

UNIVERSAL
LIBRARY

OU_156476

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 658.23 / 487B Accession No. 31879

Author Lorie J. H

Title Basic methods of . . . Research

This book should be returned on or before the date last marked

Basic Methods of Marketing Research

Basic Methods of Marketing Research

JAMES H. LORIE

*Associate Professor of Marketing
The University of Chicago*

HARRY V. ROBERTS

*Assistant Professor of Statistics
The University of Chicago*

First Edition

1951

MCGRAW-HILL BOOK COMPANY, INC.

NEW YORK TORONTO LONDON

BASIC METHODS OF MARKETING RESEARCH

Copyright, 1951, by the McGraw-Hill Book Company, Inc. Printed in the United States of America. All rights reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publishers.

PREFACE

The scope and objectives of this book are described in some detail in its first chapter. Briefly, we have attempted to explain the basic methods of marketing research in as nontechnical a manner as is compatible with an understanding of them. We use "methods" to mean the basic tools underlying marketing research, such as experimentation, sampling, and communication, rather than the numerous specific applications of these tools, such as the "radio diary" or the "pantry inventory." Specific applications are used to illustrate basic methods, but only the basic methods have been given thorough and systematic exposition.

We have freely asked our friends for help in the preparation of this manuscript, and our debt to them is very great. Mr. Eugene C. Pomerance of Foote, Cone & Belding and Mr. Peter G. Peterson of Market Facts, Inc., have read the entire first draft of the book and given us most helpful and detailed criticisms. Mr. Roe Goodman of the Survey Research Center of the University of Michigan has read our discussion of sampling and made many general and specific suggestions for improvement of the first draft. We have made frequent changes at his suggestion, and we have even inserted his comments directly into the text without special acknowledgment.

Others have read the entire manuscript and made helpful suggestions. Among these are Mr. Seymour Banks of De Paul University, Mr. Jacob Duker of the University of Toledo, Mrs. Dorothy Fromherz of Leo Burnett Co., Inc., Mr. John Jeuck of the University of Chicago, Mr. Zenon Malinowski of the University of Connecticut, Mrs. June Roberts, Mr. Joseph C. Seibert of Oxford Research Associates, Inc., and Mr. William Sell of the Minnesota Mining and Manufacturing Co. Mr. Lyle Johnson and Mr. William Wolf have helped as research assistants during the writing of the book. Special acknowledgment is due the marketing research class at the School of Business of the University of Chicago in the summer of 1950, since they read the first draft and contributed many valuable suggestions; particularly useful were the comments of K. C. Mulcahy, Donald S. Miller, Sigmund P. Zobel, and Thomas V. Cinquina. Consultation with Miss Bee Angell of Bee Angell & Associates on some of the material in Part V was very helpful.

While their influence has not been so direct, many others have left a strong impression on this book. Among these are Mr. G. Maxwell Ule of

Kenyon & Eckhardt, Mr. W. Allen Wallis and Mr. George H. Brown of the University of Chicago, Mr. Benjamin Tepping of the U.S. Bureau of the Census, and Mr. Angus Campbell of the Survey Research Center of the University of Michigan. Others in marketing research, government, and academic circles have been the source of many ideas in the book; unfortunately, an attempt to list them all would not be feasible.

During the preparation of the book, Dean Garfield Cox of the School of Business of the University of Chicago has generously extended to us the help of the clerical staff of the school; the typing of the different drafts of the manuscripts was supervised with skill and patience (with the authors) by Mrs. I. Lawrance. Mrs. June Roberts, in addition to commenting critically on the manuscript, has expedited much of the mechanical work of preparing it in its final form. Mr. Edward Werner of the University of Wisconsin helped prepare the index and deserves our thanks. The Division of Special Surveys, Bureau of Agricultural Economics, in the U.S. Department of Agriculture and its director, Dr. Forrest E. Clements, have made available the material reproduced in Chapter 20.

JAMES H. LORIE
HARRY U. ROBERTS

CHICAGO, ILL.
July, 1951

CONTENTS

PREFACE	v
---------------	---

PART I. INTRODUCTION

Chapter 1. INTRODUCTION: PURPOSE AND METHOD

1.1 The Basic Nature of Marketing Research.....	3
1.2 The General Purpose of the Book.....	4
1.3 The Audience and the Level of Discussion.....	6
1.4 The Organization of the Book.....	7

Chapter 2. A BRIEF DESCRIPTION OF MARKETING RESEARCH

2.1 Introduction.....	9
2.2 Types of Marketing Research Activity.....	9
2.3 The Agencies Performing Marketing Research.....	12
2.4 The Methods of Marketing Research.....	15
2.5 Business Environment of Marketing Research.....	18

PART II. SCIENTIFIC METHOD IN MARKETING RESEARCH

Chapter 3. SCIENTIFIC METHOD

3.1 Introduction.....	25
3.2 A Comparison of Scientific and Nonscientific Investigation.....	27
3.3 Observation.....	29
3.4 Formulation of Hypotheses.....	30
3.5 Prediction or Deduction.....	32
3.6 Testing Hypotheses or Evaluating Evidence.....	34

Chapter 4. SCIENTIFIC METHOD IN MARKETING RESEARCH

4.1 Introduction.....	40
4.2 Experimentation.....	45
<i>Controlling and Diminishing Risks of Error. Simultaneous Evaluation of the Effects of Several Variables. Further Comments on Limitations of Experimentation.</i>	
4.3 Observational Data.....	57
<i>Cross-classification. Other Statistical Tools. Other Methods of Using Observational Data.</i>	

Chapter 5. THE USE OF MARKETING RESEARCH FOR MARKETING PROBLEMS

5.1 The Recognition of Problems.	67
<i>Recognizing Problems concerning Frequently Recurring Expenditures. Recognizing Problems concerning Major Policies.</i>	
5.2 The Problem of Devising Alternatives to Established Policy. . .	78
5.3 The Use of Research To Solve Marketing Problems.	79

PART III. SAMPLING

Chapter 6. AN INTRODUCTION TO SAMPLING

6.1 The Place of Sampling in Marketing Research	83
6.2 The Nature of Sampling Theory.	85

Chapter 7. SOME FUNDAMENTALS OF SAMPLING THEORY

7.1 Introduction.	89
7.2 Populations and Samples	89
7.3 Probability and Randomness: Simple Random Sampling . .	91
7.4 Sampling Distributions	95
<i>The Mean of the Sampling Distribution of the Mean: Unbiased Estimates. The Standard Deviation of the Sampling Distribution of the Mean: The Standard Error of the Mean. (1) A Note on "Efficiency." The "Form" of the Sampling Distribution of the Mean.</i>	
7.5 Confidence Intervals.	103
7.6 Additional Comments.	108

Chapter 8. APPLICATION OF SAMPLING THEORY TO THE SAMPLING OF HUMAN POPULATIONS

8.1 Introduction	110
8.2 Cluster Sampling	115
<i>Cluster Sampling without Subsampling: Equal-sized Clusters. (1) How Big Should the Cluster Be? (2) How Should the Clusters Be Formed? Cluster Sampling without Subsampling: Unequal-sized Clusters. Cluster Sampling with Subsampling: Equal-sized Clusters. Cluster Sampling with Subsampling: Unequal-sized Clusters. Multistage Sampling with More than Two Stages. (1) Numbers of Clusters at Different Stages.</i>	
8.3 Stratification.	134
<i>Combination of Cluster Sampling and Stratified Sampling.</i>	

Chapter 9. FURTHER COMMENTS ON PROBABILITY SAMPLING FROM HUMAN POPULATIONS

9.1 Additional Refinements of Probability Sampling.	142
<i>Variable Probabilities of Selection. Disproportionate Stratified Sampling. Systematic Sampling. Double Sampling, or Two-phase Sampling. Sequential Sampling.</i>	
9.2 A Note on Methods of Estimation.	157
9.3 Practical Problems in Area Sampling.	158
<i>Identification of People with Areas. People Who Aren't at Home at the Time of Interview. Refusals. Other Problems.</i>	

Chapter 10: NONPROBABILITY SAMPLING

10.1 Introduction.	165
10.2 The Distinction between Probability and Nonprobability Sampling.	168
<i>The Use of Judgment in Selection.</i>	
10.3 A Rationale of Nonprobability Sampling.	171
<i>Quota Sampling.</i>	

Chapter 11. SOME SAMPLING PROBLEMS IN MARKETING RESEARCH

11.1 Introduction.	181
11.2 "How Big Should the Sample Be?"	181
11.3 "What Kind of Sample Do We Need?"	185
<i>Requirements for Accuracy. Cost of Obtaining Information. Recapitulation: Probability vs. Nonprobability Sampling.</i>	
11.4 Adaptations of Sampling Designs for Marketing Research.	192

PART IV. COMMUNICATION AND OBSERVATION

Chapter 12. INTRODUCTION TO THE PROBLEM OF COMMUNICATION AND OBSERVATION

12.1 Introduction.	201
12.2 Kinds of Response Errors.	203
<i>Extension of "Reliability" and "Validity" to a Group to Individuals.</i>	
12.3 Sources of Response Error.	205
12.4 The Techniques for Securing Information.	207
<i>Communication and Observation.</i>	
12.5 The Criteria for Judging Techniques.	208
12.6 A Classification of the Kinds of Information Useful in Marketing Research.	208

*Information about Behavior. Information about Intentions.
Opinions and Attitudes. Knowledge and Habits and Thought.
"Reasons Why."*

