the form

$$\sum_j q_j S_j(q) \geq \sum_j q_j S_j(R). \tag{1}$$

The expression on the left of the inequality is the individual's subjective expectation if he reports his actual belief; the expression on the right is his expectation if he reports something else.

Formula (1) defines a family of scoring (reward) systems often referred to as "reproducing scoring systems" to indicate that they motivate the estimator to reproduce his actual belief.

It is not difficult to show that the theory of such scoring systems does not depend on the interpretation of q as subjective belief; it is equally meaningful if q is interpreted as the objective probability distribution P on E. With this interpretation the estimator is being rewarded for being as accurate as possible —his objective expectation is maximized when he reports the correct probability distribution.

This is not the place to elaborate on such scoring systems (see [1], [2], [3]). Although (1) leads to a family of reward functions, it is sufficient for the purposes of this essay to select one. The logarithmic scoring system

$$S_j(R) = A \log R_j + B \tag{2}$$

has a number of desirable features. It is the only scoring system that depends solely on the estimate for the event which occurs. The expected score of the estimator is precisely the negative entropy, in the Shannon sense [4], of his forecast. It has the small practical difficulty that if the estimator is unfortunate enough to ascribe 0 probability to the alternative that occurs, his score is negatively infinite. This can usually be handled by a suitable truncation for very small probabilities.

Within this restricted framework, the Delphi design "problem" can be expressed as finding processes G which maximize the expected score of the group response. This is not a well-defined problem in this form, since the expectation may be dependent on the physical process being estimated, as well as on the group-judgment process. There are two ways to skirt this issue. One is to attempt to find G's which have some optimality property independent of the physical process. The other route is to assume that knowledge of the physical process can be replaced by knowledge about the estimators, i.e., knowledge concerning their estimation skill. The next section will deal with the second possibility.

There are two basic assumptions which underlie Delphi inquiries: (a) In situations of uncertainty (incomplete information or inadequate theories) expert judgment can be used as a surrogate for direct knowledge. I sometimes call this the "one head is better than none" rule. (b) In a wide variety of situations of uncertainty, a group judgment (amalgamating the judgments of a group of

experts) is preferable to the judgment of a typical member of the group, the "*n* heads are better than one" rule.

The second assumption is more closely associated with Delphi than the first, which has more general application in decision analysis. These two assumptions do not, of course, exhaust all the factors that enter into the use of Delphi techniques. They do appear to be fundamental, however, and most of the remaining discussion in this paper will be concerned with one or the other of the two.

Individual Estimation

Using the expert as a surrogate for direct knowledge poses no problems as long as the expert can furnish a high-confidence estimate based on firm knowledge of his own. Issues arise when existing data or theories are insufficient to support a high-confidence estimate. Under these circumstances, for example, different experts are likely to give different answers to the same questions.

Extensive "everyday experience" and what limited experimental data exist on the subject strongly support the assumption that knowledgeable individuals can make useful estimates based on incomplete information. This general assumption, then, is hardly in doubt. What is in doubt is the degree of accuracy of specific estimates. What is needed is a theory of estimation that would enable the assignment of a figure of merit to individual estimates on the basis of readily available indices.

An interesting attempt to sidestep this desideratum is to devise methods of rewarding experts so that they will be motivated to follow certain rules of rational estimation. One approach to the theory of probabilistic scoring systems described in the introduction is based on this strategem [5].

The outlines of such a theory of estimation have been delineated in the literature of decision analysis; but it is difficult to disentangle from an attendant conceptualization of a *prescriptive* theory of decisionmaking, or as sometimes characterized, the theory of rational decisionmaking. In the following I will try to do some disentangling, but the subject is complex and *is* and *ought* may still intermingle more than one might wish.

In looking over the literature on decision analysis, there appear to be about six desirable features of estimation that have been identified. The number is not sharp, since there are overlaps between the notions and some semantic difficulties plague the classification. The six desiderata are *honesty, accuracy, definiteness, realism, certainty*, and *freedom from bias*.

Honesty is a clear enough notion. In most cases of estimation, the individual has a fairly distinct perception of his "actual belief," or put another way, he has a relatively clear perception whether his reported estimate matches his actual belief. This is not always the case. In situations with ambiguous contexts, such as the group-pressure situations created by Asch [6], some individuals appear to lose the distinction. The reason for wanting honest reports from estimators is also clear. Theoretically, any report, honest or not, is valuable if the user is aware of potential distortions and can adjust for them. But normally such information is lacking.

Accuracy is also a fairly straightforward notion, and is measured by the score in most cases. It becomes somewhat cloudy in the case of probability estimates for single events, where an individual can make a good score by chance. In this case, the average score over a sequence of events is more diagnostic. But the notion of accuracy then becomes mixed with the notion of realism. Given the meaningfulness of the term, the desirability of accuracy is clear.

Definiteness measures the degree of sharpness of the estimate. In the case of probabilities on discrete event spaces, it refers to the degree to which the probabilities approach 0 or 1 and can be measured by $\sum_{j=1}^{m} R_j^2$. In the case of probability distributions on continuous quantities, it can be measured by the variance or the dispersion. In the case of selections, the comparable notion is "refinement." For discrete event spaces, one report is a refinement of another if it is logically included in the second.

The reason for desiring definiteness is less clear than for accuracy or honesty. "Risk aversion" is a well-known phenomenon in economic theory, but "risk preference" has also been postulated by some analysts [7]. In the case of discrete alternatives, the attractiveness of a report that ascribes a probability close to 1 to some alternative, and probability close to 0 to the others is intuitively "understandable." There is a general feeling that probabilistic estimates close to 0 or 1 are both harder to make, and more excellent when made, than "wishy-washy" estimates in the neighborhood of $\frac{1}{2}$. There is also the feeling that an individual who makes a prediction with a probability of .8 (and it turns out correct) knows more about the phenomenon being predicted than someone who predicts a similar event with probability .6.

All of this is a little difficult to pin down. In the experiments of Girshick, *et al.* [8], there was almost no correlation between a measure of definiteness and the accuracy of the estimates. Part of the problem here appears to be an overlap between the notion of definiteness and uncertainty, which is discussed below. At all events, there appears to be little doubt that definiteness is considered a virtue.

Realism refers to the extent that an individual's estimates are confirmed by events. It is thus closely related to accuracy. However, accuracy refers to a single estimate, whereas realism refers to a set of estimates generated by an individual. Other terms used for this notion are *calibration* [9], *precision* [10], *track record*.

Because the notion of realism is central to the first principle of Delphi stated in the introduction, namely, the substitution of expert judgment for direct knowledge, it warrants somewhat extensive discussion.

In the case of probability judgments, it is possible in theory to take a sequence of estimates from a single estimator, all with the same estimated

probability, and count the number of times the estimate was confirmed. Presumably, if the estimator is using the notion of probability correctly, the relative frequency of successes in that sequence should be approximately equal to the estimated probability. Given enough data of this sort for a wide range of different estimates, it is possible in theory to generate a realism curve for each individual, as illustrated in Fig. 1.

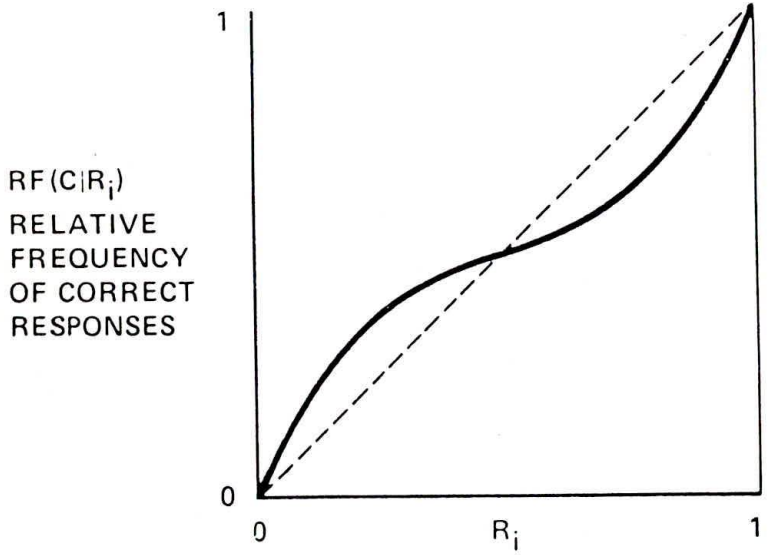


Fig. 1. Typical realism curve.

In Fig. 1 the relative frequency with which an estimate of probability R_i is verified, $RF(C|R_i)$ ("C" for "correct"), is plotted against the estimate. Realism can be defined as the degree to which the $RF(C|R_i)$ curve approximates the theoretically fully realistic curve, namely the dashed line in Fig. 1, where $RF(C|R_i) = R_i$. Figure 1 illustrates a typical realism curve where probabilities greater than $\frac{1}{2}$ are "overestimated" and probabilities less than $\frac{1}{2}$ are underestimated [11].

Various quantities can be used to measure the overall realism of an estimator. $\int_0^1 (RF(C|R_i) - R_i)^2 \; D(R_i)$ where $D(R_i)$ is the distribution of the estimator's reports R_i—roughly the relative frequency with which he uses the various reports R_i—is a reasonable measure. However, for most applications of the concept, it is the realism curve itself which is of interest.

If such a curve were available for a given individual, it could be used directly to obtain the probability of a given event based on his report. In particular, if

the individual were fully realistic, the desired probability would be R_i. At first sight, it might appear that one individual, given his realism curve, is all that is needed to obtain a desired estimate, since the curve furnishes an "objective" translation of his reports into probabilities. However, for any one specific estimate, the reports of several individuals typically differ, and in any case the realism curve is not, by itself, a measure of the expertness or knowledgeability of the individual. In particular, the frequency with which the individual reports relatively high probabilities has to be taken into account.

As a first approximation, the knowledgeability K_i of individual i can be measured by

$$K_i = \int_0^1 S(R_i) D(R_i),$$

where $S(R_i)$ is the probabilistic score awarded to each report R_i and $D(R_i)$ is, as before, the distribution of the reports R_i.

It is easy to verify two properties of K_i: (a) K_i is heavily influenced by the degree of realism of the estimator. For a given distribution of estimates, $D(R_i)$, K_i is a maximum when the individual is fully realistic. (b) K_i is also influenced by the average definiteness of the estimator. The higher the definiteness (e.g., measured by $\int R_i^2 \; D(R_i)$), the higher the expected score.

Theoretically, one might pick the individual with the highest K rating and use him exclusively. There are two caveats against this procedure. On a given question, the individual with the highest average K may not furnish the best response; and, more in the spirit of Delphi, if realism curves are available for a set of individuals, then it is sometimes feasible to derive a group report which will have a larger average score than the average score of any individual—in short, the K measure for the group can be higher than the K measure for any individual.

As far as the first principle—substitution of expert judgment for knowledge—is concerned, the question whether realism curves exist for each individual is a crucial one. Detailed realism curves have not been derived for the types of subject matter and the type of expert desired for applied studies. In fact, detailed track records for any type of subject matter are hard to come by. Basic questions are: Is there a stable realism curve for the individual for relevant subject matters? How general is the curve—i.e., is it applicable to a wide range of subject matters? How subject is the curve to training, to use of reward systems like the probabilistic score, to contextual effects such as the group pressure effect in the Asch experiments?

Certainty is a notion that is well known in the theory of economic decisionmaking. It has not played a role in the study of estimation to the same extent. In the case of economic decisionmaking, the distinction has been made between *risk* (situations that are probabilistic, but the probabilities are known) and *uncertainty* (situations where the probabilities are not known) [13]. Many analysts appear to believe that in the area of estimation this distinction breaks

down—uncertainty is sufficiently coded by reported probabilities. However, the distinction appears to be just as applicable to estimation as to any other area where probabilities are relevant. Consider, for example, the situation of two coins, where an individual is asked to estimate the probability of *heads*. Coin A is a common kind of coin where the individual has flipped it several times. In this case, he might say that the probability of heads is $\frac{1}{2}$ with a high degree of confidence. Coin B, let's say, is an exotic object with an unconventional shape, and the individual has not flipped it at all. In the case of coin B he might also estimate a probability of $\frac{1}{2}$ for heads, but he would be highly uncertain whether that is the actual probability. Probability $\frac{1}{2}$, then, cannot express the uncertainty attached to the estimate for the second coin.

A closer approximation to the notion of uncertainty can be obtained by considering a distribution on the probabilities. For example, the individual might estimate that the probability of the familiar coin has a tight distribution around $\frac{1}{2}$, whereas the distribution for the unfamiliar coin is flat, as in Fig. 2. The independent variable is labeled q to indicate that it is the individual's belief, and not necessarily his report, which is being graphed.

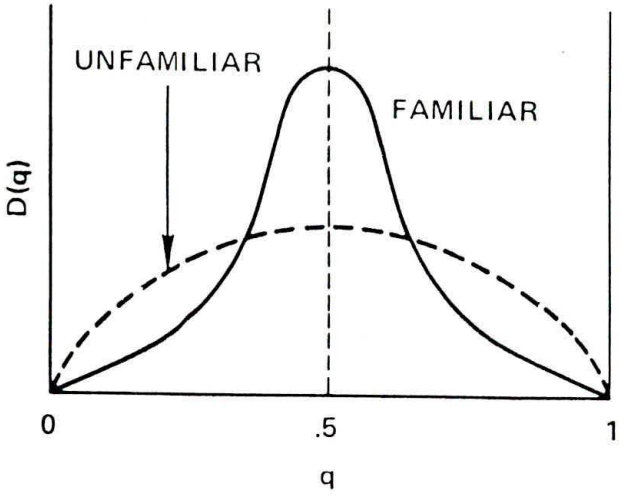


Fig. 2. Uncertainty represented as a higher-level distribution.

The use of a higher-level distribution is only an approximation to the notion of uncertainty, since the distribution itself might be uncertain, or, in more familiar language, the distribution may be "unknown." The use of additional levels has been suggested, but for practical reasons seems highly unappealing.

The problem of representing uncertainty in precise terms is closely related to past attempts to translate lack of information into probabilities by means of principles such as the "law of insufficient reason," or the rule of equal ignorance. These have invariably lead to paradoxes [13].

Using the idea of the dispersion of a second-level distribution as an approximate measure of uncertainty, there is some interaction between the notions of realism, definiteness, and certainty. It is not possible for a set of estimates to be simultaneously realistic, definite, and uncertain. Assuming that the individual will give as his first level report R_i the mean of his second-level distribution, then as R_i approaches 1 or as R_i approaches 0, the standard deviation of the distribution $D(q)$ approaches 0. Figure 3 illustrates this coupling for $R_i = .9$. If the individual is realistic and estimates a probability of .9 for a given event, then the standard deviation of his higher-level distribution for that estimate must be small.

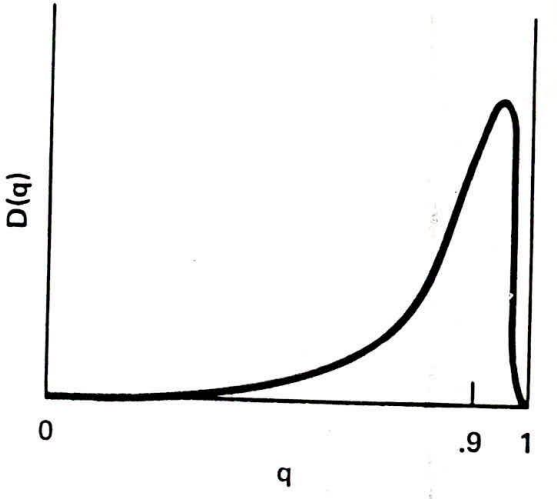


Fig. 3. Illustration of coupling between certainty and definiteness.

Unfortunately, the coupling applies only to the extremes of the 0 to 1 interval. At $q = \frac{1}{2}$, D can be about anything, and the estimator still be realistic. If the average probabilistic score for an estimate with a second-level distribution D is computed, the average score is influenced only by the mean, and otherwise is independent of D. Thus an average probabilistic score does not reflect uncertainty. It appears that something like the variance of the score will have to be included if certainty is to be reflected in a score.

At the present, the only "visible" index of certainty is the self-rating—i.e., a judgment by the individual of his competence or knowledgeability concerning the estimate. This has turned out to be a significant index for rating group estimates [14]; it is not so effective for individual estimates. Due to the lack of a theoretical definition of the self-rating, it has not been possible to include it in a formal theory of aggregation. However, the self-rating has proved to be valuable for selecting more accurate subgroups [15].

Bias is a term that has many shades of meaning in statistics and probability. I am using the term to refer to the fact that there may be subclasses of events for which $RF(C|R_i)$ may be quite different from the average relative frequency expressed by the realism curve. Of course, for this to be of interest, the subclasses involved must be identifiable by some means other than the relative frequency. It is always possible after the fact to select a subset of events for which an individual has estimated the probability R_i which has any $RF(C|R_i)$.

In the theory of test construction, e.g., for achievement tests or intelligence tests, it is common to assume an underlying scale of difficulty for the questions, where difficulty is defined as the probability that a random member of the target population can answer the question correctly [16]. This probability will range from 1 for very easy questions to 0 for very hard questions, as illustrated by the solid curve in Fig. 4. From the standpoint of the present discussion, the significant fact is that when a class of questions is identified as belonging to the very difficult group in a sample of the population, that property carries over to other members of the population—in short the property of being very difficult is relatively well defined.

At some point in the scale of difficulty, labeled d in Fig. 4, a typical member of the population could increase his score by abandoning the attempt to "answer" the question and simply flipping a coin (assuming that it is a true/false or yes/no type of question). Put another way, from point d on, the individual becomes a counterpredictor—you would be better off to disbelieve his answers.

Contrasted with this notion of difficulty is the notion that underlies theories of subjective probability that, as the individual's amount of information or skill declines, the probability of a correct estimate declines to 50 percent as illustrated by the dashed curve in Fig. 4. Ironically, it is the probabilistic notion that influences most scoring schemes, which assume that the testee can achieve 50 percent correct by "guessing," and hence the score is computed by subtracting the number of wrong answers from the number right. By definition, for the more difficult items, the testee cannot score 50 percent by "guessing" unless that means literally tossing a coin and not trusting his "best guess."

If it turns out that "difficult" questions in the applied area have this property, even for experts, then the first principle does not hold for this class. Although there are no good data on this subject, there does not appear to be a good reason why what holds for achievement and intelligence tests should not also hold for "real life" estimates. Almost by definition, the area of most interest in applications is the area of difficult questions. If so, assuming that the set of counterpredictive questions can be identified before the fact, then a good fair coin would be better than an expert. It is common in experimental design to use randomization techniques to rule out potential biases. There is no logical reason why randomization should not be equally potent in ruling out bias in the case of estimation.

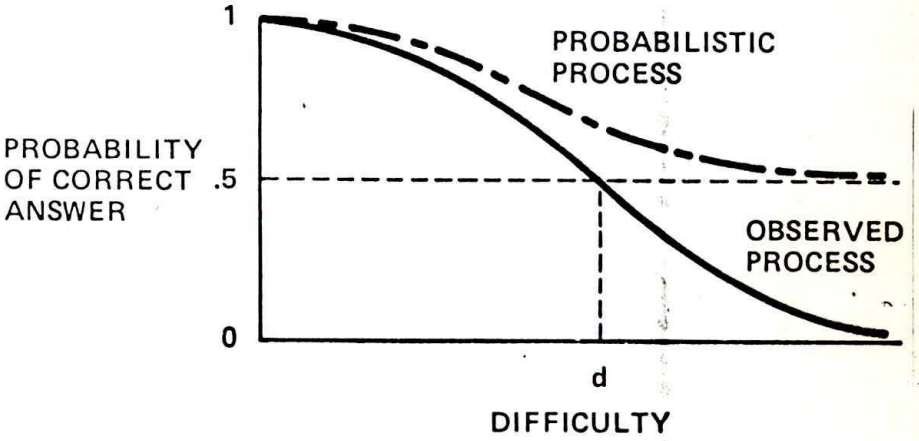


Fig. 4. Scale of difficulty in test construction.

The four notions, honesty, accuracy, definiteness, and precision, are all tied together by probabilistic scoring systems. In fact, a reproducing scoring system rewards the estimator for all four. As pointed out in the introduction, condition (1) defines the same family of scoring systems whether q is interpreted as subjective belief, or as objective probability. Thus, the scoring system rewards the estimator for both honesty and accuracy. In addition, the condition leads to the result that $\sum_j q_j S_j(q)$ is convex in q. This convex function of q can be considered as a measure of the dispersion of q; and in fact, three of the better-known scoring systems define three of the better-known measures of dispersion. Thus, if S_j is the quadratic scoring system

$$S_j(R) = 2R_j - \sum_j R_j^2,$$

then $\sum_j q_j S_j(q) = \sum_j q_j^2$ which is a measure of variance. If S_j is the spherical

scoring system

$$S_j(R) = \frac{R_j}{\sqrt{\sum_j R_j^2}},$$

then $\Sigma_j q_j S_j(q) = \sqrt{\Sigma_j q_j^2}$, a measure similar to the standard deviation. Finally, for the logarithmic scoring system, $\Sigma_j q_j S_j(q) = \Sigma_j q_j \ln q_j$, which is the negative of the Shannon entropy, another measure of definiteness.

Realism enters in a more diffuse fashion. In general, the probabilistic score for a single event is not very diagnostic, since the individual may have obtained a high (or low) score by chance. Thus, as for most scoring systems, an average (or total) score over a large set of questions is the usual basis for evaluation. But over a large set of questions, the average score is determined by the realism curve of the individual, in conjunction with the relative frequency with which he makes reports of a given probability. In general, if the estimator is not realistic, he will lose

$$\int_0^1 (RF(C|R_t)S(RF(C|R_t)) - RF(C|R_t)S(R_t))D(R_t).$$

As pointed out above, the probabilistic score does not include a penalty for uncertainty, nor does it include a penalty for bias, except where bias shows up in the realism curve. The latter case is simply the one where, for whatever reason, the individual is faced with a stream of questions in which the number of questions biased in a given direction is greater than the number biased in the opposite direction.

To sum up this rather lengthy section: The postulate that, in situations of uncertainty, it is feasible to substitute expert judgment for direct knowledge is grounded in a number of empirical hypotheses concerning the estimation process. These assumptions are, primarily, that experts are approximately realistic in the sense defined above, that the realism curve is stable over a relatively wide range of questions (freedom from bias), and that knowledgeability is a stable property of the expert. At the moment, these are hypotheses, not well-demonstrated generalizations.

Theoretical Approaches to Aggregation

Assuming that, for a given set of questions, we can accept the postulate that expert judgment is the "best information obtainable," there remains the question how the judgments of a group of experts should be amalgamated. In the present section, three approaches to this issue are discussed. The discussion is limited to elementary forms of aggregation, where the theory consists of a

mathematical rule for deriving a group response from a set of individual responses; thus, an *elementary* group estimation process can be defined as a function, $G = G(E, I, R)$.

Theory of Errors

This approach interprets the set of judgments of a group of experts as being similar to the set of readings taken with an instrument subject to random error. It seems most appropriate when applied to point estimates of a continuous quantity, but formally at least, can be applied to any type of estimate. In analogy with the theory of errors for physical measurements, a statistical measure of central tendency is considered to be the best estimate of the quantity. Some measure of dispersion is taken to represent a confidence interval about the central value.

Relevant aspects of the individual estimation process such as skill or amount of information of the expert, are interpreted as features of the "theory of the instrument."

This point of view appears to be most popular in the Soviet Union [17]; however, a rough though unexpressed version of this approach underlies much of the statistical analysis accompanying many applied Delphi studies. To my knowledge, this approach has not been developed in a coherent theory, but rather, has been employed as an informal "interpretation"—i.e., as a useful analogy.

The theory-of-errors approach has the advantages of simplicity, and similarity with well-known procedures in physical measurement theory. Much of the empirical data which have been collected with almanac and short-range prediction studies is compatible with the analogy. Thus, the distribution of estimates tends to follow a common form, namely the lognormal [18]. If the random errors postulated in the analogy are assumed to combine multiplicatively (rather than additively as in the more common Gaussian theory), then a lognormal distribution would be expected.

The geometric mean of the responses is more accurate than the average response; or more precisely, the error of the geometric mean is smaller than the average error. Since the median is equal to the geometric mean for a lognormal distribution [19], the median is a reasonable surrogate, and has been the most widely used statistic in applied studies for the representative group response.

The error of the median is, on the average, a linear function of the standard deviation [20], which would be predicted by the theory of errors. The large bias observed experimentally (bias = error/standard deviation) is on the average a constant, which again would be compatible with the assumption that experts perform like biased instruments.

Although the analogy looks fairly good, there are several open questions that prevent the approach from being a well-defined theory. There does not exist at present a "theory of the instrument" which accounts for either the observed degree of accuracy of individual estimates or for the large biases observed in experimental data. Perhaps more serious, there is no theory of errors which accounts for the presumed multiplicative combination of errors—especially since the "errors" are exemplified by judgments from different respondents.

Despite this lack of firm theoretical underpinnings, the theory-of-errors approach appears to fit the accumulated data for point estimates more fully than any other approach.

In addition, the measures of central tendency "recommended by" the theory of errors have the desirable feature that the advantage of the group response over the individual response can be demonstrated irrespective of the nature of the physical process being estimated. So far as I know, this is the only theoretical approach that has this property.

To make the demonstration useful in later sections, a somewhat more sophisticated version of the theory will be dealt with than is necessary just to display the "group effect."

Consider a set of individual estimates R_{ij} on an event space E_j, where the R_{ij} are probabilities, i.e., $\sum_j R_{ij} = 1$. We assume there is a physical process that determines objective probabilities $P = \{P_j\}$ for the event space, but P is unknown. Consider a group process G which takes the geometric mean of the individual estimates as the best estimate of the probability for each event. However, the geometric means will not be a probability, and must be normalized. This is accomplished by setting

$$G_j = \frac{\left(\prod_{i=1}^{n} R_{ij}\right)^{\frac{1}{n}}}{\sum_{j=1}^{m}\left(\prod_{i=1}^{n} R_{ij}\right)^{\frac{1}{n}}}. \tag{2}$$

We can now ask how the expected probabilistic score of the group will compare with the average expected score of the individual members of the group. It is convenient to use the abbreviation C for the reciprocal of the normalizing term

$$C = \frac{1}{\sum_{j=1}^{m}\left(\prod_{i=1}^{n} R_{ij}\right)^{\frac{1}{n}}}.$$

Using the logarithmic scoring system and setting the constants $A = 1$, $B = 0$, we

have

$$S_j(G) = \log\left(C\left(\prod_{i=1}^{n} R_{ij}\right)^{\frac{1}{n}}\right) \tag{3}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \log R_{ij} + \log C. \tag{4}$$

Taking the expected score,

$$\sum_{j=1}^{m} P_j S_j(G) = \sum_{j=1}^{m} P_j \frac{1}{n}\sum_{i=1}^{n} \log R_{ij} + \log C, \tag{5}$$

and rearranging terms, where $\bar{S}(G)$ denotes the expected score of the group, i.e., $\bar{S}(G) = \sum_{j=1}^{m} P_j S_j(G)$

$$\bar{S}(G) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} P_j \log R_{ij} + \log C, \tag{6}$$

$\log C$ appears outside the summation, because, as a constant, $\sum_{j=1}^{m} P_j \log C = \log C$. The expression $\sum_{j=1}^{m} P_j \log R_{ij}$ is just the expected score $\bar{S}(R_i)$ of individual i, and the expression on the right of (6) excluding $\log C$ is the *average* individual expected score, which we can abbreviate as $\hat{S}(R)$. Thus

$$\bar{S}(G) = \hat{S}(R) + \log C. \tag{7}$$

Since C is greater than 1, $\log C$ is positive, and the expected group score is greater than the average expected individual score by the amount $\log C$. C depends only on the individual responses R_{ij} and not on the specific events E or the objective probabilities P.

Formula (7) exemplifies a large variety of similar results that can be obtained by using different statistics as the aggregation rule and different scoring rules.[1]

Probabilistic Approach

Theoretically, joint realism curves similar to the individual realism curve of Fig. 1 can be generated, given enough data. In this case, the relative frequency $RF(C|R)$ of correct estimates would be tabulated for the joint space of

[1] Brown [21] derives a similar result for continuous distributions, the quadratic scoring system, and the mean as the group aggregation function.

esponses R for a group. Such a joint realism curve would be an empirical ggregation procedure. $RF(C|R)$ would define the group probability judgment s a function of R.

Although possible in theory (keeping in mind all the caveats that were raised with respect to individual realism curves), in practice generating joint realism urves for even a small group would be an enormous enterprise. It is conceivble that a small group of meteorologists, predicting the probability of rain for given locality many thousands of times, might cover a wide enough region of he R space to furnish stable statistics. However, for the vast majority of types of uestion where group estimation is desired, individual realism curves are ifficult to come by; group realism curves appear to be out of the question for he present.

One possible simplification at this point could be made if general rules oncerning the interdependence of individual estimates on various types of stimation tasks could be ascertained. In such a case, the joint realism curves ould be calculated from individual realism curves. Although very iffy at this point, it is conceivable that a much smaller body of data could enable the esting of various hypotheses concerning dependence. In any case, by develop- ng the mathematical relationships involved, it is possible to pursue some eoretical comparisons of probabilistic aggregation with other types of aggre- ation.

In the following, the convention will be used that whenever the name of a set of events occurs in a probability expression, it denotes the assertion of the joint ccurrence of the members of the set. For example, if X is a set of events, $X=\{X_i\}$, then $P(X)=P(X_1\cdot X_2\ldots X_n)$, where the period indicates "and." In addition, to reduce the number of subscripts, when a particular event out of a et of events is referred to, the capital letter of the index of that event will be sed to refer to the occurrence of the event. Thus $P(X_j)$ will be written $P(J)$.

The degree of dependence among a set of events X is measured by the eparture of the joint probability of the set from the product of the separate robabilities of the events. Letting D_X denote the degree of dependence within the set X, we have the definition

$$D_X=\frac{P(X)}{\prod_{i=1}^{n}P(X_i)}. \tag{8}$$

This notion is usually introduced by taking into account dependence among subsets of X as well as the more global notion defined by (8). However, for generating a probabilistic aggregation function, interactions among subsets can e ignored, proving we maintain a set fixed throughout any given computation.

A useful extension of the notion of dependence is that of dependence with respect to a particular event, say $E_j=J$.

$$D_X^J=\frac{P(X|J)}{\prod_{i=1}^{n}P(X_i|J)}. \tag{9}$$

From the rule of the product, we have

$$D_X^J=\frac{P(J\cdot X)P(J)^n}{P(J)\prod_{i=1}^{n}(J\cdot X_i)}. \tag{10}$$

The probability we want to compute is $P(J|R)$; that is, we want to know the probability of an event E_j given that the group reports R. Again, from the rule of the product, we have

$$P(J|R)=\frac{P(R\cdot J)}{P(R)}.$$

Substituting R for X in (10) and multiplying the top and bottom of the right-hand side by $P(R)/\prod_{i=1}^{n}P(R_i)$, and rearranging, gives

$$P(J|R)=\frac{D_R^J\prod_{i=1}^{n}P(J|R_i)}{D_R U(J)^{n-1}}. \tag{11}$$

Formula (11) presents the computation of the joint probability in terms of the individual reports, the dependency terms, and the "a priori" probability $U(J)$. The $P(J|R_i)$ can be derived from individual realism curves. In case the estimators are all fully realistic, then $P(J|R_i)=R_i$. $U(J)$ is the probability of the event J based on whatever information is available without knowing R.[2]

The ratio D_R^J/D_R measures the extent to which the event J influences the dependence among the estimates. If the estimates are independent "a priori," $D_R=1$. However, the fact that estimators do not interact (anonymity) or make separate estimates, does not guarantee that their estimates are independent.

[2]All of the formulations in this subsection are presumed to be appropriate for some context of information. This context could be included in the formalism, e.g., as an additional term in the reference class for all relative probabilities, or as a reference class for "absolute" probabilities. For example, if the context is labeled W, $U(J)$ would be written $P(J|W)$, $P(J|R)$ would be written $P(J|R\cdot W)$. However, since W would be constant throughout, and ubiquitous in each probability expression, it is omitted for notational simplicity.

They could have read the same book the day before. The event related dependence D_R^J is even more difficult to derive from readily available information concerning the group.

If there is reason to believe that a particular group is completely independent in their estimates, and in addition each member is completely realistic, (11) reduces to

$$P(J|R) = \frac{\prod_{i=1}^{n} R_i}{U(J)^{n-1}}. \tag{12}$$

The simplicity of (12) is rather misleading; it depends on several strong assumptions. (11) on the other hand, is exact, but contains terms which are difficult to evaluate.

An exact expression for $P(J|R)$ can be obtained which does not involve D_R by noting that

$$P(J|R) = \frac{P(J|R)}{P(J|R) + P(\bar{J}|R)}.$$

Substituting for $P(J|R)$ on the right-hand side from (11) and the corresponding expression for $P(\bar{J}|R)$ ($\bar{J}$ denotes the complement of J or "not-J") and dividing top and bottom by $D_J^R / U(J)^{n-1} D_R$ we obtain

$$P(J|R) = \frac{\prod_{i=1}^{n} P(J|R_i)}{\prod_{i=1}^{n} P(J|R_i) + D \prod_{i=1}^{n} (1 - P(J|R_i))}, \tag{13}$$

where

$$D = \frac{D_R^{\bar{J}}}{D_R^J} \left(\frac{U(J)}{U(\bar{J})} \right)^{n-1}$$

If the estimators are all fully realistic and fully independent, and the a priori probability $= \frac{1}{2}$, (13) reduces to

$$P(J|R) = \frac{\prod_{i=1}^{n} R_i}{\prod_{i=1}^{n} R_i + \prod_{i=1}^{n} (1 - R_i)}. \tag{14}$$

To complete this set of estimation formulae, if there are several alternatives in E, and it is desired to compute the group estimate for each alternative from

the individual estimates for each alternative, (13) generalizes to

$$P(E_k|R) = \frac{\prod_{i=1}^{n} P(E_k|R_i)}{\sum_{j=1}^{m} D_{jk} \prod_{i=1}^{n} P(E_j|R_i)}, \tag{15}$$

where

$$D_{jk} = \frac{D_R^{E_j}}{D_R^{E_k}} \left(\frac{U(E_k)}{U(E_j)} \right)^{n-1}.$$

(14) is similar to a formula that can be derived using the theorem of Bayes [22]. Perhaps the major difference is that (14) makes the "working" set of estimates the $P(E_j|R_i)$ which can be obtained directly from realism curves, whereas the corresponding formula derived from the theorem of Bayes involves as working estimates $P(R_i|E_j)$ which are not directly obtainable from realism curves. Of course, in the strict sense, the two formulae have to be equivalent, and the $P(R_i|E_j)$ are contained implicitly in the dependency terms. Without some technique for estimating the dependency terms separately from the estimates themselves, not much is gained by computing the group estimate with (14).

Historically, the "a priori" probabilities $U(J)$ have posed a number of conceptual and data problems to the extent that several analysts, e.g., R. A. Fisher [23], have preferred to eliminate them entirely and work only with the likelihood ratios—in the case of (14), the ratios

$$\frac{\prod_{i=1}^{n} R(E_j|R_i)}{\prod_{i=1}^{n} P(E_k|R_i)}.$$

This approach appears to be less defensible in the present case, where the a priori probabilities enter in a strong fashion, namely with the $n-1$ power.

For a rather restricted set of situations, a priori probabilities are fairly well defined, and data exist for specifying them. A good example is the case of weather forecasting, where climatological data form a good base for a priori probabilities. Similar data exist for trend forecasting, where simple extrapolation models are a reasonable source for a priori probabilities. However, in many situations where expert judgment is desired, whatever prior information exists is in a miscellaneous form unsuited for computing probabilities. In fact, it is in part for precisely this reason that experts are needed to "integrate" the miscellaneous information.

Some additional light can be thrown on the role of a priori probabilities as well as the dependency terms by looking at the expected probabilistic score. In the case of the theory-of-errors approach, it was possible to derive the result that, independent of the objective probability distribution P, the expected probabilistic score of the group estimate is higher than the average expected score of individual members of the group. This result is not generally true for probabilistic aggregation.

Since probabilistic aggregation depends upon knowing the a priori probabilities, a useful way to proceed is to define a *net* score obtained by subtracting the score that would be obtained by simply announcing the a priori probability. Letting $S^*(G)$ denote the expected net score of the group and $S^*(R_i)$ the expected net score of individual i, and $S(E)$ the score that would be obtained if $\{U(E_j)\}$ were the report, $S^*(G) = S(G) - S(E)$ and $S^*(R_i) = S(R_i) - S(E)$. The net score measures the extent to which the group estimate is better (or worse) than the a priori estimate. This appears to be a reasonable formulation, since presumably the group has added nothing if its score is no better (or is worse) than what could be obtained without it.

Many formulations of probabilistic scores include a similar consideration when they are "normalized." This is equivalent to subtracting a score for the case of equally distributed probabilities over the alternatives. Thus the score for an individual is normalized by setting $S^*(R_i) = S(R_i) - S(Q)$ where $Q_j = 1/m$ and m is the number of alternatives. In effect this is assuming that the a priori probabilities are equal.

Computing the expected group net score from (11) we have

$$\sum_{j=1}^{m} P_j G_j - S(E) = \sum_{j=1}^{m} P_j \ln \left(\frac{D_R^J \prod_{i=1}^{n} P(J|R_i)}{(J)^{n-1} D_R} \right) - S(E) \tag{16}$$

$$= -(n-1) \sum_{j=1}^{m} P_j \ln U(J) + \sum_{j=1}^{m} P_j \sum_{i=1}^{n} \ln P(J|R_i)$$

$$+ \sum_{j=1}^{m} P_j \ln D_R^J - \ln D_R - S(E) \tag{17}$$

$$= -nS(E) + n\hat{S}(R) + \sum_{j=1}^{m} P_j \ln D_R^J - \ln D_R, \tag{18}$$

whence $S^*(G) = nS^*(R) +$ Expectation of dependency terms.

If the average net score of the individual members is positive (i.e., the average member of the group does better than the a priori estimate), then the group score will be n times as good, providing the dependency terms are small

or positive. On the other hand, if the average net score of the individual members is negative, then the group will be n times as bad, still assuming the dependency terms small. Since the logarithm of D_R will be negative if $D_R < 1$, (18) shows that the most favorable situation is not independence where $D_R = 1$, $\ln D_R = 0$, but rather, the case of negative dependence, i.e., the case where it is less likely that the group will respond with R than would be expected from their independent frequencies of use of R_i.

The role of the event-related dependency term $\sum_{j=1}^{n} P_j \ln D_R^J$ is somewhat more complex. In general, it is desirable that D_R^J be greater than one for those alternatives where the objective probability P_j is high. This favorable condition would be expected if the individuals are skilled estimators, but cannot be guaranteed on logical grounds alone.

One of the more significant features of the probabilistic approach is that under favorable conditions the group response can be more accurate than any member of the group. For example, if the experts are fully realistic, agree completely on a given estimate, are independent, and finally, if it is assumed that the a priori probabilities are equal (the classic case of complete prior ignorance), then formula (14) becomes

$$P(J|R) = \frac{p^n}{p^n + (1-p)^n}, \tag{19}$$

where p is the common estimate, and n is the number of members of the group. If $p > .5$, then $P(J|R)$ rapidly approaches 1 as n increases. For example, if $p = 2/3$ and n is 5, then $P(J|R) = 32/33$. If the theory-of-errors approach were being employed, the group estimate would be 2/3 for any size group.

In this respect, it seems fair to label the probabilistic approach "risky" as compared with the theory-of-errors approach. Under favorable conditions the former can produce group estimates that are much more accurate than the individual members of the group; under less favorable conditions, it can produce answers which are much worse than any member of the group.

Axiomatic Approach

A somewhat different way to develop a theory of group estimation is to postulate a set of desired characteristics for an aggregation method and determine the process or family of processes delimited by the postulates. This approach has not been exploited up to now in Delphi research. The major reason has been the large number of nonformal procedures associated with an applied Delphi exercise—formulation of a questionnaire, selection of a panel of experts, "interpretation of results," and the like. However, if the aggregation process is defined formally as in the two preceding subsections, where

questionnaire design is interpreted as defining the event space E, and panel selection is reduced to defining the response space R, then the axiomatic approach becomes feasible.

Considering the group estimation process as a function $G=G(E,I,R)$, various properties of this function appear "reasonable" at first glance. Some of the more evident of these are:

(A) *Unanimity*. If the group is in complete agreement, then the group estimate is equal to the common individual estimate; i.e., if $R_{ij}=R_{kj}$ for all i and k, then $G(R)=R$.

(B) *Monotony*. If R and R' are such that $R_{ij}\geqslant R'_{ij}$ for all i, then $G_j(R)\geqslant G_j(R')$. If R and G are defined as real numbers then they fulfill the usual ordering axioms, and condition B implies condition A.

(C) *Nonconventionality*. G is not independent of the individual estimates; i.e., $G(R)\neq G(S)$ for every possible R and S.

(D) *Responsiveness*. G is responsive to each of the individual estimates; i.e., $G(R)\neq G'(T)$, where T is a proper subvector of R.

(E) *Preservation of Probability Rules*. If G is an aggregation function which maps a set of individual probability estimates onto a probability, then G preserves the rules of probability. For example, if $T_{ij}=R_{ij}S_{ij}$ for all i and j (as would be the case if R_{ij} is the estimated probability of E_i and S_{ij} is the estimated relative probability of an event E'_j given that E_i occurs) then

$$G(T_j)=G(R_j)G(S_j).$$

This set of conditions will be displayed more fully below.

All of these conditions have a fairly strong intuitive appeal. However, intuition appears to be a poor guide here. The first four postulates are fulfilled by any of the usual averaging techniques. But A, which is perhaps the most apparently reasonable of them all, is not fulfilled by the probabilistic aggregation techniques discussed in the previous subsection. It was pointed out there that one of the more intriguing possibilities with probabilistic aggregation is that the group estimate may be higher (or lower, depending on the interaction terms) than any individual estimate.

It can be shown that there is no function that fulfills all five of the postulates; in fact, there is no function that fulfills D and E. The proof of this impossibility theorem is given elsewhere [24]; it will only be sketched here.

Three basic properties of probabilities are (a) normalization, if p is a probability, $0\leqslant p\leqslant 1$; (b) complementation, $P(J)+P(\bar{J})=1$; and (c) multiplicative conjunction, i.e., $P(J_1\cdot J_2)=P(J_1)P(J_2/J_1)$. The last is sometimes taken as a postulate, sometimes is derived from other assumptions.

If the individual members of a group are consistent, their probability judgments will fulfill these three conditions. It would appear reasonable to require that a group estimate also fulfill the conditions, consistently with the

individual judgments. In addition, condition D, above, appears reasonable. This leads to the four postulates:

P1. $0\leqslant G(R)\leqslant 1$.
P2. $G(1-R)=1-G(R)$.
P3. $G(R\cdot S)=G(R)G(S)$.
P4. $G(R)\neq G'(S)$, where S is a subvector of R (condition D).

Here, $R\cdot S$ is the inner product of the two vectors R and S, i.e.,

$$R\cdot S=(R_1S_1,R_2S_2,\ldots,R_nS_n).$$

P1-P3 have the consequence that G is both multiplicative and additive. The multiplicative property comes directly from P3, and the additive property—i.e., $P(R+S)=P(R)+P(S)$—is derived by using the other postulates. For functions of a single variable, there is only one which is both multiplicative and additive, namely the identity function $f(x)=x$. There is no corresponding identity function for functions of several variables except the degenerative function, $G(R)=G'(R_i)=R_i$, which violates P4.

This result may seem a little upsetting at first glance. It states that probability estimates arrived at by aggregating a set of individual probability estimates cannot be manipulated as if they were direct estimates of a probability. However, there are many ways to react to an impossibility theorem. One is panic. There is the story that the logician Frege died of a heart attack shortly after he was notified by Bertrand Russell of the antinomy of the class of all classes that do not contain themselves. There was some such reaction after the more recent discovery of an impossibility theorem in the area of group preferences by Kenneth Arrow [25]. However, a quite different, and more pragmatic reaction is represented by the final disposition of the case of 0. In the 17th century, there was long controversy on the issue whether 0 could be treated as a number. Strictly speaking there is an impossibility theorem to the effect that 0 cannot be a number. As everyone knows, division by 0 can lead to contradictions. The resolution was a calm admonition, "Treat 0 as a number, but don't divide by it."

In this spirit, formulation of group probability estimates has many desirable properties. It would be a pity to forbid them because of a mere impossibility theorem. Rather, the reasonable attitude would appear to be to use group probability estimates, but at the same time not to perform manipulations with the group aggregation function which can lead to inconsistencies.

Coda

The preceding has taken a rather narrow look at some of the basic aspects of group estimation. Many significant features, such as interaction via discussion or formal feedback, the role of additional information "fed-in" to the group, the

differences between open-ended and prescribed questions, and the like, have not been considered. In addition, the role of a Delphi exercise within a broader decisionmaking process has not been assessed. What has been attempted, albeit not quite with the full neatness of a well-rounded formal theory, is the analysis of some of the basic building blocks of group estimation.

To summarize briefly: The outlines of a theory of estimation have been sketched, based on an objective definition of estimation skill—the realism curve or track record of an expert. Several approaches to methods of aggregation of individual reports into a group report have been discussed. At the moment, insufficient empirical data exist to answer several crucial questions concerning both individual and group estimation. For individual estimation, the question is open whether the realism curve is well defined and sufficiently stable so that it can be used to generate probabilities. For groups, the degree of dependency of expert estimates, and the efficacy of various techniques such as anonymity and random selection of experts in reducing dependency have not been studied.

By and large it appears that two broad attitudes can be taken toward the aggregation process. One attitude, which can be labeled conservative, assumes that expert judgment is relatively erratic and plagued with random error. Under this assumption, the theory-of-errors approach looks most appealing. At least, it offers the comfort of the theorem that the error of the group will be less than the average error of the individuals. The other attitude is that experts can be calibrated and, via training and computational assists, can attain a reasonable degree of realism. In this case it would be worthwhile to look for ways to obtain a priori probabilities and estimate the degree of dependency so that the more powerful probabilistic aggregation techniques can be used.

At the moment I am inclined to take the conservative attitude because of the gaping holes in our knowledge of the estimation process. On the other hand, the desirability of filling these gaps with extensive empirical investigations seems evident.

References

1. John McCarthy, "Measures of the Value of Information," *Proc. Nat. Acad. of Sci.* 42 (September 15, 1956), pp. 654–55.
2. Thomas Brown, "Probabilistic Forecasts and Reproducing Scoring Systems," The Rand Corporation, RM-6299-ARPA, July 1970.
3. L. J. Savage, "Elicitation of Personal Probabilities and Expectations," *J. Amer. Stat. Assoc.* 66 (December 1971), pp. 783–801.
4. E. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.
5. Savage, *op. cit.*
6. S. E. Asch, "Effects of Group Pressure upon the Modification and Distortion of Judgments," in E. E. Maccoby, T. M. Newcomb, and E. L. Hartley (eds.), *Readings in Social Psychology*, Henry Holt, New York, 1958, pp. 174–83.
7. C. H. Coombs, "A Review of the Mathematical Psychology of Risk," presented at the Conference on Subjective Optimality, University of Michigan, Ann Arbor, August 1972.
8. M. Girshick, A. Kaplan, and A. Skogstad, "The Prediction of Social and Technological Events," *Public Opinion Quarterly*, Spring 1950, pp. 93–110.

9. G. A. S. Stael von Holstein, "Assessment and Evaluation of Subjective Probability Distributions," Economic Research Institute, Stockholm School of Economics, 1970.
10. Girshick, *et al., op. cit.*
11. W. Edwards, "The Theory of Decision Making," *Psychol. Bulletin* 5 (1954), pp. 380–417.
12. F. Knight, *Risk, Uncertainty and Profit*, Houghton Mifflin, Boston, 1921.
13. Hans Reichenbach, *The Theory of Probability*, University of California Press, Berkeley, 1949, Section 68.
14. N. Dalkey, "Experimental Study of Group Opinion," *Futures* 1 (September 1969), pp. 408–26.
15. N. Dalkey, B. Brown, and S. Cochran, "The Use of Self-Ratings to Improve Group Estimates," *Technological Forecasting* 1 (1970) pp. 283–92.
16. J. P. Guilford, *Psychometric Methods*, McGraw-Hill, New York, 1936, pp. 426ff.
17. N. Moiseev, "The Present State of Futures Research in the Soviet Union," in *Trends in Mathematical Modeling*, Nigel Hawkes (ed.), Springer-Verlag, Berlin, 1973.
18. N. Dalkey, "Experimental Study of Group Opinion," *op. cit.*
19. J. Aitchison and J. A. C. Brown, *The Lognormal Distribution*, University of Cambridge Press, Cambridge, Eng., 1957.
20. N. Dalkey, "Experimental Study of Group Opinion," *op. cit.*
21. T. Brown, "An Experiment in Probabilistic Forecasting," The Rand Corporation, R-944-ARPA, March 1973.
22. Peter A. Morris, *Bayesian Expert Resolution*, doctoral dissertation, Stanford University, Stanford, California, 1971.
23. R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philos. Trans. Roy. Soc.*, London, Series A, Vol. 222, 1922.
24. N. Dalkey, "An Impossibility Theorem for Group Probability Functions," The Rand Corporation, P-4862, June 1972.
25. K. J. Arrow, *Social Choice and Individual Values*, John Wiley and Sons, New York, 1951.

C. Experiments in Delphi Methodology*

M. SCHEIBE, M. SKUTSCH, and J. SCHOFER

Introduction

The emphasis in the Delphi literature to date has been on results rather than on methodology and evaluation of design features. The other articles in this chapter do address the latter aspects. Still, quite a number of issues remain unsolved, particularly those concerned with the details of the internal structure of the Delphi. For example, the way in which subjective evaluation is measured may affect the final output of the Delphi. A number of variables enter here. Ostrom and Upshaw [1] have noted that the range of the scale provided has a marked effect on judgment. Persons playing the role of judges who estimated themselves as "relatively harsh" assigned average "sentences" of four years to "criminals" when presented with a one-to-five-year scale, and twenty-one years when presented with a 1 to 25-year scale. The difficulties involved with the selection of a suitable scale range can be solved by the employment of an *abstract* scale rather than one representing, for example, hard dollars or years. An abstract scale allows relative measures to be made. Abstract scales are particularly suited to the measurement of values, as for example in the development of goal weights to represent relative priorities for goal attainment.

A number of psychological scaling techniques which result in abstract scales are available. This study reports on the comparison of several scaling techniques which were tested in the context of an experimental Goals Delphi.

Another issue is that of the effects of feedback input, which form the sole means of internal group communications in the Delphi process. It is important to the design of Goals Delphis to determine the nature and strength of the feedback influence. In the experiment reported below, the impact of feedback was identified by providing participants with modified feedback data. The resulting shifts of opinion were then used as measures of feedback effectiveness.

Methods for the measurement of consensus are also considered and a redefinition of the endpoint of a Delphi is offered. Instead of consensus, the *stability* of group opinion is measured. This allows much more information to be derived from the Delphi, and in particular, preserves opinion distributions that achieve a multimodal consensus.

*This study was supported by the Urban Systems Engineering Center, Northwestern University, NSF Grant GU-3851.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

A number of Delphi studies have used high/low self-ratings of participant confidence. Evidence of the value of such confidence ratings in improving the results of the Delphi is somewhat limited, except under certain conditions of group composition [4]. In this study, the use of high/low self-confidence ratings is again evaluated, and the influence of a number of other personal descriptive variables is tested.

Other design features include the application of short turn-around times using a computerized system for supporting inter-round analysis of the Delphi data. Although Turoff (Chapter V, C) has used a more complex, interactive computer system for this purpose, a simpler program is used here merely to accelerate accounting tasks.

Description of Procedure

The objective of this experimental Delphi study was the development and weighting of a hierarchy of goals and objectives for use in evaluating a number of hypothetical transportation facility alternatives. In the terminology suggested by Wachs and Schofer [2], *goals* are long-term horizonal aims derived directly from unwritten community values; *objectives* are specific, directional, measurable ends which relate directly to goals. Previous experiments by Rutherford, *et al.* [3] had indicated that Goals Delphis should be initiated by the development of objectives rather than goals, for the tendency toward upward drift in generality can be minimized if the Delphi participants are first asked to work at the more specific level. The development of goals, once the objectives have been defined, can be accomplished with much greater ease than can the reverse process.

The flow chart of Fig. 1 illustrates graphically the process by which the Goals Delphi and the design experiments were carried out. First, the initial list of objectives was generated. The process administrators presented the hypothetical transportation situation to the participants by means of a verbal description (Appendix I of this article), and a map (Appendix II of this article). The participants were then given a set of five blank 3 × 5 cards and asked to list no more than five objectives which they felt were applicable in the hypothetical situation. In all, seventy-seven objectives were submitted.

To derive goals from the list of objectives, and to eliminate the overlaps between them, a grouping procedure was followed. The process administrators first rejected those objectives that were exact duplicates of others, and assigned the remaining ones to sets. Each set represented objectives tending toward a common goal. Nine major goals were established, and these were "named" appropriately. Statements which were not strictly objectives were left out of the grouping process. The complete list of goals and objectives is given in Table 1. These goals were then returned to the participants for their evaluation. They were given the opportunity to add new objectives, and several were received and incorporated into the goal set.

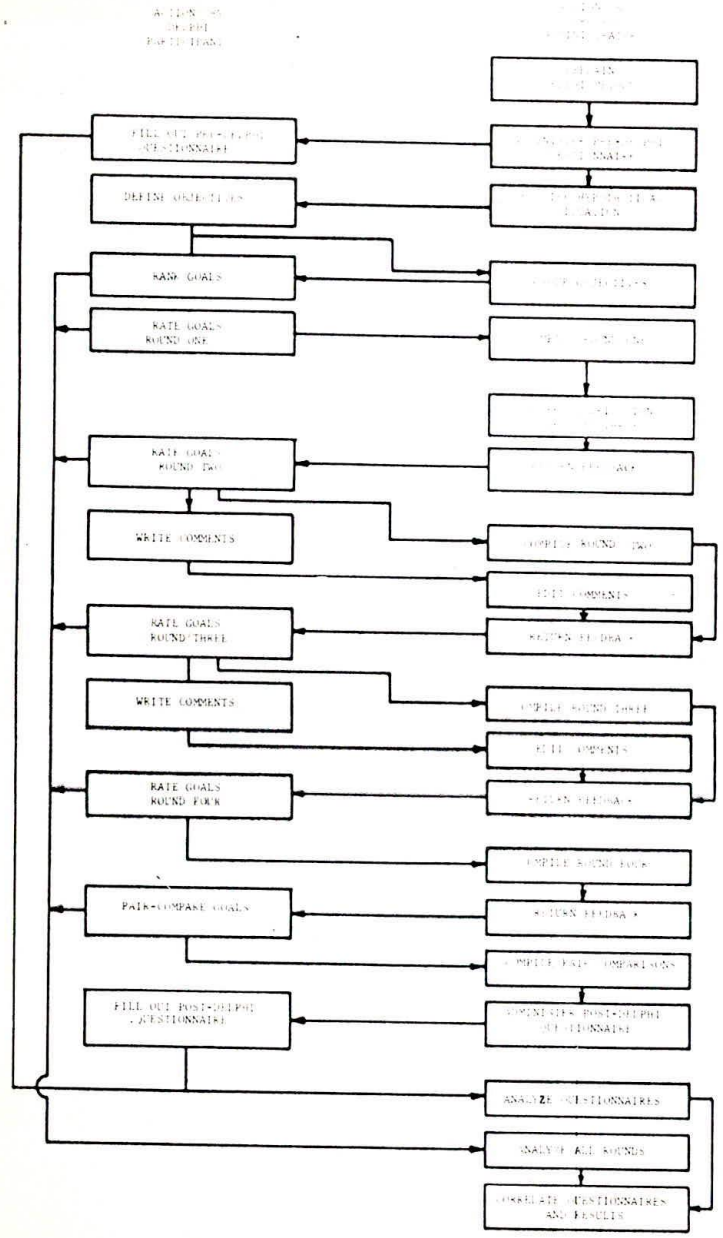


Fig. 1. The Goals Delphi experimental design.

Table 1: Goals and Objectives Suggested for the Hypothetical Transportation Scenario

GOAL I:

Minimize the adverse environmental impacts of the system.

Objectives:

Minimize air, noise, and water pollution.

Minimize negative illumination and vibration effects.

GOAL II:

Preserve the recreational and social environment.

Objectives:

Provide compensation for parkland removed.

Minimize the amount of parkland taken.

Minimize the demolition of historic buildings.

Minimize the amount of non-urban land required for the new facility.

Minimize the adverse social consequences of the transportation facility location by providing adequate compensation to families and businesses (and employees).

Minimize the amount of urban land required for the new facility.

Minimize the number of residences relocated.

Minimize the disruption of existing neighborhoods, people and businesses.

GOAL III:

Minimize operating and construction costs.

Objective:

Minimize operating and construction costs.

GOAL IV:

Maximize mobility and accessibility for the urban area residents.

Objectives:

Minimize travel time.

Minimize monetary travel cost.

Reduce congestion

Locate the facility to increase mobility and accessibility in the area.

Provide sufficient mobility for all members of society.

Table 1 (cont.)

GOAL V:

Design the facility to operate as safely as possible.

Objective:

Design the facility to operate as safely as possible.

GOAL VI:

Coordinate the transportation system with land use development.

Objectives:

Maximize the flexibility of the facility so as not to hinder possible further urban development in the area.

Encourage the desired land use development pattern as stated in the land use plan.

GOAL VII:

Design the facility so that its aesthetic appeal is in harmony with the environment.

Objectives:

Use transport facilities to highlight the character of the city to increase awareness, interest, and participation by the facility users.

Design the facility to be visually pleasing to the surrounding community.

GOAL VIII:

Maximize positive economic benefits for the entire area.

Objective:

Maximize positive economic benefits for the entire area.

GOAL IX:

Minimize the adverse impacts occurring during construction.

Objective:

Minimize the adverse impacts occurring during construction.

Following the development of the goals hierarchy, a decision was made (largely because of time constraints) to concentrate attention at the goal level. The objectives, therefore, were not included in the weighting procedure. Objectives, however, were at all times appended to the goals related to each, so that participants would always be aware of the specific meaning of each goal.

Participants were first asked to do a simple ranking of goals. As discussed elsewhere in this paper, one of the purposes of the Delphi was to compare different scaling methods. Participants were therefore asked to follow the ranking with a rating analysis. Nine-point Likert scales were used (0 = unimportant, 9 = very important). This type of scale was felt to be easily understood by the participants. In addition, when the ends are anchored adjectively, as in semantic differential scales, this scale is commonly found to have interval properties. Using the computer program developed for this study, the results of this first round were analyzed. The program was processed using a remote terminal; goal weights served as inputs, and histograms and various distributional statistics were produced as outputs. Frequency distributions of scores for each goal prepared by the computer were presented to the participants, along with the mean for each distribution.

Participants were asked to once again rate each goal on a nine-point scale, using the information from the previous round as feedback. In addition, those participants whose score on any goal was significantly distant from the group mean value for that goal were asked to write a few words explaining the reasons behind their positions. These statements were edited and returned in the next round. This procedure continued for a total of four rounds. The results are given in Fig. 2 and 3, which show the histograms produced in the first and the final rounds.

After the fourth weighting round, the participants were asked to perform a pair-comparison rating of all the goals. This was done to compare this scaling method with the nine-point rating scale and the ranking methods.

The initial development of the goals was accomplished during one two-hour class period. The rank ordering of the goals and the first three weighting rounds were conducted in a second two-hour period two days after the first. The fourth round took place an additional five days later, while the pair comparisons were made a week after the first weighting round.

An Experiment on the Effect of Feedback

There has been quite a bit written about the uses of feedback in the Delphi technique. Most of this, such as the work of Dalkey, Brown, and Cochran [4] at Rand, has concentrated on the effects of different types of feedback, such as written statements and various statistical measures. The effects of this feedback, particularly in the almanac-type Delphis, have been measured by comparing

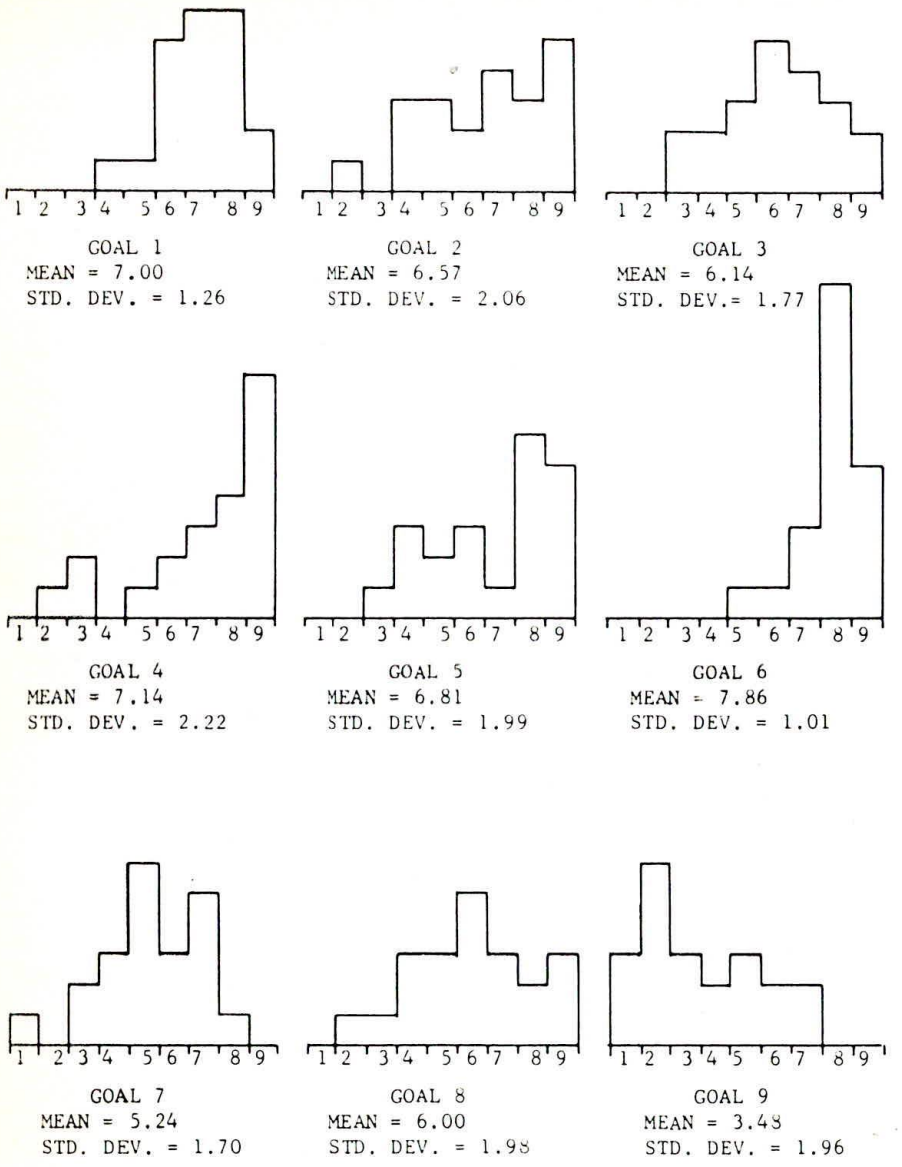


Fig. 2. Results of round one rating analysis.

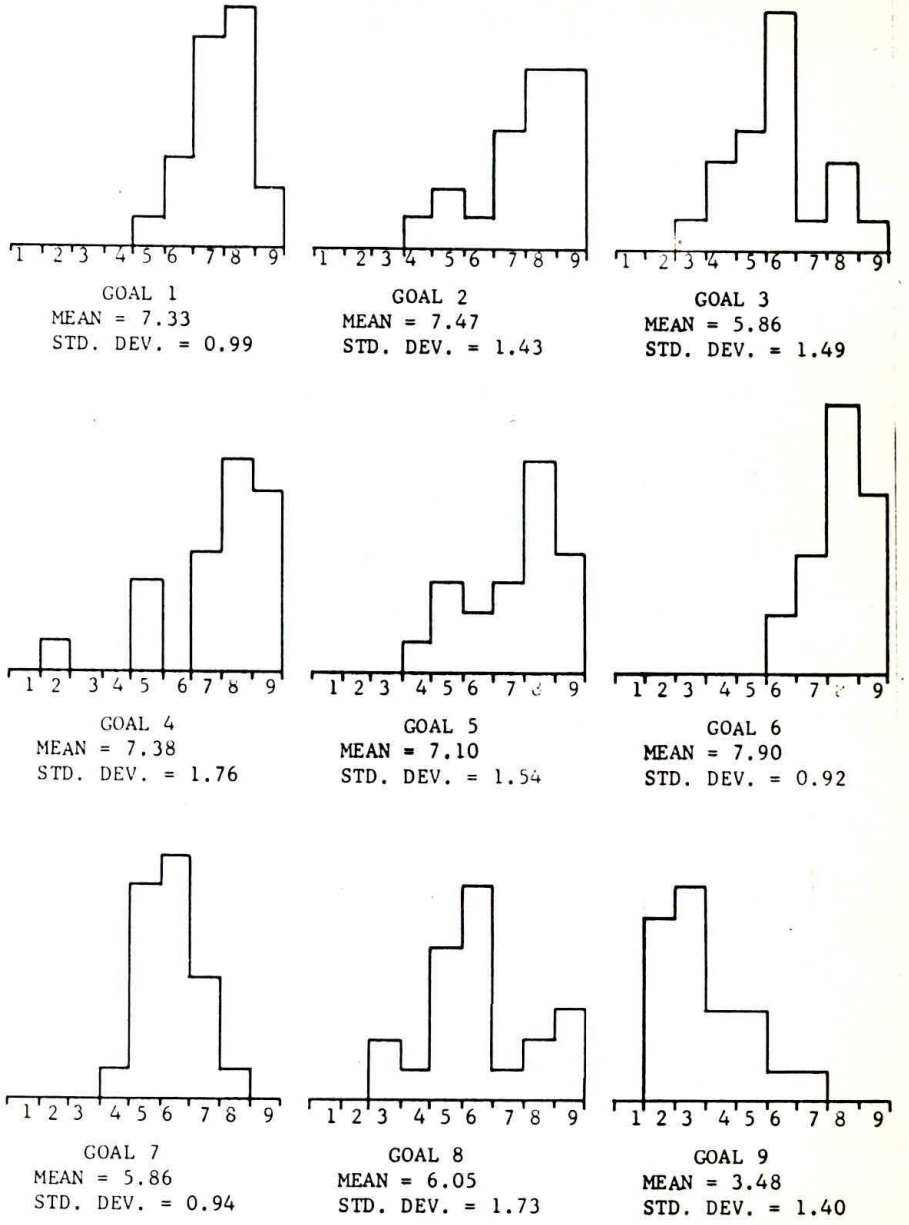


Fig. 3. Results of round four rating analysis.

the accuracy of the opinions of a group given a certain feedback with that of a group given different feedback or no feedback at all. In Policy and Value Delphis the effect of feedback is evaluated by measuring the degree of consensus which is reached and the speed with which it is reached.

There seems to be very little in the literature, however, which examines the round-by-round effect of feedback or investigates the manner in which the feedback affects the distribution of scores in a particular round. In this study, it was decided to investigate these aspects of feedback, since the kind and amount of feedback used in the Delphi may be an important variable in its results. A greater understanding of the impacts of feedback might lead to better Delphi design. The method employed was to provide participants with false feedback data, and then to observe the effect of this on the distribution of priority-weight scores.

Two types of feedback were used in this study. The first was a graphical representation of the distribution of scores together with a listing of the mean of the distribution. In addition, in later rounds edited, anonymous comments by the participants concerning the importance of the various goals were distributed. During the experiment on feedback, one goal was chosen and the distribution was altered by the administrators so as to change markedly the position of the mean. Since this was done after the first weighting round, no written feedback accompanied the altered distribution. The goal chosen for this test was Number 3 ("Minimize the operating and construction costs"). This goal was chosen because it appeared to have a good consensus after the first iteration. In addition, it was judged to be substantively important. It was felt that most participants would be very surprised by the altered distribution.

The second-round distribution showed that the feedback had had an effect, since a number of persons shifted their positions away from the true mean. By the third round, the distribution was once again similar to what it had been in the first round, although the distribution was shifted slightly to the lower end of the scale and remained that way permanently, showing residual effects of the gerrymandering. Figure 4 shows the actual distributions and the altered feedback used.

In attempting to explain the reasons behind these changes, the following hypothesis is offered. Upon seeing the first round of feedback information, the respondents had three options: they could ignore the feedback and keep their votes constant; they could rebel against the feedback and move their votes to the right, in the interest of moving the group mean closer to their true desire; or they could acknowledge the feedback and move their votes nearer the false mean. If they had followed either of the first two options, it would indicate that the feedback was not effective in changing individual attitudes. That the third option was in fact taken, however, indicates that the feedback did have an effect on the participants.

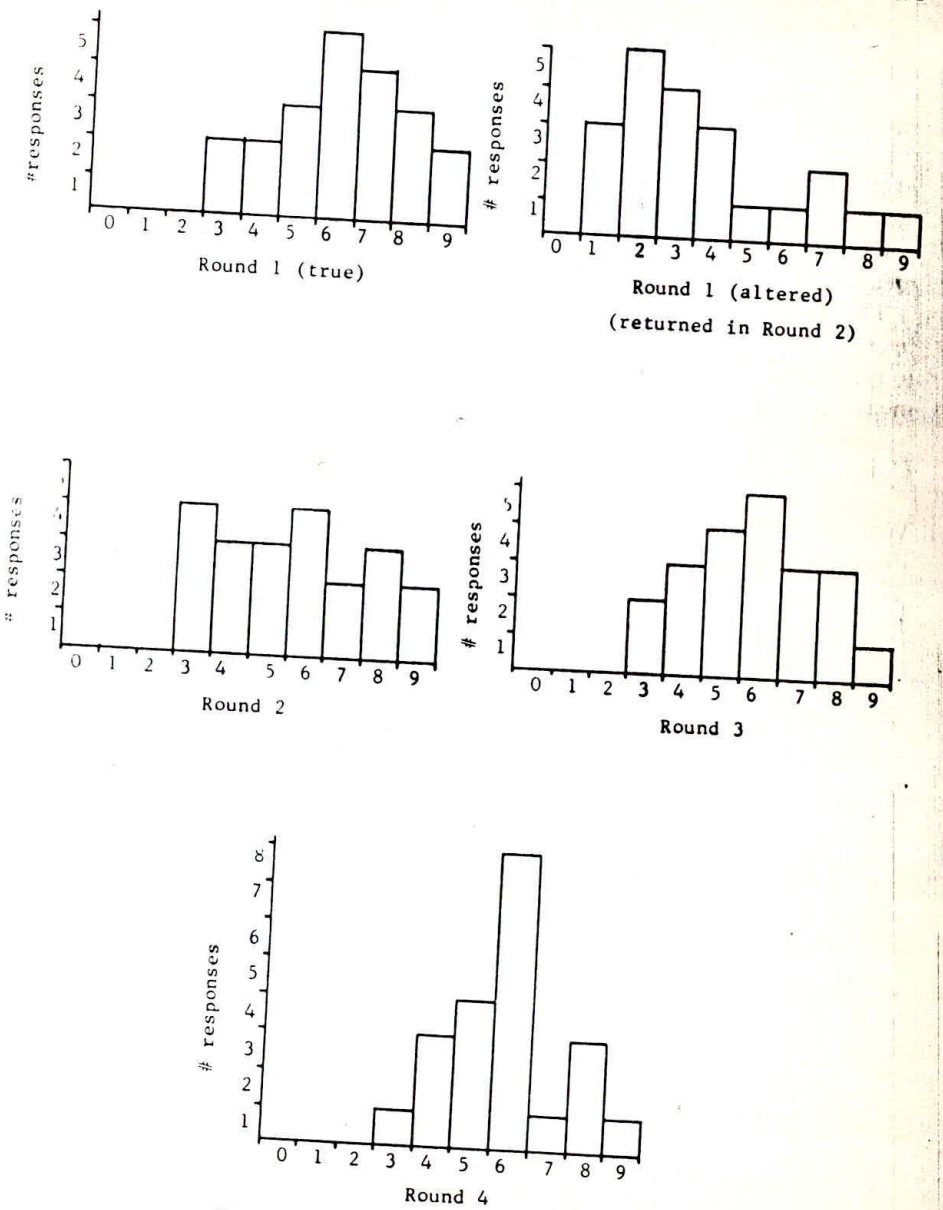


Fig. 4. Distributions generated for Goal 3.