12.7 The Organization of Part IV.. . . .	214
--	-----

Chapter 13. SECURING INFORMATION ON BEHAVIOR

13.1 Introduction.	215
13.2 Observation.	217
<i>Accuracy of Observation. Unambiguity of Observation. Cost. Ability and Willingness of Respondents To Permit Observation.</i>	
13.3 Self-observation.	221
<i>→Accuracy and Unambiguity.</i>	
13.4 The Communicative Techniques.....	224
<i>Memory. (1) Seeking Information about Recent Events. (2) Stimuli to Memory—The Recognition Method—Aided Recall— Unaided Recall—Types of Events and the Ability To Remember. Selecting the Person Best Able To Provide the Information. Am- biguity in Communication.</i>	
13.5 The Willingness of Respondents To Provide Information.	243
<i>Unwillingness Because of Anticipation of Reward or Punish- ment. Unwillingness for Emotional Reasons. Unwillingness To Take Time To Provide Thoughtful Answers.</i>	
13.6 Cost.	252

Chapter 14. SECURING INFORMATION ON INTENTIONS

14.1 Introduction.	253
14.2 The Usefulness of Statements of Intention in Forecasting	256
14.3 Accuracy and Ambiguity...	262

Chapter 15. SECURING INFORMATION ON ATTITUDES

15.1 Introduction.	265
15.2 Getting Accurate and Unambiguous Expressions of Attitude	266
<i>Techniques for Obtaining Attitudes. (1) The Single-question Technique. (2) Ranking Individuals According to Attitudes to- ward a Single Stimulus. (3) Ranking Stimuli According to At- titudes toward Them. Willingness of Respondents. The Ability of Respondents To Provide Accurate and Unambiguous State- ments.</i>	

Chapter 16. SECURING INFORMATION ON KNOWLEDGE AND HABITS OF THOUGHT

16.1 Introduction.....	289
16.2 Types of Knowledge.....	291

Knowledge of an Already Adopted Stimulus. Knowledge of Stimuli Introduced under Test Conditions. Understanding Sales Messages.

16.3 Habits of Thought.....	294
<i>Associations. Other Reactions to Stimuli.</i>	
16.4 Ambiguity.....	296
16.5 Accuracy.....	297

Chapter 17. SECURING INFORMATION ON "REASONS WHY"

17.1 Introduction.....	299
17.2 Ambiguity.....	301
<i>Getting the Respondent To Understand the Exact Meaning of the Question concerning Motivation. Enabling the Respondent To Give a Coherent and Complete Answer. Understanding the Implicit As Well as the Explicit Assumptions Underlying the Respondent's Answer.</i>	
17.3 Accuracy.....	308
<i>Inaccuracy Attributable to Unwillingness. Inaccuracy Due to Inability.</i>	

PART V. ADMINISTRATION

Chapter 18. SOME GENERAL COMMENTS ON THE ADMINISTRATION OF MARKETING RESEARCH

18.1 Introduction.....	319
18.2 Organization of a Marketing Research Group.....	320
18.3 What Functions Should Be Performed Internally? ..	320
18.4 Two Instruments of Control..	325

Chapter 19. COORDINATING RESEARCH WITH THE NEEDS OF MANAGEMENT

19.1 Introduction.....	330
19.2 Exploration: Defining the Problem.....	330
<i>Knowledge of the Enterprise, the Industry, and the General Business Environment. Relationship between Research and Management. Sources of Secondary Data. Exploration in the Field. (1) Exploratory Interviews with (or Observation of) Consumers, Dealers, Salesmen, etc. Decision as to Whether Formal Research Is Needed.</i>	
19.3 Design of Projects.....	338
<i>Rough Sketch of the Sampling Design. Developing the Questionnaire. Pretesting. Final Selection of the Sample. Scheduling</i>	

<i>Fieldwork. The Preliminary Code. Detailed Specifications for Tabulation. Some Organizational Problems in Project Design.</i>	
19.4 Reporting to Management.....	345
<i>Interim Reports to Management. The Written Report. Different Methods of Reporting to Management. Following Up.</i>	
Chapter 20. INTERVIEWING.	
20.1 Introduction..	349
20.2 Interviewing.....	349
<i>Qualifications for Interviewers. How To Interview. Introduction. Setting the Stage for the Interview. Asking the Questions. Getting Adequate Answers. Recording the Interview. Gathering the Face-sheet Data. Closing the Interview. Sampling Problems in Fieldwork.</i>	
Chapter 21. SELECTING, TRAINING, AND SUPERVISING INTERVIEWERS	
21.1 Selection of Interviewers.....	378
<i>Sampling Design and Selection of Interviewers. Finding Interviewers. Selection of Interviewers. (1) Application Blanks. (2) The Employment Interview. (3) References. (4) Personnel Tests.</i>	
21.2 Training of Interviewers.....	390
<i>Initial Training. Training for Individual Studies.</i>	
21.3 Supervision of Interviewers.....	396
<i>Evaluation of the Quality of Fieldwork. Relations with Fieldworkers.</i>	
Chapter 22. PROCESSING MARKETING RESEARCH DATA	
22.1 Introduction.....	406
22.2 Coding.....	406
<i>Precoding. Editing. Final Coding.</i>	
22.3 Tabulation.....	417
<i>Hand Tabulation. Machine Tabulation. (1) Punching. (2) Running. Other Methods of Tabulation.</i>	
22.4 Calculations.....	429
22.5 Mechanical Preparation of Reports.....	430
<i>Some Suggestions for Further Reading</i>	433
INDEX.....	439

PART I

INTRODUCTION

CHAPTER 1

INTRODUCTION: PURPOSE AND METHOD

1.1 The Basic Nature of Marketing Research

Consumers' attitudes and behavior in a competitive economy influence to a large extent the survival and profitability of business enterprises. When economic life was less specialized and individual firms were, for the most part, smaller than now, the "manufacturers" of the day often knew their customers personally. As a result, the task of determining consumers' attributes and desires was probably relatively simple—though still often not easy. When a shoemaker started to work on a pair of shoes he had the advantage of having talked directly to the purchaser to ascertain his needs. If the shoes were not satisfactory, the shoemaker heard about it directly. Today, in contrast, the typical shoemaker seldom knows much about the people who purchase his product. These people are usually scattered so widely that they can be reached only by a complex selling organization. The number of people and agencies intervening between the shoemaker and the salesmen who talk to the purchasers is so great that communication from purchaser to producer is almost inevitably meager and unsatisfactory. The manufacturer may not know the magnitude of his firm's sales to consumers, where these sales are being made, the kinds of people who are buying his product, and what these people like and dislike about it. Since the attributes and desires and reactions of these consumers are of such crucial importance to the survival of the manufacturer, it seems reasonable to believe that manufacturers could profit by enabling consumers to transmit more adequately information about these things.

Since consumers are frequently spread over thousands of miles and since information about them may in each instance be difficult to ascertain, it is not surprising that people with specialized skills and interests are often employed to establish and maintain the channel of communication from consumers to the businessmen who serve them. People who have specialized in this task call their specialty "marketing research."¹ Marketing research

¹ Marketing research is not limited to this task of facilitating communication from consumer to businessman, but in this book the discussion will be limited to this aspect of marketing research. Many agricultural economists do marketing research which is focused on the plans and attitudes of producers rather than consumers. There is a further discussion of this difference in types of marketing research in the second chapter, which is intended to be a brief description of marketing research in this country.

provides information on a wide variety of subjects such as the effectiveness of advertisements, the preferences of consumers for different product designs or packages, current rates of sale, the intentions of consumers regarding their future consumption plans, etc. All these types of information involve consumers and their attitudes and behavior.

1.2 General Purpose of the Book¹

In this book we are concerned exclusively with the problems involved in getting this information accurately and efficiently. Unfortunately, it is often easy to forget that getting this information is not an end but merely a means to help make possible marketing decisions which will best achieve the objectives of the business enterprise. Nor should one forget that research, which is merely systematic and objective investigation, is only one process by which this information can be secured. The shoemaker of the seventeenth century got his information about consumers by a personal and simple process for which the term "research" seems inappropriate. Similarly, in the twentieth century, information about consumers is very often secured by relatively unsystematic observation of and conversation with consumers. Also, information about consumers can be secured by executives through the intuitive and nonformal interpretation of their own experience. It is undoubtedly true that the vast majority of business decisions are made without the assistance of research, and it is equally true that many of the most talented sales executives have had little appreciation of or skill in the systematic and formal procedures of marketing research. Many marketing executives believe—and with some justification—that all that is required to establish adequate communication with consumers is some sort of informal liaison with the people who buy or might buy the company's product. For example, such executives might rely on occasional conversations with retailers for information about the purchasers of their product.

In spite of the wide reliance on and great value of common sense, good intentions, and unsystematic procedures, often the task of understanding consumers is sufficiently important to justify the costs of more specialized, technical training and more systematic procedures. Common sense may not prevent one from making invalid generalizations about all consumers on the basis of information regarding a small and atypical group of the consuming public; training in statistical inference may be necessary. Honesty and good intentions may not be sufficient to prevent investigators from

¹ Much of the material in the remainder of this chapter is of the sort that is ordinarily found in prefaces or forewords. We are including it in the body of the book because (1) more people will read it if it is there rather than in a prefatory statement and (2) the remainder of the book will be more easily understood if this introductory material is read.

phrasing a question so as to suggest an answer or imply that prestige or social approval is associated with a particular response; training in psychology may be necessary. The determination of attitudes is difficult; psychometrics may be needed. Reasons of this sort often make formal and systematic research a more fruitful source of useful information than other types of investigation; reasons of this sort are the justification for books on marketing research.

Someone may object at this point that the sort of training (statistical inference, psychology, psychometrics, etc.) which has been indicated as being necessary for marketing research is training in the basic disciplines of the social sciences rather than in anything called "marketing research." Since the problems encountered in marketing research are usually problems common to the various social sciences, would it not be better to become thoroughly trained in these basic disciplines rather than in marketing research? The answer is "yes." If one could become a competently trained economist, statistician, psychometrician, sociologist, etc., and also become familiar with the structure and functioning of marketing institutions, the need for marketing research as a separate field of study would disappear. The study of marketing research is not justified because it produces better marketing researchers than the indicated alternative course of study. It is justified because training in marketing research is *possible*, while the alternative—in spite of the size and quality of our educational system—usually is not. The study of marketing research can be viewed, then, as a compromise between the unattainable ideal of persons thoroughly trained in all the social sciences and the opposite extreme of persons unfamiliar with the requirements of valid research in even a limited field of application of the social sciences. By excluding that part of economic theory which is unimportant for the needs of marketing research, by restricting the study of statistics and psychology to those concepts and techniques which are useful in solving the most important and frequently encountered marketing problems, and by focusing only the relevant segments of the other disciplines on the limited range of social behavior with which marketing is concerned, it may be possible to achieve this compromise.

Some people have suggested another possible alternative to a thorough training in all the social sciences. They suggest the intensive training of specialists who can be coordinated in achieving some goal requiring the exercise of several specialized skills or the possession of several specialized types of knowledge.¹ Specialists are certainly needed in marketing re-

¹ During World War II, teams of specialists were attached to various high military commands in America and Britain with the objective of studying important military problems. These "operations research" or "operations analysis" teams were composed of such diverse specialists as mathematicians, lawyers, economists, engineers, and statisticians. Up to now, however, operations analysis has not been widely

search, but their effective coordination requires either a coordinator familiar with the fundamental concepts and practical difficulties in the specialties which are to be coordinated, or, failing this, specialists with a reasonable knowledge of the related fields. This book is intended to facilitate coordination of both sorts. Achieving coordination among specialists and achieving the objective referred to in the preceding paragraph are the same thing.

1.3 The Audience and the Level of Discussion

A book conceived in the way indicated above might be useful to various groups in various ways. First, students just beginning their study of marketing research may come to recognize the demands of a career in marketing research, may gain a first understanding of the principles of the underlying disciplines, and may see outlined together both the areas of their deficiencies and references to sources which may remedy them. Second, those who buy and use marketing research may acquire an understanding of its requirements and may therefore become more competent critics and more efficient and wiser purchasers and users. Third, practitioners, according to their background and experience, may find a new way of thinking about the subject, may be introduced to new ideas and techniques, may find their particular specialties placed in a useful context, or may simply find a new organization of already familiar material.

In keeping with the above conceptions and objectives, the discussion of the book is not technical and does not assume extensive specialized training. The technical prerequisites are simple. The reader should be familiar with the contents of the usual elementary course in descriptive statistics and the rudiments of statistical inference. An elementary course in economics would be helpful though not essential. Finally, some familiarity with marketing problems would help.

The modest list of technical prerequisites does not mean, however, that the ideas presented in this book are always self-evident or easy. Even a nontechnical discussion of the theory underlying marketing research requires close attention.

The emphasis is strongly upon a study of fundamental methods of research rather than upon a description of the techniques, gadgets, and devices currently used. There are two reasons for this emphasis on analysis and slighting of description. First, any manuscript attempting a comprehensive description of current practice would be out of date before its publication. We attempt, however, to indicate the nature and relevance

applied in civilian enterprise. The academic specializations encountered in market research are largely those of the social sciences and statistics.

of current developments in the underlying disciplines (*e.g.*, statistics, psychology, etc.) for marketing research. These developments are usually less frequent and are almost always of more fundamental and enduring importance than are the changes in the practice of commercial research. The second reason for foregoing extensive description is that such knowledge can be acquired from experience more efficiently than from studying. An understanding of the fundamental principles of scientific research, on the other hand, can probably be acquired more efficiently through formal instruction.

Usually the discussion is at high levels of generalization. This level of generality is maintained largely in the belief that the understanding of very general principles of analysis is an essential prerequisite to the solution of particular problems. The relevance of these generalizations for a particular problem is usually indicated by the use of examples of marketing research projects.

1.4 The Organization of the Book

The organization of the book has been foreshadowed by the preceding discussion. A short chapter following this one describes various types of marketing research projects, the agencies conducting them, and the methods used. This descriptive chapter is intended to provide a minimum background of institutional knowledge for those who must learn principles useful in analyzing problems before actually encountering them in practice. The present introductory chapter and the following descriptive one constitute the first part of the book.

The second part of the book deals with the nature of scientific research and its application to marketing. There are three chapters in this part. The first discusses scientific method in general. The second discusses the applicability of scientific method to the so-called social sciences and to marketing research in particular. This chapter contains a fairly extended discussion of experimentation concerning social behavior and of the difficulties involved in making predictions based upon the analysis of historical data. The third and final chapter of the second part is an attempt to indicate a systematic procedure for bringing marketing problems to light and for allocating research resources to their solution.

The second part deals with the requirements of scientific investigation and indicates the need, before beginning any investigation, of a plan for getting certain necessary information. The execution of this plan involves two major operational problems. The first of these, the problem of sampling, or, more generally, the specification of the people or institutions from which information will be obtained, is discussed in the third part of the book. Both the underlying principles of probability sampling and the

major steps in applying these principles to particular problems are discussed. Nonprobability sampling is also covered, and an attempt is made to indicate the rationale underlying this kind of sampling and to indicate the circumstances in which it is useful in marketing research.

The second operational problem involves getting information from the sources which are specified by the sampling plan. The fourth part of the book deals with that problem. Securing information involves either communication or observation. Criteria are suggested for judging the relative efficiency of the various forms of communication and observation in obtaining information on past behavior, intentions, attitudes, motivation, and knowledge and habits of thought.

In the concluding part, the administration of marketing research projects is discussed. The task of administration is the task of doing what is planned. Administration cannot be discussed in any comprehensive fashion, but maybe it is possible to indicate the administrative problems peculiar to marketing research and to suggest some principles useful in solving the problems.

This organization of the book has, we believe, certain advantages, but these advantages are achieved only at certain costs. The organization, for example, makes it difficult to use the book as a reference book. There is no single place in the book in which there is a discussion of questionnaire writing, for example. By reading all of Part IV the reader may come to understand what is known about people's memories and motivations and about difficulties of avoiding bias and securing interpretable information and may thus learn much of value in writing questionnaires. This discussion covers over 100 pages, however, and is not a convenient source of rules of thumb or other abbreviated and semimechanical guides to the actual process of writing a questionnaire. In a sense, this may be a deficiency in the organization. Such guides are sometimes useful and certainly facilitate quick reference. It is hoped and believed, however, that the organization of the book succeeds in usefully organizing the fields of knowledge basic to marketing research and thus provides a sound basis for the practice or use of marketing research.

Because of the somewhat restricted scope and rather general level of the discussion, some additional readings may be useful. There are many books on marketing research which differ markedly from this one, and these may be useful in providing a discussion from a different point of view. In addition, we have included at the end of the book a highly selective bibliography of books and articles which consider with greater thoroughness some of the theoretical issues which we have raised. These listings do not exhaust the good sources available on any subject. They include only those works which the authors have read and found useful.

CHAPTER 2

A BRIEF DESCRIPTION OF MARKETING RESEARCH

2.1 Introduction

A discussion of basic principles of research, such as we intend this book to be, is most meaningful for those who have had experience with the practice of marketing research. For those who are inexperienced, it seems desirable to provide a very brief description of marketing research. A vicarious and insubstantial experience is probably better than none.

2.2 Types of Marketing Research Activity

Many attempts have been made to list the various activities of marketing research personnel. A fairly comprehensive and not unusual kind of listing is reproduced below.

Research Relating to Product or Service

- New-product development (nonengineering)
- Improvement of present products
- New uses for old products
- Competitive position of company's products
- Packaging
- Pretesting of new products
- Customer preferences
- Product elimination or simplification of lines
- Relative distribution costs and profits of products
- Naming of products

Research Relating to Markets

- Analysis of consumer market (including industrial)
- Analysis of wholesale market
- Relative profitableness of markets
- Analysis and interpretation of general market data
- Estimating potential sales (general)
- Estimating demand for new products
- Market analysis by customer
- Market analysis by areas
- Establishment of sales territories
- Establishment of sales quotas
- Competitive conditions in markets

- General business forecasting
- Analysis of potential new market areas
- Analysis of potential of old market areas

Research Relating to Policy (not new-product policy but general policy)

- Price structure
- Sales methods or distribution policy
- Advertising policy
- Cancellations
- Returned goods
- Credit
- Discounts
- Public relations
- Dealer relations
- Transportation
- Sales compensation
- Wholesale and dealer margins
- Premiums
- Raw-material price changes
- Inventory

Research Relating to Sales Methods

- Advertising and selling practices in relation to competitors
- Distribution costs
- Sales methods or devices
- Choice of advertising media
- Selection of distributive channels
- Sales performance tests
- Sales training¹

The above listing illustrates well the diversity of marketing research activity but fails to indicate adequately the purpose common to these activities. All the activities in the list presented above have one thing in common; they represent implicit or explicit attempts to get information from or about customers or potential customers so that wise marketing decisions can be made. It is probably not a simplification to say that almost all the activities in this list represent attempts to estimate the willingness of people to purchase goods and services under rigorously specified conditions. Such a conception of marketing research embraces practically the entire list and gives it unity. For instance, research relating to the development of new products is research devoted to getting information from or about consumers so that their willingness to purchase proposed new products can be

¹ This listing is from a report by the National Industrial Conference Board, "Organization for Market Research," *Studies in Business Policy*, No. 12, Part I, "Industry Experience," p. 4, New York, 1945.

estimated. Similarly, research concerning improvements in present products aims at estimating the demand for products with various improvements. Moreover, it can be seen that research concerning new uses for old products seeks to estimate the demand for a product in various uses; research on the competitive position of a company's product seeks to find the relative demand for a product and its close substitutes. The same underlying purpose can be discerned in the rest of the activities above. This purpose is not so apparent, perhaps, in research devoted to the establishment of the sales quota, but the purpose is there. Sales quotas indicate the relative difficulty of achieving a specified volume of sales in different markets, and this difficulty can be indicated only by measuring the willingness and ability of people to buy in these various markets. Similarly, research on fluctuations in general business conditions represents an attempt to assess the impact of such fluctuations on the willingness of people to buy a particular product or type of product. Just as the ultimate purpose of marketing activity is the sale of goods or services, so marketing research ultimately attempts to appraise accurately the effects of different practices and policies on the volume of sales. As far as the sales executive is concerned, the determinants of sales volume are the willingness and ability of consumers to make purchases, and it is this willingness and ability which must be constantly assessed if marketing resources are to be efficiently used.