The third round, as a result of the feedback of the second round, also shows the effect of feedback. The second-round distribution showed that participants were attempting to increase the priority for Goal 3, although with respect to their true initial opinions, they were actually decreasing this priority. It seems likely that many respondents, upon seeing this, felt that the group was moving closer to their original position and they decided to return to their first-round vote, since it no longer appeared that this position would be far distant from the mean value of the group.

This experiment suggests that the respondents are, in fact, sensitive to the feedback of distributions of scores from the group as a whole. These results seem to indicate that most repondents are both interested in the opinions of the other members of the group and desirous of moving closer to the perceived consensus.

Comparison of Scaling Techniques

With the exception of Dalkey and Rourke [5], there is little discussion in the literature of the different methods of scaling which could be used in a Delphi. The two most common methods which are used are simple ranking and a Likert-type rating scale. Even when these methods are used, there have seldom been attempts to ensure that the scales developed are, in fact, interval scales.

The necessity of having an interval scale is seldom emphasized in Delphis. There is the suspicion that on some occasions the scales derived are ordinal scales. An ordinal scale merely shows the rank order of terms on the scale, and no statement can be made concerning the distance between quantities.

Presumably, the primary reasons for using a Delphi, especially when comparing policies or measuring values, include the determination of not only which policies are considered most important, but also the degree to which each policy is preferred over the other possibilities. In order to assure that this can be determined, an interval scale must be obtained.

In this study, three methods which usually yield interval scales were tested. These methods were simple ranking, a rating-scale method, and pair comparisons. The purpose in trying three scales was to determine if all three methods yielded approximately equivalent interval scales. If this is found to be the case, then in future designs any one of these scales could be used. In this situation, it would probably be wisest to choose that scaling method which was considered easiest to perform by the participants in the Delphi. In this study it was found that the rating-scale method was considered by the participants as the most comfortable to perform. The limitation of the pair-comparison method is that it is time consuming. For example, to apply this method to a set of ten objectives, each participant must make forty-five judgments. The ranking method is fairly easy for a small number of goals, but becomes increasingly difficult as the number of goals increases, for it essentially requires the participant to order the entire list of items in his mind. In addition, many participants felt uncomfortable performing this method because they were prevented from giving two goals an equal ranking (i.e., forced ranking). While this dilemma might possibly have encouraged more thought concerning underlying priorities, it was felt that the frustration caused had a negative effect on the end result. The rating-scale method was found to be quick, easy to comprehend, and psychologically comforting. The participant's task is easy, since he must rate only one item at a time. The problem that remained was to determine whether such a scaling procedure would yield a scale with interval properties.

In this experiment it was found that each of the three methods yielded somewhat different scales. Using the Law of Comparative Judgment [6], scale values for each goal for each round were derived. These values were then translated onto a scale from one to nine. Graphical representations of these scales are given in Fig. 5. Because of the presence of feedback, the four rating rounds are not independent. Each one depends on those previous to it. The scales derived in each successive round should not be identical, for if the scale remains constant from round to round, the justification for using an iterative approach is lost. In addition, because of the order in which the scales were developed, the ranking scale can only be compared with the first-round rating scale and the pair-comparison scale can only be compared with the fourth-round rating scale. Because of four rounds of feedback between them, the ranking scale and pair-comparison scale should be compared only cautiously, and should not be expected to be identical

The interscale comparison shown in Fig. 5 is not especially encouraging. The pair-comparison method is known to produce interval scales, and the similarity in results of this approach and the round-four rating results is not strong. The scales produced by ranking and round-one rating are, however, not too different from each other. It is possible to interpret the progression of rating scales from round one to round four as a movement in the direction of the pair-comparison scale. This experiment did not pursue further weighting rounds, but, as discussed below, major changes in weights beyond round four do not seem likely. In addition, later pair-comparison responses might differ from that shown in Fig. 5. Given the complexity of the pair-comparison method for participants, however, it may not be unreasonable to accept cautiously the results of simple rating methods as fair approximations to an interval scale.

The Effect of Personal Variables on Participant Behavior

Dalkey, Brown, and others have considered and used the confidence of participants in their responses to reach more accurate estimates of quantitative phenomena in Delphi exercises. Working with almanac-type data not available to the participants, they found that by selecting for inclusion in the feedback only those responses considered "highly confident" by their proponents, a

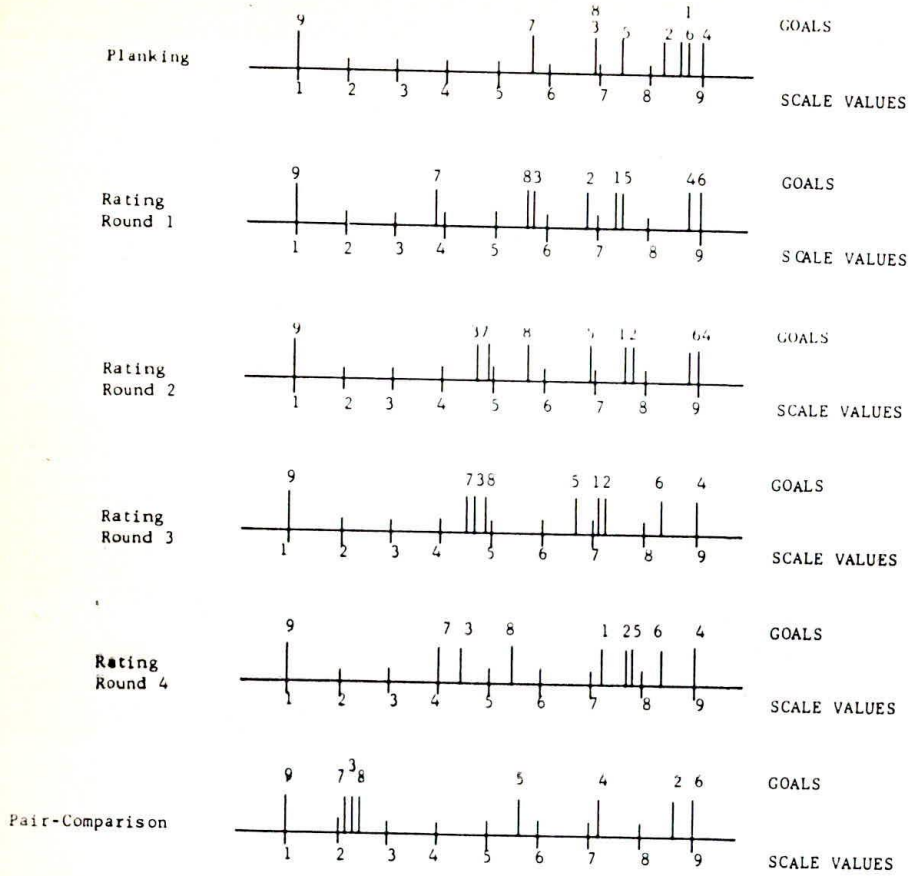


Fig. 5. Comparisons of goal weights on various categorical judgment scales.

slightly superior result was achieved [7]. Later, they found that in situations in which relative confidence was measured and in which the "highly confident" group was reasonably large, a definitely superior result could be expected [4].

Studies in the psychology of small groups, however, indicate that highly confident persons should be less influenced by group pressure than those with less confidence, and therefore it would be expected that highly confident individuals move less toward consensus than do others in the Delphi context. Later, Dalkey *et al.* [4] showed that "over consensus" may occur, and the ratio of average error to standard deviation may actually increase, if consensus is forced too quickly. In order to reach some greater understanding of theory and

observation, therefore, several simple hypotheses were tested. It was felt that confidence might involve more than simply confidence in individual answers, and therefore a selection of variables representing various aspects of personal confidence were sought, as well as high/low confidence in each response. Appendix III of this article shows the questionnaire issued to all participants before the Delphi. The variables measured were as follows:

Pre-Delphi Survey Question Number	Independent Variable
1	A. Perceived academic standing relative to other participants
2	B. At-oneness with other participants
3	C. Familiarity with other participants
4	D. Confidence of own capacity in the subject area of the Delphi
5 and 6	E. Experience in the field
7	F. Expectation of the Delphi process (confidence in Delphi)

Each of these confidence variables was then correlated against dependent variables describing the amount of movement actually made by each participant toward the center of the distribution. It was, of course, not possible in this value-judgment Delphi to test accuracy as well, as was done in Dalkey's experiments.

The dependent variables were summed for every participant over each of his nine responses, and were as follows:

(a) Total amount of change from round one to round four
(b) Total number of monotonic (as opposed to oscillating) changes made
(c) Amount of change made in second round
(d) Number of responses exactly on the mode in round three
(e) Number of responses within three places of the mode in round three

It was hypothesized that high confidence would be associated with small amounts of total change, monotonic rather than oscillating change, and low confidence with a high degree of change in round two and a high conformity to the consensus in round three.

The response with regard to confidence in individual answers (measured by high/low ratings of confidence on the Delphi answer sheets) was correlated significantly, negatively, but not very highly with the amount of change in the second round, although not with the total change between rounds one and four.

Percent of "highly confident" answers, however, was cross-correlated positively with perceived academic status, although this was not significantly connected with either movement variable. Amount of change in the second round was also just correlated (positively) with the "at-oneness" variable, although there was no relationship at all between "at-oneness" and percent highly confident. These represented the only significant correlations found.

The evidence for the effect of confidence on the tendency to converge is somewhat sketchy. The only conclusions that can be drawn from the experience is that the initial surprise on being confronted with some distribution of group opinion may to some extent cause the less confident members who believe that they associate with the rest of the group to move toward the center of opinion, but that this tendency is certainly not an overwhelming one.

At the end of the Delphi a second questionnaire (Appendix IV of this article) was used to determine whether the kind of feedback provided had any conscious effect on movement in the Delphi. The variables were as follows:

Post-Delphi Survey Question Number	Independent Variable
1	G. Satisfaction with the results
2,3 and 4	H. Success of feedback in the learning process
5 and 6	I. Frustration with communication levels
7	J. Optimism for Delphi
8	K. Feeling of being rushed

These variables were correlated with the same dependent variables. Both optimism for the future of Delphi and satisfaction with the process correlated significantly and positively with the number of monotonic changes made, perhaps indicating that those people who were not caused to change their opinion radically were in better spirits after the Delphi than the others. However, the success-of-feedback variable was strongly and negatively correlated with the propensity to conform to the mode in round three. In other words, those who did conform to the visible majority had difficulty in giving and taking ideas from the feedback. This is interesting in that it indicates the different kinds of feedback that may affect people in different ways. The tendency to converge strongly has elsewhere been shown by Schofer and Skutsch [8] to be due to emphasis in the visible consensus and on the need to create consensus. Satisfaction with the process was also negatively correlated with the conformity variable. (Satisfaction and agreement with feedback were

strongly cross-correlated.) Clearly, the people who were strongly conforming were not happy with the Delphi at all. The question of what is cause and what is effect, however, remains to be answered. Yet one might speculate that, especially in a value-oriented Delphi, the group pressure from some forms of feedback can be overly strong, forcing participants to take positions which they find uncomfortable. While compromise may be uncomfortable in any situation, the real danger here is that participants may leave the process without really compromising their feelings at all. That is, perhaps the anonymity of the Delphi itself may have encouraged participants to capitulate, but only on paper. They may later hold to their original views, and, if the results of the Goals Delphi are used to develop programs to meet their needs, participants might ultimately be quite dissatisfied with the results.

A cautionary note is relevant at this point. Another study by Skutsch [9] has shown that the form of the feedback itself influences consensus development. Despite the fact that participants in this experiment were encouraged to report their verbal rationale for their positions, the rapidity with which the process was carried out tended to discourage such responses. As a result, histograms of value weights formed the bulk of the feedback. It is just this kind of limited, "hard" feedback which tends to force what might be an irrational consensus, one which might be only temporary.

Opinion Stability as a Method of Consensus Measurement

In most Delphis, consensus is assumed to have been achieved when a certain percentage of the votes fall within a prescribed range—for example, when the interquartile range is no larger than two units on a ten-unit scale. Measures of this sort do not take full advantage of the information available in the distributions. For example, a bimodal distribution may occur which will not be registered as a consensus, but indicates an important and apparently insoluble cleft of opinion. Less dramatically, the distribution may flatten out and not reach any strongly peaked shape at all. The results of the Delphi are no less important for this, however. Indeed, considering that there is a strong natural tendency in the Delphi for opinion to centralize, resistance in the form of unconsensual distributions should be viewed with special interest.

A measure which takes into account such variations from the norm is one that measures not consensus as such, but *stability* of the respondents' vote distribution curve over successive rounds of the Delphi. Because the interest lies in the opinion of the group rather than in that of individuals, this method is preferable to one that would measure the amount of change in each individual's vote between rounds.

To compare the distributions of opinion between rounds, the histograms may be subtracted columnwise and the absolute value of the result taken. In Table 2

this approach is applied to the weight histograms reported for Goal 5 in the first three rounds of the Delphi. Columnwise subtraction between the first and second, and the second and third, rounds gives the results shown in Table 2. The absolute values of the differences between histograms are aggregated to form total units of change; but since any one participant's change of opinion is reflected in the histogram differences by two units of change, net person-changes must be computed by dividing total units of change by two. Finally, the percentage change is determined by dividing net changes by the number of participants. Clearly, in the example shown in Table 2, the distribution of value weights for Goal 5 became more stable between rounds one and three.

The question of what represents a reasonable cut-off point at which the response may be said to be unchanged, and therefore finally in its stable position, poses some problems, however. Since there is no underlying statistical theory in what has so far been proposed, no true statistical level may be set, as might, for example, be possible with a statistical change in variance test.[1]

Empirical examination of the responses in the Delphi, however, showed that at any point in time a certain amount of oscillatory movement and change within the group is inevitable. This might be conceptualized as a sort of underlying error function, a type of internal system noise. What is needed is a "confidence" measure which allows the distinction to be drawn between this kind of movement and strong group movements that represent real changing opinion. Such an estimate has tentatively been made from studies of observed probability of movement.

Leaving aside objective 3 (for reasons made apparent earlier), the propensity of the individual to alter his score as a function of distance from the center point was measured. This was done by calculating the proportion of respondents at each scale distance from the mode that moved toward the mode between rounds. The results, displayed in Fig. 6, show a strong tendency for increased amounts of movement with distance from the center point. They also show that a percentage change is to be expected among respondents who are already dead on the mode itself. The amount of movement at the mode (about 15%) has therefore been taken to represent the base of oscillatory movement to be expected, and this is supported by the fact that the amount of change at the centroid does not alter appreciably between rounds.

Using the 15% change level to represent a state of equilibrium, any two distributions that show marginal changes of less than 15% may be said to have reached stability; any successive distributions with more than 15% change should be included in later rounds of the Delphi, since they have not come to the equilibrium position.

[1] Conventional variance tests were found to be unsuited to the case of change in histogram shape in this context. Most rely on independent samples; none is strong enough to pick up small changes of shape, and none robust enough to deal with non-normal distributions.

Table 2: Example of Stability Measurement Computations for Goal 5

Rating	1	2	3	4	5	6	7	8	9
Absolute difference in number selecting rating, rounds 1-2 (a)	0	0	0	2	2	2	4	0	2
Total units of change (b)	12								
Net person-changes (c)	6								
Number of participants	21								
Percent change (d)	28.6%								

Rating	1	2	3	4	5	6	7	8	9
Absolute difference in number selecting rating, rounds 2-3 (a)	0	0	1	0	1		0	0	1
Total units of change (b)	4								
Net person-changes (c)	2								
Number of participants	21								
Percent change (d)	9.5%								

(a) These numbers are the absolute differences in the histograms for the two successive rounds.

(b) These numbers are the sums of the absolute differences in the histograms.

(c) Net changes are total units of change divided by 2.

(d) Percent change is net change divided by the number of participants.

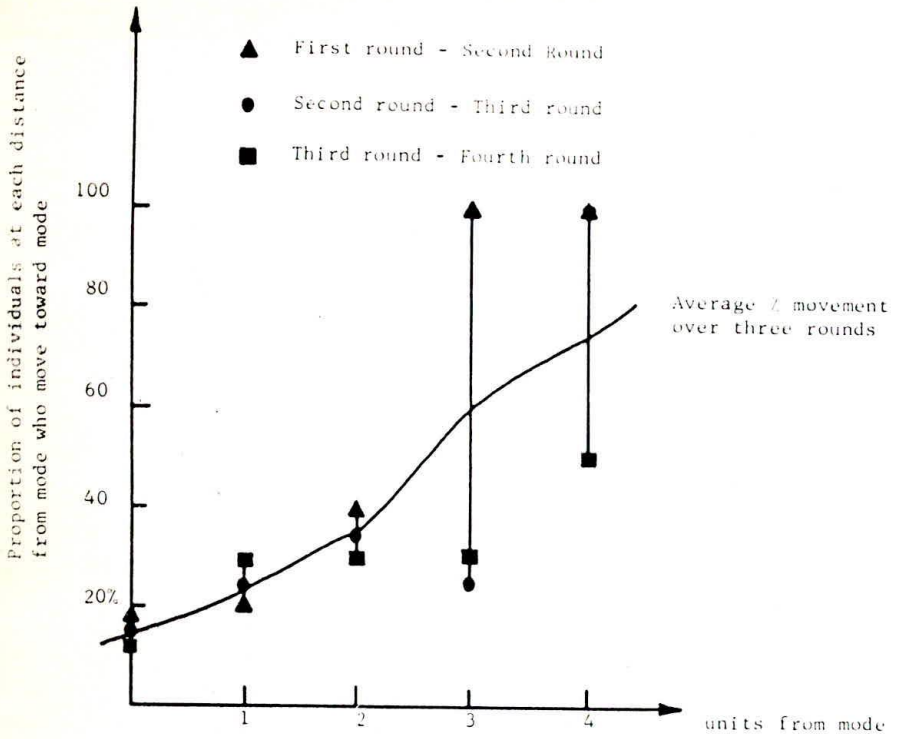


Fig. 6. Proportion of respondents shifting their positions, as a function of distance from the mode.

The results for all nine goals included in this experimental Delphi using this analysis are shown in Table 3. From these data, there can be no doubt as to the general tendency toward stabilization. Only one goal, 7, had not reached a stable position by the end of the third round, although 3, 8, and 9 were all only just stable.

In general, this method seems to have a number of advantages. Firstly, it allows the use of more of the information contained in the distributions. There are applications in which, at the end of the Delphi process, the entire distribution may be used, as for example in linear-weighting evaluation models where goal-weight distributions are treated stochastically, such as that by Goodman [10]. In addition, this stability measure is relatively simple to calculate, and has much greater power and validity than parametric tests of variance.

Perhaps most important, one of the original objectives of Delphi was the identification of areas of difference as well as areas of agreement within the

Table 3: Results of Stability Analysis

Goal	Amount change Rounds 1-2	Amount change Rounds 2-3
1	13% stable	10%
2	17%	10% stable
3	38%	14% just stable
4	10% stable	10%
5	26%	10% stable
6	2% stable	2%
7	24%	33% increasingly unstable
8	22%	14% just stable
9	24%	14% just stable

participating group. Use of this stability measure to develop a stopping criterion preserves any well-defined disagreements which may exist. To the organizer of a Goals Delphi, this information can be especially useful.

Delphi Service Program

In order to make several iterations possible in the space of a very short time period, a computer time-sharing terminal was used to process the results of this Delphi experiment. Unlike the systems described by Price (see Chapter VII, B), the program used in this Delphi was an accounting device only; verbal feedback was compiled and read to participants by the organizers.

In this application, histograms produced by the computer terminal were copied by hand onto an overhead projector transparency to provide immediate feedback to participants, who themselves determined their positions in the distributions relative to the group. It is anticipated that, for future experiments, computer-generated histograms will be produced in multiple copies, one of which will be provided to each participant.

This type of computer support, oriented toward the use of a single terminal for all participants, may be especially desirable for Goals Delphi applications, where, because of the lay nature of the respondents, it seems especially desirable to keep all of those involved in a single room, and to maintain a relatively high rate of progress throughout the survey.

Closure

The potential applicability of the Delphi method to goal formulation and priority determination for public systems is very great. Yet, because the detailed characteristics of the design of the process can have important effects on the nature of the outcomes, it will be important to tailor the Goals Delphi to the problems at hand. The structuring of internal characteristics which are appropriate to a Goals Delphi should be based on a rather complete understanding of the linkages between form and function in the Delphi environment. While considerable experience must be gained before Delphi can be offered as a routine goal-formulation process, this discussion has suggested some structural and process features relevant to this important application of the Delphi method.

Appendix I: Hypothetical Decision Scenario

The following transportation-facility-location problem is offered as an appropriate context for developing local-scale transportation planning objectives. Within this context, there is a need to establish an objective set, and to evaluate *quantitatively* several alternative plans in the context of the objectives.

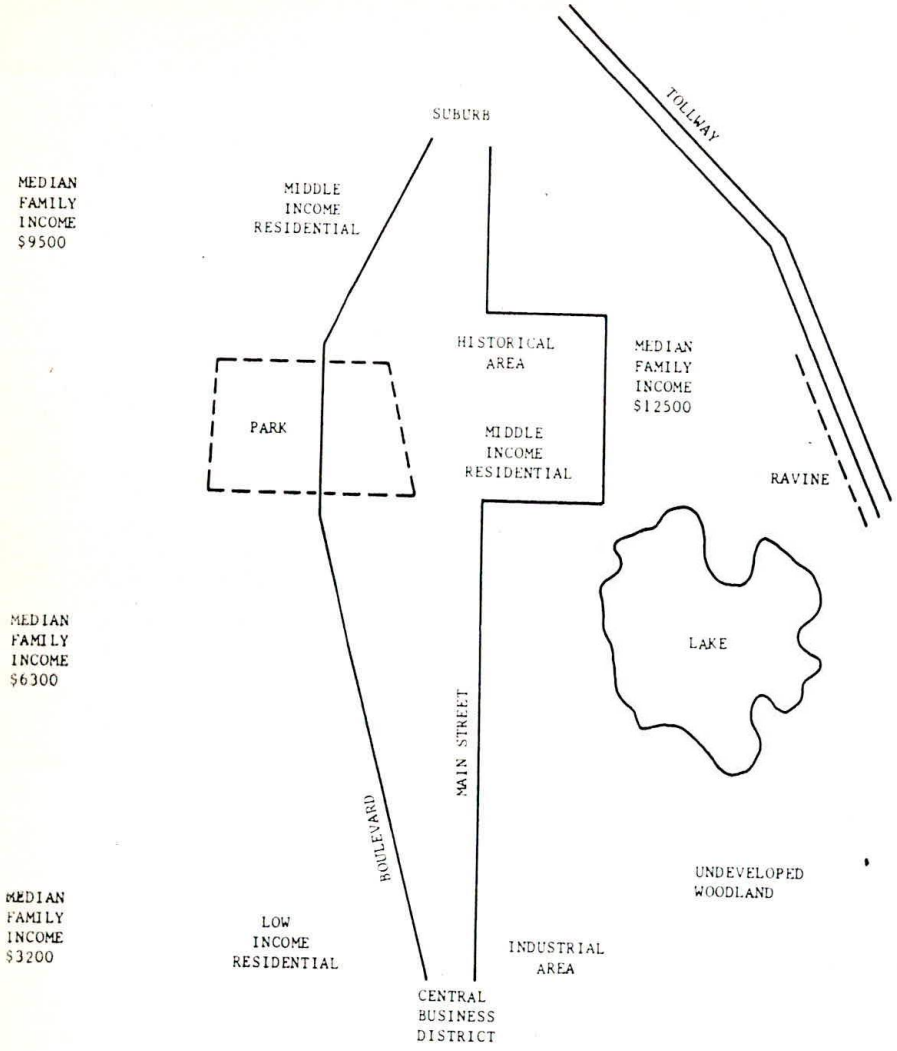
A two-mile transportation link is proposed in an urban area. It is to run from the Central Business District (CBD) to new, developing suburbs to the north. This area is presently served by a four-lane boulevard with an average daily traffic (ADT) of forty thousand vehicles, and by a four-lane street with an ADT of twenty thousand vehicles. This street, however, circles an historical area by means of four 90° turns, and traffic must travel this section at twenty-five mph. The southeast corner of the historical area comes within five hundred yards of the edge of a lake, and the main street presently is only one block from the lake at this point. A tollway also passes the suburb and proceeds in a southeasterly direction. The tollway passes within one mile of the CBD, with its alignment located in a ravine. The elevation of the ravine is such that to build a connector to the tollway from the CBD would require a great deal of earthwork, and even with this the grade would be about 3%.

The alignment of the boulevard is such that it begins in the CBD, proceeding northwesterly through a low-income area to a large park, where it turns and continues in a northeasterly direction through a middle-income residential area to the suburbs.

The four-lane street heads due north from the CBD, passes through an industrial park, and then makes four sharp turns around the historical area and proceeds directly into the suburbs.

Citizen opposition can be expected in four areas. Public opinion has long been against any changes in the historical area. A citizen group can be expected to form opposing an alignment through the park. One can also be expected to form opposing removal of houses in the middle-income areas north and east of the park. Problems can also be expected if the alignment goes through the low-income area, requiring relocation of some households.

Appendix II: Map of Hypothetical Transportation Scenario



Appendix III: Pre-Delphi Self-Rating Form

Code Number_____

PRE-DELPHI SURVEY

1.	As a transportation planner, in a class of transportation planners, my skills in planning would put me about here, relative to the others.	Very Highly Skilled	No Skill at all
		1 2 **3** 4 5 6 7	
2.	I think my ideas are, in essence, in agreement with the rest of the class.	Yes, absolutely	No, not at all
		1 2 **3** 4 5 6 7	
3.	I know most of the people in the class very well.	Yes, pretty much	No, none at all
		1 2 **3** 4 5 6 7	
4.	I have some definite ideas about what the goals in transportation planning are and should be.	Yes, lots	No, none
		1 2 **3** 4 5 6 7	
5.	I have been in transportation for longer than most of the other people here.	Yes	No
		1 2 **3** 4 5 6 7	
6.	I have a lot of experience in planning outside of school.	Yes	No
		1 2 **3** 4 5 6 7	
7.	I am anticipating that the Delphi is going to be a good thing for goal setting.	Yes, I think it will be	No, I think it may be a waste of time
		1 2 **3** 4 5 6 7	

Appendix IV: Post-Delphi Evaluation Form

Code Number_____

POST-DELPHI SURVEY

1.	I feel satisfied with with the results in general.	1 2 3 4 5 6 7	I'm not really happy with the results at all.
2.	I learned ideas from the feedback.	1 2 3 4 5 6 7	I didn't learn a thing from the feedback.
3.	In general, I agreed with the ideas in the feedback.	1 2 3 4 5 6 7	I disagreed with everything in the feedback.
4.	I could express my ideas OK this way.	1 2 3 4 5 6 7	I couldn't really write what I wanted to say.
5.	I feel as if I really wanted to talk to people.	1 2 3 4 5 6 7	I didn't feel the need to talk at all.
6.	I have a feeling people didn't understand or think about my reasons.	1 2 3 4 5 6 7	I think people understood my reasons pretty well.
7.	I think the Delphi could be operational in goal setting more generally.	1 2 3 4 5 6 7	I don't think it could be operational, really.
8.	I think it went too fast.	1 2 3 4 5 6 7	I think it went too slowly.

References

1. T. H. Ostrom and H. S. Upshaw, "Psychological Perspective and Attitude Change," in A. G. Greenwald *et al.*, *Psychological Foundations of Attitudes*, Academic Press, New York, 1968.
2. M. Wachs and J. L. Schofer, "Abstract Values and Concrete Highways," *Traffic Quarterly*, **23**, No. 1 (1969).
3. G. S. Rutherford, J. L. Schofer, M. Wachs, and M. Skutsch, "Goal Formulation for Socio-Technical Systems," *Journal of the Urban Planning and Development Division, Proceedings of the American Society of Civil Engineers*, forthcoming.
4. N. C. Dalkey, B. Brown, and S. Cochran, "The Delphi Method IV: Effect of Percentile Feedback and Feed-in of Relevant Facts," Rand Corporation, RM-6118-PR, 1970.
5. N. S. Dalkey and D. L. Rourke, "Experimental Assessment of Delphi Procedures with Group Value Judgment," Rand Corporation, R612-ARPA, 1971.
6. W. S. Torgerson, *Theory and Methods of Scaling*, Wiley, New York, 1958.
7. N. C. Dalkey, "The Delphi Method: An Experimental Study of Group Opinion," Rand Corporation, RM-5888-PR, 1969.
8. M. Skutsch and J. L. Schofer, "Goals Delphis for Urban Planning: Concepts in their Designs," *Socio-Economic Planning Sciences*, **7**, pp. 305-313, 1973.
9. M. Skutsch, "Goals and Goal Setting: A Delphi Approach," unpublished master's thesis, Department of Industrial Engineering and Management Science, Northwestern University, 1972.
10. K. M. Goodman, "A Computerized Investigation and Sensitivity Analysis of a Linear Weighting Scheme for the Evaluation of Alternatvies for Complex Urban Systems," unpublished master's thesis, Department of Civil Engineering, Northwestern University, 1971.

D. Propensity to Change Responses in a Delphi Round as a Function of Dogmatism

NORMAN W. MULGRAVE and ALEX J. DUCANIS

Since one of the assumptions of Delphi is that it "reduces the influence of certain psychological factors such as ... the unwillingness to abandon publicly expressed opinions, and the bandwagon effect of majority opinion" (Helmer, 1966), it would seem of interest to examine the effect of personality upon an individual's performance during several Delphi rounds. Specifically, the question can be raised concerning the willingness of a more dogmatic individual to change his answer in a Delphi round (whether he is an expert or a nonexpert). Since dogmatic thinking is characterized by resistance to change (Rokeach, 1960), it might be posited that the dogmatic individual would be less likely to change his position when confronted by the opinions of others. It might be further presumed that the type of question asked, i.e., those upon which the individual could be considered either more or less expert, might also affect performance of highly dogmatic as opposed to less dogmatic individuals. It was therefore predicted that the number of changes made by a high dogmatism group (D_H) would be less than a low dogmatism group (D_L), and that the (D_H) group would change less on questions on which they might be considered expert than on questions on which they would be considered less expert.

Method

The subjects for the study were ninety-eight graduate students enrolled in a class in Educational Psychology, most of whom were school teachers.

Berger's (1967) revision of Rokeach's Dogmatism Scale (the FCD Scale) was administered on the first day of class. Subsequently the class was used as a Delphi panel and asked to make certain estimates. Four types of questions were utilized. Ten questions defined the subjects as nonexpert, such as the number of farms in the United States. Ten other questions concerning class size, teachers' salaries, length of the school year, and similar items defined the subjects as experts.

Eighteen other questions were value-oriented items. Subjects were to respond in terms of what certain values in the United States *will* be in 1980, and also

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

what they *should* be. The latter set made each person a fully qualified "expert."

With these question categories as a base, it was possible to use the questions to define the respondents as expert or nonexpert.

The Delphi procedure was continued through three rounds. During rounds two and three each subject was given the group median, interquartile range and his own response to the previous round for each item. He was asked to "review [his] projection on the basis of the information provided" and to change his answer if he wished to do so.

Results

The D_H and D_L groups were identified as those scoring in the upper or lower 27 percent of the class on the FCD Scale (Berger, 1967). Second- and third-round changes were tabulated for those who were both inside and outside the interquartile range for each of the four sets of questions. The results of that tabulation are shown in Tables 1 and 2.

Table 1
Proportion of Individuals Changing Answers Round Two

Type of Item	Percent Changing High Dogmatism		Percent Changing Low Dogmatism	
	Inside Range	Outside Range	Inside Range	Outside Range
Almanac	29.5	76.2	13.6	63.1
"School" Questions	20.3	64.7	6.1	54.0
Predictions of Values	13.4	74.8	4.5	50.8
What Values Should Be	12.8	71.9	7.2	59.3

Table 2
Proportion of Individuals Changing Answers Round Three

Type of Item	Percent Changing High Dogmatism		Percent Changing Low Dogmatism	
	Inside Range	Outside Range	Inside Range	Outside Range
Almanac	6.9	29.2	6.8	14.5
"School" Questions	4.1	18.6	3.4	14.8
Predictions of Values	8.9	19.7	1.9	6.7
What Values Should Be	9.2	15.8	1.9	6.0

A significant difference was found between the groups in the number of times they changed their answers. For round two the value for chi-square was computed as 18.48 with 7 degrees of freedom. This value is significant at the .01 level. The corresponding value for the third round was 14.78, which is significant at the .05 level.

Discussion

It would seem that personality characteristics of the individual involved in the Delphi panel have some effect upon his propensity to change. Of interest as well was the finding that the High Dogmatism group exhibits significantly more changes. Thus the prediction that they would be less likely to change is not upheld. A possible explanation for this may be that if the dogmatic individual looks to authority for his support, then in the absence of any clearly defined authority the dogmatic individual would tend to seek the support of whatever authority seems present. In this case authority would be the median of the group response.

The second prediction that the High Dogmatism group would change less on questions where they could be considered expert, i.e., "school questions" and "what values should be" than on questions where they could not be considered expert, i.e., "almanac questions" and "what values will be," was upheld on the second round (chi-square 6.622 with one degree of freedom) but not on the final round. There were no significant differences on either round for the Low Dogmatism group.

These results seem to indicate that the High Dogmatism group is less likely to change an answer to a question on which they consider themselves expert than one on which they consider themselves less expert, but that in the presence of some "perceived" authority such as the group median, High Dogmatism groups will exhibit more change than Low Dogmatism groups.

References

V. F. Berger, "Effects of Repression and Acquiescence Response Set on Scales of Authoritarianism and Dogmatism," unpublished master's thesis, University of Oklahoma, 1967.

N. C. Dalkey, "The Delphi Method: An Experimental Study of Group Opinion," Rand Corporation, 1969.

O. Helmer, "The Use of the Delphi Technique in Problems of Educational Innovations," The Rand Corporation, 1966, P-3499.

O. Helmer and N. Rescher, "On the Epistemology of the Inexact Sciences," Rand Corporation, 1960, R-353.

J. Martino, "The Consistency of Delphi Forecasts," *The Futurist* 4 (1970), p. 63.

M. Rokeach, *The Open and Closed Mind*, Basic Books, New York, 1960.

IV. E. The Performance of Forecasting Groups in Computer Dialogue and Face-to-face Discussion

KLAUS BROCKHOFF*

The Problem

Advances in mathematical and statistical techniques, the availability of efficient computers as well as ideas and attempts to utilize certain organizational structures in the compilation of expertise are basic elements for an intensified discussion of the problems of forecasting specific future developments and events. As J. Wild has shown, those conditional forecasts which are derived from models by certain statistical techniques are based as much on empirical knowledge as on ad-hoc extrapolations, projections, or expert opinions.[1] However, the isolation of independent variables from the surrounding conditions and the intra-personal process of information processing often do not become clearly visible in the latter. Thus there is a danger that uncontrolled, or uncontrollably, misinterpretations and false judgments may occur.

We do not want to infringe upon the controversy on the superiority of "forecasts" as compared with "projections," which is being carried out in the theory of science[2] as well as on an empirical-pragmatic[3] level. It has by no means been settled for the forecasts. This seems particularly true when the comparison is drawn on the basis of a benefit-cost relationship.[4] If only for this reason we are interested in the question whether or not the utilization of the empirical knowledge of groups of experts in the derivation of statements about future developments or events can be improved upon by organizational arrangements. Improvement is meant as an increase in the accuracy of these statements. This is one reason for the development of the Delphi method.[5]

*In collaboration with D. Kaerger and H. Rehder.

[1] J. Wild, "Probleme der theoretischen Deduktion von Prognosen," *Zeitschrift für die gesamte Staatswissenschaft (ZfgS)* 126 (1970), pp. 553–75.

[2] *Ibid.* and E. v. Knorring, "Probleme der theoretischen Deduktion von Prognosen," *ZfgS* 128 (1972), pp. 145–48; J. Wild, "Zur prinzipiellen Überlegenheit theoretisch deduzierter Prognosen," *ZfgS* 128 (1972), pp. 149–55.

[3] E.g., R. M. Copeland, R. J. Marioni, "Executives' Forecasts of Earnings per Share versus Forecasts of Naïve Models," *Journal of Business* 45 (1972), pp. 497–512, and the literature quoted there.

[4] Thus the suggestion in H. A. Simon, D. W. Smithburg, V. A. Thompson, *Public Administration*, New York, 1961, p. 493.

[5] O. Helmer, N. Rescher, "On the Epistemology of the Inexact Sciences," *Management Science* 6 (1959), pp. 25–52.

The Delphi Method: Techniques and Applications, Harold A. Linstone and Murray Turoff (eds.)
ISBN 0-201-04294-0; 0-201-04293-2

The conditions of group performance have been investigated in thousands of studies. Only few studies have been devoted to the question whether the peculiar organizational structure of the Delphi group[6] leads to higher group performance than the face-to-face discussion in a group in which the each-to-all pattern of the communication system can be activated.[7] Beyond this the unwieldy, nonhomogeneous, and inaccurate definition of types of tasks by which the performance of groups is judged[8] and thus the classification of concrete formulations of the question make it very difficult to derive statements as to the particular capacity of groups in forecasting.[9]

A large proportion of the statements as to the superiority of certain forms of group organization compared with others was obtained by observing group performance in solving certain kinds of problems and by assuming that the results would apply to tasks which appeared comparable. Thus references to the ability of groups to forecast particular future events were judged on the basis of their performance in responding to almanac-type questions. Martino has demonstrated that the answers to almanac-type questions observe the same type of distribution as the answers to forecasting questions.[10] However, it has not been investigated whether the parameters of the distribution vary significantly from one type of task to the other under conditions which are comparable otherwise. Thus it appears desirable to reconsider the original assumption.

In the following report we try to investigate some of these questions experimentally.

2. *Initial Hypotheses on the Performance of Forecasting Groups*

2.1 Group Performance and Group Size

2.1.1. Measurement of Variables It has been shown in various studies that the performance of a group may depend on its size.[11]

[6]Cf. below, section 2.3.1.

[7]On the restriction of the each-to-all pattern cf. M. E. Shaw, "Some Effects of Varying Amounts of Information Exclusively Possessed by a Group Member upon His Behavior in the Group," *Journal of General Psychology* 68 (1963), pp. 71–79.

[8]For the differentiation of types of tasks cf., e.g., M. E. Shaw, *Group Dynamics: The Psychology of Small Group Behavior*, New York, 1971, pp. 59, 403ff.; A. P. Hare, *Handbook of Small Group Research*, New York, 1962, pp. 246ff; C. G. Morris,"Task Effects on Group Interaction," *Journal of Personality and Social Psychology* 4, (1966), pp. 545–54. For problems of definition also R. Ziegler, *Kommunikationsstruktur und Leistung sozialer Systeme*, Meisenheim a. Glan, 1968, pp. 96ff.

[9]The limited choice of type of tasks and the strict assumptions on experimental group problem solving are most criticized. H. Franke, *Gruppenproblemlösen, Problemlösen in oder durch Gruppen? Problem und Entscheidung*, Heft 7, München, 1972, pp. 1–36, here p. 26 *et seq.*

[10]J. P. Martino, "The Lognormality of Delphi Estimates," *Technological Forecasting*, 1, (1970) pp. 355–358.

[11]Cf., e.g., J. D. Steiner, "Models for Inferring Relationships between Group Size and Potential Group Productivity," *Behavioral Science* 11 (1966) pp. 273–83; F. Frank and L. R. Anderson, "Effects of Task and Group Size upon Group Productivity and Member Satisfaction," *Sociometry* 34 (1971), pp. 135–49.

The *group size* is determined by the number of members of a group. This measure refers solely to formal criteria. Thus a person who does not contribute to the activity of the group, either because of his own reticence or because of a formal system of communication which does not accept his contributions, is still considered a member of the group.

Group performance can be "synthesized" by a statistical aggregation of individual performances.[12] However, such groups lack the essential characteristic of communication which exists in natural groups. In the following we will report entirely about experiments with natural groups.[13]

It may seem natural to study group performance of groups with a considerable number of members. However, in the experiments to follow we have deliberately concentrated on small groups of four to eleven people. One reason for this is that very many small and medium-sized organizations are applying Delphi. They can call in only small groups of experts. Even so they may wonder about their performance, and how to measure it. With regard to the possible objection that the results from the observation of small groups may be subject to considerable "noise," we may say that basic to most evaluations is the use of the median of individual responses. The median, however, is not sensitive to large dispersions, even if they are one-sided. On the other hand, it goes without question that it would be desirable to repeat our experiments in order to check on the reliability of the results.

Very different things can be understood by the "performance" of a group. The tripel, number of pieces of information exchanged, time needed for solving a problem and number of mistakes, can be considered a "classical" yardstick of performance. Ziegler ascribes the origin of this tripel to a paper written by Bavelas in 1950[14]. As Barnard's definition of performance—"the accomplishment of the recognized objectives of cooperative action"[15]—makes clear, however, this classical tripel is not compulsory. Indeed, it generally remains unclear whether performance refers to the goals (recognized objectives) of the members of the group, to those of the group, or to those presented to the group.

The task given to the groups is to find an answer A to a question which deviates as little as possible from the answer A' which can be verified now or in the future. Increasing performance then means that $|A - A'|$ approaches 0. In order to make comparisons between different questions or different groups a

[12]For a chronology of the publications on statistically "synthesized" performance, cf. J. Lorge, D. Fox, J. Davitz, and M. Brenner, "A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920–1957," *Psychological Bulletin* 55 (1958), pp. 337–72, here pp. 367f.

[13]Here natural group does not mean only a natural group with a: ach-to-all pattern of the communication system. With this I depart from the narrower definition in my earlier publication. See K. Brockhoff, "Zur Erfassung der Ungewissheit bei der Planung von Forschungsprojekten (zugleich ein Ansatz zur Bildung optimaler Gutachtergruppen)," in H. Hax, *Entscheidung bei unsicheren Erwartungen*, Köln, 1970, pp. 159–88, here pp. 167f.

[14]R. Ziegler, *op. cit.* p. 18; see also p. 55.

[15]C. J. Barnard, *The Functions of the Executive*, Cambridge, Mass., 1962, p. 55.

standardization is necessary. Thus the relative deviation of an estimate from the correct answer is used:

$$\frac{|A - A'|}{A'}.$$

We call this expression the "error." If the error refers to a person, we speak of an individual error. If the error refers to group performance, we speak of a group error. If A is the median of the estimates of all members of a group, we speak of a median group error (MGE).

The "mean group error" as used by Dalkey[16] is not identical with the MGE as given here. The basic difference is that Dalkey uses the logarithm of the quotient A/A' in order to test hypotheses about the distribution of his "mean group error." The distribution is of secondary importance for our present considerations regarding performance. For this reason we do not use logarithms here.

The MGE is used here directly as a measure of performance. It is not put into relation to the expenditures made for its derivation.

Further measures of group performance which are mentioned are the ability of the group to survive in a changing environment, its satisfaction, and the habitual change of its members.[17] We do not intend to study group performance in such a broad context (although we have unsystematically collected remarks on member satisfaction).

2.1.2 The Relationship of Performance to Group Size The study of hypotheses about a relationship between group size and group performance occupies a prominent position in small-group research. A brief survey of the diversity of empirical results was compiled by Türk.[18] A uniform result cannot truly be expected, because the individual studies were carried out under different conditions (types of tasks, performance measures, etc.). With respect to forecasting it has been hypothesized that the mean group error decreases with increasing group size.[19] It should be taken into account, however, that this statement has been formulated only for synthetic groups, i.e., a statistical aggregation of individual judgments, and with reference to the performance of the group in answering fact-finding questions.[20]

[16]Cf. N. C. Dalkey, "The Delphi Method: An Experimental Study of Group Opinion," Rand Corp., RM 5888 PR, 1969. Also H. Albach, "Informationsgewinnung durch strukturierte Gruppenbefragung—Die Delphi-Methode", *Zeitschrift für Betriebswirtschaft* (*ZfB*), Suppl. 40, Yr. 1970, pp. 11–26, here p. 20.

[17]Summarized, e.g., by M. Deutsch, "Group Behavior," in D. L. Sills (ed., *International Encyclopedia of the Social Sciences* 6, New York, 1968, pp. 265–75, here p. 274.

[18]K. Türk, "Gruppenentscheidungen. Sozialpsychologische Aspekte der Organization kollektiver Entscheidungsprozesse," *ZfB* 43. (1973), pp. 295–322, here p. 302.

[19]N. C. Dalkey, *op. cit.* pp. 9f.

[20]"These were questions where the experimenters knew the answer but the subjects did not": N. C. Dalkey, *op. cit.*, p. 10, fn.

The assumption of a decreasing error with increasing group size is based on the probability model of search.[21] From this model one deduces the possibility of compensating individual errors by calculating the mean for the group.

In natural groups, however, the rigid conditions on which alone the statements of the probability model are valid cannot always be fulfilled. Since, however, a consistent system of other factors that influence group performance as well (as the direction of their influence cannot be given), we may formulate:

Hypothesis G: With increasing group size, the group performance increases ceteris paribus.

It becomes clear in section 2.5 to what extent the restriction *ceteris paribus* can be repealed in our experiments.

2.2 Group Performance and Expertise

2.2.1 The Measuring of Expertise H. A. Simon represents "expertise" as a possible basis of authority or as a form of authority.[22] By expertise we mean expert knowledge upon which professional authority can be founded. This expert knowledge can be "proven"[23] by demonstration or by recourse to confirmation through third parties. A "proof" by recourse to third parties can hardly confirm more than a refutable conjecture as to the expertise of a person. If influencing variables of group performance are being sought, the experiments which make these variables visible can hardly contain an analyzable test of expertise at the same time. It is necessary to measure expertise as an independent variable by some other means.

One could proceed by testing which persons demonstrate expert knowledge in solving fact-finding questions. When such persons have been found, they can be engaged in forecasting. This takes for granted that the answering of both types of questions can be considered to be identical types of tasks.[24] Until now, however, no empirically tested statement to this effect exists.

[21]Cf., e. g., P. R. Hofstätter, *Gruppendynamik. Die Kritik der Massenpsychologie*, 11th ed., Reinbek, 1970, pp. 35ff., 160ff.

[22]H. A. Simon, *Administrative Behavior*, 3rd ed., New York and London, 1965, p. 76.

[23]F. Landwehrmann, "Autorität," in E. Grochla (ed.), *Handwörterbuch der Organisation*, Stuttgart, 1969, col. 269–73, here col. 270, refers to H. Hartmann, *Funktionale Autorität*, Stuttgart, 1964.

[24]For a procedure oriented thus, cf. M. A. Jolson, G. L. Rossow, "The Delphi Process in Marketing Decision Making," *Journal of Marketing Research*, 8 (1971), pp. 443–48. Another procedure, based on the solution of test questions and a test of the understanding of professional terminology, is described by A. J. Lipinski, H. M. Lipinski, R. H. Randolph, "Computer-Assisted Expert Interrogation: A Report on Current Methods Development," *Technological Forecasting and Social Change* 5 (1973), pp. 3–18, here pp. 9f. (The same in S. Winkler [ed.], *Computer Communications, Impact and Implications*, New York, 1973, pp. 147–54. The authors also test the "quality of respondents' comments" [presumably on factual questions], the degree of attention and the degree of optimism [with the aid of a price list for old phonograph records] and inquire from this a rank order of expertise.)

Further, one could consider whether expertise ratings by third parties can be used. However, it may be of interest to maintain the anonymity of all persons who may possibly participate in a forecasting group.[25] Another problem that arises is what criteria should be used in choosing those persons who are to judge the expertise of others.

Thus there remains the possibility of determining expertise by self-rating. For this purpose an ordinal scale is generally used, from which one value can be chosen to indicate expertise. We worked with a scale of real numbers graded from 1 to 5, in which low numbers must be used to express a low degree of expertise while high numbers may be used to express a high degree of expertise. Such a determination of expertise is employed already in some forecasting groups by their management.[26] These results of individual forecasts are weighted according to the self-ratings when a group judgment is derived.

Measurements of expertise which are obtained for individuals should also permit statements as to the expertise of the total group. Since self-ratings are measured on an ordinal scale it is not permissible to form an arithmetic mean of all the self-ratings of the members of the group. We therefore characterize the expertise of a group by the median of the individual self-ratings.

Whether such self-ratings have a high positive correlation with ratings by third parties has not yet been studied in realistic situations. For this reason it remains an open question whether corresponding confirmatory results of psychological tests[27] can be applied to real situations, which generally are not free of conflicting interests.

Even if no significant positive correlation exists between ratings by third parties and self-ratings concerning expert knowledge, it is not determined which of the ratings is more correct. Thus in the approach used here, which involves self-ratings, the question remains whether the participants in the experiment rate themselves correctly when compared with an (inapplicable) objective standard.

2.2.2. The Relationship between Expertise and Group Performance We assume that groups with high self-ratings of expertise perform better than groups whose members rate themselves as less qualified. With this assumption we follow Dalkey, Brown, and Cochran[28]. Their results must, however, be examined with care insofar as they were obtained from answering fact-finding questions. Furthermore, the subjects were able to compare all questions to one another

[25]Cf. Section 2.3.1.

[26]D. L. Pyke, "A Practical Aporoach to Delphi: TRW's Probe II," *Futures* 2 (1970), pp. 143–52; H. P. North, D. L. Pyke, "Probes of the Technological Future," *Harvard Business Review 3 (1969), pp. 68–76; A. J. Lipinski, L. M. Lipinski, R. H. Randolph, op. cit.*, pp. 11ff.

[27]M. A. Wallach, N. Kogan, J. Bem, "Group Influence on Individual Risk Taking," *Journal of Abnormal and Social Psychology* 65 (1962), pp. 75–86, here p. 83.

[28]N. C. Dalkey, B. Brown, S. Cochran, "The Use of Self-Ratings to Improve Group Estimates," *Technological Forecasting*, 1 (1970) pp. 283–292.

before rating their expertise with respect to each question. In many applications it is not possible to present all questions at once. We shall therefore follow a different procedure by presenting tasks in a sequential manner. Even so, we assume that the basic relationship is still valid. Thus, we arrive at

Hypothesis E: With increasing expertise, group performance increases ceteris paribus.

2.3 Group Performance and Communication System

2.3.1 The Characteristics of Face-to-Face Discussion Groups vs. Delphi Groups The question whether decentralized or centralized organizations exhibit higher performance is another of the classical questions in organization research. The attempt to set up a universally applicable rule of organization to answer this question had to be dropped because little by little conditions become known on which first the one organizational structure and then the other seemed more advantageous for accomplishing specific tasks. Fundamentally, it is assumed in these studies that all organizational structures considered are capable of accomplishing certain tasks. A formal organization is characterized essentially by its communication system and the distribution of competence among its members.[29] It is further assumed that fact-finding questions as well as forecasting questions can be answered more accurately by groups than by individuals, if the (expected) error[30] is taken as the measure of accuracy.

Fact-finding questions and forecasting questions can be discussed in natural groups with an each-to-all pattern of communication. If it is desired to have a group judgment, this task can be left up to the group, or a rule for aggregating the group judgment from the individual judgments of the participants can be given. Depending on the situation, the use of such a rule can be restricted to the case where the group does not agree on a group judgment within a given period of time.[31]

Particularly Carzo's results indicate that in natural groups in which communication is not limited, solutions to complex tasks are reached rapidly, with few errors, and to the satisfaction of the group members.[32] One objection to this is that this form of organization may also produce dysfunctional effects.

A first dysfunctionality may arise as certain group members consciously or unconsciously influence the group result to a greater degree than their expertise

[29]This is made particularly clear by H. Albach, "Organisation, betriebliche," in *Handwörterbuch d. Sozialwiss.* 8, Stuttgart, Tübingen, Göttingen, 1961, pp. 111–17.

[30]Cf. H. Schöllhammer, "Die Delphi-Methode als betriebliches Prognose-und Planungsverfahren," *Zeitschrift für betriebswirtschaftliche Forschung (ZfbF)*, N. S., 22nd Yr. (1970), pp. 128–37, here particularly fn. 8.

[31]Cf. the experiments by B. Contini, "The Value of Time in Bargaining Negotiations—Some Empirical Evidence," *American Economic Review* 58 (1968), pp. 374–93.

[32]Cf. R. Carzo, Jr., "Some Effects of Organization Structure on Group Effectiveness," *Administrative Science Quarterly* 7 (1963), pp. 393–424.

warrants.[33] A further dysfunctionality arises if the exchange of information is interfered with by "noise."[34] A more far-reaching possibility for the occurrence of dysfunctional effects is that the transmission of information necessary for task accomplishment from some group members to others is blocked by somebody interested[35] or that the transmission of information from some group members in the time period given for accomplishing the task is altogether impossible. In the latter case the participation of the individual group members in the exchange of information may be independent of the degree of expertise. Then participants with a high degree of expertise cannot necessarily influence group performance.

One can summarize that possible dysfunctionalities in a natural group with an each-to-all pattern of communication result from utilizing the communication system and the system of competence in a manner which does not correlate positively with the degree of expertise. Assuming that these effects often interfere with group performance, rules should be set up which reduce the effects of dysfunctionality or prevent their appearance. A set of such rules was suggested and introduced by Helmer and Dalkey[36] and given the name "Delphi." In spite of diverse variations in procedure, the applications known to date have as their primary objective: "...the establishment of a meaningful group communication structure."[37]

2.3.2. The Relationships According to Dalkey's studies, Delphi groups demonstrate a certain, though not significant superiority when compared with certain other groups in solving fact-finding questions.[38] We apply this perception to our

Hypothesis D: The performance of Delphi groups is ceteris paribus higher than the performance of natural groups with an each-to-all pattern of communication.

The diverse arrangements of Delphi experiments make it necessary to investigate some of their special features as well. Of particular importance is the question whether the performance of Delphi groups is the same in each round, or whether it increases with increasing number of rounds, at least up to

[33]This very abridged presentation must be viewed on the basis of the entire discussion concerning the question: which conditions promote behavioral conformity in individuals in groups; cf. A. P. Hare, *Handbook ..., op. cit.*, chapters, 2, 13 (there in reference to status rivalry); L. Festinger, E. Aronson, "The Arousal and Reduction of Dissonance in Social Contexts," in D. Cartwright, A. Zander (eds.), *Group Dynamics: Research and Theory*, Evanston, Ill., 1960, pp. 214–31.

[34]A. P. Hare, *op. cit.*, chapter 10, pp. 272ff.

[35]On this broad field, see H. H. Kelley, J. W. Thibaut, "Group Problem Solving," in G. Lindzey, E. Aronson (eds.), *Handbook of Social Psychology* 4, Reading, Mass., 1968, pp. 1–101, here pp. 6ff., 26ff.

[36]N. C. Dalkey, O. Helmer, "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science* 9 (1963), pp. 458–67.

[37]M. Turoff, "Delphi and its Potential Impact on Information Systems," *AFIPS Conference Proceedings*, Fall Joint Computer Conference (Fall 1971), 39, pp. 317–26, here p. 317.

[38]N. C. Dalkey, *op. cit. passim*, particularly p. 22.

the fourth round.[39] This leads to

Hypothesis R': The performance of Delphi groups increases ceteris paribus with increasing number of rounds.

We test this hypothesis here for up to five rounds. It cannot be expected that hypothesis R' is valid for an unlimited number of rounds, since growing dissatisfaction of the participants and increasing time requirements make it seem senseless to continue the consultations indefinitely. Therefore, we modify hypothesis R' to

Hypothesis R: The performance of the Delphi groups increases ceteris paribus with increasing number of rounds only at first. Finally, the increase in performance can be reduced and inverted. The question in which round the "performance inversion" begins must be answered empirically.

Finally, Helmer and Dalkey[40] showed an interest in the observation that the variance of the responses around the median decreases with increasing number of rounds. The reduction of variance is not in itself a criterion for increased performance. One must view this observation on the basis of hypothesis *R*: together with it, variance reduction gains importance in the sense that it means increasing certainty and accuracy of the answers.[41] We therefore also test this question by investigating

Hypothesis V: The variance of answers around the median decreases ceteris paribus with increasing number of rounds.

Two statistical measures of variance are at our disposal for testing this hypothesis: average quartile difference and average variance from the median. The latter measure offers certain advantages for a comparison between groups of different sizes. For this reason it is given preference here.

2.4 Group Performance and Type of Task: Fact-finding Questions and Forecasting

Only a few of the generally available results of forecasts by Delphi groups can be tested against reality, because they mainly refer to events which are

[39]Cf. also J. B. Martino, "An Experiment with the Delphi Procedure for Long-Range Forecasting," *IEEE Trans. on Engineering Management* 15 (1965), pp. 138–44.

[40]O. Helmer, *Social Technology*, New York and London, 1966, pp. 101ff.; N. C. Dalkey, *op. cit.*, p. 20.

[41]They are based according to the Delphi method, on renewed intrapersonal conflict solution and problem solving, after being provided with additional data. On the interpersonal process which is to be eliminated here, cf. J. Hall and M. S. Williams, "A Comparison of Decision-making Performances in Established and Ad Hoc Groups," *Journal of Personality and Social Psychology* 3 (1966), pp. 214–22. On the practical organization of the elimination of dysfunctionality and pressure to conform, cf. 3.2, below.

expected to take place in the distant future. For this reason, the comparison of performance of face-to-face discussion groups and Delphi groups is made by observing tasks which appear similar.[42] The problem of solving fact-finding or almanac-type questions is assumed to be similar to forecasting. The answers of such questions are, as a matter of principle, unknown to the participants but known to the experimenters. These two applications of Delphi, its use in forecasting and its simulation with only subjectively unknown bits of knowledge, exist as yet side-by-side without comparison. Since the complexity of a task is an important determinant of group performance, but the criteria for determining tasks of varying degrees of complexity are not clear enough, we want to test directly

Hypothesis F: The performance of a group in answering fact-finding questions is ceteris paribus equal to that in forecasting.

The hypothesis is deduced from the assumption that both types of tasks exhibit the same degree of complexity. The tests should be carried out separately for face-to-face groups and Delphi groups of the same size. If the hypothesis is refuted, many of the statements about Delphi-groups which we rederived using fact-finding questions cannot be maintained. In order to test hypothesis *F*, we chose only facts referring to events that had occurred, on the average, six months before the experiments. The forecasts, on the other hand, refer mainly to a period of time which did not exceed six months after the tests were carried out.

2.5 The Relationship between the Hypotheses

The repeated use of *ceteris paribus* in the hypotheses leads us to assume that they are in fact related to each other. It seems senseless to describe the great variety of possible combinations. The rejection of certain hypotheses can reduce the test program considerably as it leaves only a small number of relationships which need to be tested. Hypothesis *E*, e.g., can be tested for a given group size, a given type of question, a given type of group, and, in Delphi groups, with regard to the results of any round. No matter what the result is, it is irrelevant for the further experiments, if expertise should be distributed equally in the different groups. Since the distribution of the actual expertise becomes known from the results of the experiments, it is advisable to bring about the necessary clarifications at first.

The situation is similar when discussion groups and Delphi groups are compared with each other. If hypothesis *R* is not tested for the latter, it cannot

[42] N. C. Dalkey, *op. cit.*, pp. 9f. Dalkey cites (p. 23) a paper by Campbell, in which similar methods evidently were used to the ones planned here. The original publication was not available.

be determined from which round the results should be taken if compared with the performance of the face-to-face discussion groups.

The presentation of our results in Section 4 is organized according to such reasoning.

3. *The Experiments*

3.1 Participants, Group Formation, Place, and Time

All experiments were carried out as part of a lab. course listed in the University of Kiel catalogue. It was planned to have "students" and practitioners work in separate groups and to compare the results. However, since we could not give credits for the course, too few students registered to be able to form even *one* small group.

Practitioners were designated and chosen from the permanent staffs of the local banks with the assistance of the bank managers. Bank employees were chosen because hardly any other line of business is represented in the area by enough individual organizations with personnel trained in economics and with relatively uniform fields of business. At the same time, the size of the participating organizations is generally so large that persons who perform specialized functions (long-term credits, short-term credits, investment brokerage, etc.) and who differ with respect to the lengths of their employment could be chosen. This seemed desirable in order that definite differences could show up in the self-ratings of expertise in reference to the individual tasks.

The thirty-two participants were randomly assigned to four groups, having five, seven, nine, and eleven participants, respectively. At the face-to-face discussions, however, registered participants were absent for various personal reasons, so that the groups had four, seven, eight, and ten participants only.

The experiments began with an introductory lecture about forecasting methods and an exercise in the use of the displays which were used in the experiments with the Delphi groups. After that, each subject participated in a session in which the members were organized as a Delphi group. At a later session a face-to-face discussion took place. After the experiments were concluded, an opportunity for criticism was given. Eight months later, the results were communicated. We have tried to motivate our participants to cooperate well in the preparatory lecture and demonstration. Besides, we offered book prizes for outstanding performances with regard to different types of questions and the two basic group structures.

The experiments were conducted in May and July 1973. With three exceptions they were scheduled for Thursdays to conform with late closing time of banks. The Delphi groups worked in the computer center of the University of Kiel; the face-to-face discussions were carried out in a library room.

3.2 The Organization of the Delphi Groups and the Face-to-Face Groups

The Delphi groups were set up so that the participants received all information from the experimenters on a computer-generated display.[43]

The participants in a group were not supposed to establish immediate contact with each other. They responded to all questions by writing an alpha-numeric text in their normal language. The responses can be divided into three classes: (1) responses that were known only to the experimenters; (2) responses which, after the responses of all participants had been received by the experimenter, became objects of computing procedures, the results of which were made known to all participants; (3) responses which were recorded and made known to all participants without any changes. The first class includes the name of the participant and the degree of expertise that he expresses with regard to each question. This is handled differently in the face-to-face discussions groups. A response of the second category is an individual estimate. After the computation of the median of the responses of the group members, this figure is made known to all participants. The third category includes all arguments for divergent opinions of those whose responses lay outside of the lower or upper quartiles.

Computer communication has been praised as a means to enable experts to communicate with each other even though they are separated from each other by large distances. In the real world this could mean savings in travel expenses and in the efforts expended in coordinating dates for groups of experts. Beyond this it is of importance for the experiments that the computation of quartiles, and the preparation, distribution, collection, reproduction, and renewed distribution of questionnaires do not have to be carried out by hand during the sessions. This gives one the chance to shorten the experiments considerably.[44]

The entire exchange of information between the participants as well as between the experimenter and the participants during the sessions, with the exception of certain recurrent standard formulations, was stored on tape as a record of the experiments.

Figure 1 shows part of a record. A separate data file, an abstract from the records, is kept on tape. It serves as the data base for the diverse computations.

The abstract from the record shows the beginning of a session of the Delphi group with seven participants. Vertical lines on the left edge of the text signal those portions of the texts which appear in the same form on the display of each participant. In section 1 the names of the participants are given to the experimenter. Section 2 contains the first fact-finding question for the group. In

[43] The programming was done by D. Kaerger, who will report separately on problems that arose herewith. The program had to be in FORTRAN IV. It was run on a PDP 10.

[44] For a brief discussion of the advantages and disadvantages of "computer communication" compared with direct communication, cf. A. I. Lipinski, H. M. Lipinski, R. H. Randolph, *op. cit.*, particularly pp. 11–12.


```
GEBEN SIE BITTE IHREN NAMEN AN

BRIDSTRUP
CARL MOELLER
BOLTE
JOERG BISKUP
THOMAS M. RIECKEN
GLOCKNER
KAEMPFER                                                        1

FUER WELCHEN BETRAG KAUFTEN DIE KREDITINSTITUTE
DER BRD 1972 FESTVERZINSLICHE WERTPAPIERE                       2
IN MRD.DM

FAKTEN... KAUF VON FESTVERZINSLICHEN WERTPAPIEREN
DURCH KREDITINSTITUTE IN DEN JAHREN 1967 BIS 1969
JEWEILS ZWISCHEN 1Ø UND 15 MRD.                                 3
BRUTTOABSATZ FESTVERZINSLICHER WERTPAPIERE 1971
3Ø,3 MRD. DM

GEBEN SIE IHREN EXPERTENGRAD--ZIFFER 1. BIS 5.-- ZU DIESER FRAGE AN..

EXPERTENGRAD  FRAGE   1 TEILNEHMER  3          3
EXPERTENGRAD  FRAGE   1 TEILNEHMER  4          2
EXPERTENGRAD  FRAGE   1 TEILNEHMER  5          1
EXPERTENGRAD  FRAGE   1 TEILNEHMER  6          1                4
EXPERTENGRAD  FRAGE   1 TEILNEHMER  7          3
EXPERTENGRAD  FRAGE   1 TEILNEHMER  8          4
EXPERTENGRAD  FRAGE   1 TEILNEHMER  9          4

GEBEN SIE IHRE SCHAETZUNG FUER RUNDE   1 AN..

SCHAETZUNG  1  FRAGE  1   TEILNEHMER 3              5Ø.ØØ
SCHAETZUNG  1  FRAGE  1   TEILNEHMER 4              45.ØØ
SCHAETZUNG  1  FRAGE  1   TEILNEHMER 5              35.ØØ
SCHAETZUNG  1  FRAGE  1   TEILNEHMER 6              5Ø.ØØ        5
SCHAETZUNG  1  FRAGE  1   TEILNEHMER 7              48.ØØ
SCHAETZUNG  1  FRAGE  1   TEILNEHMER 8              4Ø.ØØ
SCHAETZUNG  1  FRAGE  1   TEILNEHMER 9              21.5Ø

ABWEICHUNGSFRAGE AN TEILNEHMER   9   GESTELLT
 TEILNEHMER   9 -INFORMATION..
KEINE
ABWEICHUNGSFRAGE AN TEILNEHMER   6   GESTELLT                    6
 TEILNEHMER   6 -INFORMATION..
1972 WAR EIN BESONDERS GUTES JAHR FUER DEN ABSATZ VON WERTPAPIEREN

FUER WELCHEN BETRAG KAUFTEN DIE KREDITINSTITUTE
DER BRD 1972 FESTVERZINSLICHE WERTPAPIERE                       7
IN MRD.DM
```

Fig. 1. Abstract from a record

```
|FAKTEN... KAUF VON FESTVERZINSLICHEN WERTPAPIEREN

                                                                    8
|DURCH KREDITINSTITUTE IN DEN JAHREN 1967 BIS 1969
 JEWEILS ZWISCHEN 10 UND 15 MRD.
 BRUTTOABSATZ FESTVERZINSLICHER WERTPAPIERE 1971
 30,3 MRD. DM
|ZUSATZINFORMATION..   MITTELWERT =      45.00
 KEINE INFORMATION
 1972 WAR EIN BESONDERS GUTES JAHR FUER DEN ABSATZ VON WERTPAPIEREN  9

|GEBEN SIE IHRE SCHAETZUNG FUER RUNDE  2 AN..
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 3          50.00
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 4          45.00
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 5          35.00          10
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 6          50.00
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 7          48.00
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 8          35.00
 SCHAETZUNG  2  FRAGE   1   TEILNEHMER 9          42.30
 ABWEICHUNGSFRAGE AN TEILNEHMER   5   GESTELLT
  TEILNEHMER   5 -INFORMATION..
```

FIGURE 1 : ABSTRACT FROM A RECORD

Fig. 1. (continued)

section 3 you see information which is considered relevant for the judgment of the problem and which is given to each participant. Additional information of this sort is not given to participants when forecasting questions are asked. This difference is justified by the assumption that the subjects need some information to refresh their memories with respect to judging "facts" which are about six months old. It is expected that they do not need help in evaluating present-day facts as a basis for their forecasts.

In the following section, 4, participants are asked to give their degree of expertise. It is given only once for each question. In section 5, we enter the response portion of the first round.

The numbering of the participants in sections 4 and 5 serves only as an internal identification. It begins with "3", because "1" and "2" are reserved for the experimenter and the tape on which we store the record.

After the estimates have been made (in section 5) the 0.25, 0.5, and 0.75 quartiles are calculated. The participants whose estimates lie outside the 0.25

and 0.75 quartiles are shown the quartile values which they exceeded or fell short of, and are asked to give reasons for this divergence, in case they think they possess particular additional information. The text of the questions, which is repeated at this place, is not put out in the record shown here. In section 6, data necessary for the analysis are recorded.

The second round begins with section 7. First, the question is repeated, then the relevant information. In section 9, additional data are given, namely, group responses from the previous round and any additional information which was collected in section 6 of the previous round. This information is presented in the following order: first, additional information from those participants whose estimates fell short of the lower quartile; then, information from those whose estimates exceeded the upper quartile. In the present case, there is only one item of additional information. The original text accompanying this information, which does not refer to its sources, is not reproduced here.

Beginning with section 10, the process which was described for section 5 and the succeeding sections is repeated. Five rounds are carried out for each question. This scope was chosen in compliance with the observation that after the fourth round generally the results do not improve (see section 2.3.2). An additional round is added here to test this statement.

After the five rounds are completed the next question is asked. It is a forecasting question. The two types of questions are asked alternately so that possible effects of learning or fatigue do not influence only one type of question.

In face-to-face discussion groups, the group members are asked to introduce themselves to each other by their name, field of employment, official position, and the number of years spent in banking. The idea of this was to provide each participant with a basis for judging the experience of the discussion partners in the following discussions. To what extent this information was taken into account in the formation of the group judgments could not be registered explicitly. Furthermore, the participants were asked to specify their degree of expertise for each question on a record. They noted their personal estimates for each question before any discussion took place. A discussion of the problem was expected to follow and a unanimous group estimate was demanded. A discussion leader was not appointed.

3.3 The Questions

The fact-finding questions and the forecasting questions refer to finance, banking, stock quotations, and foreign trade. We could choose from a list of ninety fact-finding questions and thirty forecasting questions that were kept on a separate tape. The questions were chosen at random from this stock. Each question of the two different types had the same chance of being chosen. However, no question appeared twice in a group.