In slightly different terminology, all the foregoing types of marketing research represent attempts to estimate the magnitude of demand¹ and to understand the causes that determine it. There is another type of marketing research that is not directed toward the study of demand but rather toward the conditions that determine supply. This type of research is particularly important for two groups. First, agricultural producers are necessarily concerned about fluctuations in supply because such fluctuations affect importantly the prices of the goods that are sold and the incomes of the sellers. The people who sell wheat certainly do not control the supply of wheat in the same way that the manufacturer of a particular brand of cigarettes or toothpaste or razor blades controls the supply of these commodities. The supply of these latter commodities is less a matter of speculation since it is administratively determined. The supply of wheat, on the other hand, is the result of the decisions of and climatic influences affecting thousands of independent producers, and these factors must be studied if fluctuations in supply are to be forecast accurately. Second, fluctuations in supply are important for the processors and manu-

¹ "Demand" is the quantity of a specified product that a specified group of individuals is willing and able to buy under specified circumstances during a specified period of time.

facturers whose purchases of raw materials and parts importantly affect fluctuations in costs and profits. The meat packers and flour millers, for example, are vitally concerned with fluctuations in supplies of livestock and wheat. Fluctuations in the prices of these raw materials are one of the major factors in determining the profitability of these types of enterprise.

In this book we do not consider the problems of agricultural marketing research or other research focused on an explanation of the forces determining supply, although most of the methods of analysis which we do discuss would be appropriate for problems of this type. As a consequence of this arbitrary limitation in the scope of the book, we are left with the task of discussing the methods of measuring demand. Unlike many discussions of this subject in economic theory, a discussion of marketing research must consider things other than price. Marketing research is concerned with the effects of changes in advertising, packages, channels of distribution, methods of paying salesmen, competitors' practices, credit policy, etc., on demand—that is, on the rate at which people are willing to buy specified goods or services.

2.3 The Agencies Performing Marketing Research

Marketing research is any systematic attempt to get information useful in solving marketing problems (making marketing decisions). It is reasonable to believe, therefore, that most enterprises which sell goods and services have at some time engaged in marketing research. In most enterprises, especially the small ones, this research activity is probably not called research, may not be recognized as research, and almost certainly does not involve the efforts of personnel specialized in research. Many enterprises, however, particularly the larger ones, have explicitly recognized the continuing usefulness of marketing research and have employed specialized personnel to conduct it. Sometimes these personnel are employed continually and exclusively by the firm whose products are the subject of the research. At other times the research personnel are employed by specialized institutions (research companies, advertising agencies, etc.) whose services are purchased by businesses with marketing problems.

Marketing research is conducted in many types of enterprises, including manufacturers, department stores, research agencies, universities, governmental agencies, banks, etc. Almost all commercial marketing research, however, is performed in four types of organization: (1) the research departments of manufacturing and mercantile enterprises, (2) the independent marketing research agencies, (3) the research departments of advertising agencies, and (4) the research departments of publishing companies.

Quite a few large manufacturing and distributive enterprises have a marketing research department. The director of this department frequently

reports to a sales manager, to a vice president in charge of sales, or even to the president of the company or his assistant, though there are many other variations in the way such departments are integrated into the business organization. In all cases, however, it is the job of personnel in the "internal" research department to provide relevant information to the executives responsible for the formulation of marketing policy.

In recent years there has been a great increase in the number of firms having such specialized research departments and an equally great increase in the number of personnel in these departments and in their budgetary importance.¹ Since these increases have taken place during increasingly prosperous times and, for the most part, in a period of increasingly high corporate-income-tax rates, it is possible that the increased emphasis on marketing research is but another manifestation of the corporate "extravagance" which is to be expected during such periods. There is reason to believe, however, that marketing research has achieved enough successes—modest though they may be—to ensure its continued secular growth, even though this growth may be obscured from time to time by reversals.

Many of the large corporations in the United States today have marketing research departments, but it is not unusual for these corporations to employ outside research personnel as well. The most important sources of such outside expert aid are the marketing research agencies and the research departments of advertising agencies. The relationships between the company and outside research personnel vary according to the competence of each and their status in their respective organizations. In some cases, the company personnel serve as little more than a liaison between company executives and the outside personnel who design and execute research projects. On the other hand, some company research departments assume the major initiative in designing and interpreting research projects and delegate merely the technical, routine tasks to external agencies. Of course, many intermediate variations between these extremes of relative responsibility are found in practice.

The independent marketing research agencies, the second major type of research organization, are probably the most important single source of external research. These organizations vary widely in their organization and functions. Many of these agencies are equipped to perform all types of research operations, but some are specialized according to the technical skills possessed. For example, some agencies specialize in psychological measurements; others in performing the fieldwork which is involved in many consumer surveys; others in tabulation; etc. Other agencies are specialized not according to skills possessed but according to the type of research performed. Some agencies specialize—though not completely—in

¹ During the period 1948-1950, however, the rate of growth greatly diminished.

measuring the size of radio audiences; some, the readership of advertisements; others, the reactions of movie audiences; others, the movement of branded drugs and groceries through retail outlets, etc. Marketing research agencies receive their revenue in the form of fees for services performed. These agencies are numerous and widely scattered, although heavy concentrations are, of course, found in the largest cities.

The advertising agencies are the third important type of organization doing marketing research and the second important source of outside research personnel. Advertising agencies originally made their profit by buying advertising space "wholesale" and selling it "retail" to advertisers. Gradually their function and orientation shifted, and these agencies came primarily to represent advertisers in purchasing advertising space and time. Technically the major source of agency remuneration is still the same—that is, revenue arises from selling space and time to advertisers, which the agencies are able to buy at lower rates than the advertisers could. It would be less misleading, however, to think of agencies as being paid by clients for performing services and for offering expert advice. The agencies perform many services, their number and importance having increased through the years. Agencies create copy, do art work, plan advertising campaigns, select media, buy space and time, etc. As in many American industries, competition has not taken the form of price competition but rather competition in offering more and better services. Almost inevitably one of these services—and one of the most important—came to be providing research which would enable advertisers to expend their advertising dollars more wisely. Marketing research departments of advertising agencies now provide information regarding the characteristics of different segments of the populations which might be customers of the advertiser, information concerning the characteristics of the audiences of various media, information concerning the effects of various advertising expenditures, etc. At first, the research activities of these agencies were narrowly restricted to advertising research. Gradually, however, these agencies began to undertake research not directly connected with advertising, and now it is not uncommon to find advertising agencies doing research on product development, packaging, sales promotion, etc. Clients receive much research without charge, the agency compensation coming from the sale of space and time for advertising. Extraordinary research for clients involves additional remuneration.¹

The fourth important type of organization doing marketing research is the publishing company. Most mass media—newspapers, radio stations, magazines, etc.—employ research personnel whose job it is to provide

¹ The definition of "extraordinary" varies from agency to agency and from advertiser to advertiser.

information for the owners of the media and for advertisers in the media concerning the nature of the audience that is reached, the reactions of this audience to advertisements and editorial material, and the purchases and attitudes of the audience. Some large metropolitan newspapers do continuing research on the characteristics of the metropolitan audience: The *Chicago Tribune*, for example, has a continuing consumer panel which reports at frequent intervals expenditures for various designated types of commodities. Many magazines conduct periodic surveys concerning the activities, attributes, and attitudes of their audiences.

There are many other organizations performing marketing research. Marketing personnel in universities do research on their own, and also frequently act as consultants to business enterprises. General management-consultant firms are more and more coming to employ specialized marketing research personnel. Banks, insurance companies, and other institutions, too, frequently engage in marketing research although they typically do not sell such research. Their findings, however, are frequently made public. Finally, a large number of governmental agencies do work which is closely allied to marketing research and which is frequently useful to marketing researchers.

2.4 The Methods of Marketing Research

In order to provide some broad perspective on the methods of research before we get involved in basic methodological problems, we shall now indicate in a very general way the different kinds of basic research data: (1) historical data, (2) survey data, and (3) experimental data.¹ We shall discuss these in the order named.

The routine operations of the business enterprise frequently are recorded, and these records, often extending back through the years, provide a valuable source of information. Sales data, classified by territory, product, class of customer, salesman, etc., are frequently available from routine records without the initiation of research procedures. Also, it is possible to get from other sources a great deal of information concerning the characteristics of the various groups which constitute the consumers and potential consumers of any product. The government makes available through the publications of the U.S. Bureau of the Census and the Department of Commerce much demographic and economic data of this sort. Similarly

¹ This listing of types of data is not perfectly logical: e.g., surveys may provide data essential to an experiment. What the listing is meant to convey is that some data (historical) are available from existing records, some data (survey) are collected by an extraordinary effort without deliberate alteration of the environment which determines the data, and some data (experimental) are collected with some extraordinary effort and with the deliberate alteration of the environment that determines the data.

many local chambers of commerce and many state governments provide data concerning localized segments of the population. There are many other sources of such data. And often, when data are not available concerning the income, education, and other attributes of importance for marketing, indices which are highly correlated with these attributes can be had. Data on income-tax returns, magazine circulation, telephone ownership, automobile registration are widely available and are frequently highly correlated with the data of more direct interest in marketing research.

The statistical analysis, elaborate or simple, of the sales data of the individual enterprise and of the relationships of these data to "external" variables such as income, population, etc., is known as "sales analysis." Such analysis is probably more widely used than any other method of research as a basis for marketing decisions. And this is not surprising. Such data are usually available without any large expenditure of energy or money and can be made to yield, with relatively elementary statistical procedures, suggestive hypotheses concerning the consequences of alternative marketing decisions and can also be made to provide some evidence concerning the validity of these hypotheses. For example, classification of sales and sales effort by type of customer might indicate that a relatively large percentage of sales effort is being expended on types of customers who produce a very small percentage of total sales. Such information possibly could lead to a more efficient allocation of sales effort by type of customer. Or, a consideration of sales by sales districts and of the purchasing power and consumption habits of the people in these districts might suggest that the sales effort in some districts was substantially less productive or less efficient than in other districts. This information might lead to an investigation of differences in sales practices and a subsequent improvement in sales in those districts which had produced below expectations. Quite obviously the number of relationships which can be studied is great.

The second major source of information is not the data produced in a routine manner by the operation of the business enterprise or provided by governmental or other agencies, but rather data which can be secured only by an expenditure of resources for research. One of the major devices for obtaining information is the survey. The survey involves an effort to reach a "representative" group of the population under study (e.g., consumers or dealers) in order to extract directly from them particular information of interest. Surveys that are designed to get information from segments of the voting population on voting intentions have become familiar in recent years. Marketing surveys designed to get information on buying intentions, or behavior, or motivation, or on other subjects are also common though less widely publicized.

These surveys have two major variants with many subdivisions in each.

The first variant involves establishing "one-time" communication with a segment of the population in order to secure information. Communication can be established in various ways. Questionnaires can be sent out through the mail. People can be contacted over the telephone, as has been done in getting the data for the Hooperatings of radio programs. Or, perhaps more commonly, communication is established through personal contact between fieldworkers and the segment of the population from which data is sought. These different techniques of establishing communication involve problems which will be discussed in subsequent sections of the book.

The "one-time" survey contrasts with a type which involves continuing periodic communication between researcher and various individuals in the population under study. The second basic type of survey involves the use of a "panel," or stable group of individuals from the relevant population. The elements of the panel are sometimes families, sometimes individuals, and sometimes institutions, such as retail grocery stores. Originally, the research organization must explain the purpose of the survey and procedures by which it will operate. From then on, respondents use standardized forms to report on the particular activities being investigated or are the subject of routine periodic investigation by fieldworkers. The panel technique is widely used to measure the movement of branded articles from retail outlets into the hands of the consumer. Sometimes the panel is of individuals or families who periodically report their purchases of these branded items. In other cases the panel consists of retail stores which are periodically checked in order to determine the volume of their sales of the specified items. The relative advantages and limitations of the panel technique as compared with the "one-time" survey are discussed in subsequent chapters.

The third major type of data which can be used as evidence of the probable effects of alternative marketing decisions are data arising from experiments. The word "experiment" has a specialized meaning when used in the context of a discussion on scientific research. An experiment is a process created or conditioned by an investigator so that observations can be made which will yield unambiguous conclusions concerning the relationships between the observed variables. For example, different methods of shelf display may be instituted in a number of self-service retail stores in order to observe the sales resulting from each method. Experiments of this kind are frequently called "sales tests." The motivation underlying the experiment is the desire to create data that lead to unambiguous interpretation. Survey data or historical data almost always contain elements of ambiguity because of the difficulty of disentangling the desired relationships from the observed data. Properly designed experiments are not subject to this difficulty. Experiments are, in many situations, the most powerful tool of

investigation that the researcher possesses. There are peculiar difficulties involved in using experimental techniques to study the social behavior of human beings, and there are costs involved which frequently are not present in research based on other techniques. The nature of these peculiar difficulties and the circumstances in which the relative costliness of experiments seem justified are discussed in subsequent chapters.

These three basic types of data are secured by many specialized techniques, some of which have acquired brand names and can be purchased through commercial agencies. It probably would only be confusing and it is almost certainly superfluous to list in this brief introductory chapter the names of these specialized techniques and their characteristics. In the illustrations which are used to clarify the principles of research some of these branded techniques appear and are explained.

2.5 Business Environment of Marketing Research

From what has already been said about the growing importance of marketing research and the variety and number of research organizations, a reader might get the impression that marketing research has been long and well established in the business community as a means to enlightened marketing decisions. And, in the subsequent discussion of the principles of scientific marketing research, one's impression that such research is widely encouraged and heeded would undoubtedly be reinforced. As a matter of fact the recognition of the usefulness of marketing research has been fairly recent. Further, this research is frequently regarded as a marginal activity whose usefulness is often not easily demonstrable and whose expense can easily become unbearable under the pressure of adverse business conditions. It has been estimated that in recent years only about 5 per cent of the total research funds expended by industry are for marketing research. And even this small amount of research is frequently conducted under unfavorable conditions in which answers to impossible questions are demanded on short notice and with negligible expenditure.

This book is, in one sense, restricted in scope. It gives relatively little emphasis to the very important problems of establishing favorable relationships between the research organization and the rest of the business enterprise. Little attention is given to the vexing and often crucially important problems of "selling" management on the wisdom of research expenditures and on the validity of research results. Only in the last section of the book, and in particular in Chapter 18, do we treat these questions at all. We frequently discuss "ideal" solutions to basic research problems even though it is recognized that such ideals are frequently unattainable. The reason for this neglect of problems created by the business environment is not a deprecation of their importance, but rather a determination to brush away

all distractions from the important and difficult tasks of recognizing where research is needed, in deciding what evidence is relevant to the solution of the recognized problem, in marshaling the available evidence, and in interpreting it. Before going on to this "ivory tower" discussion of research, it might be well to present a brief description—as a sort of backdrop for the analytical discussion—of the business environment as the marketing researcher might wish it to be and as it actually is.

The marketing researcher helps marketing executives by providing information relevant to the intelligent choice among available alternatives of marketing practices or policies. An ideal environment from the researcher's point of view would be one in which (1) line executives clearly understood the value of research and its purposes, (2) the information needed to make a particular decision was precisely specified, and (3) the resources and time necessary for the acquisition of this information were made available without argument. Winston Churchill in his writings has indicated that he, as a minister in the British government, was an executive who attempted to create such an environment. The following memorandum¹ indicates his appreciation of the value and purpose of research:

Surely the account you give of all these various disconnected Statistical Branches constitutes the case for a central body which should grip together all Admiralty statistics, and present them to me in a form increasingly simplified and graphic.

I want to know at the end of each week everything we have got, all the people we are employing, the progress of all vessels, works of construction, the progress of all munitions affecting us, the state of our merchant tonnage, together with losses, and numbers of every branch of the R. N. and R. M. The whole should be presented in a small book such as was kept for me by Sir Walter Layton when he was my statistical officer at the Ministry of Munitions in 1917 and 1918. Every week I had this book, which showed the past and the weekly progress, and also drew attention to what was lagging. In an hour or two I was able to cover the whole ground, as I knew exactly what to look for and where.

How do you propose this want of mine should be met?

In Churchill's other writings about the war he has shown with amazing consistency a far-sighted recognition of the need for information of various kinds. He provided the means for securing such information and usually weighed it exhaustively and objectively in reaching his conclusion. War, with its relatively frequent surprises and calls for quick decisions and urgent action, often does not permit the amassing of much relevant evidence and its prolonged analysis. Business, too, frequently and probably necessarily involves decisions based on something other than a rigorous analysis of all available objective evidence. Someone once said that businessmen

¹ Winston Churchill, *The Gathering Storm*, p. 730, Houghton Mifflin Company, Boston, 1948.

do not have time to think. The person who made this statement was probably an academician who intended to imply that businessmen do not indulge often in the time-consuming process of exhaustively, explicitly, and unemotionally weighing evidence in order to reach conclusions. Such procedures are appropriately deemed "academic," because it is seldom that any but academic institutions can, in ordinary times, provide the time and opportunity for such formalization of the problem-finding and problem-solving processes. The businessman under the pressure of the necessity for taking action can seldom do more than quickly and intuitively appraise some of the relevant evidence. Further, frequently this appraisal is colored by emotional factors which a competent researcher would not consider relevant to the problem being decided. In spite of the very prevalent need in business for nonscientific procedures in reaching decisions, there are probably many instances in which scientific procedures are feasible and offer better chances of reaching "correct" decisions than do the emotional and subjective methods which are so common. At least an understanding of scientific analysis and of its potentialities would reduce somewhat the areas of uncertainty and doubt.

The smooth amalgamation of scientific analysis and subjective methods for arriving at decisions is not easy. Most executives are unable to say, as did Winston Churchill, "Here is the information I want. How do you propose this want of mine should be met?" As a result the marketing researcher not only has to supply information, but he must play some role in deciding what information is to be obtained and, once the information has been obtained, in helping to decide the practical actions dictated or suggested by the evidence. As a result, very few marketing research specialists are engaged solely in research. Most become involved in questions of policy. The quality and character of marketing research depends therefore not only on the technical skills and analytical abilities of the specialists but on their ability to ensure that research activities are well conceived and wisely used. In practice there are wide variations in both the technical qualifications and the nonresearch skills of marketing research personnel, with a corresponding variation in the usefulness of marketing research and the esteem in which it is held by businessmen. In some companies the marketing research department will be completely unheeded by executives because of deficiencies in the executives or in the research personnel. Often research is used only to justify postponing decisions that should be made at once. Sometimes, unfortunately, "research" may be conducted solely to provide dishonest propaganda for advertising or public relations. In other companies, by contrast, the research department may be consulted

There is one characteristic of much marketing research that distinguishes it from many other types of research carried on in business and elsewhere. Much marketing research that is undertaken is performed in order to provide evidence quickly for decisions which must be made quickly. The consequent necessity for getting answers in a hurry often causes marketing research to differ substantially from the sort of research that would be possible under less stringent time requirements. As a result of this deviation of marketing research from pure research or applied research devoted to long-run problems, much marketing research seems unscientific and superficial. Usually these characteristics seem inevitable, however, when formal methods of securing and analyzing data are applied to the solution of problems which must be solved soon after they are recognized. The quality of marketing research could probably be improved if it were used routinely in the solution of problems and if research could be used not only to solve problems but also to identify them. In this way the interval between the recognition of problems and the time at which a solution must be found could be increased. Chapter 5 discusses the possibilities of actively using research to discover problems before they would otherwise be recognized. By using research for this end and by increasing the competence with which research is conducted, the deviations between actual marketing research and ideal marketing research can be lessened.