All questions refer to items which are reported in the monthly statistics of the German Federal Bank (Deutsche Bundesbank), the daily stock market quotations of the Frankfurt Stock Exchange, and the market reports of the big banks. Only very few of the fact-finding questions refer to facts which are reported in the foreign trade statistics.

In all cases the correct responses can be verified objectively at the time of the experiments or at a later date. In the opening lecture it was called to the attention of the participants that, for example, the questions about certain past or future interest rates did not refer to the rates of the respective local institutions, which may depend largely on effects of local competition. Rather, they refer to the rates which are listed as averages in the statistics of the Federal Bank.

In five cases the wording of the questions was unclear to the participants.[45] This resulted partly from an inexact formulation on our part and partly from imperfect knowledge of the definitions as used by the Federal Bank on the part of the participants. In the face-to-face discussions, clarifications could be made immediately. In two cases in the Delphi groups, the "correct answer" in the sense in which it was understood by the participants was carried on rather than the answer to the original formulation of the question. Further cases of general misunderstanding did not become evident.

In a final discussion many of the participants expressed the feeling that the fact-finding questions were rather irrelevant and annoying: all the facts could be looked up with no trouble. This point of view was not expressed with regard to the forecasting questions. It should be recorded, however, that both sets of questions refer to the same objects, although at different points in time. (Thus, for example, one question asks for the price of a share of RWE common stocks six months before the experiments and another for the same quotation six months afterward).

3.4 The Volume of the Tests

It was originally planned that each group of different size and organization should be asked to give ten forecasts and to answer ten fact-finding questions. This plan was carried out with one exception. The Delphi group with eleven participants was able to handle only eight questions of each kind.

The time spent on the discussions amounts to between 140 and 200 minutes. In the Delphi rounds between 200 and 240 minutes of connect time were spent per participant.[46] (This length of time does not correspond to the CPU time,

[45] On the significance of the wording of questions and its influence on the level of estimates, cf. J. R. Salancik, W. Wenger, and E. Helfer, "The Construction of Delphi Event Statements," *Technological Forecasting and Social Change* 3 No. 1 (1971), pp. 65–73.

[46] With 20 questions, that is 10 to 12 minutes; with 16 questions, 12 to 15 minutes. Lipinski, Lipinski, and Randolph, *op. cit.*, report 15 minutes per question, on comparable hardware.

however.[47] The following points can be considered as possible explanations for the greater length of time spent on the Delphi rounds: (1) Participants write more slowly than they speak. (2) Communications between the experimenter and the participants takes place "sequentially," i.e., if participants j and k are involved, $j < k$, the message of participant k to the experimenter or back can be exchanged only after the same kind of message has been exchanged between participant j and the experimenter. This pattern of sequential communication is determined by the available technology. (3) Communication among the participants and between the participants and the experimenter can take place only during the periods of time in which the computer, which operates in a time-sharing mode, is available for the job. Although the CPU was not busy with batch operation during the experiments, the demand for memory space for other jobs which were also initiated at remote terminals was noticeable during the experiments. The participants considered such delays very disturbing. Finally let us point out that since the available teletype terminals type more slowly than the displays, preliminary experiments indicated that the former were not suitable for the experiments.

The operating system allowed for the connection of 13 displays at one time. This determined a possible maximum of group size. However, as each participant was supposed to join each type of group only once, we have limited maximum size to 11. The minimum size of groups was determined by the consideration that we wanted to have two clearly determinable participants whose estimates lay outside the quartile values. This can be achieved with five people. The fact that we had one man less in the discussion group did not interfere with this principle.

4. The Results[48]

4.1 The Distribution of Expertise

We want to investigate whether the expertise of the group members is distributed evenly or unevenly in the groups. It can be assumed that the distribution is even, because the participants and the questions were assigned to the groups at random.

However, the preliminary question whether expertise is rated differently with regard to fact-finding questions and forecasts must be clarified. Only if the

[47] D. Kaerger will contribute to the further analysis. See also, Institute for the Future, "Development of a Computer-Based System to Improve Interaction among Experts," First Annual Report to the National Service Foundation, 8/1/73, p. 6, Table 2. The relationship of CPU time to connect time varies from 1:110 to 1:135.

[48] The tests quoted in the following are described, e.g., by S. Siegel, *Non-Parametric Statistics for the Behavioral Sciences*, New York, and London, 1956; G. A. Lienert, *Verteilungsfreie Methoden in der Biostatistik*, Meisenheim a. Glan, 1962.

hypothesis of an uneven distribution is rejected can a comparison between the groups be made with aggregated data from fact-finding questions and forecasts. Therefore, we first test "auxiliary hypothesis 1": *expertise is rated differently in each group with regard to fact-finding questions and forecasts.*

The comparison of the quartiles given in Table 1 reveals different results in only five out of twenty-four cases. However, the narrow limits of a scale from 1 to 5 does not allow this result to appear sufficiently reliable.[49] We therefore compare the differences between the cumulative relative frequencies of the expertise ratings within each group for fact-finding questions and forecasts. The Kolmogorov-Smirnov Test shows no significant difference of the distributions on the 5 percent level.

Table 1
Quartiles of Expertise in Fact-Finding Questions and Forecasting Questions

		Quartiles					
		Lower Quartile		Median		Upper Quartile	
Type of Group	Group Size	Fact-finding questions	Forecasting questions	Fact-finding questions	Forecasting questions	Fact-finding questions	Forecasting questions
Face-to-Face Discussion Group	4	1	1	1	2	2	3
	7	2	2	2	2	3	3
	8	1	1	2	2	2	3
	10	1	1	2	1	2	2
Delphi Group	5	1	1	2	2	3	3
	7	2	2	2	3	3	3
	9	1	1	2	2	3	3
	11	1	1	2	2	3	3

Thus, auxiliary hypothesis 1 is rejected. We can use the entire set of data to investigate auxiliary hypothesis 2. It says that *between the groups no differences occur in the relative frequencies with which the different scale values of expertise are chosen.*

Table 2 shows the relative frequencies of the distributions of expertise.

[49] This conjecture is based on general reflections and empirical results on the correct construction of a scale. On the inferiority of the scale with five divisions compared with scales with more graduations in estimates of decisionmaking groups, see G. Huber and A. Delbecq, "Guidelines for Combining the Judgments of Individual Group Members in Decision Conferences," manuscript, TIMS-meeting, Detroit, 1971, pp. 5ff. It could not be investigated whether these statements can be applied to our results, because of the difference in the task and because American subjects may be less familiar with a scale using five divisions than are Germans.

Table 2
Distribution of Expertise (%)

		Degree of Expertise				
Type of Group	Group Size	1	2	3	4	5
Face-to-Face Discussion Groups	4	44	32	13	5	6
	7	21	38	34	6	1
	8	30	44	19	4	3
	10	45	33	15	6	1
Delphi Groups	5	39	32	16	9	4
	7	15	35	37	9	4
	9	29	28	25	13	5
	11	39	30	18	9	4

Within each type of group comparable results appear, as expected.[50] The distribution of expertise is in both cases indistinguishable between the smallest and the largest groups; the distribution varies greatly between the group with seven participants and the largest and smallest groups, respectively (cf. Table 3).

Table 3
Maximum Absolute Difference between Cumulative Distributions of the Relative Frequencies of the Degree of Expertise between Pairs of Groups (in %)

Face-to-Face Discussion Groups				Delphi Groups			
Group Size	7	8	10	Group Size	7	9	11
4	23	14	5	5	24	14	2
7		15	24	7		14	24
8			15	9			12

[50] A comparison with the groups with the same rank of size within the type of group shows no significant differences at $p > 5$ percent, neither with the Kolmogorov-Smirnov Test nor with the sign test. The latter test was carried out in compliance with the possible objection that the samples were interrelated.

An investigation of the observations for all groups of a given type leads us to reject auxiliary hypothesis 2. If one were to formulate auxiliary hypothesis 2 for a pairwise comparison between groups it would be refuted in all cases except when the smallest group is compared with the largest group or the second largest group respectively (Kolmogorov-Smirnov Test on the 5 percent level). A consideration of the data in Table 2 now gains increased importance. The significant differences are due to the fact that in the middle-sized groups, and particularly in the groups with seven participants, the ratings of expertise seem higher than in the smallest or largest groups. It is obvious, and is not examined more closely here, that the lower degrees of expertise are chosen much more frequently than the higher degrees. Whether this is an illustration of the often assumed pyramid of qualifications and abilities, or whether it only reflects a fear of using the higher values on the scale, cannot be determined definitively. The second assumption is supported by the observation that in the Delphi groups, where greater anonymity is guaranteed, 9.6 percent of the self-ratings fall into the categories four and five, whereas in the face-to-face discussion groups, where the self-ratings were occasionally asked for directly by other members of a group, only 5.5 percent fall into these categories. The results of the next section contribute to the extension of these reflections.

4.2 The Significance of the Self-Ratings

A first indication as to the validity of hypothesis E can be gained by observing individual errors and self-ratings. We test whether the lowest individual errors in each group and in relation to each question are attained by those persons who rate their expertise highest. Individual errors (see section 2.1.1) are taken as absolute values. In the Delphi groups we can determine this error in every round. We restrict ourselves here to the first and the last round. In the face-to-face discussion groups the only data for judging individual errors is from the questionnaires filled out before entering the discussion of each question. We test auxiliary hypothesis 3:

The distributions of the highest self-ratings with regard to each question and the self-ratings of those who attain the highest level of performance (i.e., the lowest individual error taken as an absolute value) coincide.

Auxiliary hypothesis 3 is tested separately for fact-finding questions and forecasting questions in face-to-face discussion groups and Delphi groups. In the latter case it is also tested separately for the first and the last rounds. In all cases, auxiliary hypothesis 3 is refuted at a high level of significance in a χ^2 test (0.01). Thus, one must assume that in the situations investigated here, self-ratings with regard to expertise do not give enough information as to which persons actually possess expertise. For this reason either the ability to give a

self-rating which corresponds to actual expertise must be studied more closely and, if possible, promoted or other methods of determining expertise before the beginning of the questioning must be tested for their effectiveness. The Institute for the Future emphasizes the latter problem[51] in its recent studies.

Auxiliary hypothesis 3 was based on individual performance. Hypothesis E, on the contrary, refers to the performance of the entire group. We attempt to operationalize this viewpoint by testing the rank correlation[52] between group performance with respect to each question and the average expertise of the group with respect to the same question, separately for each type of group, each group size, and each of the two types of questions. The skew distribution of the degrees of expertise already lets us expect that important information cannot be gained from such a test. Indeed, we do not find any significant rank correlation coefficient in the relationships tested for fact-finding questions. A classification of the data in a 2×2 contingency table according to the criteria: low and high degree of expertise *vs.* upper and lower half of the scale of the rank figures for group performance, does not lead to significant relationships. With regard to forecasts, significant relationships (at the 5 percent level) show up in Delphi groups with five and nine participants in the third as well as in the fifth round. However, as is shown in Table 6, these groups are not noted for particularly good overall results. In this respect the correlation seems unimportant.

Since the conclusion that objectively existent expertise is not an essential factor of individual or group performance contradicts the definition of expertise and thus is not tolerable as an explanation of the results, it can only be concluded that the self-rating of expertise by practitioners for tasks of the present type does not coincide with their objective expertise.

After the experiments were concluded, this unsatisfactory result led to the question whether explanations can be found for the choice of the rank figures of expertise by the individuals. We attempt to explain the behavior of our subjects by the following auxiliary hypothesis 4:

The degree of expertise with regard to a question is related to the number of years that a subject spent in banking. Furthermore, it is higher whenever the subject matter of the question coincides with one of the fields handled during the years in the profession.

The necessary data were collected by questionnaire. Answers from up to twenty-eight participants were available. In the analyses a high positive correlation showed up between age and number of years in the profession, so that separate hypotheses for these two variables were not tested. Further, a positive correlation showed up between the number of fields one had experience in (a list of possible fields was presented which could, however, be supplemented)

[51] Cf. Institute for the Future, "Development of a Computer-Based System to Improve Interaction among Experts," *op. cit.*

[52] Spearman rank correlation with correction for ties.

and the number of years in the profession. Thus the two components of hypothesis 4 can no longer serve as mutually independent variables for explaining the degree of expertise. Therefore, we tested the hypothesis in a simplified form, once for the influence of the number of years in the profession and once for the influence of the fields of employment on the choice of the degree of expertise. In neither case, a Kolmogorov-Smirnov Test refutes the hypothesis (0.001 level).

Thus an observation that was gained in the face-to-face discussion groups is confirmed. Once in a while during the discussions the number of years of "banking experience" was brought into play to decide points of controversy, obviously with the view that this is a criterion for measuring expertise objectively. The same is true for the present field of employment of the persons in question. Obviously, however, these criteria for judging on the expertise are not sufficient in the light of very special questions. It would probably be better to consider, for example, the regular observation of special sections of the bank statistics.

4.3 The Performance of the Delphi Groups

We formulated two hypotheses regarding the performance of the Delphi groups. We first test hypothesis *V*. To do so, we determine how frequently the measure of variance for the last round is smaller than that for the first round (cf. Table 4).

Hypothesis *V* cannot be refuted. When up to five rounds are carried out in Delphi groups a reduction of variance of the estimates takes place.

A closer examination shows that it cannot be rejected that variance reduction appears with equal frequency in all groups, and that variance reduction occurs independent of the type of question (chi-square test, 5 percent level).

Table 4
Frequency of Variance Reduction in Delphi Groups, Round Five compared with Round One (% of all possible cases)

Group Size	Fact-Finding Questions	Forecasting Questions
5	100	80
7	90	80
9	90	100
11	100	100

Hypothesis *R* will be tested now for judging performance. In order to represent the performance of a group for a certain type of question, individual performances must be aggregated. Since the individual performances are "index numbers" only the geometric mean can be chosen for this. At first we turn to the fact-finding questions. If considered individually, it becomes evident that the lowest and the highest median group error (taken as an absolute value) lie with approximately equal frequency in the first two rounds (cf. Table 5). In case of identical figures for the observed variable, its first appearance was considered.

Table 5
Relative Frequency of the Lowest (and Highest) Median Group Error by Number of Rounds, Group Size, and Type of Question

Group Size	Round (Fact-Finding Questions)				
	1	2	3	4	5
5	0.70(0.40)	0.20(0.40)	0.10(0.20)	0.0(0.0)	0.0(0.0)
7	0.40(0.70)	0.30(0.20)	0.10(0.0)	0.10(0.10)	0.10(0.0)
9	0.40(0.80)	0.30(0.0)	0.30(0.0)	0.0(0.20)	0.0(0.0)
11	0.635(0.40)	0.125(0.20)	0.125(0.10)	0.0(0.10)	0.125(0.0)

Group Size	Round (Forecasting Questions)				
	1	2	3	4	5
5	0.50(0.70)	0.40(0.20)	0.0(0.10)	0.0(0.0)	0.10(0.10)
7	0.70(0.60)	0.10(0.20)	0.0(0.20)	0.10(0.0)	0.10(0.0)
9	0.70(0.70)	0.30(0.10)	0.0(0.0)	0.0(0.10)	0.0(0.10)
11	0.375(0.865)	0.50(0.0)	0.0(0.125)	0.0(0.0)	0.125(0.0)

The median values for each group, which are easily read from Table 5, all lie in the first or second round. If we consider the lowest median group errors (taken absolutely), we observe that round two evidently is of greater importance concerning forecasting questions than concerning fact-finding questions, whereas rounds one and three are of much greater importance for fact-finding questions than for forecasting questions. On the other hand, the highest median group errors (taken absolutely) are much more heavily concentrated in the first round for forecasts than for fact-finding questions.

This analysis does not, however, take into consideration the degree to which the medians deviate from the correct values. This may be evaluated by looking at the geometric mean of the individual errors for each round, each group, and each type of question. For it could be possible that good results deteriorate not at all or only very little, while poor results improve greatly with an increasing number of rounds.

Data to judge on this question are presented in Table 6.

Table 6
Geometric Mean of the Individual Errors (taken as absolute values)

	Round (Fact-Finding Questions)				
Group Size	1	2	3	4	5
5	0.20	0.19	0.18	0.23	0.27
7	0.13	0.09	0.07	0.10	0.11
9	0.22	0.14	0.09	0.17	0.16
11	0.24	0.29	0.23	0.28	0.22

	Round (Forecasting Questions)				
Group Size	1	2	3	4	5
5	0.28	0.27	0.28	0.35	0.35
7	0.44	0.42	0.43	0.36	0.36
9	0.25	0.20	0.16	0.16	0.16
11	0.12	0.09	0.10	0.10	0.10

Data are analyzed by Friedman's two-way analysis of variance by ranks.

A significant difference for the entire set of data for fact-finding questions barely fails to be demonstrated at the 10 percent level ($\chi_r^2 = 7.6$ compared with the tabulated value of 7.78). The computed value of χ_r^2 for forecasting questions is considerably lower (5.75) and fails the 20 percent level. The phenomenon that for fact-finding questions in all groups except the one with eleven participants the highest performance is attained in the third round does not influence the tests significantly. This is even less so for forecasting questions, where the best performance is achieved twice in the second round and otherwise in the fourth or fifth rounds.

If one assumes that people get bored with answering the same questions over and over again and if one therefore cuts out the results of the last round, we observe a minor increase in the test statistics. The calculated value for fact-finding questions exceeds the significance level of 10 percent. For forecasts the increase is so slight ($\chi_r^2 = 5.77$) that no further consequences can be drawn from it. Taking everything together, our results seem to indicate that it is not reasonable to extend the number of rounds in Delphi groups beyond the third round.

This would support hypothesis R while it refutes hypothesis R'. Since in all cases investigated it is not assumed that the self-rating of expertise varies with the number of rounds, the result cannot be tested further in this respect.

The observations of group performance are not supported by the results of the best individual performance. In no case, i.e., neither when all rounds are considered nor when the number of rounds taken into consideration is limited, can a significant difference in the best individual performance be observed which would vary with the number of rounds. The best individual performance is very often maintained over consecutive rounds. So, if one would have objective criteria by which one could pick real experts, one might expect a greater stability of their judgments as compared with that of all members of our present groups.

4.4 The Size of the Groups

We test hypothesis G directly, separately for face-to-face discussion groups and Delphi groups. We do not correct the hypothesis to include the distribution of expertise (as determined subjectively by self-ratings) between groups, as it has practically a random influence.

When the data in Table 6 are analyzed by the rows, significant differences (on the 0.1 percent level) show up in Friedman's analysis of variance by ranks ($\chi_r^2 = 12.84$ for fact-finding questions and $\chi_r^2 = 15$ for forecasting questions). It is clear that the group performance in all rounds may be rank ordered as follows:

Fact-finding questions: Group with seven participants on top, followed by the groups with nine, five, and eleven participants.

Forecasting questions: Group with eleven participants on top, followed by the groups with nine, five, and seven participants.

The groups of seven and eleven reverse their rank order of performance in fact-finding questions as compared with forecasting questions.

This observation cannot be explained by varying self-ratings of expertise, as can be read from the refutation of the hypothesis concerning a different distribution of expertise in fact-finding questions and forecasting questions (see section 4.1).

If one considers the best individual performances directly, the result for the groups is confirmed, except for a shift in the rank orders of the groups with five and eleven participants for fact-finding questions, and of the groups with seven and five participants for forecasting questions. The individual estimates of the participants can thus be considered to be one factor which influences the result.

The quality of the estimates could be determined by the frequency of information exchange between the participants.[53] The frequency of information exchange is shown in Table 7.

Table 7
Frequency of Information Exchange between the Participants in Delphi Groups

	Absolute Number of Information Exchanges				% of All Possible Exchanges of Information			
	After rounds				After rounds			
Group Size	1	2	3	4	1	2	3	4
5	7	9	4	2	17.5	22.5	10.0	5.0
7	10	13	11	15	25.0	32.5	26.5	37.5
9	43	33	17	20	53.8	41.3	21.3	25.0
11	26	17	11	11	40.7	26.5	17.2	17.2

We have aggregated data for fact-finding questions and forecasting questions as we have found that the frequency of information exchange does not vary significantly with the type of question (binomial test, 5 percent level).

We find that the absolute frequency of information exchange between the participants varies significantly between the individual groups ($\chi_r^2 = 19.8$ is significant on the 5 percent level). Nevertheless the group with nine participants clearly leads the sequence, followed by those groups with eleven, seven, and five participants. If the frequency of information exchange is related to the number of possibilities for information exchange (which depends on the number of questions as well as on the number of participants outside of the quartiles, of course) a significant difference between the groups can likewise be determined ($\chi_r^2 = 15.0$). In this case only the groups with seven and eleven participants change their positions in the rank order as compared with the rank order of absolute frequencies of information exchange. One could assume that

[53] See I. Lorge *et al.*, "Solutions by Teams and Individuals to a Field Problem at Different Levels of Reliability," *Journal of Educational Psychology* 46 (1955), pp. 17–24.

the participants themselves, as keepers of information, channel varying amounts of information into the group. If so, it should be possible to find a rank correlation between group performance and the frequency of information exchange with regard to each question within each group. It turns out, however, that a significant result (5 percent level) can be identified only for the group with nine participants.

Low, and in part negative, values for the rank correlation, particularly in the group with eleven participants, can probably be explained by the fact that the opportunities for information exchange were used to transmit signs of impatience toward the end of each session. These were not considered to contribute to group performance in the stipulated sense, and thus were not counted in selecting the data of Table 7.

In the face-to-face discussions group performance is distributed differently (see Table 8). For the fact-finding questions we discover the following:

If performance is measured again by a median group error which is composed of individual estimates given before entering discussion, we find that the group with seven participants attains the highest level of performance. The groups with ten, four, and eight participants follow in that order. It should be noted that these data are only approximately comparable to those for round one in Table 7 because the initial data used here are given before the start of the exchange of information. If we take the geometric mean of the absolute values of the differences between the group estimate and the correct values, the group with ten participants appears at the top of the scale of performance. The groups with seven, four, and eight participants follow.

Let us now turn to the forecasting questions.

As before, a corresponding rank order of group performance cannot be

Table 8
Geometric Mean of the Median of Individual Relative Errors (taken as absolute values) before Discussion and the Relative Error of the Group Estimate in Face-to-Face Discussion Groups

	Fact-Finding Questions		Forecasting Questions	
Group Size	Geometric Mean Individual Errors	Group Errors	Geometric Mean Individual Errors	Group Errors
4	0.171	0.163	0.17[illegible]	0.184
7	0.116	0.154	0.103	0.147
8	0.257	0.220	0.072	0.121
10	0.141	0.139	0.187	0.212

demonstrated for the two types of group structure. Here the group with eight participants leads, and the lowest level of performance is attained by the group with ten participants (see Table 8).

When these results are interpreted it should be noted that the fact that in several cases one member of the Delphi groups did not participate in the face-to-face discussion groups does not affect the results of the latter in a uniform fashion.

Moreover, the direct observation of the group activity suggests that the number of members present in a discussion group cannot be considered the decisive variable. It is more important to consider the number of members of a face-to-face discussion group who actively take part in the discussion, as the nonparticipants do not influence the group judgment in any way. This situation occurs in our face-to-face discussion groups. According to our observations, in the group with eight participants, one member of the group only very rarely took part in the discussions. In the group with ten members two persons did not express any opinions during the entire experiment. When the group average was determined by voting, they conformed to the majority opinion, which was apparent. Another person only rarely joined in the consultations. Thus the "active" part of the group is almost always reduced to seven persons.[54]

Since in the two smaller groups the group itself required that each member participate on each question, there the "active" group corresponds in size with the entire group.[55] Thus our observation material is reduced practically to "active" discussion groups with four, seven, and possibly eight participants.

After these reservations as to the interpretation of the results, it is not surprising that the rank order of performances, if partitioned with respect to fact-finding questions and forecasting questions, as well as with respect to both types of group structures, does not generally exhibit comparable results in groups of varying size. Results coincide in the smallest group only. This, however, cannot be considered significant.

4.5 Hypothesis *D*

The problems of drawing a comparison between the results from the Delphi groups and the face-to-face discussion groups have already been mentioned. Besides, one has to find a generally accepted criterion on which to base the

[54]Unfortunately, only short notices about the impressions of the author after each session of the face-to-face discussion groups exist as a proof of these statements. Video tapes of the discussions do not exist; they would certainly have contributed to the interpretation of the results.

[55]Our observations coincide with the statements of Alter *et al.*, who confirm clique formation with disturbing internal discussion, a higher absolute proportion of inactive group members, and poorer agreements with dominant participants as dysfunctional effects in large groups in brainstorming sessions. Cf. U. Alter, H. Geschka, H. Schlicksupp, "Methoden und Organisation der Ideenfindung," report on a group project of Battelle Institute, Frankfurt (July 1972), Methodological Appendix, p. 20.

comparison. Dalkey's statements are based on the number of cases of superior performance. However, the interest in group performance can be based with at least the same right on the amount of the errors that occur. We take up this latter point.

The geometric average of all group performances for fact-finding questions in face-to-face discussion groups (0.167) is higher than the corresponding value for the third round of Delphi groups, which is as low as 0.1[illegible]. However, the result is reversed for forecasts. Here the corresponding value of 0.209 for the third round in Delphi groups is higher than the result of 0.162 for the face-to-face discussion groups. Thus an unequivocal relationship cannot be established. The performance of the only group with identical size under different organizational structures also differs greatly.

It is further noteworthy that with regard to the forecasts, the discussion in the discussion groups shows no progress in performance if the results are compared with the geometric means of individual errors before discussion. On the other hand, the result of the discussion in all the Delphi groups is that the mean error is reduced up to the third round. Furthermore, we find that the performance level of the face-to-face discussion groups is approximately equal for fact-finding questions and for forecasts, whereas the performance level of the Delphi groups with regard to forecasts is much lower as compared with fact-finding questions.

4.6 Hypothesis *F*

The preceding statements have already made it clear that different results are definitely produced when group performance with regard to fact-finding questions is differentiated from group performance with regard to forecasts (cf. Tables 6, 8). A general confirmation or refutation of hypothesis *F* on the basis of group performance is not possible. We therefore formulate and test the auxiliary hypothesis 5 in addition to what we have found until now. It says:

There exists a positive relationship between the rank order of performance of individuals in each group with regard to fact-finding questions and forecasts. Performance is defined as the geometric mean of the individual errors (taken as absolute values) in answering fact-finding questions and forecasting questions.

The auxiliary hypothesis is refuted in each of the groups. With increasing group size we calculate rank correlations of -0.300, -0.391, $+0.357$ and -0.150. In the majority of cases not even the sign corresponds to the expectations of the auxiliary hypothesis.

5. *Summary*

Although one should not overestimate the results from the very few experiments presented here, they do lead to doubts as to the efficiency of the Delphi method. Of course, it must be admitted that the attempt to use the Delphi

method for short-term forecasting is a comparatively tough test, for it was originally designed for long-range forecasting. At another occasion (sales estimates) it was also found that the errors of short-term forecasts can be very much higher than those of long-term forecasts.[56] If one assumes that the results of the forecasts could be interpreted as an attempt to estimate an unknown status quo at the time when the experiments took place as well, then they should be corrected by the average value of the relative difference between the status quo and the realization of each topic. These corrections vary between 0.042 and 0.098 in the individual groups, according to the choice of questions. Their application does not alter the rank order of the results as compiled in Tables 6 and 8. Therefore we may refer directly to the text with our summary:

(1) It cannot be discerned that fact-finding questions are suitable test material for recognizing expertise or appropriate organizational structures for forecasting groups.

(2) A general positive relationship between group size and group performance cannot be recognized.

(3) In face-to-face discussion groups the measure of the group size must be determined by the number of active participants. Appropriate precautions should be developed.

(4) Variance reduction almost always occurs in Delphi groups between the first and the fifth rounds, but the best results are as a rule already known in the third round. Further rounds may impair the results.

(5) Self-ratings of expertise show a positive relationship to the performance of the persons questioned in only two of four Delphi groups. They tend to be lower in face-to-face discussion groups than in the Delphi groups, and are determined substantially by the extent of professional experience rather than being set with regard to the questions in case. It is important to employ and develop better methods for the determination of expertise.

(6) A direct comparison of Delphi groups and face-to-face discussion groups was not possible because several participants dropped out. However, the results, if separated for fact-finding questions and forecasts, do not point in one direction.

(7) Proponents of the Delphi method will point out that our subjects, being banking experts, are better able to express themselves in a face-to-face discussion than on a display, even if its use has been explained to them. This should apply particularly when the space for exchanging information among participants is limited. The fear of making mistakes in the operation of the display could lead to exaggerated caution. However, if one agrees with this argument, the first point of this summary has to be explained also, as the results are not uniform in this respect. Anyhow, it appears to be important to avoid a "new

[56]J. Berthel, D. Moews, *Information und Planung in industriellen Unternehmungen*, Berlin, 1970, pp. 158 ff. (with data from fifteen firms).

Taylorism," on which some *mis*-concepts are founded.[57] However, it must be granted that the originators of the Delphi method did not say that it has to be operated as a computer dialogue.

(8) Only in the Delphi group with the greatest exchange of information did we observe a positive relationship to group performance. The results indicate that in small Delphi groups more opportunities for information exchange should be given. However, it probably must be tested whether the information given by the participants does coincide with what others would want to know, i.e., whether it adds to their knowledge.[58] How can the "confusion effect" in the majority of the discussions, which is recognized when the "reference values" mentioned there are compared (cf. Table 8), be explained without this distinction and the assumption of a difference between information supplied and information demanded?

(9) It must be admitted that in our strongly discipline-oriented group there has been relatively little opportunity for improving estimate by sharing information as compared to interdisciplinary groups concerned with other tasks. However, this affects all our groups in the same way. This criticism would be valid if it could be demonstrated that different groups react differently to these two types of tasks.

[57]See W. Kirsch, "Auf dem Wege zu einem neuen Taylorismus?" in H. R. Hansen, M. P. Wahl (eds.), *Probleme bein Aufbau betrieblicher Informationssysteme*, Müchen, 1973, pp. 338–48.

[58]E. Witte (ed.), *Das Informationsverhalten in Entscheidungsprozessen*, Tübingen, 1972, especially, pp. 44ff.; R. Bronner, E. Witte, B. R. Wossidlo, "Betriebswirtschaftliche Experimente zum Informations Verhalten in Entscheidungsprozessen," in E. Witte, *op. cit.*, pp. 186ff.

publications in the Delphi area is also included. Another group, spun off from IFF, is the Futures Group, a profit concern willing to do Delphi studies on a proprietary basis. While some of its work is not available to the public, we have included a list of its recent efforts to indicate the scope of application for Delphi in this sort of environment.

In the next section we find essentially all other articles on Delphi not encompassed by these journals or organizational headings. One can see from this list the beginnings of university, corporate, and governmental use of the technique. Unfortunately, many users have not published their results and are not, therefore, represented in this list.

For a great many years Delphi has been associated with the subject of Technological Forecasting. We have provided a separate list of articles on technological forecasting which do discuss Delphi. Our caution to the reader about this set of references is that some of these writings tend to define and consider Delphi solely within the scope of that one application area. It has been our intent in this book to present Delphi in the wider context of an alternative communication form.

There is evident in the literature a set of new directions having to do with the application of the Delphi concept in such areas as computerized conferencing, management information systems, citizen feedback, participatory democracy, etc. For convenience we have placed these references in a separate list.

For those seeking references in other languages we have compiled a separate list of some foreign language articles on Delphi.

The final section of this bibliography is perhaps the most important to the potential Delphi designer. It was gathered by perusing the papers on Delphi and asking what writings in other disciplines have been referenced there. The papers in this list represent work in the fields of economics, operations research, philosophy, planning, psychology, sociology, and statistics. In essence, this provides a guide to the techniques and knowledge in other areas of utility to the study and use of Delphi. As one would expect, psychology and sociology make up a large part of the referenced material.

The lists which follow represent a search of materials readily available to the editors through 1974. We apologize in advance to those authors we have missed. If anyone has additions to make, and so informs us, we will compile them for an update in a future edition of this book.

Distribution of Bibliography

The Distribution of articles in the bibliography is expressed in the following table. While Delphi has been around since the early fifties, much more material has been published on the subject in the last four years than in all the years before 1970.

	Before 1970	1970–1974	Total
Journal Articles	14	54	68
Rand Papers	34	4	48
IFF and Futures Group	13	67	80
Other Delphi Articles	45	95	140
Technological Forecasting	19	25	44
New Directions	3	61	64
Foreign Language Articles	6	39	45
Total of Delphi References	134	355	489
Related Work	136	45	181
Total	270	400	670

Journal Articles

Alderson, R. C. *et al.*, "Requirements Analysis, Need Forecasting, and Technology Planning Using the Honeywell PATTERN Technique," *TFSC* 3, No. 2 (1972).

Amara, Roy C., "A Note on Cross-Impact Analysis," *FUTURES* 4, No. 3 (1972).

——, and Andrew J. Lipinski, "Some Views on the Use of Expert Judgment," *TFSC* 3, No. 3 (1972).

——, and Gerald R. Salancik, "Forecasting: From Conjectural Art Toward Science," *TFSC* 3, No. 4 (1972).

Ament, R., "Comparison of Delphi Forecasting Studies in 1964 and 1969," *FUTURES* 2, No. 1 (March 1970).

Beaton, A. E., "Scaling Criterion of Questionnaire Items," *Socio-Economic Planning Sciences* 2, Nos. 2, 3, 4 (1969).

Bernstein, G. B., and Marvin Cetron, "SEER: A Delphic Approach Applied to Information Processing," *TFSC* 1, No. 1 (June 1969).

Bender, Strack, Ebright, and von Haunalter, "Delphic Study Examines Developments in Medicine," *FUTURES* 1, No. 4 (June 1969).

Bjerrum, C. A., "Forecast of Computer Development and Applications, 1968-2000," *FUTURES* 1, No. 4 (June 1969).

Blackman, A. Wade, "A Cross-Impact Model Applicable to Forecasting for Long Range Planning," *TFSC* 5, No. 3 (1973).

——, "The Use of Bayesian Techniques in Delphi Forecasts," *TFSC* 2, No. 3/4 (1971).

Campbell, George S., "Relevance of Signal Monitoring to Delphi Cross-Impact Studies," *FUTURES* 3, No. 4 (December 1971).

Cetron, M. J., and G. B. Bernstein, "SEER: A Delphic Approach Applied to Information Processing," *TFSC* 1, No. 1 (Spring 1969).

Cole, H. S. D., and J. Metcalf, "Model Dependent Scale Values for Attitude Questionnaire Items," *Socio-Economic Planning Sciences* 5, No. 9 (1971).
Currill, D. L., "Technological Forecasting in Six Major U. K. Companies," *Long Range Planning* 5, No. 1 (March 1972).
Dalkey, Norman C., "An Elementary Cross-Impact Model," *TFSC* 3, No. 3 (1972).
——, "An Experimental Study of Group Opinion: The Delphic Method," *FUTURES* 2, No. 3 (September 1969).
——, "Analyses from a Group Opinion Study," *FUTURES* 1, No. 6 (December 1969).
——, B. Brown, and S. Cochran, "Use of Self-Ratings to Improve Group Estimates," *TFSC* 1, No. 3 (March 1970).
Day, Lawrence H., "Long Term Planning in Bell Canada," *Long Range Planning* London, U. K. (submitted for publication, 1973).
Decker, Robert L., "A Delphi Survey of Economic Development," *FUTURES* 6, No. 2 (April 1974).
Derian, J. C., and F. Morize, "Delphi in the Assessment of R&D Projects," *FUTURES* 5, No. 5 (October 1973).
Dobrov, G. M., and L. P. Smirnov, "Forecasting as a Means for Scientific and Technological Policy Control," *TFSC* 4, No. 1 (1972).
Doyon, L. R., T. V. Sheehan, and H. I. Zayor, "Classroom Exercises in Applying the Delphi Method for Decision-Making," *Socio-Economic Planning Sciences* 5, No. 4 (1971).
Dressler, Fritz R. S., "Subject Methodology in Forecasting," *TFSC* 3, No. 4 (1972).
Enzer, Selwyn, "A Case Study Using Forecasting as a Decision-Making Aid," *FUTURES* 2, No. 4 (December 1970).
——, "Cross-Impact Techniques in Technology Assessment," *FUTURES* 4, No. 1 (March 1972).
——, "Delphi and Cross-Impact Techniques—An Effective Combination for Systematic Futures Analysis," *FUTURES* 3, No. 1 (March 1971).
Goodman, Joel, "Delphi and the Law of Diminishing Returns," *TFSC* 2, No. 2 (1970).
Goodwill, Daniel F., "A Look at the Future Impact of Computer-Communications on Everyday Life," *TFSC* 4, No. 2 (1972).
Gordon, T. J., "Cross-Impact Matrices: An Illustration of Their Use for Policy Analysis," *FUTURES* 1, No. 6 (1969).
——, "Potential Changes in Employee Benefits," *FUTURES* 2, No. 2 (1970).
——, S. Enzer, and R. Rochberg, "An Experiment in Simulation Gaming for Social Policy Studies," *TFSC* 1, No. 3 (March 1970).
——, and H. Hayward, "Initial Experiments with the Cross-Impact Matrix Method of Forecasting," *FUTURES* 1, No. 2 (December 1968).
——, R. J. Wolfson, and R. C. Sahr, "A Method for Rapid Reproduction of Group Data and Individual Estimates in the Use of the Delphic Method," *FUTURES* 1, No. 6 (December 1969).
Grabbe, Eugene M., and Donald L. Pyke, "An Evaluation of the Forecasting of Information Processing Technology and Applications," *TFSC* 4, No. 2 (1972).
Helmer, O., "Cross-Impact Gaming," *FUTURES* 4, No. 2 (June 1972).
Huckfeldt, Vaughn E., and Robert C. Judd, "Issues in Large Scale Delphi Studies," *TFSC* 6, No. 1 (1974).
Johnson, Howard E., "Some Computational Aspects of Cross-Impact Matrix Forecasting," *FUTURES* 2, No. 2 (June 1970).
Judd, Robert C., "Use of Delphi Methods in Higher Education," *TFSC* 4, No. 2 (1972).
Kane, Julius, "A Primer for a New Cross-Impact Language—KSIM," *TFSC* 4, No. 2 (1972).
Lachmann, Ole, "Personnel Administration in 1980: A Delphi Study," *Long Range Planning* 5, No. 2 (June 1972).
Martino, Joseph P., "The Lognormality of Delphi Estimates," *TFSC* 1, No. 4 (Spring 1970).
——, "The Optimism/Pessimism Consistency of Delphi Panelists," *TFSC* 2, No. 2 (1970).
——, "The Precision of Delphi Estimates," *TFSC* 1, No. 3 (March 1970).
Overbury, R. E., "Technological Forecasting: A Criticism of the Delphi Technique," *Long Range Planning* 1, No. 4 (June 1969).
Pavitt, Keith, "Analytical Techniques in Government Science Policy," *FUTURES* 4, No. 1 (March 1972).

Pill, Juri, "The Delphi Method: Substance, Context, A Critique and An Annotated Bibliography," *Socio-Economic Planning Science* 5, No. 1 (1971).
Pyke, Donald L., "A Practical Approach to Delphi," *FUTURES* 3, No. 2 (June 1970).
Rochberg, Richard, "Convergence and Viability because of Random Numbers in Cross-Impact Analyses," *FUTURES* 2, No. 3 (September 1970).
——, "Information Theory, Cross-Impact Matrices, and Pivotal Events," *TFSC* 2, No. 1 (1970).
Rowland, D. G., "Technological Forecasting and the Delphi Technique—A Reply," *Long Range Planning* 2, (December 1969).
Salancik, J. R., "Assimilation of Aggregated Inputs into Delphi Forecasts: A Regression Analysis," *TFSC* 5, No. 3 (1973).
——, William Wenger, and Ellen Helfer, "The Construction of Delphi Event Statements," *TFSC* 3, No. 1 (1971).
Sandow, Stuart, "The Pedagogy of Planning: Defining Sufficient Futures," *FUTURES* 3, No. 4 (December 1971).
Schneider, Jerry B., "The Policy Delphi: A Regional Planning Application," *TFSC* 3, No. 4 (1972).
Smil, Vaclav, "Energy and the Environment—A Delphi Forecast," *Long Range Planning* 5, No. 4 (December 1972).
Stover, John, "Improvements to Delphi/Cross-Impact," *FUTURES* 5, No. 3 (June 1973).
Sulc, Oto, "A Methodological Approach to the Integration of Technological and Social Forecasts," *TFSC* 2, No. 1 (June 1969).
——, "Interaction between Technological and Social Changes: A Forecasting Model," *FUTURES* 1, No. 5 (September 1969).
Teeling-Smith, George, "Medicine in the 1990's: Experience with a Delphi Forecast," *Long Range Planning* 3, No. 4 (June 1971).
Turoff, Murray, "An Alternative Approach to Cross-Impact Analysis," *TFSC* 3, No. 3 (1972).
——, "Delphi Conferencing: Computer-Based Conferencing with Anonymity," *TFSC* 3, No. 2 (1970).
——, "The Design of a Policy Delphi," *TFSC* 2, No. 2 (1970).
——, "Meeting of the Council on Cybernetic Stability: A Scenario," *TFSC* 4, No. 2 (1972).
Umpleby, Stuart, "Is Greater Citizen Participation in Planning Possible and Desirable?" *TFSC* 4, No. 1 (1971).
Welty, Gordon, "Plato and Delphi," *FUTURES* 5, No 3 (June 1973).
Wills, Gordon, "The Preparation and Development of Technological Forecasts," *Long Range Planning* 2, No. 3 (March 1970).

Rand Papers in Chronological Order

Sackman, H., *Delphi Assessment: Expert Opinion, Forecasting and Group Process*, R-1283-PR, April 1974.
Brown, Thomas, *An Experiment in Probabilistic Forecasting*, R-944-ARPA, 1972.
Dalkey, Norman C., *An Elementary Cross-Impact Model*, R-677-ARPA, May 1971.
——, and Bernice Brown, *Comparison of Group Judgment Techniques With Short-Range Predictions and Almanac Questions*, R-678-ARPA, May 1971.
Schmidt, D. L., *Creativity in Industrial Engineering*, P-4601, March 1971.
Dalkey, Norman C., and Daniel L. Rourke, *Experimental Assessment of Delphi Procedures With Group Value Judgments*, R-612-ARPA, February 1971.
Quade, E. S., *On the Limitations of Quantitative Analysis*, P-4530, December 1970.
Farquhar, J. A., *A Preliminary Inquiry into the Software Estimation Process*, RM-6271-PR, August 1970.
——, R. J. Lewis, and D. Snyder, *Measurement and Analysis of the Quality of Life: With Exploratory Illustrations of Applications to Career and Transportation Choices*, RM-6228-DOT, August 1970.
Quade, E. S., *An Extended Concept of "Model,"* P-4427, July 1970.

Brown, T. A., *Probabilistic Forecasts and Reproducing Scoring Systems*, RM-6299-ARPA, June 1970.
Quade, E. S., *Cost Effectiveness: Some Trends in Analysis*, P-3529-1, March 1970.
Dalkey, Norman C., H. Brown, and S. W. Cochran, *The Delphi Method, IV: Effect of Percentile Feedback and Feed-In of Relevant Facts*, RM-6118-PR, March 1970.
Paxson, Edwin, *A Delphi Examination of Civil Defense*, RM-6247-ARPA, February 1970.
Dalkey, Norman C., H. Brown, and S. W. Cochran, *The Delphi Method, III: Use of Self Rating to Improve Group Estimates*, RM-6115-PR, November 1969.
Rescher, N., *Delphi and Values*, P-4182, September 1969.
Dole, S. H., G. H. Fischer, E. D. Harris, and J. String, Jr., *Establishment of a Long-Range Planning Capability*, P-3540, September 1969.
Brown, Bernice, S. Cochran, and N. Dalkey, *The Delphi Method, II: Structure of Experiments*, RM-5957-PR, June 1969.
Pinkel, B., *On the Decision Matrix and the Judgment Process: A Developmental Decision Example*, P-3620, June 1969.
Dalkey, Norman C., *The Delphi Method: An Experimental Study of Group Opinion*, RM-5888-PR, June 1969.
Quade, E. S., *The Systems Approach and Public Policy*, P-4053, March 1969.
Dalkey, Norman C., *Predicting the Future*, P-3948, October 1968.
Brown, Bernice B., *DELPHI Process: A Methodology Used for the Elicitation of Opinions of Experts*, California, P-3925, September 1968.
Quade, E. S., and W. I. Boucher, *Systems Analysis and Policy Planning: Applications in Defense*, California, June 1968.
Dalkey, Norman C., *Experiments in Group Prediction*, P-3820, March 1968.
——, *Quality of Life*, P-3805, March 1968.
Rochberg, R., *Some Comments on the Problem of Self Affecting Predictions*, P-3735, December 1967.
Helmer, Olaf, *Systematic Use of Expert Opinions*, P-3721, November 1967.
Dalkey, Norman C., *DELPHI*, P-3704, October 1967.
Helmer, Olaf, *Prospects of Technological Progress*, California, P-3643, August 1967.
——, *Methodology of Societal Studies*, California, P-3611, June 1967.
——, *The Future of Science*, P-3607, May 1967.
Rescher, N., *The Future as an Object of Research*, P-3593, April 1967.
Helmer, O., *New Developments in Early Forecasting of Public Problems: A New Intellectual Climate*, P-3576, April 1967.
Haydon, Brownlee, *The Year 2000*, P-3571, March 1967.
Helmer, Olaf, *Analysis of the Future: The Delphi Method*, P-3558, March 1967.
Quade, E. S., *Cost-Effectiveness: Some Trends in Analyses*, P-3529, March 1967.
Neiswender, R., *"The Exploration of the Future," Réalités*, No. 245, June 1966. (Translated in P-3540, February 1967.)
Helmer, Olaf, *The Use of the Delphi Technique in Problems of Educational Innovations*, P-3499, December 1966.
Downs, Anthony, *Bureaucratic Structure and Decision Making*, RM-4646-1-PR, October 1966.
Jouvenel, B. de, *Futuribles*, P-3045, January 1965.
Helmer, Olaf, and Bernice Brown, *Improving the Reliability of Estimates Obtained from a Consensus of Experts*, P-2986, September 1964.
Gordon, Theodore J., and Olaf Helmer, *Report on a Long-Range Forecasting Study*, P-2982, September 1964.
Helmer, Olaf, *Convergence of Expert Consensus Through Feedback*, P-2973, September 1964.
Quade, E. S., *Analysis for Military Decisions*, R-387-PR, November 1964.
Helmer, Olaf, *The Systematic Use of Expert Judgment in Operations Research*, P-2795, September 1963.
, and E. S. Quade, *An Approach to the Study of a Developing Economy by Operational Gaming*, P-2718, March 1963.
Dalkey, Norman C., and Olaf Helmer, *The Use of Experts for the Estimation of Bombing Requirements—A Project Delphi Experiment*, RM-727-PR, November 1951.
Kaplan, A., A. L. Skogstad, and M. A. Girshick, *The Prediction of Social and Technological Events*, P-93, April 1949.

Publications of the Institute for the Future*

Amara, Roy C., "A Framework for National Science Policy Analysis," P-18 (published in *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-2, No. 1, January 1972).
——, "The Social Responsibilities," P-16, April 1972.
——, "Toward a Framework for National Goals and Policy Research," P-20 (published in *Policy Sciences*, Vol. 3, 1972).
——, and Andrew J. Lipinski, "Some Views on the Use of Expert Judgment," P-19 (published in *Technological Forecasting and Social Change*, Vol. 3, No. 3, 1972).
——, and J. R. Salancik, "Forecasting from Conjectural Art Toward Science," P-14 (published in *Technological Forecasting and Social Change*, Vol. 3, No. 4, 1972).
Ament, Robert H., "Comparison of Delphi Forecasting Studies in 1964 and 1969," P-9 (published in *FUTURES*, Vol. 2, No. 1, March 1970) (reprint).
Baran, Paul, *The Future of Newsprint, 1970–2000*, R-16, December 1971.
——, *Japanese Competition in the Information Industry*, P-17, May 1972.
——, *Notes on a Seminar on Future Broad-Band Communications*, WP-1, February 1970.
——, "On the Impact of the New Communications Media upon Social Values," P-3 (published in *Law and Contemporary Problems*, Vol. 34, No. 2, Spring 1969) (reprint).
——, *Potential Market Demand for Two-Way Information Services to the Home, 1970–1990*, R-2[illegible], December 1971.
——, and Andrew J. Lipinski, *The Future of the Telephone Industry, 1970–1985*, R-20, September 1971, a special industry report.
Becker, Harold S., *A Framework for Community Development Action Planning, Volume I: An Approach to the Planning Process*, R-18, February 1971.
——, *A Framework for Community Development Action Planning, Volume II: Study Procedure, Conclusions, and Recommendations for Future Research*, R-19, February 1971.
——, *A Method of Obtaining Forecasts for Long-Range Aerospace Program Planning*, WP-7, April 1970.
——, and Raul de Brigard, *Considerations on a Framework for Community Action Planning*, WP-9, June 1970.
Brigard, Raul de, and Olaf Helmer, *Some Potential Societal Developments: 1970-2000*, R-7, April 1970.
Enzer, Selwyn, *A Case Study Using Forecasting as a Decision-Making Aid*, WP-2, December 1969.
——, *Cross-Impact Techniques in Technology Assessment*, P-15.
——, *Delphi and Cross-Impact Techniques: An Effective Combination for Systematic Futures Analysis*, WP-8, June 1970.
——, *Federal/State Science Policy and Connecticut: A Futures Research Workshop*, R-24, October 1971.
——, *Some Commercial and Industrial Aspects of the Space Program*, P-5, November 1969.
——, *Some Developments in Plastics and Competing Materials by 1985*, R-17, January 1971.
——, *Some Prospects for Residential Housing by 1985*, R-13, January 1971.
——, Wayne I. Boucher, and Frederick D. Lazar, *Futures Research as an Aid to Government Planning in Canada: Four Workshop Demonstrations*, R-22, August 1971.
——, Wayne I. Boucher, and Frederick D. Lazar, *Futures Research as an Aid to Government Planning in Canada: Four Workshop Demonstrations—Supporting Appendices*, R-22A, August 1971.
——, and Raul de Brigard, *Issues and Opportunities in the State of Connecticut: 1970-2000*, R-8, March 1970.
——, Raul de Brigard, and Frederick D. Lazar, *Some Considerations Concerning Bankruptcy Reform*, IFF Report R-28, March 1973.
——, Dennis L. Little, and Frederick D. Lazar, *Some Prospects for Social Change by 1985 and Their Impact on Time/Money Budgets*, R-25, March 1972.
——, Gordon J. Slotsky, Dennis L. Little, James E. Doggart, and David A. Long, *The Automobile Insurance System: Current Status and Some Proposed Revisions*, WP-18, December 1971.

**Reports*—Formal documentation of Institute Studies
Working Papers—Preliminary contributions to the Institute's work for its sponsors
Papers—Individual contributions by Institute staff members to the professional literature

Gordon, Theodore J., *A Study of Potential Changes in Employee Benefits, Volume I: Summary and Conclusions*, R-1, April 1969.
——, *A Study of Potential Changes in Employee Benefits, Volume II: National and International Patterns*, R-2, April 1969.
——, *A Study of Potential Changes in Employee Benefits, Volume III: Delphi Study*, R-3, April 1969.
——, *A Study of Potential Changes in Employee Benefits, Volume IV: Appendices to the Delphi Study*, R-4, April 1969.
——, *Potential Institutional Arrangements of Organizations Involved in the Exploitation of Remotely Sensed Earth Resources Data*, P-6 (published as AIAA Paper No. 70-334, March 1970).
——, *Some Possible Futures of American Religion*, P-4, May 1970.
——, *The Current Methods of Futures Research*, P-11, August 1971.
——, and Robert H. Ament, *Forecasts of Some Technological and Scientific Developments and their Societal Consequences*, R-6, September 1969.
——, Selwyn Enzer, Richard Rochberg, and Robert Buchele, *A Simulation Game for the Study of State Policies*, R-9, September 1969.
——, Olaf Helmer, Selwyn Enzer, Raul de Brigard, and Richard Rochberg, *Development of Long-Range Forecasting Methods for Connecticut: A Summary*, R-5, September 1969.
——, Dennis L. Little, Harold L. Strudler, and Donna D. Lustgarten, *A Forecast of the Interaction between Business and Society in the Next Five Years*, R-21, April 1971.
——, Richard Rochberg, and Selwyn Enzer, *Research on Cross-Impact Techniques with Applications to Selected Problems in Economics, Political Science, and Technology Assessment*, R-12, August 1970.
Helmer, Olaf, *Long-Range Forecasting—Roles and Methods*, P-7, May 1970.
——, *Multipurpose Planning Games*, WP-17, December 1971.
——, *On the Future State of the Union*, R-27, May 1972.
——, *Political Analysis of the Future*, P-1, August 1969.
——, *Report on the Future of the Future-State-of-the-Union Reports*, R-14, October 1970.
——, *Toward the Automation of Delphi*, International Technical Memorandum, Institute for the Future, March 1970.
——, *et al., Development of Long-Range Forecasting Methods for Connecticut: A Summary*, IFF Report, R-5, September 1969.
——, and Helen Helmer, *Future Opportunities for Foundation Support*, R-11, June 1970.
Institute for the Future, "Development of a Computer-Based System to Improve Interaction Among Experts," First Annual Report, National Science Foundation, Grant GJ-35 326X, August 1973.
Institute for the Future, "Social Assessment of Mediated Group Communication: A Workshop Summary," March 1974.
Johansen, Robert, Richard H. Miller, Jacques Vallee, "Group Communication through Electronic Media: Fundamental Choices and Social Effects," Working Paper, March 1974.
Kramish, Arnold, *The Non-Proliferation Treaty at the Crossroads*, P-2, October 1969.
LeCompte, Gare, *Factors in the Introduction of a New Communications Technology into Syria and Turkey: Background Data*, WP-10, August 1970.
Lipinski, Andrew J., *Toward a Framework for Communications Policy Planning*, WP-19, December 1971.
Little, Dennis L., *Models and Simulation: Some Definitions*, WP-6, April 1970.
——, *STAPOL: Appendix to the Simulation Game Manual*, WP-14, October 1970.
——, and Raul de Brigard, *Simulation, Futurism, and the Liberal Arts College: A Case Study*, WP-15, April 1971.
——, and Richard Feller, *STAPOL: A Simulation of the Impact of Policy, Values, and Technological and Societal Developments upon the Quality of Life*, WP-12, October 1970.
——, and Theodore J. Gordon, *Some Trends Likely to Affect American Society in the Next Several Decades*, WP-16, April 1971.
——, Richard Rochberg, and Richard Feller, *STAPOL: Simulation-Game Manual*, WP-13, March 1971.
Nagelberg, Mark, *Simulation of Urban Systems: A Selected Bibliography*, WP-3, January 1970.
——, and Dennis L. Little, *Selected Urban Simulations and Games*, WP-4, April 1970.
Rochberg, Richard, *Information Theory, Cross-Impact Matrices, and Pivotal Events*, P-8 (published in *Technological Forecasting and Social Change* 2, No. 2, 1970)(reprint).
——, Theodore J. Gordon, and Olaf Helmer, *The Use of Cross-Impact Matrices for Forecasting and Planning*, R-10, April 1970.

Sahr, Robert C., *A Collation of Similar Delphi Forecasts*, WP-5, April 1970.
Salancik, J. R., Theodore J. Gordon, and Neale Adams, *On the Nature of Economic Losses Arising from Computer-Based Systems in the Next Fifteen Years*, R-23, March 1972.
Vallee, J. *et al.*, *Group Communication Through Computers*, R-32 and R-33, July 1974.
Wilson, Albert, and Donna Wilson, *Toward the Institutionalization of Change*, WP-11, August 1970.

Delphi Studies by the Futures Group

Since some of these studies are not available for general public distribution, a short description is included of some of the items.

Boucher, W. I., *Quantifiable Goal Statements for the U.S. Criminal Justice System: A Preliminary Assessment*, Report 53-30-01, May 1972.
Done for the National Advisory Commission on Criminal Justice Standards and Goals; members of the Commission served as respondents. May eventually be published as part of the report on the Commission's work. Time horizon: to 1980–1985.
——, *Report on a Hypothetical Focused Planning Effort (FPE)*, Report 43-01-06, February 1972.
——, and Theodore J. Gordon, *The Literature of Cross-Impact Analysis: A Survey*, Report 41-24-01/2, February 1972.
——, and J. Talbot, *Report on a Delphi Study of Future Telephone Interconnect Systems*, Report 49-26-01, April 1972.
The first Delphi conducted entirely through interviews; also the first (to our knowledge) to have as its goal the conceptual design of a physical system.
Becker, H. S., and J. Incerti, *Future Technological and Social Changes: A Delphi Study of Opportunities and Threats for an Industrial Firm*, Report 76-35-02, January 1973.
Done for a multidivisional, hard-technology corporation; the aim was to identify prospective social and technological developments relevant to the diversions, determine their probability, indicate who might be most instrumental in making them happen, assess their impact if they did happen, and define policies appropriate for the company to take. Time horizon: to 1982 +.
Gordon, T. J., and Harold A. Becker, *Analysis of Ailing Products—It's Decisions That Count (A Case Study of R & D Methodology)*, Report 24-15-01, February 1972.
——, and J. Cohen, *An Investigation of Future Technological and Social Developments and Their Implications for Entry into the Produce Market*, Report 66-36-01, October 1972.
The third "interview Delphi"; here the attempt was to determine whether, in view of possible changes in technology and society generally, it made sense for a particular company to consider becoming a produce supplier. Time horizon: to 1982.
——, William L. Renfro, and W. I. Boucher, *Challenges and Opportunities in the Photographic Industry: Report on a Focused Planning Effort*, Report 85-42-03, June 1973.
Weaver, W. T., *A Delphi Study of the Potential of a New Communications System*, Report 58-26-02, August 1972.
The second "interview Delphi": again, the focus fell on the conceptual design of a specific physical product. Time horizon: to 1975–76.

Other Delphi Articles

Adelman, I., and Morris, C. T., "Society, Politics and Economic Development: A Quantitative Approach," Johns Hopkins Press, Baltimore, 1967.
Adelson, M., M. Alkin, C. Carey, and O. Helmer, "The Education Innovation Study," *American Behavioral Scientist* 10, No. 7 (1967).

...inut, Bruce, *A Study of Consensus on Psychological Factors Related to Recovery from Nuclear Attack*, Human Sciences Research, Inc., HSR-RR-71/3 Office of Civil Defense, May 1971.

—, *A Review of the Development and Application of the Delphi Method*, Human Sciences Research, Inc., September 1970.

...hary, E. *et al.*, *Communication Needs of the Seventies and Eighties*. AT&T Marketing and Western Electric Engineering, December 1971.

...rach, J. L., and R. J. Twery, "The Application of Forecasting Techniques in Research and Development Planning," *Presented at AICHE meeting*, November 1969.

...yne, C. Dudley, Jr., and Walter Price, "Technological Forecasting," in Richard F. Vancil (ed.), *Formal Planning Systems—1971*, Harvard Business School, Cambridge, Mass.

...ford, Michael, T., "The Future of Communications Services in the Home," *Bell Canada*, September 1972.

...il, Daniel, "The Oracles at Delphi," *The Public Interest*, Fall 1965.

...ender, A. D., A. E. Strack, G. W. Ebright, and G. von Haunalter, *A Delphic Study of the Future of Medicine*, Smith, Kline, and French Laboratories, January 1969.

...rghofer, D. E., "General Education in Post-Secondary Non-University Institutions in Alberta," Research Studies in Post-Secondary Education, No. 9 (Edmonton, Alberta: Alberta College Commission, April 1970).

...rnstein, George, *et al.*, *A Fifteen Year Forecast of Information Processing Technology*, Naval Supply Systems, AD681752, January 1969.

...rko, Harold, "Predicting Needs in Librarianship and Information Science Education," Proceedings of the ASIA Annual Meeting, Vol. 7, 1970.

...ch, Ronald, *Britain 1984: Unilever's Forecast: An Experiment in Economic History of the Future*, Dartan, Longman & Todd Ltd., London, 1963.

...wn, Bernice, "Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts," *ASTME Vectors* 3, No. 1 (1968).

—, "Technological Forecasting by Iterative Guesstimation," *Product Design and Value Engineering* 13, (October 1968).

...mpbell, Robert M., "A Methodological Study of the Utilization of Experts in Business Forecasting," Ph.D. dissertation, University of California, Los Angeles, California, 1966.

—, and David Hitchin, "The Delphi Technique: Implementation in the Corporate Environment," *Management Services*, Report Number 11, Graduate School of Business Administration, U. C., March 1969.

...tron, Marvin J., and Don Overly, "Toward a Consensus on the Future," *Innovation*, No. 31, May 1972.

...ambers, J. C. *et al.*, "Catalytic Agency for Effective Planning," *Harvard Business Review*, January-February 1971.

...apman, R. L., and F. N. Cleaveland, *Meeting the Needs of Tomorrow's Public Service*, Report of the National Academy of Public Administration, January 1973.

...arke, S. C. T., and H. T. Coutts, "The Future of Teacher Education," unpublished manuscript, Edmonton, Alberta: Faculty of Education, University of Alberta, 1971.

...aver, J. Kenneth, "A Practical Method for Technology Assessment," paper presented at the First International Conference on Technology Assessment, held in The Hague, The Netherlands, May 27-June 2, 1973.

—, "Technology Assessment by Cross-Impact," paper presented at the Engineering Foundation Conference on Technology Assessment held at New England College, Henniker, N. H., July 1972.

...phert, Frederick, and Walter L. Gant, "The Delphi Technique," *Journal of Teacher Education* 21, No. 3 (1970).

—, and Walter L. Gant, "The Delphi Technique: A Case Study," *PHI DELTA KAPPAN*, January 1971.

... and Walter L. Gant, "The Delphi Technique: A Tool for collecting Opinions in Teacher Education," *Journal of Teacher Education* 21, No. 3, (1970).

...oy, James F., "Practical Refinements to the Cross-Impact Matrix Technique," in *Industrial Applications of Technological Forecasting*, Cetron and Bartocha (eds.), Wiley-Interscience, New York, 1971.

Dalkey, Norman C., *Studies in the Quality of Life* (Delphi and Decision-making), Lexington Books, 1972.

——, and Olaf Helmer, "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science* 9, (1968).

Darling, Charles M., III, *An Experimental 1968 Public Affairs Forecast*, Available from: National Industrial Conference Board, New York, N. Y.

Davidson, Frank P., "Futures Research: A New Scientific Discipline," *Proceedings of the Social Statistics Section*, American Statistical Association, 1969.

Day, Lawrence H., "The Future of Computer and Communications Services," AFIPS Conference *Proceedings* 43, (1973): 723–34.

Dean, B. V., S. Mathis, *et al.*, *Analysis of the Exploratory Development Project Evaluation Experiment*, Case Western Reserve University School of Management, TM 165, November 1969.

Delbecq, A. C., and A. Van DeVen, *A Group Process Model for Problem Identification and Program Planning*, University of Wisconsin, Madison, October 1970.

Doak, Nick, "Future Rubber Processing," *Rubber Journal*, November 1972.

Doyle, Frank J., and Daniel Z. Goodwill, "An Exploration of the Future in Educational Technology," *Bell Canada*, January 1971.

——, and Daniel Z. Goodwill, "An Exploration of the Future in Educational Technology," in H. A. Stevenson, R. M. Stamp, J. D. Wilson (eds.), *The Best of Times, The Worst of Times—Contemporary Issues in Canadian Education*, Holt, Rinehart and Winston, Montreal, 1972.

——, and Daniel Z. Goodwill, "An Exploration of the Future of Medical Technology," *Bell Canada*, 1971.

Dyck, H. J., and G. J. Emery, *Social Futures for Alberta*, Westrede Institute, Edmonton, Alberta, 1970.

Emlet, H. E., and John Williamson, "Alternative Methods for Eliciting Estimates for Health-Care Benefits and Required Resources," presented at Joint ORSA-TIMS meeting, Atlantic City, November 1972.

Enzer, Selwyn, "Applications of Futures Research to Society's Problems," in Nigel Hawkes (ed.), *International Seminar on Trends in Mathematical Modeling*, Venice, December 13–18, 1971 (New York: Springer-Verlag, 1973), pp. 243–87.

Feldman, Philip A., "Technology Assessment of Computer-Assisted-Instruction," *Bell Canada*, Business Planning, Montreal, August 1972 (internal document).

Firschler, Martin A., and L. S. Coles, "Forecasting and Assessing the Impact of Artificial Intelligence on Society," Proceedings from International Joint Conference on Artificial Intelligence, Stanford Research Institute, August 1973.

Folk, Michael, "A Critical Look at the Cross-Impact Method," Research Report RR-5, Educational Policy Research Center, Syracuse University Research Corporation, August 1971.

——, "Computer and Educational Futures Research," *The Potential of Educational Futures*, Warren Ziegler and Michael Marien (eds.). Part of NSSE Series on Contemporary Education. Charles A. Jones, Worthington, Ohio, 1972.

Fulmer, Robert M., "Forecasting the Brave New World of 1984," *Manage*, March 1971.

——, "Forecasting the Future," *Managerial Planning*, July/August 1972.

——, *Managing Associations for the 1980's*, report for the American Society of Association Executives, January 1972.

——, "On Forecasting the Future," Inspection News, May 1971.

Glazier, Frederick, P., "A Multi-Industry Experience in Forecasting Optimum Technological Approach," in James R. Bright and M. E. F. Shoeman (eds.), *A Guide to Practical Technological Forecasting*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

Golding, Edwin, "A Technological Forecast in Transportation 1975–2000," 39th ORSA meeting, May 1971.

Goodwill, D. Z., "An Exploration of the Future in Business Information," *Bell Canada*, October 1971.

Gordon, P., and MacReynolds, W., "Optimal Urban Forms," *Journal of Regional Science* 14, No. 2, August 1974.

Gordon, Theodore J., "Forecasters Turn to Delphi," *Futurist*, February 1967.

——, "Forecasting Policy Impacts," in Walter A. Hahn and Kenneth F. Gordon (eds.), *Assessing the Future and Policy Planning*, Gordon and Breach Science Publishers, New York, 1973.

——, "New Approaches to Delphi," in *Technological Forecasting for Industry and Government*, James Bright (ed.), Prentice-Hall, Englewood Cliffs, N. J., 1968.

, and Harold S. Becker, "The Cross Impact Matrix Approach to Technology Assessment," *Research Management*, July 1972.
, and R. E. LeBleu, "Employee Benefits 1970-1985," *Harvard Business Review*, January February 1970.
Gustafson, D., and R. Ludke, *Information Manual on Computer-Aided Medical Diagnosis Study for Physician Likelihood Ratio Estimators*, University of Wisconsin, Madison, February 1970.
——, R. Shulka, A. Delbecq, and G. Walster, *A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals*, University of Wisconsin, Madison, January 1971.
Hallan, Jerome B., and Benjamin Harriss III, "Estimation of a Potential Hemodialysis Population," *Medical Care* 8, No. 3 (May-June 1970).
Hammonds, Timothy M., and David L. Call, *Utilization of Protein Ingredients in the U. S. Food Industry*, Part II, Department of Agricultural Economics, Cornell University, July 1970.
Helmer, Olaf, *Accomplishments and Prospects of Futures Research*, Center for Futures Research, Graduate School of Business Administration, University of Southern California, Los Angeles, California, August 1973.
——, "Analysis of the Future: The Delphi Method," in *Technological Forecasting for Industry and Government*, James Bright (ed.), Prentice-Hall, Englewood Cliffs, New Jersey, 1968.
——, "The Delphi Method—An Illustration," in *Technological Forecasting for Industry and Government*, James Bright (ed.), London, 1968.
——, *The Delphi Method for Systematizing Judgments about the Future*, Institute of Government and Public Affairs, University of California, Los Angeles, MR-61, April 1966.
——, "The Delphi Technique and Educational Innovation," *Inventing Education for the Future*, W. Hirsch (ed.), Chandler, 1967.
——, "New Development in Early Forecasting on Public Problems, a New Intellectual Climate," *Vital Speeches* 33, (June 1967).
——, *Social Technology*, Basic Books, New York, 1966.
——, "The Use of Expert Opinion in International Relations Forecasting," Center for Futures Research, Graduate School of Business Administration, University of Southern California, Los Angeles, July 1973.
——, and T. J. Gordon, "Probing the Future," *News Front Magazine* 9, (April 1965).
Huckfeldt, Vaughn E., "A Forecast of Changes in Postsecondary Education," Western Interstate Commission for Higher Education, Boulder, Colorado, 1972.
——, and Robert C. Judd, "Methods for Large Scale Delphi Studies," Western Interstate Commission for Higher Education, Boulder, Colorado, 1972.
Hudson, Lawrence, "Uses of the Cross-Impact Matrix in Exploring Alternatives for the Future," Educational Policy Research Center, Syracuse University Research Corporation, December 1969.
Hudspeth, Delayne, and Stuart Sandow, *A Long-Range Planning Tool for Education: The Focus Delphi*, Educational Policy Research Center, Syracuse University, April 1970.
Johnson, Howard E., "An Experience in Teaching the Delphi Technique," Graduate School of Business, the University of Texas at Austin, 1971.
Jolson, Marvin, and Gerald L. Rossow, "The Delphi Process in Marketing Decision Making," *Journal of Marketing Research* 8, (November 1971).
Judd, Robert C., "Delphi Applications in Decision Making," *Planning and Changing* 2, No. 3, 1971.
——, "Delphi Method: Computerized 'Oracle' Accelerates Consensus Formation," *College & University Business* 53, No. 1 (July 1972).
——, "Forecasting to Consensus Gathering, Delphi Grows Up to College Needs," *College and University Business* 53, No. 1 (July 1972).
Kiefer, David, "Rubber Industry: A Glimpse of the Future," *Chemical and Engineering News*, April 1972.
Kimble, Robert, *Delphic Forecasting of Critical Personnel Requirements*, U. S. Army Electronics Command, Fort Monmouth, N. J., November 1969.
Knock, Richard T., *Electro-Optical Technology*, Long Range Planning Service Report 468, Stanford Research Institute, Menlo Park, Calif., August 1972 (Client Proprietary).
Lamont, Valarie, W. Pearson, and S. Umpleby, *The Future of the University*, Educational Research Laboratory, University of Illinois, May 1971.
Locke, H. B., "NRDC and Industrial Chemistry," *Inventions for Industry*, No. 32, July 1968.
Maddox, M. M., "Delphi Survey of Critical Coal Issues," Hanover: Dartmouth College, 1973.