PART II

SCIENTIFIC METHOD IN MARKETING RESEARCH

CHAPTER 3

SCIENTIFIC METHOD

3.1 Introduction

Although business decisions frequently have to be made without recourse to scientific investigation, this book is written in the belief that scientific methods provide a foundation upon which useful procedures can be developed for marketing research. The general nature of scientific method must be understood before the application of scientific method in marketing research can be understood. This chapter therefore considers scientific method in very general terms in order to provide insight into its scope, potentialities, and limitations. The theory of scientific inquiry is broken down in order to show the separate processes involved, and the roles of such disciplines as logic, mathematics, and statistics are sketched briefly. The subject is too complex to enable anyone to become proficient either in scientific method or any of its associated disciplines by reading a chapter or a book. Despite this complexity, however, it is hoped that this chapter can help provide an understanding of the basic ideas of scientific method.

The great prestige of the word "scientific" seems to arise primarily from admiration of the results produced by science—such as television sets and atomic bombs—rather than from regard for the process, scientific method, which made these results possible. The labors of scientists are frequently regarded not only as scholarly but also as mysterious, remote, and incomprehensible. This attitude is undoubtedly fortified by the startling differences between the everyday world and the world revealed by the scientific investigations of physicists and chemists. The everyday world is composed of loosely organized, imprecisely defined sensations of colors, noises, odors, tastes, and tactile feelings. In the world of the scientist, abstract, precisely defined relationships often explain the workings of reality in a way that is alien to ordinary experience.

Scientific method, too, is to some extent foreign to everyday experience; it is for the most part logical and objective and frequently makes extensive use of mathematical reasoning and complicated experiments. Despite this, the basic principles of scientific research are from one viewpoint neither complicated nor mysterious. In fact, the embryo of scientific method can be found in the way in which we learn about the world in everyday life. There is something of scientific method in the process by which an infant

learns to speak and understand language. The maxims of common sense are the product of a process that resembles scientific method. There is, indeed, no sharp distinction between the methods by which knowledge is obtained in everyday living and the process by which scientific knowledge is obtained. Scientific inquiry is merely the high refinement and formalization of the elements of logic and objectivity which frequently mark informal analysis and solution of everyday problems.

Both scientific and common-sense processes of analysis represent attempts to increase knowledge. The differences between them are differences in detail rather than in purpose or basic method. In the search for knowledge, scientists have developed techniques that increase the speed and efficiency of the investigative process, and it is for this reason that these techniques are worth studying. These techniques are not based on a set of commandments; they are an organized, flexible group of procedures which have been found useful in increasing knowledge of the world of the senses. The procedures vary greatly from science to science from investigation to investigation within each science. These variations are so great, in fact, that some writers (*e.g.*, James Conant)¹ have said (1) that general, elementary discussions of scientific method are necessarily foolish and (2) that scientific method, if it is to be understood at all by the nonscientist, can be understood only by studying the history of particular scientific investigations. We disagree with this position and shall indicate our reasons for disagreement by a general, elementary discussion of scientific method. We shall start with a familiar example of a simple investigation by a nonscientist.

A middle-aged man feels a sharp pain in his stomach after a bountiful meal. Somewhat alarmed, he quickly thinks of possible explanations for the discomfort: appendicitis, ulcer, cancer, simple indigestion, etc. Since he has never previously been troubled with any of the more serious ailments but has previously experienced discomfort after heavy eating, he tentatively concludes that the trouble is simple indigestion "caused" by overeating. He reasons that the pain will go away and will not recur so long as he is more careful about his diet. The pain does go away, and, so long as he practices greater moderation in his eating, it does not recur. This simple illustration brings out the main steps in scientific method. These steps can be listed under the four following headings:

1. Observation: the sensation of pain.
2. Formulation of an explanation for what was observed (formulation of hypothesis): the pain was due to overeating.

¹ James Bryant Conant, *On Understanding Science*, Yale University Press, New Haven, 1947).

3. Prediction of the future: the pain will soon go away and will not recur if overeating is avoided in the future.

4. Testing of the hypothesis: the pain does go away and, so long as overeating is avoided, does not recur.

Such an enumeration of the steps of scientific method is an oversimplification but perhaps a useful one. In broad outline, at least, this same pattern can be detected in most scientific investigations. Neither the scientist nor the man in our example would think explicitly in terms of a step-by-step sequence. In most examples, these steps would be blurred and overlapping and the entire sequence might be repeated several times. Always, as in the simple illustration given here, the application of scientific method builds on previous knowledge and experience and often provides new knowledge which can serve as a starting point for further investigation.

Interest in scientific investigation derives from the desire to understand the world and thereby predict the future. In other words, step 3, prediction, is of primary interest. In the example just given, there were really two predictions: first, that the pain would soon go away, and second, that it would not recur so long as overeating was avoided.

In order to forecast social behavior accurately it is helpful to be able to perform well all the steps of the scientific process, either formally (as the scientist does in scientific work) or informally (as the man did in reasoning about the gastric disturbance). As we pointed out earlier, scientists have evolved formal procedures for improving the quality of predictions. It is necessary to emphasize, however, that the formal methods are often not sharply different from the informal ones. The formal methods are too expensive and time-consuming to be used exclusively. Because of this the formal procedures of scientific method must be economized and reserved for special occasions in which (1) the issues at stake seem very important, and (2) the formal procedures seem markedly superior to informal ones. One of the most serious and perplexing problems confronting scientists and ordinary citizens alike is the decision between formal and informal methods. Both have special advantages and disadvantages. Frequently both are used in securing information relevant to a particular decision.

3.2 A Comparison of Scientific and Nonscientific Investigation

There are several important distinctions between scientific and nonscientific investigations, although in every case these distinctions are differences of degree rather than of kind. All these differences are, in most circumstances, important advantages for the formal procedures.

1. Scientific method uses more precise definitions of terms than do nonscientific methods. The use of numerical concepts—counting and measurement—may be regarded as the height of precision, but qualitative

concepts, precisely defined, also play an important role in scientific method. The psychoanalytical investigations of Freud, for example, are concerned primarily with qualitative concepts. The precise definition of terms permits easier communication among different scientists and assists greatly in the development of objective procedures for the evaluation of evidence. The ultimate of precision in definition is quantification. An extreme statement of the desirability of quantification was given by Lord Kelvin: "When you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind." This statement is, as we have said, extreme. Qualitative definitions are used fruitfully in scientific procedures as well as elsewhere. Nonetheless, numerical expression frequently permits not only more precise communication but also more precise evaluation of the significance of evidence.

2. Scientific method makes greater use of existing knowledge, both as a starting point for investigations and as a check on the results of the investigations. Frequently results obtained in one field of scientific investigation have unexpected applications to other fields far removed from the original one. The systemization of knowledge that is characteristic of scientific method increases the access of scientists to each other's ideas. This is strongly conducive to the advancement of knowledge.

3. Scientific method has evolved much more objective procedures for verifying hypotheses or evaluating evidence. Both scientific and nonscientific methods usually leave room for subjectivity in the evaluation of evidence, but there is less subjectivity in scientific method. This is perhaps the most important advantage of all for scientific procedures. This fact more than any other distinguishes scientific method from nonscientific. Ideally, at least, different scientists confronted with the same evidence should come to the same conclusion, regardless of their emotions or feelings. Of course it is not enough that procedures for the evaluation of evidence be objective; they must also frequently lead to the right conclusion. But an essential preliminary to the right conclusion in the evaluation of evidence is objectivity. A great deal of thought and experience has gone into the development of scientific procedures that meet these requirements.

Having indicated the most important characteristics of scientific method, we can now introduce more detail. The discussion is organized about the four stages characteristic of scientific investigation: observation, formulation of hypotheses, prediction, and testing of hypotheses.

In order to make the theoretical discussion more meaningful, we will introduce an example of an actual problem in marketing research. A manufacturer of paper containers (for milk, primarily) encountered difficulty

in selling a container of a certain design. The trouble arose at the consumer level originally, consumer complaints causing several local dairies to refuse to buy the containers. There was considerable evidence to indicate that a part of the consumer dissatisfaction was because of defects in the product which had later been corrected. For example, the container was originally too thin and too easily punctured, and the directions for opening the seal were inadequate. Moreover, production problems had arisen in filling and sealing the bottles. These defects had been remedied, but apparently a serious problem still remained. Some preliminary attempts at consumer interviewing had been made by the sales officers of the company together with representatives of the local dairies, and this had unearthed some usable information of the types of consumer complaints which still seemed to be important. It was at this stage that the problem was given to a marketing research agency for further exploration and recommendations as to the research, if any, which might be needed. A report of the progress of this research project is interwoven with the abstract discussion of scientific method in the following pages.

3.3 Observation

Observation is usually considered to be the first step in the scientific procedure because all speculation, or formulation of hypotheses, about the world is ultimately based on experience. Frequently hypotheses are based on experience which was not consciously sought for the purpose of scientific inquiry. In these cases it would be fair to say that experience had to precede the inquiry and was not itself part of the formal process of investigation. Often, however, investigators deliberately seek experience either directly through new observation or vicariously through reference to the recorded observations of others in order to get ideas about patterns of events and causal relationships.