Marien, Michael, *The Hot List Delphi*, The Educational Policy Research Center at Syracuse, ER-6, 1972.
Martino, Joseph P., *An Experiment with the Delphi Procedure for Long-Range Forecasting*, IEEE, February 1967.
——, "What Computers May Do Tomorrow," *Futurist*, October 1969.
McGlauchlin, Laurence D., "Technological Audits: An Aid to Research Planning," in James R. Bright and M. E. F. Schoeman (eds.), *A Guide to Practical Technological Forecasting*, Prentice-Hall, Englewood Cliffs, N. J., 1973.
Milholland, Arthur V., *et al.*, "Medical Assessment by a Delphi Group Opinion Technique," *The New England Journal of Medicine*, June 14, 1973.
Milkovich, George T., Anthony Annoni, and Thomas A. Mahoney, "The Use of Delphi Procedures in Manpower Forecasting," *Management Science* 19, No. 4 (December 1972).
Moore, C. G., and H. P. Pomrehn, *Technical Forecasting of Marine Transportation Systems 1970-2000*, U. S. C., June 1969, [reprinted in *Technological Forecasting and Social Change* 3, (1971), p. 99].
Morrow, James Earl, *A Delphi Approach to the Future of Management*, Ph.D. Dissertation, Georgia State University, Atlanta, 1971.
Naill, R. F., "Long Term Natural Resource Availability: Environmental and Political Implications in the United States," Washington NSF-RANN, 1973.
Nanus, Burt, Michael Wooton, Harold Borko, "Social Impacts of the Multinational Computer," National Computer Conference Exposition, New York City, June 1973.
North, Harper Q., *A Probe of TRW's Future*, TRW Corporation, July 1966.
——, "Delphi Forecasting in Electronics," Thompson Ramo Woodridge, Corporate Report, 1968 (unpublished).
——, and Donald L. Pyke, "Probes of the Technological Future," *Harvard Business Review*, May-June 1969.
——, and Donald L. Pyke, "Technological Forecasting in Planning for Company Growth," *IEEE Spectrum*, January 1969.
Parker, E. F., "British Chemical Industry in the 1980's—A Delphi Method Profile," *Chemistry and Industry*, January 1970.
——, "Some Experience with the Application of the Delphi Method," *Chemistry and Industry*, September 1969.
Pill, J., *The Delphi Method: Substance, Context, a Critique, and an Annotated Bibliography*, Technical Memorandum No. 183, Case-Western Reserve University School of Management, May 1970.
Pyke, D. L., *A Practical Approach to Delphi, Technological Forecasting and Long-Range Planning*, American Institute of Chemical Engineers, November 1969.
——, *Long-Range Technological Forecasting: An Application of the Delphi Process*, Fifth Summer Symposium of Energy Economic Division of ASEE, U. C. L. A., June 21, 1968.
——, *The Role of Probe in TRW's Planning, Technological Forecasting—An Academic Inquiry*, Graduate School of Business, University of Texas, April 1969.
——, *TRW's Probe II*, presented at Advanced Technological Forecasting: Experiences and Workshops, Industrial Management Center, Inc., September 2-5, 1969.
Reilly, Kevin D., "Prospects for Use of the Delphi Method in Information Science Research," *Proceedings of the American Society for Information Services*, July 25-26, 1970.
Reisman, A., and Martin Taft, *On a Computer-Aided System Approach to Personnel Administration*, Dept. of OR, Case Western Reserve University, December 1968.
——, *et al.*, *Evaluation and Budgeting Model for a System of Social Agencies*, Dept. of OR, Case Western Reserve University, TM-167, November 1969.
Rescher, Nicholas, "A Questionnaire Study of American Values by 2000 A.D.," *Values and the Future*, Baier and Rescher (eds.), Free Press, Glencoe, Ill., 1969.
Riesenfeld, W., *et al.*, "Perceptions of Public Service Needs," *The Gerontologist*, Summer 1972.
Rosove, Perry E., "A Trend Impact Matrix for Societal Impact Assessment," Center for Futures Research, Graduate School of Business Administration, University of Southern California, Los Angeles, April 1973.
Sandow, Stuart, *A Survey of Continuing Education Goals for Degree Granting Post-Secondary Institutions in New York State Research with the Focus Delphi Process*, EPRC, Syracuse, 1972.
——, and Michael Folk, *Report of a Planning Study of the United Ministers in Higher Education. Research with the Focus Delphi Process*, Educational Policy Research Center, Syracuse, 1972.
Scheele, S., "The Future of Leisure and Recreation (Proprietary Study), SET, Inc., 1972.

——, "The Role of the Mentally Retarded in Society" (Proprietary Study), SET, Inc., 1970.
Scott, Randall F., and Dick B. Simmons, "Programmer Productivity and the Delphi Technique," *DATAMATION*, May 1974.
Sheldon, A., and Curtis McLaughlin, "The Future of Medical Care: A Delphi Study," Meeting of the Society for General Systems Research, December 1970.
Shimada, Syozo, *A Note on the Cross-Impact Matrix Method*, Hitachi Ltd., Research Report No. HC-70-029, March 1971.
Shimmen, Toru, "A Short Paper on Cross-Impact Analysis: A Basic Cross-Impact Model," Institute for Future Technology, Tokyo, 1973.
Teeling-Smith, J., *Medicine in the 1990's*, Office of Health Economics, London, England, October 1969.
Turoff, Murray, *A Summary of the Rand Civil Defense Delphi*, Office of Emergency Preparedness, Systems Evaluation Division, Technical Memorandum-122, February 1970.
——, "A Synopsis of Innovation by the Delphi Method," ORSA Meeting in Detroit, 1970.
Voyer, Robert, *Future Developments in Canada*, Science Forum, Science Council of Canada, October 1969.
Waldron, J. S., *An Investigation into the Relationship among Conceptual Level, Time Delay of Information Feedback, and Performance in the Delphi Process*, Doctoral Dissertation, Syracuse University, 1970.
Weaver, Timothy, "Delphi: A Forecasting Tool," *Notes on the Future of Education*, 1 Issue 3 (1970), Educational Policy Research Center, Syracuse University.
——, "The Delphi Forecasting Method," *Phi Delta Kappan*, January 1971.
——, "The Delphi Forecasting Method: Some Theoretical Considerations," in *The Potential of Educational Futures*, Ziegler and Marien (eds.), Charles A. Jones, Worthington, Ohio, 1972.
——, *The Delphi Method: Background and Critique*, Educational Policy Research Center, Syracuse University, June 1970.
——, *The Delphi Method: Research Design and Findings*, Educational Policy Research Center, Syracuse University, June 1970.
——, *An Exploration into the Relationship between Conceptual Level and Forecasting Future Events*, Doctoral Dissertation, Syracuse University, 1969
Welty, Gordon, "A Critique of the Delphi Technique," *Proceedings of the American Statistical Association*, Washington, D. C., 1972.
——, "A Critique of Some Long-Range Forecasting Developments," 38th Session of the International Statistical Institute, Washington, D. C., August 1971.
——, "Problems of Selecting Experts for Delphi Exercises," *Academy of Management Journal* 15, (1972):1.
——, "The Selection and Deflection of Expertise in Delphi Exercises," joint ORSA-TIMS meeting, Atlantic City, November 1972.
Wennerberg, Ulf, "Using the Delphi Technique for Planning the Future Libraries," Unesco Bulletin for Libraries, V. 26 (5), 242-246, September October 1972.
Williamson, John, *Prognostic Epidemiology of Absenteeism*, Department of Medical Care, Johns Hopkins University, March 1971.
——, *Prognostic Epidemiology of Breast Cancer*, Department of Medical Care, Johns Hopkins University, 1971.

Technological Forecasting

Arnfield, R. V. (ed.), *Technological Forecasting*, Edinburgh University Press, Edinburgh, 1969.
Barach, J. L., and R. J. Twery, "The Application of Forecasting Techniques in Research and Development Planning," presented at the AIChe meeting, November 1969.
Bayne, C. Dudley, Jr., and Walter Price, "Technological Forecasting," in *Formal Planning Systems—1971*, Richard F. Vancil (ed.), Harvard Business School, Cambridge, Mass., 1971.
Bestuzhev-Lada, I. V., "Bourgeois 'Futurology' and the Future of Mankind," *Political Affairs*, September 1970.

——, "How to Forecast the World's Future: Asimov Disputed," *The Current Digest of the Soviet Press* 19, No. 20 (June 1967).
——, "Social Forecasting," *Soviet Cybernetics Review* 3, No. 5 (May 1969).
——, *Window into the Future*, Mysl, Moscow, 1970.
Bright, James R., *A Brief Introduction to Technological Forecasting—Concepts and Exercises*, Pemaquid Press, Austin, Texas, 1972.
——, "Evaluating Signals of Technological Change," *Harvard Business Review* 48 (1970):1.
——(ed.), *Technological Forecasting for Industry and Government, Methods and Applications*, Prentice-Hall, Englewood Cliffs, New Jersey, 1968.
——and M. E. F. Schoeman (ed.), "A Guide to Practical Technological Forecasting" (a second collection of papers), Prentice-Hall, Englewood Cliffs, N. J., 1973.
Chambers, J. C., *et al.*, "How to Choose the Right Forecasting Technique," *Harvard Business Review*, July–August 1971.
Davidson, Frank, "Futures Research: A New Scientific Discipline?" *Proceedings of the Social Statistics Section*, American Statistical Association meeting, August 1969.
Davis, Richard C., "Organizing and Conducting Technological Forecasting in a Consumer Goods Firm," in James R. Bright and M. E. F. Schoeman (eds.), *A Guide to Practical Technological Forecasting*, Prentice-Hall, Englewood Cliffs, N. J., 1973.
De Houghton, Charles, William Page, and Guy Streatfeild, *...and Now the Future: A PEP Survey of Future Studies*, PEP Broadsheet 529, 37 (London: PEP, August 1971).
Ehin, I., "Some Methodological Problems of Forecasting," *Transactions*, AS Estonian SSR, 20, No. 4 (1971).
Gerstenfeld, Arthur, "Technological Forecasting," *The Journal of Business* 44, No. 1 (January 1971).
Girshick, M., A. Kaplan, and A. Skogstad, "The Prediction of Social and Technological Events," *Public Opinion Quarterly* 13, (Spring 1953); Rand P-93, April 1949.
Hall, P. D., "Technological Forecasting for Computer Systems," in *Technological Forecasting and Corporate Strategy*, G. S. C. Wills, *et al.* (eds.), American Elsevier, New York, 1969.
Hayden, Spencer, "How Industry Is Using Technological Forecasting," *Management Review*, May 1970.
Heiss, K., K. Knorr, and O. Morgenstern, *Long Term Projections of Political and Military Power*, Mathematica, Princeton, N. J., January 1973.
Kiefer, David M. (ed.), "The Futures Business" (survey article), *Chemical and Engineering News*, August 11, 1969.
Lanford, H. W., *A Synthesis of Technological Forecasting Methodologies* (Wright-Patterson Air Force Base, Ohio: Headquarters, Foreign Technology Division, Air Force Systems Command, U. S. Air Force, May 1970).
Lätti, V. I., *A Survey of Methods of Futures Research*, Report NORD-3 (Pretoria: Institute for Research Development, South African Human Sciences Research Council, 1973).
Martino, Joseph, "The Paradox of Forecasting," *The Futurist*, February 1969.
——, *Technological Forecasting for Decision Making*, American Elsevier, New York, 1972.
——, "Technological Forecasting Is Alive and Well in Industry," *The Futurist*, August 1972.
——, "What Computers May Do Tomorrow," *The Futurist*, October 1969.
——, "Tools for Looking Ahead," *IEEE Spectrum*, October 1972.
——, *et al.*, *Long Range Forecasting Methodology*, A Symposium Held at Holman AFB, Alamogordo, New Mexico, October 11, 1967.
Massenet, M., "Methods of Forecasting in the Social Sciences," in *Three Papers Translated from the Original French for the Commission on the Year 2000*, Brookline, Mass., American Academy of Arts and Sciences.
McLoughlin, William G., "Technological Forecasting LTV," *Science and Technology*, February 1970.
Mitroff, Ian, and Murray Turoff, "Technological Forecasting and Assessment: Science and/or Mythology?" *IEEE Spectrum*, March 1973.
North, Harper Q., "Technological Forecasting in Planning for Company Growth," *IEEE Spectrum*, January 1969.
——, "Technological Forecasting to Aid Research and Development Planning," *Research Management* 12, No. 4 (1969).
——, "Technology, the Chicken—Corporate Goals, the Egg," *Technology Forecasting for Industry and Government*, James R. Bright (ed.), Prentice-Hall, Englewood Cliffs, N. J., 1968.

——, and Donald L. Pyke, *Technological Forecasting in Industry*, NATO Defense Research Group Conference, National Physical Laboratory, England, November 1968.
Pry, Robert, *et al.*, *Technological Forecasting and Its Application to Engineering Materials*, National Materials Advisory Board, NMAB-279, March 1971.
Quinn, J. B., "Technological Forecasting," *Harvard Business Review*, March–April 1967.
Rosen, Stephen, "Inside the Future," *Innovation Magazine*, No. 18, 1971.
Thiesmeyer, Lincoln R., "The Art of Forecasting the Future Is Losing Its Amateur Status," *Financial Post*, June 26, 1971.
——, "How to Avoid Bandwagon Effect in Forecasting: The Delphi Conference Keeps Crystal Ball Clear," *Financial Post*, October 23, 1971.
Turoff, Murray, "Communication Procedures in Technological Forecasting," *IEEE 73 INTERCON Proceedings*, Session 28: Technological Forecasting Methodologies, March 1973.
Wills, Gordon, *et al.*, *Technological Forecasting*, Penguin Books, Baltimore, 1972.

New Directions

Amara, Roy, and Jacques Vallee, "FORUM: A Computer-Based System to Support Interaction among People," IFIP Congress, *1974 Proceedings*, August 1974.
Bahm, Archie J., "Demo-Speci-ocracy," *Policy Sciences* 3, No. 1 (1972).
Baran, Paul, "Voice-Conferencing Arrangement for an On-Line Interrogation," Institute for the Future, March 1973.
——, Hubert M. Lipinski, Richard H. Miller, and Robert H. Randolph, "ARPA Policy-Formulation Interrogation Network," Semiannual Technical Report, April 1973.
Baver, Raymond, "Societal Feedback," *The Annals* 2, September 1967.
Billingsley, Ray V., "System Simulation as an Interdisciplinary Interface in Rural Development Research," unpublished, Texas Agricultural Experiment Station Technical Article 9915, College Station, Texas (April 1972).
Carter, George E., "Computer-Based Community Communications," mimeographed, Computer-Based Education Research Laboratory, University of Illinois, Urbana, Illinois (1973).
——, "Second Generation Conferencing Systems," Computer-Based Educational Research Laboratory, University of Illinois (November 1973).
Conrath, David W., "Teleconferencing: The Computer, Communication, and Organization," *Proceedings* of the First International Conference on Computer Communication (1972).
——, and J. H. Bair, "The Computer as an Interpersonal Communication Device: A Study of Augmentation Technology and its Apparent Impact on Organizational Communication," Second International Conference on Computers and Communications, Stockholm, August 1974.
Englebart, D. C., "The Augmented Knowledge Workshop," *AFIPS Conference Proceedings* 42, 1973 Joint Computer Conference.
——, "Coordinated Information Services for a Discipline-or-Mission-Oriented Community," Second Annual Computer Communications Conference, San Jose, Calif. (January 1973).
——, "Intellectual Implications of Multi-Access Computer Networks," Interdisciplinary Conference on Multi-Access Computer Networks, Austin, Texas (April 1970).
Etzioni, Amitai, "Minerva: An Electronic Town Hall," *Policy Sciences* 3, No. 4 (1972).
Ewald, William R., Jr., *ACCESS To Regional Policymaking*. A report to the National Science Foundation, July 27, 1973.
——, *GRAPHICS for Regional Policymaking, A Preliminary Study*. A report to the National Science Foundation, August 17, 1973.
Gusdorf, *et al.*, *Puerto Rico's Citizen Feedback System*, Technical Report 59, OR Center, M. I. T. (April 1971).
Hall, T. W., "Implementation of an Interactive Conference System," Spring Joint Computer Conference, Vol. 38, AFIPS Press, 1971.
Havron, M. Dean, and Anna E. Casey Stahmer, "Planning Research in Teleconference Systems," Department of Communications, Canada, November 1973.

Appendix

Johansen, Robert, and Richard H. Miller, "The Design of Social Research to Evaluate a New Medium of Communication," a working paper prepared for the American Sociological Association Annual Meeting, February 1974.
Johnson, Norman, and Edward Ward, "Citizen Information Systems," *Management Science* 19, No. 4, Part II (December 1972).
Kimbel, Dieter, "Computers and Telecommunication Technology," International Institute for Management of Technology, 1973.
Krendel, Ezra S., "A Case Study of Citizen Complaints as Social Indicators," *IEEE Transactions of Systems, Science, and Cybernetics*, October 1970.
Kupperman, R. H., and R. H. Wilcox, "Interactive Computer Communications Systems—New Tools for International Conflict Deterrence and Resolution," Second International Conference on Computers and Communications, Stockholm, August 1974.
Leonard, Eugene, *et al.*, "Minerva: A Participatory Technology System," *Bulletin of Atomic Scientists*, November 1971.
Licklider, J. C. R., "The Computer as a Communication Device," *International Science and Technology*, April 1968.
Linstone, H. A., "Communications in Futures Research," paper presented at the Conference of Research Needs in Futures Research, January 10–11, 1974, sponsored by the Futures Group, to be published in W. I. Boucher (ed.), *The Study of the Future: An Agenda for Research*.
Lipinski, A. J., H. M. Lipinski, and R. H. Randolph, "Computer-Assisted Interrogation: A Report on Current Methods Development," International Conference on Computer Communications, IEEE, October 1972.
Lipinski, Hubert M., and Robert H. Randolph, "Conferencing in an On-line Environment: Some Problems and Possible Solutions," *Proceedings* of the Second Annual Computer Communications Conference, 1973.
Little, J. K., C. H. Stevens, and P. Tropp, "Citizen Feedback System: The Puerto Rico Model," *National Civic Review*, April 1971.
Macon, N., and J. D. McKendree, "EMISARI Revisited: The Resource Interruption Monitoring System," Second International Conference on Computers and Communications, Stockholm, August 1974.
Miller, Richard H., "Trends in Teleconferencing by Computer," *Proceedings* of the Second Annual Computer Communications Conference, 1973.
Noel, Robert, "Polis: A Computer-Based System for Simulation and Gaming in the Social Sciences" (September 1972), Department of Political Sciences, University of California at Santa Barbara.
Renner, Rod, *Conference Systems Users Guide*, TM-225, Office of Emergency Preparedness, AD 756–813, November 1972.
——, *Delphi Users Guide*, TM-219, Office of Emergency Preparedness, September 1971.
——, *et al.*, "EMISARI: A Management Information System Designed to Aid and Involve People," *COINS* (International Symposium on Computer and Information Science), December 1972.
Santi, V. De., *et al.*, *Genie*, Office of Economic Opportunity, June 1971.
Schuyler, James, and Robert Johansen, "ORACLE: Computer Conferencing in a Computer-Assisted Instruction System," International Conference on Computer Communications, IEEE, October 1972.
Sheridan, Thomas B., "Citizen Feedback: New Technology for Social Choice," *Technology Review*, M. I. T., January 1971.
——, "Technology for Group Dialogue and Social Choice," *AFIPS Conference Proceedings* 39 (1971), 1971 Fall Joint Computer Conferences.
Stevens, Chandler H., "Citizen Feedback and Societal Systems," *Technology Review*, M. I. T., January 1971.
Thompson, Gordon, "Moloch or Aquarius," *THE* 4 (February 1970), Bell Northern Research.
——, "A Socially Responsible Approach to Comn.unication System Design," Second International Conference on Computers and Communications, Stockholm, August 1974.
Turoff, Murray, "Computerized Conferencing and Real Time Delphis," Second International Conference on Computers and Communication, August 1974, Stockholm.
——, "Conferencing via Computer," Proceeding of *NERM*. IEEE, November 1972.

——, "Conferencing via Networks," *Computer Decisions*, January 1973.
——, "DELPHI and Its Potential Impact on Information Systems," Conference Proceedings *FJCC* 39, AFIPS Press, 1971.
——, "DELPHI + Computers + Communications = ?" in *Industrial Applications of Technological Forecasting*, Cetron and Ralph (eds.), Wiley-Interscience, New York, 1971.
——, "The DELPHI Conference," *Futurist*, April 1971.
——, "Human Communication," *EKISTICS* 35, No. 211 (June 1973).
——, "*Party-Line* and *Discussion*, Computerized Conference Systems," International Conference on Computer/Communications, IEEE, October 1972.
——, "Potential Applications of Computerized Conferencing in Developing Countries," Rome Special Conference on Futures Research, 1973.
——, "Potential Applications of Computerized Conferencing in Developing Countries," *Herald of Library Sciences* 12, No. 2–3 (April–July 1973).
Umpleby, Stuart, *Citizen Sampling Simulations: A Method for Involving the Public Social Planning* ERL, University of Illinois, June 1970.
——, "Citizen Sampling Simulations: A Method for Involving the Public in Social Planning," *Policy Sciences* 1, No. 3 (1970).
——, *The DELPHII Exploration, A Computer-Based System for Obtaining Subjective Judgements on Alternative Futures*, University of Illinois, CERL Report F-1, August 1969.
——, "Structuring Information for a Computer-Based Information System," *FJCC* Conference Proceedings 39, AFIPS Press, 1971.
——, and Valarie C. Lamont, "Final Report—Social Cybernetics and Computer-Based Communications Media," mimeographed, Computer-Based Education Research Laboratory, University of Illinois, Urbana, Illinois (1973).
Vallee, Jacques, "Network Conferencing," *Datamation*, May 1974.
Weston, J. R., and C. Kristen, "Teleconferencing: A Comparison of Attitudes, Uncertainty and Interpersonal Atmospheres in Mediated and Face-to-Face Group Interaction," prepared for the Social Policy and Programs Board, the Department of Communications, Canada, December 1973.
Wilcox, Richard, "Computerized Communications, Directives, and Reporting Systems," Information and Records Administration Conference (IRAC), May 19, 1972.
——, and R. Kupperman, "EMISARI: An On-Line Management System in a Dynamic Environment," International Conference on Computer Communications, IEEE, October 1972.
Wilson, Stan, "A Test of the Technical Feasibility of On-Line Consulting Using APL," APL Laboratory, Texas A&M University, 1973.

Some Foreign-Language Articles

Albach, H., "Informationsgewinnung durch strukturierte Gruppenbefragung—Die Delphi-Methode,"ZfB, Suppl., 40th Year, (1970), pp. 11–26.
Atlas Copco MCT AB, "The Driving of Tunnels—Needs and Technologies," Nacka, Sweden, 1972.
Dvořák, K. *et al.*, "The Application of the Delphi Method to the Forecasting of Machinery Technology," *Trend* 3, (1972).
Ehriemark, G., and L-G Remstrand, "A Modified Delphi-Experiment—Performance and Feasibility," from *Some Papers on Futures Studies*, Konsultkollegiets kompendier 1, Beckmans, Stockholm, Sweden, 1971.
Ejerhed, A., G. Hulten, M. Johnsson, Y. Karlsson, M. Olander, "Futures Studies in Public Health," report from the Swedish Planning and Rationalization Institute of Health and Social Services (SPRI), Stockholm, Sweden, 1971.
The Federation of Swedish Industries, "Political Values and Structural Developments in the Environment of the Federation of Swedish Industries," Stockholm, Sweden, 1971.
Gardborn, I., "An Experiment with the Delphi-Technique," Stockholm School of Economics and Business Administration, Stockholm, Sweden, 1971.

Gidai, E., "On the Practical Use of Intuitive Methods," paper presented at the Summer School of Prognostics, Warsaw, September 1972.
Grojer, J-E, "A Futures-Study with Delphi-Technique Applied to the Urbanization Process," University of Stockholm, Sweden, 1971.
Halberštátová, R., and I. Žádný, "The Application of the Delphi Method in the Research of Iron Metallurgy," *Trend* 4, (1972).
Horn, J., "The World in the Future—A Delphi-Experiment," Naturvetenskapliga Forskningsradet/The Swedish National Science Research Council, Stockholm, Sweden, 1972.
Husen, T., "Education in the Year 2000—A Futures Study," Stockholm, Sweden, 1971.
Institute of Management, "The Use of some Delphi Method Principles for Research in the Development of a Management System of Socialist Industrial and Business Organizations," Research Report, Prague, 1972.
Jacobsson, S., "Build for the Future—The Future for the Builders," Byggnadsindustrin (Building Industry) journal, No. 18 (1969), Stockholm, Sweden.
Krauch, Helmut, and Georg Rüdinger, "Ein Rückkoppelungs—System zwischen Rundfunkanstalt and Zuhörerschaft," Studiengruppe für Systemforschung, Heidelberg, Germany, November 1970.
Lachman, O., "Personnel Administration in 1980: A Delphi-Study," Danish Institute of Graduate Engineers, Naerum (Copenhagen), Denmark, 1971.
Landergren, U., S. Fredriksson, and A. Littke-Persson, "An Application of the Delphi Technique to a Forecast of the Field of Ceramics—A Methodological Experiment," Forsvarets Forskningsanstalt/The Swedish National Defense Research Institute (FOA), Stockholm, Sweden, 1968.
Liden T., and L. O. Teinvall, "Convenience 1985—The Prediction of a Selling-Argument," University of Lund, Faculty of Business Administration, Lund, Sweden, 1972.
Machálek, J., and O. Šulc, "A Combination of Delphi and Cross-Impact Matrix Methods," *Trend* 5 (1970).
The Nordic Museum, "The Future of the Museums on the History of Civilization," Nordic Museum, Stockholm, Sweden, 1973.
The Nordic Summer-University, "New Sciences? —A Delphi Study," *Information Fra Nordisk Sommeruniversitet*, No. 2 (1969), Copenhagen, Denmark.
Oxelösund Iron Works AB, "Delphi 1985," Oxelösund, Sweden.
Perstorp AB, "The Future and Its Importance for Perstorp," Perstorp, Sweden, 1969.
Platmanufaktur AB (producers of cans, bottles, and other packaging), "Technological Forecasting in the Beer Field," Malmö, Sweden.
——, "Values and Their Changes in Sweden," Malmö, Sweden.
Research Institute for Construction Materials, "A Forecast of Key Tendencies in the Development of Construction Materials," Research Report, Brno, Czechoslavia, 1973–74.
The Royal College of Forestry, "The Future Role of the Forest—A Delphi Study." Stockholm, Sweden, 1973.
Salancik, G. R., T. J. Gordon, and N. Adams, for Skandia AB, "On the Nature of Economic Losses Arising from Computer-Based Systems in the Next Fifteen Years," Institute for the Future, Menlo Park, Calif., Report R-23, March 1972.
Schollhammer H., "Die Delphi-Methode als betriebliches Prognose- und Planungsverfahren," ZfB, M.S., 22nd Year (1970), pp. 128–137.
Sjölind, S., "A Future Study of the Medicine Market (Drug Market) in Sweden," Stockholm School of Economics and Business Administration, Stockholm, Sweden, 1971.
Skandia AB (Insurance Company), "The Internal Delphi-Study of Skandia," Stockholm, Sweden, 1971.
Svenska Kullagerfabriken AB (SKF), "Technology Evolution in Bearing Manufacturing for the Coming 15 Years," Göteborg, Sweden, 1970.
Swedish Agency for Administrative Development (SAFAD), "Information, Documentation and Media—Report on a Delphi Study," Allmanna Förlaget, Stockholm, Sweden, 1971.
The Swedish Board for Technical Development, "Priorities for Research in Medical Techniques in Sweden," Stockholm, Sweden, 1969.
The Swedish Central Bureau of Statistics, "The Future Development of Computer Usage and Its Effects on the Labour Market," Stockholm, Sweden, 1973.
Swedish Lloyd, "The Future Transportation Environment," Göteborg, Sweden, 1969.

The Swedish Newspaper Publishers Training Board, "Swedish Newspaper Industry—A Futuristic Vision by the Press," Stockholm, Sweden.
The Swedish Savings Banks Association, "The Future Environment of the Swedish Savings Banks," Stockholm, Sweden.
The Swedish Shipbuilding Experimental Tank, "The Gasturbine Technique in Merchant Ships," Göteborg, Sweden, 1972.
Swedish Unilever AB, "Scandinavian Environmental Study," Stockholm, Sweden, 1971.
Ullman, A., "Research and Development in the Field of Materials," Naval Material Department, Stockholm, Sweden, 1971.
Volvo AB, "Delphi—An Experiment with a Futuristic Methodology," Göteborg, Sweden, 1970.
Westerlund, S., "The Delphi-Technique—Swedish Practice," *Ekonomen*, No. 20 (1968), Stockholm, Sweden.
Wikstrom, S., "The Distribution of Our Every Day Commodities—A Futures Study," University of Stockholm, Faculty of Business Administration, Bonniers Publishers, Stockholm, Sweden, 1973.
Ytongbolagen, Hallabrottet, "Product Innovations at Ytong," Hallabrottet, Sweden, 1968.

Acknowledgment

The editors wish to thank Göran Axelsson and Jan Wisén of the Swedish Agency for Administrative Development and Ota Šulc of the Czechoslavak Academy of Sciences, Prague, for their assistance in compiling portions of this bibliography.

Related Work

Ackoff, Russell L., "Towards a Behavioral Theory of Communication," *Management Sciences* 4, No. 3 (1958).
——, "Towards a System of Systems Concepts," *Management Sciences* 17, No. 11 (July 1971).
Adelson, M., *et al.*, "Planning Education for the Future," *American Behavioral Scientist*, March 1967.
Adorno, T. W., E. Frenkel-Brunswik, D. J. Levinson, and R. N. Sanford, "The Authoritation Personality, Part One," Wiley, New York, 1950.
Allen, Allen D., "Scientific Versus Judicial Fact Finding in the United States," *IEEE Transactions on Systems, Man & Cybernetics*, SMC-2, No. 4, 1972.
Allen, Vernon L., "Situational Factors in Conformity," *Advances in Social Psychology*, in Vol. II by Berkowitz, Academic Press, New York, 1965.
Arrow, K. J., "Alternative Approaches to the Theory of Choice in Risk-Taking Situations," *Econometrica* 19, No. 4 (1951).
Asch, S. E., "Effects of Group Pressure upon the Modification and Distortion of Judgments," *Readings in Social Psychology*, Holt, Rinehart and Winston, New York, 1958.
——, "Studies of Independence and Conformity: A Minority of one against a Unanimous Majority," *Psychological Monographs* 70 (1956).
Back, K. W., "Time Perspective and Social Attitudes," paper presented at APA annual meeting, symposium on Human Time Structure, Chicago, September 1965.
Baier, K., and N. Rescher, *Values and the Future*, Free Press, New York, 1969.
Bailey, Gerald, Peter Nordlie, and Frank Sistrunk, "Literature Review, Field Studies, and Working Papers," (from IDA Teleconferencing Studies), Human Sciences Research, October 1963, Institute for Defense Analyses, Research Paper P-113. NTS #AD-480 695.
Bass, B. M., "Authoritatianism or Acquiescence?" *Journal of Abnormal and Social Psychology* 51, (1955), pp. 616–23.

Bavelas, Alex, "Teleconferencing: Background Information," Institute for Defense Analyses, Research Paper P-106, 1963.
Bell, D., "Twelve Modes of Prediction," in Julius Gould (ed.), *Penguin Survey of the Social Sciences*, Penguin Books, Baltimore, 1965.
Belnap, Nuel, *An Analysis of Questions*, System Development Corporation Technical Memorandum 1287/000/00, Santa Monica, Calif., 1963.
Berelson, Bernard, and Gary Steiner, *Human Behavior*, Harcourt and Brace, New York, 1964.
Berger, R. M., J. P. Guilford, and P. R. Christensen, "A Factor-Analytic Study of Planning Abilities," *Psychological Monographs* 71 (Whole No. 435), 1957.
Beum, Corlin D., and Everett G. Brundage, "A Method for Analyzing the Sociomatrix," *Sociometry* 13, No. 2 (May 1950).
Bonier, R. J., *A Study of the Relationship between Time Perspective and Open-Closed Belief Systems*, M.S. thesis, Michigan State University Library, 1957.
Boocock, Saran S., and E. O. Shild, *Simulation Games in Learning*, Sage Publishing Co., Beverly Hills, Calif., 1968.
Bottenberg, R. A., and R. E. Christal, "Grouping Criteria—A Method which Retains Maximum Predictive Efficiency," *Journal of Experimental Education* 36, No. 4 (Summer 1968), pp. 28–34.
——, and J. H. Ward, Jr., *Applied Multiple Linear Regression*, PRL-TDR-63-6, AD-413 128, Lackland AFB, Texas: Personnel Research Laboratory, Aerospace Medical Division, March 1963.
Bretz, Rudy, "The Selection of Appropriate Communication Media for Instruction: A Guide for Designers of Air Force Technical Training Programs," Rand Report R-601-PR, February 1971, pp. 30 ff.
Brockhoff, K., "On Determining Relative Values," *Zeitschrift für Operations Research* 16 (1972), pp. 221–32.
Brown, Judson S., "Gradients of Approach and Avoidance Responses and Their Relation to Levels of Motivations," *Journal of Comparative and Physiological Psychology* 41, No. 6 (1948).
Caldwell, C. H. Coombs, Schoeffler, and R. M. Thrall, "A Model for Evaluating the Output of Intelligence Systems," *Naval Research Logistics Quarterly* 8 (1961).
——, and R. M. Thrall, "Linear Model for Evaluating Complex Systems," *Naval Research Logistics Quarterly* 5 (1958).
Campbell, Donald T., and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago, 1963.
Cantril, H., "The Prediction of Social Events," *Journal of Abnormal Psychology* 33 (1938), pp. 364–89.
——, "The World in 1952: Some Predictions," *Journal of Abnormal and Social Psychology* 38 (1943), pp. 6-47.
Carbonell, Jaime R., "On Man-Computer Interaction: A Model and Some Related Issues," *IEEE Transactions on Systems Science and Cybernetics*, January 1969.
Cavert, C. Edward, "Procedures for the Design of Mediated Instruction," State University of Nebraska Project, 1972.
Chapanis, Alphonse, "The Communication of Factual Information through Various Channels," *Information Storage and Retrieval* 9.
Christal, R. E., "JAN: A Technique for Analyzing Group Judgment," *Journal of Experimental Education* 36, No. 4 (Summer 1968), pp. 25–7.
——, *Officer Grade Requirements Project: I. Overview*, PRL-TDR-65-15, AD-622 806, Lackland AFB, Texas: Personnel Research Laboratory, Aerospace Medical Division, September 1965.
——, "Selecting a Harem—And Other Applications of the Policy-Capturing Model," *Journal of Experimental Education* 36, No. 4 (Summer 1968), pp. 35–41.
Churchman, C. W., *Challenge to Reason*, McGraw-Hill, New York, 1968.
——, *The Design of Inquiring Systems*, Basic Books, New York, 1971.
——, *Theory of Experimental Inference*, Macmillan, New York, 1948.
Clark, Charles H., *Brainstorming*, Doubleday, Garden City, N. Y., 1958.
Cohen, Arthur R., "Some Implications of Self-Esteem for Social Situations," in Persuasion and Persuasibility, Hovland and Janis (eds.), Yale University Press, New Haven, 1959.
Coleman, J., E. Katz, and H. Menzel, "The Diffusion of an Innovation," *Sociometry* 20 (1957), pp. 253–70.
Collaros, P. A., and L. R. Anderson, "Effect of Perceived Expertness upon Creativity of Members in Brainstorming Groups," *Journal of Applied Psycnology* 53, No. 2 (1969).

Collins, B. E., and H. Guetzkow, *A Social Psychology of Group Processes for Decision-Making*, Wiley, New York, 1964.
Delbecq, André, and Andrew Van de Ven, "A Group Process Model for Problem Identification and Program Planning," *Journal of Applied Behavioral Science* 7, No. 4 (1971).
Dickson, Paul, *Think Tanks*, Atheneum, New York, 1971.
Drucker, Peter, *Technology, Management, and Society*, McGraw-Hill, New York, 1970.
Dunnette, M. D., "Are Meetings Any Good for Solving Problems?", *Personnel Administration*, 1964.
Eckenrode, R. T., "Weighting Multiple Criteria," *Management Science* 12, No. 3 (1965).
Edwards, A. L., *Techniques of Attitude Scale Construction*, Appleton-Century-Crofts, New York, 1957.
——, and L. L. Thurstone, "An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals," *Psychometrika* 17, No. 2 (1952).
Emlet, Harry, *et al.*, "Selection of Experts," 39th ORSA meeting, May 1971.
Eysenck, H. J., "The Validity of Judgments as a Function of the Number of Judges," *Journal of Experimental Psychology* 25, No. 6 (1939).
Farmer, Richard, and Barry M. Richman, *Comparative Management and Economic Progress*, Richard Irwin Inc., 1965.
Fisher, Lloyd, "The Behavior of Bayesians in Bunches," *The American Statistician*, December 1972.
Garner, W. R., *Structure and Uncertainty as Psychological Concepts*, Wiley, New York, 1962.
Gerardin, Lucien A., "Topological Structural Systems Analysis: An Aid for Selecting Policy and Actions for Complex Sociotechnological Systems," paper presented at the IFAC/IFORS Conference on Systems Approaches to Developing Countries, Algiers, May 1973.
Gordon, William, "Operational Approach to Creativity," *Harvard Business Review* 34, No. 6 (November–December 1956).
Guttman, L. A., "A Basis for Scaling Qualitative Data," *American Sociological Review* 9, No. 2 (1944).
Hainer, Raymond M., Sherman Kingsbury, and David Gleicher, *Uncertainty in Research, Management and New Product Development*, Reinhold, New York, 1967.
Hall, E. J., "Decisions, Decisions, Decisions," *Psychology Today*, November 1971.
——, Jane S. Mouton, and Robert R. Blake, "Group Problem Solving Effectiveness under Conditions of Pooling vs. Interaction," *Journal of Social Psychology* 59 (1963), pp. 147–57.
Hare, A. P., "A Study of Interaction and Consensus in Different Sized Groups," *American Sociological Review* 17, No. 3 (1952).
——, "Handbook of Small Group Research," Free Press, Glencoe, Illinois, 1962.
Harvey, O. J., "Conceptual Systems and Attitude Change," in M. Sherif and C. Sherif (eds.), *Attitude, Ego Involvement and Change*, Wiley, New York, 1967.
—— (ed.), *Experience, Structure and Adaptability*, Springer Publishing Co., 1966.
——, "System Structure, Flexibility, and Creativity," in O. J. Harvey (ed.), *Experience, Structure, and Adaptability*, Springer Publishing Co., 1966.
——, D. E. Hunt, and H. M. Schroder, *Conceptual Systems and Personality Organization*, Wiley, New York, 1961.
——, and J. A. Kline, *Some Situational and Cognitive Determinants of Role Playing: A Replication and Extension*, Tech. Rep. No. 15, Contract Nonr. (07), University of Colorado, 1965 (cited by Harvey [1967]).
Havron, M. Dean, and Mike Averill, "Questionnaire and Plan for Survey of Teleconference Needs among Government Managers" for the Socio-Economic Branch, Department of Communications, Canada, Contract OGR2-0303, November 30, 1972.
Haythorn, William, "The Influence of Individual Members on the Characteristics of Small Groups," *Journal of Abnormal and Social Psychology* 48, No. 2 (1953).
Helmer, Olaf, and N. Rescher, "On the Epistemology of the Inexact Sciences," *Management Science* 6, No. 1 (1959).
Hoffman, L. R., and G. G. Smith, "Some Factors Affecting the Behaviors of Members of Problem Solving Groups," *Sociometry* 23, No. 3 (1960).
Hoffman, P. J., "The Paramorphic Representation of Clinical Judgments," *Psychological Bulletin* 47 (1960), pp. 116–31.
House, Robert J., "Merging Management and Behavioral Theory: The Interaction between Span of Control and Group Size," *Administrative Science Quarterly* 14, No. 3 (September 1969).
Huber, George, and André Delbecq, "Guidelines for Combining the Judgements of Individual Members in Decision Conferences," *Academy of Management Journal* (June 1972).

——, V. Sahney and D. L. Ford, "A Study of Subjective Evaluation Models," *Behavioral Science* 14, No. 6 (1969).
——, J. Wardrop, and T. Herr, *An Analysis of Errors Obtained When Aggregating Judgments*, Paper #6704, Social Systems Research Institute, University of Wisconsin, Madison, 1967.
Israeli, N., *Abnormal Personality and Time*, Science Press, New York, (1936).
——, "Attitudes to the Decline of the West," *Journal of Social Psychology* 4 (1933).
——, "Group Estimates of the Divorce Rate for the Years 1935–1975," *Journal of Social Psychology* 4, No. 1 (1933).
——, "The Psychopathology of Time," *Psychological Review* 39, No. 5 (1932).
——, "Some Aspects of the Social Psychology of Futurism," *Journal of Abnormal and Social Psychology* 25, No. 2 (1930).
——, "Wishes Concerning Improbable Future Events: A Study of Reactions to the Future," *Journal of Applied Psychology* 16 (1932), pp. 584–88.
Janis, Irving L., "Groupthink," *Psychology Today*, November 1971.
Johnson, S. C., "Hierarchial Clustering Schemes," *Psychometrika* 32, No. 3 (1967).
Jouvenel, B. de, "L'Art de la Conjecture," in the *Futuribles Series*, edition du Roucher, Monaco, 1964. (English translation: *The Art of Conjecture*, Basic Books, New York, 1967.)
Kahneman, O., and A. Tversky, "Subjective Probability: A Judgment of Representativeness," Oregon Research Institute, *Research Bulletin II*, No. 2 (1971).
Kaplan, A., A. Skogstad, and M. Girschick, "The Prediction of Social and Technological Events," *Public Opinion Quarterly* 14, No. 1 (1950).
Kastenbaum, R. J., "A Preliminary Study of the Dimensions of Future Time Perspective," Doctoral Dissertation, University of Southern California. Ann Arbor, Mich., University Microfilms, No. 60 394, 1960.
Katz E., "Communication Research and the Image of Society: Convergence of Two Traditions," *American Journal of Sociology* 65 (1960).
Kelly, H., and J. Thibaut, "Experimental Studies of Group Problem Solving and Process," *Handbook of Social Psychology* 2, Addison-Wesley, Reading, Mass., 1954.
Kite, Richard W., Paul C. Vity, "Teleconferencing: Effects of Communication Medium, Network and Distribution of Resources," IDA Study S-233.
Kleinmuntz, B., *Formal Representation of Human Judgment*, Wiley, New York, 1968.
Kotler, Philip, "A Guide to Gathering Expert Estimates: The Treatment of Unscientific Data," *Business Horizons*, October 1970.
Lassey, William R., *Leadership and Social Change*, University Associates Publishers, Iowa City, 1971.
Leavitt, H. J., "Some Effects of Certain Communication Patterns on Group Performance," *Journal of Abnormal and Social Psychology* 46, No. 1 (1951).
Levinson, D. J., "An Approach to the Theory and Measurement of Ethnocentric Ideology," *Journal of Psychology* 28 (1949), (cited by Rokeach [1960]).
Lichtenstein, Sarah, and J. Robert Newman, "Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities," *Psychometric Science* 9, No. 10 (1967).
Likert, R. A., "A Technique for the Measurement of Attitudes," *Archives of Psychology* 22 (1932), pp. 55.
Linstone, H. A., "On Discounting the Future," *Technological Forecasting and Social Change* 4 (1973), pp. 335–38.
——, "Planning: Toy or Tool," *IEEE Spectrum* 11, No. 4 (April 1974).
Lund, F. H., "The Psychology of Belief," *Journal of Abnormal Psychology* 20, No. 1 (1925).
Mackay, D. M., "Towards an Information Flow Model of Human Behavior," *British Journal of Psychology* 47 (1956), pp. 30.
Madden, J. M., "An Application to Job Evaluation of a Policy-Capturing Model for Analyzing Individual and Group Judgment," *Journal of Industrial Psychology* 2, No. 2 (1964), pp. 36–42.
Madden, J. M., and M. J. Giorgia, "Identification of Job Requirement Factors by Use of Simulated Jobs," *Personnel Psychology* 18, No. 3 (Autumn 1965), pp. 321–31.
Maier, Norman R. F., "Assets and Liabilities in Group Problem Solving," *Psychological Review* 74, No. 4 (July 1967).
——, *Problem Solving and Creativity in Individuals and Groups*, Brooks/Cole Press, 1970.
Mansfield, Edwin, "The Speed of Responses of Firms to New Techniques," *Quarterly Journal of Economics* 77 (May 1963).
Martin, James, *Design of Man-Computer Dialogues*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

Maslow, A., "The Further Reaches of Human Nature," Viking Press, New York, 1972.
Mason, Richard O., "A Dialectical Approach to Strategic Planning," *Management Science*, April 1969.
McGrath, Joseph E., and Irwin Altman, *Small Group Research*, Holt, Rinehart and Winston, New York, 1966.
McGregor, D., "The Major Determinants of the Prediction of Social Events," *Journal of Abnormal Psychology* 33, No. 2 (1938).
Meadow, Arnold, "Evaluation of Training in Creative Problem Solving," *Journal of Applied Psychology* 43, No. 3 (June 1959).
——, "Influence of Brainstorming and Problem Sequence in a Creative Problem Solving Test," *Journal of Applied Psychology* 43, No. 6 (December 1959).
Meister, David, and Gerald F. Rabideau, *Human Factors Evaluation in System Development*, Wiley, New York, 1965.
Menzel, H., "Innovation, Integration and Marginality," *American Sociological Review* 25, No. 5 (1960).
Miller, D. W., and M. K. Starr, *The Structure of Human Decision*, Prentice-Hall, Englewood Cliffs, N. J., 1967.
Mitroff, Ian I., "A Communication Model of Dialectrical Inquiring Systems—A Strategy for Strategic Planning," *Management Science* 14, No. 10 (June 1971).
——, and Frederick Betz, "Dialectical Decision Theory: A Meta-Theory of Decision Making," *Management Science* 19, No. 1 (September 1972).
Mulder, Mauk, and Henk Wilke, "Participation and Power Equalization," *Organizational Behavior and Human Performance* 5, No. 5 (1970).
Naylor, J. C., and R. J. Wherry, Sr., *Feasibility of Distinguishing Supervisors' Policies in Evaluation of Subordinates by Using Ratings of Simulated Job Incumbents*, PRL-TR-64-25, AD-610 812, Lackland AFB, Texas: Personnel Research Laboratory, Aerospace Medical Division, October 1964.
Newell, Allen, and Herbert Simon, "Computer Simulation of Human Thinking," *Science*, December 22, 1961.
Nowakowska, Maria, "Perception of Questions and Variability of Answers," *Behavioral Science* 18, No. 2 (March 1973).
Nylen, Donald, Robert Mitchell, and Anthony Stout, *Handbook of Staff Development and Human Relations Training*, Institute for Applied Behavioral Sciences, National Education Association, Washington, D. C., 1967.
Osborn, A. F., *Applied Imagination*, Scribners, New York, 1957.
Peters, William S., and George Summers, *Statistical Analysis for Business Decisions*, Prentice-Hall, Englewood Cliffs, N. J. 1968.
Peterson, C., and L. R. Beach, "Man as an Intuitive Statistician," *Psychological Bulletin* 68, No. 1 (1967).
Pettigrew, T. F., "The Measurement and Correlates of Category Width as a Cognitive Variable," *Journal of Personality* 26, No. 4 (1958).
Pfeiffer, J. L., "Preliminary Draft Essays and Discussion Papers on a Conceptual Approach to Designing Simulation Gaming Exercises," Technical Memorandum #1 (preliminary draft), Syracuse, N. Y.: Educational Policy Research Center, October 1968.
Pfeiffer, J. William, and John E. Jones, *A Handbook of Structured Experiences for Human Relations Training*, Vols. I, II, and III, University Associates Press, Iowa City, 1972.
Pierce, J. R., and J. E. Karlin, "Reading Rates and the Information Rate of a Human Channel," *Bell System Technical Journal* 36, No. 2 (1957).
Reid, Alex, "New Directions in Telecommunications Research," a report prepared for the Sloan Commission on Cable Communications, June 1971.
Rescher, A., "The Inclusion of the Probability of Unforeseen Occurrences in Decision Analysis," *IEEE Transactions on Engineering Management*, Vol EM-14, December 1967.
Roberts, A., "Dogmatism and Time Perspective: Attitudes Concerning Certainty and Prediction of the Future," unpublished doctoral dissertation, University of Denver, 1959.
——, and R. S. Herrmann, "Dogmatism, Time Perspective, and Anomie," *Journal of Individual Psychology* 16 (1960), pp. 67–72.
Rogers, E. M., Diffusion of Innovations, Free Press, New York, 1962.

Rokeach, M., *Beliefs, Attitudes, and Values*, Jossey Bass Inc., San Francisco, 1968.
——, "A Method for Studying Individual Differences in 'Narrowmindedness,'" *Journal of Personal*
20 (1951), pp. 219–33.
——, *The Open and Closed Mind, Investigations into the Nature of Belief Systems and Personality Systems*, Bas
Books, New York, 1960.
——, and B. Fruchter, "A Factorial Study of Dogmatism and Related Concepts," *Journal*
Abnormal and Social Psychology 53, No. 3 (1956).
Romney, A. Kimball, *et al.*, *Multidimensional Scaling*, Vol. I (Theory), Vol. II (Applications), Semin
Press, New York, 1972.
Rubenstein, Albert H., "Research on Research: The State of the Art in 1968," *Research Manageme*
11, No. 5 (September 1968).
Samuelson, P. A., "Consumption Theory in Terms of Revealed Preference," *Economics* 15 (1948
——, "Probability and the Attempts to Measure Utility," *The Economic Review*, Hitotsubsh
University, Tokyo, 1950.
Schroder, H. M., M. J. Driver, and S. Streufert, *Human Information Processing Individuals and Grou*
Functioning in Complex Social Situations, Holt, Rinehard and Winston, 1967.
Shepard, R. N., and Tegtsoonian, "Retention of Information Under Conditions Approaching
Steady State," *Journal of Experimental Psychology* 62 (1961).
Sherif, M., *The Psychology of Social Norms*, Harper Brothers, New York, 1936.
Shull, Fremong, André Delbecq, and L. Cummings, *Organizational Decision Making*, McGraw-Hil
New York, 1970.
Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, 1956.
Sigford, J. V., and R. H. Parvin, "Project PATTERN: A Methodology for Determining Relevanc
in Complex Decision Making," *IEEE Transactions on Engineering Management*, Vol. EM-12, No. 1
March 1965.
Souder, W. E., "The Validity of Subjective Probability of Success Forecasts by R&D Managers,'
IEEE Transactions on Engineering Management, Vol. EM-16, No. 1, February 1969.
Steiner, Ivan D., *Group Process and Productivity*, Academic Press, New York, 1972.
Stevens, S. S., "Ratio Scales of Opinion," in D. K. Whitla (ed.), *Handbook of Measurement an*
Assessment in Behavioral Sciences.
Stolber, Walter B., "The Objective Function in Program Budgeting, So Basic Outlines,'
Zeitschrift für die Gesamte Staatswissenschaft, 127 Band/Z Heft, May 1971.
Taylor, C. W., Widening Horizons in Creativity, Wiley, New York, 1964.
——, and R. Ellision, "Biographical Predictors of Scientific Performance," *Science* 155, No. 376
(March 3, 1967).
Thomas, Hugh, "On the Feedback Regulation of Choice Behavior," *International Journal of Man*
Machine Studies 2 (1970).
Thurstone, L. L., and E. J. Chave, *The Measurement of Attitude*, University of Chicago Press, Chicago
1959.
Torgeson, W. G., *Theory and Methods of Scaling*, Wiley, New York, 1958.
Treisman, A. M., "Monitoring and Storage of Irrelevant Message in Selective Attention," *Journal o*
Verbal Learning Behavior 3 (1964).
——, "Selective Attention in Man," *Brain Medical Bulletin* 20, No. 12 (1964).
Van de Ven, Andrew, and André L. Delbecq, "The Comparative Effectiveness of Applied Group
Decision-Making Processes," *Academy of Management Journal* (1974).
——, and André Delbecq, "Nominal and Interacting Group Processes for Committee Decision
Making Effectiveness," *Academy of Management Journal* 14, No. 2 (1971).
Vickers, Sir Geoffrey, *The Art of Judgment: A Study in Policy Making*, Basic Books, New York, 1965.
Walker, Evan H., "The Nature of Consciousness," *Mathematical Bioscience* 7, No. 1/2 (February
1970).
Wallach, M. A., and N. Kogan, "The Roles of Information, Discussion, Consensus in Group-Risk
Taking," *Experimental Social Psychology* 1 (1965).
Ward, J. H., Jr., *Hierarchical Grouping to Maximize Payoff*, Lackland AFB, Texas: Personnel
Laboratory, Wright Air Development Division, March 1961.
——, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical*
Association 58 (March 1963), pp. 236–44.

Warfield, John N., "An Assault on Complexity," a Battelle Monograph, No. 3, April 1973.
Webb, E. J., and J. R. Salancik, *The Interview, or the Only Wheel in Town*, University of Texas Press, Austin, 1966.
Webber, Melvin, *Societal Contexts of Transportation and Communication*, Working Paper No. 220, Institute of Urban and Regional Development, University of California-Berkeley, November 1973.
——, "The Urban Place and the Nonplace Urban Realm," in *Explorations into Urban Structure*, University of Pennsylvania Press, Philadelphia, 1963, pp. 79ff.
Whisler, Thomas L., *The Impact of Computers on Organizations*, Praeger, New York, 1970.
White, Ralph K., "Selective Inattention," *Psychology Today*, November 1971.
Wilson, A. G., "Research for Regional Planning," *Regional Studies* 3 (1969).
Winkler, Robert, "The consensus of Subjective Probability Distributions," *Management Science* 15, No. 2 (October 1968).
——, "The Quantification of Judgment: Some Methodological Suggestions," *American Statistical Association* 62, No. 320 (December 1967).
Zipf, George Kingsley, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Reading, Mass., 1949.

Subject Index

abuses of Delphi, 101, *see also* pitfalls
accuracy, 231–232, 241
accuracy measurement, 238–239
alternatives to Delphi, 60
anonymity, consequences of, 59, 545, 585
applications, 4 (list), 73–225, 158
  appropriateness of, 4
  business and industry, 76–78, 168–225
  of contributory Delphi, 28
  corporate, 168–226
  cross-impact, 361–365, 370–375, 375–380
  fields of, 11 (list)
  goal formulation, 262–282
  government planning, 75–76, 84–166
  to information management, 364–365
  on-line conferencing, 362–363
  as organizing tool, 361–362
  properties of, 4 (list)
  recommended, 119 (table)
  scenario-building, 466–477
  specialized techniques, 383–487
  see also examples, uses
appropriateness of Delphi, 4, 23
ARPANET, 501
Arrow's theorem, 538
audience for results, 69

bias, 246–248
bias by time period, 164–165
biased responses, 160, 230, 231

caveats, 40, 93–94, 158, 223, *see also* suggestions
certainty, 243–246
characteristics 5, 8 (table), 103, 236
  conventional, 5
  real-time, 5
cluster analysis, 385, 390–391
comments, collecting and editing, 216
commitment, 59
committees, problems with, 86
committee system, 85
committee work, via computer conferencing, 555–557
communication
  comparison of modes, 522
  computer-aided, *see* computerized conferencing
  of Delphi results, 70
  electronic, 517–534
  group, *see* group communication
  research, 520–527
  of statistical results, 221 (fig.)
  taxonomies, 528–532
  types of, 8–9 (table)
  vs. transportation, 510, 513–515
comparison of Delphi
  to normal group modes, 7–9 (table)
  to panel studies, 222–226
computer, as a communications medium, 499
computer-based learning, 550–562
computerized conferencing, 489–492, 497–515
  advantages, 499–500, 502–503, 504–505
  analysis of, 524 (fig.)
  applications, 491 (list), 510–515
  commercial application, 501–502
  cost, 491, 503–504
  and Delphi, 490–491
  in educational system, 550–562
  forecast, 550–562
  future of, 509–515
  impact on society, 514
  origins, 505–509
  possible misuses, 560–561
  reactions to, 504–505, 508–509
  utility, 491
  versions available, 500–502
computerizations, time-sharing, 281–282
computer management of society, 563–569
computer misuse, 563–569
computer utilization, 302
computers and Delphi, 487–570
conditional probability estimates, 106–107
conflictual synthetic systems, 30
conference Delphi, defined, 5
confidence, 79
confidence scale, 91–92, 274–276
conflict management and diagnosis, 557–558
consensual systems, 21
consensus, 435, 466
  artificial, 100
  operationally measured, 108
  as opinion stability, 277–281
consensus-oriented Delphi, 28, 463–486
  to facilitate, 42
  utility of, 23
consensus range, 105
consistency, 329–333, 470, 472
context, sources of, 65
conventional Delphi, defined, 5
convergence, 162–163, 167, 229
correlation between questions, 163–164
costs for a Delphi, 158
credibility, 160
criticisms, 5, 573–586
cross-impact analysis, 325–382
  annotated bibliography, 365–368
  applications, 361–365
  Bayesian-based theory, 329–335
  causal probability, 342

correlation coefficient, 344–346
difference equation method, 346–348
discrete event, 325–368
early models, 327–328
event set, 339
example, 357–361
information theory method, 353–357
KSIM, 369–382
likelihood measure method, 348–349
maximum information added, 351–353
non-Bayesian theory, 341–357
resolution of inconsistencies, 333–335
scenarios, 335–336
via simulation, 369–382
see also KSIM

deception, 585–586
definiteness, 241
Delphi
advantages, 158
as analysis tool, 100–101
applicability, 59
caveats, *see* caveats
characteristics, *see* characteristics
classical, 505
cost, 158
definition, 3, 38, 489
design considerations, *see* design considerations
as educational device, 100–101
effort involved, 213 (fig.)
effectiveness, 115–118
evaluation, *see* evaluation
examples, *see* examples
flow, 39 (fig.)
future for, 487–570, 496
goals, 96
graphically aided, 40, 403
as heterogenistic tool, 494–495
impact of monitor, 57
information flow, 214 (fig.)
interpretation of results, 70
manpower needs, 212
mathematically treated, *see* group estimation
modifications, 59, 60, 62, 103, *see also* cross impact, computer conferencing, policy Delphi
objectives, 125–126
panels, *see* panel
performance, 312–315
philosophy, *see* philosophy
pictorially aided, 435
pitfalls, *see* pitfalls
potential improvements, 35
problems, *see* problems
procedure, *see* study design
questionnaire design, *see* questionnaire design
reasons for using, 114
refinements, 103
reliability, 116
results, *see* results
roles of, 76, 86
statements, 232–233
suggestions, *see* suggestions
underlying assumptions, 239–240
uses, *see* uses
vs. face-to-face discussion, 291–321
design considerations, 64–70, 120–121
context, 65
communication of results, 70
interpretation and summation of responses, 70
leadership roles, 545–546
orchestrating interaction, 69–70
panel creation, *see* panel, panel selection
questionnaire design, *see* questionnaire design
stimulating response, 68–69
time, 65–67
design considerations, *see also* study design
design-monitor team, 93
effort required of, 213 (fig.)
desirability scale, 90, 136 (table), 472
dialectical inquiring system, 29–33
contrasted to adversary procedure, 32
see also Hegelian IS
Dialectical Policy Inquirer, 31–32
differential salience, 416–417
discounting the future, 574–578
DISCUSS, 500–501
dispersion, 229–230
dogmatism
effects on Delphi, 288–290
dominance judgement, 409
dyadic, 408

effectiveness of Delphi, 117 (table)
encouraging involvement, 115, 118
facilitating communication, 115, 118
electronic media communication, 517–534, 539–540
EMISARI, 500, 506–507
empirical science, 21
error measurement, median group error, 294
estimation
desirable features of, 240–248
multidimensional, 399–400, 402–431
group, *see* group estimation
individual, *see* individual estimation
evaluation, 89–90, 114–119, 159, 227–322
comparative, 116 (table)
empirical, 115, 118
of feedback, 270–272, 276–277
of impact of study, 159
by panelists, 115–118
performance, 291–327
reliability, 116
replicability, 115
summary, 320–321
see also validity
evaluation criterion for Delphi, 54
evaluation matrix, 105
example, 113 (fig.)
event probability distribution, 469 (fig.), 475 (fig.)
evolution of Delphi, 10–11
examples
aircraft competition, 370–375
business, 174–175
chemical, 171
civil defense policy, 96–97
communications, 171
computer, 169
county government, 96
cross impact, 357–361
drug abuse, 124–159
education, 174–175
exploring human relationships, 52
forecast government employment, 99
genetic counseling services, 100
glass, 171
Hegelian, 96–97
home computer uses, 175, 182
housing, 170
images of future, 436–439
Kantian, 95–96
land-use planning, 97
leisure and recreation, 169
market opportunities, 66–67
market research, 170
medical, 78–81, 174–175
metropolitan transit, 375–380
national priorities, 95–96
newsprint, 169
personnel management, 170
pharmaceutical, 171
philosophical orientation of, 76, 77
planning, 81–83
plastics, 195–209
pollution sources, 111 (fig.)
quality of life, 387–401
reaction to retarded, 41
regional planning, 112–114
social change, 170
steel and ferroalloy, 210–226
telephone, 169
water resources management, 102–123
see also uses, applications
experimental consensual systems, 21
expertise
effect on group performance, 295–297
illusory, 581–582
role of, 30, 84
vs. non-experts, 112, 144–145 (table)
extrapolative scenario, 469

face-to-face discussion
computer-aided, 502
vs. Delphi, 291–321
failure of Delphi, reasons for, 6
factor analysis, 396–398
fallacies, 584–585
feasibility scale, 90–91, 135 (table), 472
feedback, effects of, 270–272, 276–277
forced choice, 510
forecasting vs. fact-finding, 299–300
formal symbolic systems, 24
format of Delphi, *see* study design
FORUM, 517
foundations of Delphi, 17–36
Lockean basis, 22
future prospects of Delphi, 191–192, 496, 498–570
fuzzy set theory, 478, 494

goals, defined, 263
goals, Delphi, 263–267
group communication, 535–549
large widespread groups, 548–549
steps in feedback session, 540–544
using portable equipment, 546–547
via electronic media, 517–534, 539–540
group estimation, definition, 236
group estimation theory, 236–261
axiomatic approach, 257–259
error theory, 249–251
probabilistic approach, 251–257
score, 237–239
see also, estimation
group feedback session, 540–544
group ordering problems, 538

habitus mentalis, 48–49
Hegelian Inquiring System, 29–33
characteristic questions, 19
characteristics, 29
suitability, 31
heterogeneity, 494
heuristic process, 482
importance of, in forecasting, 581–582
history of Delphi, 10

honesty, 240
HYPERTUTOR, 501

idiomergent, 49–51
  vs. industrial, 51 (tables), 54 (table)
importance scale, 91, 137 (table)
indices, *see* scales
individual estimation, 240–247
INDSCAL, 414–426
inductive representation, 21
information
  flow in Delphi, 214 (fig.)
  from a dialectic viewpoint, 30
information content
  of Leibnizian IS, 24
  of Lockean IS, 21
in-house vs. consultants, 190
innovation, 512
inquiring systems (IS), 19–35, 43–44, 566–567
  definition, 21
  differentiated, 18–19
  goal orientation of, 29
  Hegelian, 29–33
  Kantian, 25–29
  Leibnizian, 23–25
  list of, 15, 19
  Lockean, 21–23
  Merleau-Ponty, 43–44
  Singerian, 33–35
interaction matrix, 371 (fig.), 376 (fig.)
interaction modes, 57–64
  affairs, 62
  comparison, 58 (table)
  episodes, 60–61
  events, 61–62
  experiences, 59–60
  occurrences, 63–64
  transactions, 57–59
interquartile spread, 163 (fig.), 216
interrelated events, 106, *see also* cross-impact analysis
interval scale, importance of, 272

justifications of Delphi, 10

Kantian inquiry
  examples of, 28
Kantian Inquiring System, 25–29
  characteristic questions, 19
  characteristics, 26
  suitability, 29
KSIM, 369–382
  computer output, 377–380
  examples, 370–375, 375–380
  interaction matrix, 371 (fig.), 376 (fig.)
  interactions, 376–377
  mathematics, 372–373
  policy testing, 377–380
  variables, 370, 375

Law of Comparative Judgment, 273
learning curve, 469 (fig.)
Leibnizian Inquiring System, 23–25
  characteristic questions, 19
  characteristics, 23
  suitability, 25
letter of invitation, 93–94
Likert scale, 272
Lockean Inquiring System, 21–23
  basis of Delphi, 22
  characteristic questions, 19
  characteristics, 21
  methodological example, 22
  problems, 22
  strengths, 23
  suitability, 23
  weaknesses, 23
long-range planning via computer conferencing, 555–557

MAILBOX, 501–502, 507–508
management, aided by computer conferencing, 512–513
materials, 69
MDSCAL, 411
measure of polarization, 92
median group error (MGE), 294
Merleau-Ponty Inquiring System, 43–44
  applications, 43
methodology, 262–287
misusing Delphi results in business, 189–190
models
  of Delphi, 25
  empirical, 21
  inquiring system, 21
motivation, 69
multidimensional scaling, 385, 402–431
  advantages, 414
  annotated bibliography, 426–431
  definition, 402
  example, 410–414, 418–426
  INDSCAL, 414–426
  ordinary, 409–414
  preference analysis, 422–423
  two-way, 409–414
  uniqueness property, 419
  use in Delphi, 403, 426
multi-model synthetic systems, 27
multiple studies, 190

normalization, *see* multidimensional scaling
normative scenario, 469
normative system-building, 463–486 (fig. 465, 485)

  clustering, 473–476
  learning loops, 482–484
  policy generation, 480–482
  simulation, 481
  utility, 470
normative systems, 464
NUCLEUS, 550, 558
number of rounds, 88, 94, 212, 229, 320
  stopping criterion, 281

objectives, defined, 263
opinion stability measure, 277–281
optimism-pessimism consistency, 230, 231, 584
ORACLE, 501, 550–562
ordinal scale, 272
overselling, 584–585

panel
  compared to Delphi, 222–226 (fig. 224)
  contributions required of, 213 (fig.)
  creation, 68
  mix, 68, 102
  respondent's objectives, 133 (table)
  self-definition, 57
  size, *see* panel size
  subdivision, 103
panel selection, 94, 127, 210–211, 582–583
panel size, 86
  effect on performance, 292–295, 315–318, 320
participation statistics, 130 (table)
participatory democracy, 493
PARTY LINE, 500, 506
payment for Delphi by corporations, 187–188
payments, 69
perception, 414–426, 576–577
performance evaluation, 291–327
phases of Delphi, 5–6
philosophical modes, 18–35
philosophy, 15–71
pictorial aids, 435
pitfalls, 571–586
  deception, 585–586
  discounting the future, 574–578
  illusory expertise, 581–582
  optimism-pessimism bias, 584
  overselling, 584–585
  prediction urge, 578
  simplification urge, 579–581
  sloppy execution, 582–584
PLATO, 500–501
policy Delphi, 84–101
  as precursor to committees, 86
  examples, 95–100, 124–159
  guidelines, 93–94
  measure of polarization, 92
  mechanics of, 87–95
  participants, 88–89
  phases of, 88
  problems, 92, 100–101
  rating scales, 89
  role of, 86–87, 100
  sample questions, 97–98
  size, 86
  uses, 94–95
policymaking, 463–464, 480–482
policy question, 75, 84
prediction urge, 578
PROBE, 170
problem identification, 477–480
problems, 6, 22, 100–101, 189–190
  biased responses, 160, 166–167
  criticisms, 6–7, 573–586
  dogmatic drive for conformity, 582
  reasons for failure, 6
  group ordering, 538
  see also pitfalls
probability estimates
  difficulty with, 107
  guide for, 107–108
problem solving via computer conferencing, 510–511
proprietary problems, 191

quality of life, 495
  difficulty in measuring, 582
  factor analysis, 396–398
  models, 388–390, 389 (fig.)
  rankings, 392–393 (tables)
  results, 391–396
question design, 5[illegible]41, 543 (fig.)
questionnaire design, 93, 127–134, 172–174, 196, 198, 232–233
  examples, 106 (fig.), 154–155, 198 (fig.), 200–201 (fig.), 204 (fig.), 441–442, 448–462
  mailing delays, 129
  sample questions, 97, 162
  time for completion, 129, 132
  for trend extrapolation, 219 (fig.)

realism, 241–243
reality construction, 37–71
real-time Delphi, definition, 5
regression analysis vs. subjective judgment, 395–396
reliability, 116
respondents, *see* panel
response distribution, 230
response rate, 132
results of Delphi, 40, 134–158, 437–439
  create an audience for, 69
  interpretation of, 70, 583
  used as data, 188

risk analysis by Delphi, 77–78
roles of Delphi, 76
round one, 212–216
  design, 212–214
  information package, 104
  results, 214–216
rounds, number of, 88, 94, 212, 229, 320
rounds, stopping criterion, 281
round three, 220–222
round two, 216–220
  design, 216–219
  feedback, 104, 105
  results, 220
rule of the triangle, 330

scales, 89, 90–92 (table), 135–137, 218 (fig.), 262, 467, 471
  confidence, 91–92, 218 (fig.)
  desirability, 90, 136 (table)
  feasibility, 90–91, 135 (table)
  importance, 91, 137 (table)
  see also multi-dimensional scaling
scaling techniques, comparison, 272–273
scenario-building logic, 464–470
scenario-building technology, 470–477
scenarios, 335–336, 386, 469
  definition, 470
  examples, 550–561, 563–569
  use of, in Delphi, 79–80
self-rating, 129, 131 (table), 233–234, 296
  significance of, 310–312, 320
simplification urge, 579–581
Singerian Delphi, 35
Singerian Inquiring System, 33–35
  characteristic question, 19
  distinctive features, 34
  main features, 33
  potential for Delphi, 35
  strengths, 35
  weaknesses, 35
sloppiness in Delphi, 582–584
social choice, 537–539
sociometry, 421
specialized techniques, 383–487, *see also* cross impact analysis, policy Delphi, computer conferencing
statistical methods, re Lockean IS, 22
statistical summaries, 109 (fig.), 221 (fig.), 438, 444–447 (fig.)
stimulating response, 68–69
strengths of Delphi, 23, 586
study design, 126–127, 172–174, 182 (table), 210, 390–391, 436–437, *see also* design considerations
study sophistication, 160
subjective probability, 580
suggestions for Delphi, 64–70, 226
  introduce ambiguities, 44–45
  recommendations, 120–121
  introduce *what if* items, 46
  see also caveats, design considerations
syncretic scenario, 464
synergistic thinking, 435

technology assessment, 29
teleconferencing, 517–520
teleological, 33
time, 129, 132
time needed by participants, 306–307
time domain, 65–67
trend extrapolation, 237
trend extrapolation form, 219 (fig.)
truth content of Delphi, 24
types of Delphi, 5
  conventional, 5
  real time, 5

uncertainty, 229, 243–246
usefulness of Delphi, 54
uses of Delphi, 46–47, 51, 121, 125
  administrative planning, 82
  budget estimation, 89
  combine opinions, 160
  as educational process, 94
  educational area, 82
  elicit a hot list, 83
  encourage participation, 46
  establishing priorities, 161
  estimating historical data, 78–79
  examining the past, 82–83
  filter out noise, 83
  health-care planning, 81
  investigate past performance, 95
  probe insights, 47
  problem identification, 81
  regional planning, 81–82
  risk analysis, 77–78
  see also examples, applications
using Delphi results in business, 188–189
utility, 538

validity, 18
  as consensus, 22
  comparison to real data, 79
  Lockean, 21
  philosophical positions, 19–20
  see also evaluation
validity scale, 218 (fig.)
variance, as function of rounds, 299
variations in Delphi, 59, 60, 62