It is hard to generalize about this searching process—either about how it is usually done or how it can most efficiently be done. Obviously the observer should in so far as possible free himself from preconceptions about what will be found in any exploratory process and should seek to record relevant phenomena objectively. This requires that the objectives of his search must be sufficiently clear that he can distinguish relevant from irrelevant phenomena. Obviously the process of observation should be thorough, objective, and accurate, and this is about all that can be said in general.

The recording of observations itself sometimes offers special problems. The most precise and objective record is usually a numerical one, and the attempt to make quantification possible frequently requires ingenuity and special skills. Examples of these difficulties are given in the discussion in

Part IV of the measurement of attitudes. When the record can be made numerical there are convenient means of summarizing and transmitting it precisely through the use of descriptive statistics. These means will not be discussed, since a familiarity with them is assumed (page 6). When experience is sought vicariously through the record of the experience of others, it is important that the accuracy of the record be known if the record is to be appraised and used safely and intelligently. The history of human inquiry is full of examples of wasted time and effort because of the uncritical acceptance of recorded observations.

The observational stage of the project involving the container manufacturer was very quickly carried through. It consisted chiefly of discussions between the analysts of the research firm and officers of the container manufacturer and gathering, in a rather unsystematic way, all the information which seemed easily available. A much more thorough exploration, of the type discussed in Chapter 18, would have been desirable. Since the manufacturer wanted only an opinion as to the kind of research that might be necessary and since this opinion was needed quickly, there was no possibility of carrying this investigation further. At the end, there was only a collection of rather disjointed and ill-connected facts, the most important of which have already been given. What was done with these facts will be seen as the development unfolds.

3.4 Formulation of Hypotheses

The formulation of hypotheses is not a purely formal process. Of all stages of scientific inquiry, this one most completely eludes description. Hypotheses (generalizations about reality that permit predictions) do not come to the mind of the scientist automatically or through the application of a fixed and learnable set of procedures. The scientist cannot observe a set of facts and see in them automatically or in some mechanical way an explanatory generalization. The perception of explanatory generalization is essential to scientific method, but such perceptions are not achieved solely by scientific method. When the apple hit Newton on the head—as legend has it—there was not any automatic process which led Newton to conceive of the law of motion. Other men of intelligence have been hit on the head by pieces of fruit without such gratifying by-products. Newton's imagination played on the experience with the apple and upon other experiences and achieved some generalization which seemed to provide an explanation of them. Newton's process of thinking cannot be reduced to rules which can be memorized and applied with equally spectacular results.

Critics of scientific method and in particular critics of quantitative scientific method have been quick to point out these facts. They have con-

tended that scientific method is inappropriate for the study of many questions in the social sciences (the criticism has not, so far as we know, been leveled against the natural sciences). The fact that formal scientific methods do not help much in one stage of the scientific process has nothing to do with their applicability for the other stages. It is contended further, however, that people who work too closely with formal scientific methods often lose touch with the imaginative processes needed to formulate important hypotheses. The danger is certainly a real one. In the social sciences there are elaborate scientific studies of trivial and unimportant hypotheses, and in marketing research it is easy to find examples of "scientific nose-counting" that is of little use to anyone. Yet these facts do not really condemn scientific method. For imagination in the invention of hypotheses without some objective and accurate test of their validity is as sterile as objective and accurate tests of unimaginative and unimportant hypotheses.

While the ability to formulate hypotheses must be relegated in part to the realm of "genius" and "inspiration", it is also true that extensive and precise observation and general scientific knowledge assist genius and inspiration. In so far as observation is helped by scientific method, and scientific knowledge comes from previous application of scientific method, it can be said that formal scientific procedures play a role in aiding the formulation of hypotheses.

In marketing research terminology, the formulation of hypotheses is frequently called the "definition of the problem." The problem arises, as in the milk-container example, as a collection of observations and facts which indicate that some difficulty exists. Before anything can be done in marketing research, it is necessary to define what the problem is or, in general scientific terminology, to formulate hypotheses which can be usefully tested. In the example of the milk-container manufacturer, four important hypotheses emerged. Each of these will be listed and discussed in succeeding paragraphs. Before the detailed discussion, however, it is interesting to note that the rather confused situation which originally confronted the marketing research firm was crystallized, and, rightly or wrongly, a small number of hypotheses believed to be of predominant importance emerged. The process by which this transformation was brought about, though less profound than the process in Newton's mind which related the falling apple to the laws of motion, can be described no better.

1. *The company's milk container provided materially improved sanitation.* The company felt that the basic advantage of its product was improved sanitation over conventional glass bottles and other types of paper containers. This hypothesis, however, had never been rigorously tested. It

was felt by the research agency that all the other issues involved in the problem would be viewed in quite a different light, according to whether or not this hypothesis was or was not supported by evidence.

There seemed also to be two subhypotheses involved. In the first place, there was the question of whether there was any technical difference in the sanitation provided by this container and that provided by ordinary containers. Secondly, if there was a difference in sanitation, according to some technical definition of the word, was the difference large enough to be of any practical importance? For example, did the use of bottles involve a danger comparable to that involved in drinking unpasteurized milk?

2. *The improvement in sanitation is important to the consumer.* Assuming that pasteurized milk in the company's container was actually safer than other pasteurized milk, the question immediately occurs of whether or not this fact makes any difference to consumers. Do users of pasteurized milk have reservations about its cleanliness aside from the fact that they might be expected to be "for" cleanliness in much the same way that they are "for" safety and yet drive cars dangerously?

3. *The functioning of the new milk container can be explained satisfactorily to the consumer.* The opening and closing of the milk container was somewhat tricky, and difficulty in accomplishing this had been at the root of almost all consumer complaints which had not already been met by revisions in the product. For the container to be accepted by consumers, it seemed absolutely necessary that consumers understand how to use it. This hypothesis therefore seemed a critical one.

4. *Consumers, once they understand how to use the container, will prefer it to more conventional ones.* Once these hypotheses were explored, it would be possible to take a clearly defined course of action with reasonable confidence that the decision concerning the containers was the best of the possible alternatives. These suggested hypotheses are not the only ones which might have arisen from the original data, but they seem satisfactorily to unify the original data so as to point toward practical solutions of the manufacturer's problem.

3.5 Prediction or Deduction

A discussion of certain basic ideas must precede the consideration of prediction. Scientific terms are joined together, for the purpose of communication, into sentences. These sentences, when they assert something about the nature of some particular objects or events or class of objects or events, are called "propositions." In particular, hypotheses are expressed as propositions. The validity of any particular proposition is determined by examining the objects or events to which the proposition refers and noting the extent to which the observation corresponds with the propo-

sition. The way this correspondence is judged is discussed at length later in this chapter under Testing Hypotheses or Evaluating Evidence.

First, however, it is necessary to examine the relationships between propositions. The discipline which is concerned with such relationships is logic. The characteristic nature of logic is best illustrated, perhaps, by the syllogism. A syllogism typically is a set of three propositions, the first two of which are based on evidence or speculation which is not contained within the syllogism. Propositions of this sort are called "premises." The third proposition is one which is derived from the premises. The process by which the conclusion—the third proposition—is derived from the premises is called "deduction." The syllogism typically is one of the following forms:

Premise I: If A is true, then B will occur.

Premise II: A is true.

Conclusion: B will occur.

The rules of logic indicate that one and only one conclusion is compatible with or implied by the premises. In the example given here the relationship is so simple that no rules seem necessary. More complex examples would indicate the need for some procedural rules or, at least, that such rules would be helpful. In economic theory, for example, it is possible to start from a small number of premises and deduce a very large number of particular conclusions. In mathematical logic a small number of premises (usually called "axioms" or "postulates") lead to vast systems of conclusions.

The premises contain as much information as the conclusion, or more, and deduction proceeds from the general to specific instances of the general case. If the human mind were capable of immediately grasping the implications of a set of assumptions, the study of logic would be unnecessary. But the human mind is not capable of such feats, as a moment's reflection on the above examples about economics and mathematics will show. Theoretical textbooks on science are deductive in nature: that is, they gradually unfold the logical implications of a small number of premises. These premises do not receive their justification in the textbook; their empirical validity rests upon the results of earlier research.

The connection between deductive logic and prediction must now be made more clear. As stated earlier, a hypothesis is one kind of proposition. The logical implications of hypotheses are, in fact, predictions as to what will happen. If the predictions are not fulfilled, it is concluded that the hypothesis was not correct. While this is somewhat oversimplified, it is best to postpone further complications until later in this chapter.

Reference to the example of the container manufacturer indicates that the relation between hypothesis and prediction is not so simple as the theoretical discussion indicated. Before any of the suggested hypotheses could

be tested through comparing observed events with the deduced consequences of the hypotheses, it would first be necessary to make several assumptions which could not be tested empirically. For example, the second of the hypotheses enumerated above, namely, "The improvement in sanitation is important to the consumer." Other hypotheses would have to be joined to this in order to make predictions which could be tested. For example, it might be assumed that *if* people attributed any illness to drinking of milk, any improvement in sanitation would be important to them. It is possible, logically, that sanitation could be important even if apparent lack of cleanliness had never been blamed for disease. If, however, a substantial number of people attributed illnesses to pasteurized milk, the hypothesis about sanitation can be regarded as verified at least partially. Many other supplementary hypotheses would be brought in to explore different facets of the original hypothesis.

3.6 Testing Hypotheses or Evaluating Evidence

"The so-called 'logical' method permitted the methodist to proceed from inadequate cognizance to a specious and useless conclusion; these methodists then took great pleasure in thinking that they had moved in a straight instead of a crooked line between these unfortunate states."¹ The discipline which enables us to proceed in a "straight line" from assumptions to conclusions is deductive logic. Logic alone, however, does not constitute scientific method.

The scientist is interested not only in exploring the implications of a set of assumptions but also in testing their validity, for the soundness of a conclusion is dependant both upon the logic of an argument and upon the validity of the assumptions from which it proceeds. Logic alone is powerless to determine the validity or invalidity of an assumption. For the scientist, the validity of an assumption or an assertion is determined by the extent of its agreement with empirical evidence. Empirical evidence is that which comes to us through the senses. The scientist thus must always rely upon the powers of human observation.

This means that the scientist is faced with a different problem from that of the logician. The logician must derive conclusions from generalizations by purely deductive means. The logician reaches his conclusions with absolute certainty.² The scientist, on the other hand, must attempt to appraise the validity of assumptions on the basis of a limited number of

¹ Ezra Pound, *Polite Essays*, p. 106, Faber and Faber, Ltd. (The Scientific Press), London, 1937.

² While technically correct, this statement may be misleading. The annals of science are filled with examples of incorrect deductions, and in complicated logical systems it may be exceedingly hard to detect errors. Even famous mathematical proofs have been found to contain errors.

facts, never proceeds solely with deductive analysis, and reaches conclusions which are held only tentatively. The reason for the lack of certainty is simple: any useful generalization, hypothesis, or assumption—one that states more than is already known—will imply that certain unobserved events or relationships are true in the “real” world. As more and more of these events and relationships are tested by observation, it is possible that the generalization will be found to require revocation or modification.

Once a generalization or hypothesis has been made, it will be necessary to test it before it can be tentatively accepted. “The” hypothesis, along with other hypotheses, may be thought of as a group of propositions which logically imply certain things about the sensible world. The scientist deduces that if his generalization is valid, events A, B, and C (as yet unobserved) must occur. Observation can determine whether A, B, and C actually do occur. It is necessary sometimes to make extremely elaborate observations and computations or to perform complicated experiments to find out whether A, B, and C do occur, but once their occurrence or non-occurrence has been determined, there is some evidence for accepting or rejecting the generalization being tested. The occurrence of A, B, and C does not necessarily mean that the generalization is valid, for there are always other generalizations with which the occurrence of A, B, and C would be consistent. And these latter generalizations may not be consistent with the one being tested. Nor is it necessarily true that A, B, and C actually occurred merely because observation seemed to indicate that they did. For many centuries men believed that light travels in a straight line. Subsequent more precise observations have indicated the fallacy of this belief. Once observations have been made, however, the results are used either to substantiate or to indicate the invalidity of the hypothesis being tested. After a hypothesis has withstood repeated tests of this kind it may be given the dignified title “scientific law.” The dignity and impressiveness of the title should not disguise the fact that such laws are continually being repealed and modified as new evidence comes in.

An example of this process may clarify it. During the latter part of the nineteenth century many factories prohibited smoking by production workers. Men who had become used to tobacco frequently began to chew it instead of smoke it under these circumstances. The men would smoke both before and after working hours and chew in between. Two kinds of tobacco were needed: plug tobacco for chewing and fine-cut tobacco for smoking. A tobacco manufacturer observing the situation reasoned that these workers would prefer a single tobacco adaptable for both chewing and smoking over the two varieties which were then being carried. Although the tobacco manufacturer probably didn't explicitly formulate a hypothesis, he could have. He could have made the hypothesis, “more than 20 (or

any other number) per cent of the factory workers in the United States would prefer his combination tobacco to the two varieties now being purchased, if the price per ounce were the same, etc." The manufacturer implicitly assumed that a certain percentage of workers would prefer tobacco that could be both smoked and chewed satisfactorily and that his tobacco had these attributes. These assumptions (together with the usual assumptions about economic behavior) logically implied the conclusion that a certain percentage of the workers would buy the manufacturer's tobacco. This conclusion could then be tested by marketing research, perhaps through the use of sales tests.

Statistical inference would enable the manufacturer to evaluate the information obtained from the sales test in order to decide whether the evidence was consistent with the hypothesis or whether the hypothesis was contradicted. It will be helpful to indicate briefly the nature of this process. Probability theory is the basis of statistical inference. Probability theory is formal just as is deductive logic, and the conclusions of probability theory follow just as certainly from a set of assumptions. A typical example of the form in which statements of probability are made follows:

Assumption: The probability of obtaining a head on a single toss of a coin is equal to $1/2$.

Assumption: The coin is tossed 2 times.

Conclusion: The probability of obtaining a head on both tosses is $1/4$; the probability of obtaining a tail on both tosses is $1/4$; the probability of obtaining one head and one tail is $1/2$.

If the two tosses were to result in two heads, the evidence would not seem inconsistent with the hypothesis that the probability of obtaining a head is $1/2$. For if the hypothesis were true, the probability of getting two heads in succession would be $1/4$. If, however, 10 heads appeared in succession, the evidence would seem strongly inconsistent with the hypothesis that the probability of a head is $1/2$. For if this hypothesis were so, the probability of obtaining 10 heads in succession would be only $1/1024$. Since the probability of 10 heads in succession is so small, one would be inclined to discard the hypothesis that the probability was only $1/2$. The hypothesis would not be discarded because the observed sequence of events would be impossible if the hypothesis were true, but because the events would be highly improbable. The theory of statistical inference—based on probability theory—enables one to evaluate the risk (probability) of discarding a hypothesis (such as the assertion that the probability of a head equals $1/2$) when it is really true. Similarly, it is possible to evaluate the risk of accepting a hypothesis when it is actually false. Statistical inference would thus enable the tobacco manufacturer to interpret his evidence (the results of the sales test) in such a way that he could evaluate the

probability of rejecting a hypothesis that was actually true or of accepting a hypothesis that was false.

The discussion of testing hypotheses has been, up to this point, a little unrealistic. In practice everyone does and must appraise the validity of hypotheses even when the evidence relevant for this appraisal is somewhat contradictory and not subject to numerical expression and statistical analysis. For example, in examining hypotheses about the structure and functioning of ancient civilizations, historians must examine much evidence and, "on balance," decide whether the tentatively held explanations can be accepted or must be rejected. Judges and juries in evaluating evidence in the courtroom must weigh evidence which does not all point the same way and must reach some conclusion about the guilt or innocence of the defendant. There are legal rules for evaluating evidence just as there are accepted procedures of historiography. There is no formal body of rules for weighing conflicting nonnumerical evidence in marketing research. The weighing is almost inevitably subjective, but the principles of procedure which apply to numerical evidence offer some guide to the interpretations of evidence which is not numerical.

The testing of the hypotheses referred to in the milk-container example may be worthy of some comment to indicate again the relationship between an abstract presentation of scientific method and the problems actually encountered in marketing research. Actually, the study referred to in this chapter (which has been somewhat disguised) met the fate of many marketing research studies: it was never carried beyond the planning stage. The plan, however, outlined the types of tests which might have been applied. It will be helpful to indicate, without going into the details of implementation, what was actually planned.

The first hypothesis, regarding the effect of types of milk containers on actual sanitation, involves some engineering rather than marketing research. The second hypothesis, regarding the importance of cleanliness to the consumer, was to have been tested by interviewing users of pasteurized milk on such questions as the following: Has the user ever had fears about the cleanliness of pasteurized milk? Has the user ever attributed any illness to the drinking of pasteurized milk? These and other questions would, undoubtedly, have been carefully phrased and tested if the study had gone beyond the blueprint stage. The third hypothesis, that the functioning of the container could actually be explained, was to have been tested experimentally. Different methods were to be tried—advertising, instructions on the container, word of mouth by the delivery man. The success of these methods was to be ascertained by interviewing a sample of consumers and actually going so far, in addition to the ordinary questions about attitudes toward the product, as to ask users to demonstrate how they opened

the container, closed it, stored it in the refrigerator, poured milk out of it, etc. The final hypothesis, that a sample of consumers would prefer the container to other containers once they understood its functioning, was to be determined by a preference test—in which, in effect, people would be asked whether they liked this container more than others with which they were familiar.

The description of the tests of the hypotheses in this example has been made deliberately sketchy because of our desire to concentrate on the main approach rather than to become involved too deeply in details of implementation. The questions of sampling, questioning, experimentation, etc., which are barely mentioned in the few brief sentences above, are the subject of the major part of this book. The function of the example has been to show the connection between abstract scientific method and an actual marketing research problem. It shows quite clearly that the general outline of scientific reasoning can be discerned even in an apparently chaotic practical problem, but that in practice the reasoning is more complex and the steps are blurred.

Rather than pursue this example further, then, it will be more helpful to consider a deliberately simplified example of another marketing research project. This example has been simplified to illustrate and summarize the salient steps in scientific method.

The vice president in charge of sales of a small food manufacturer was speculating one day on ways in which his company could increase its volume of sales and its profits. In thinking of his experience with his own company and of the experience of other companies he noted that many companies selling the same type of food as his company had achieved substantial increases in sales following their adoption of cellophane packages. Since the increases in sales seemed to follow the adoption of cellophane in many different cases, the vice president began to think that manufacturers of certain type food products could increase their sales by adopting cellophane packages. The vice president assumed that food of a certain type would be purchased at a greater rate in cellophane than in opaque packages and further assumed that his product was food of that certain type. These assumptions logically implied the proposition that the vice president's product would sell at a greater rate in cellophane packages. Rather than make a major change in package solely on the basis of such armchair speculation, the vice president decided to test his conclusion by getting some empirical evidence. To this end he hired a marketing research agency to design and conduct a sales test in which it would be possible to compare the volume of sales of his product in the old and in the new packages. This test showed that sales did increase with the adoption of cellophane; the vice president felt that his conclusion had been sound and adopted cellophane for all the packages of his company's product.

This example contains implicitly the various steps of the scientific process. The vice president, on the basis of his experience (observation), made a generalization (formulation of a hypothesis) which seemed to explain the phenomena which he had observed. In this instance the vice president formed the hypothesis that food of a specified type would be sold at a greater rate in cellophane packages. From this generalization he deduced (predicted) the particular conclusion that his company, manufacturing the specified type of food, could increase its sales by adopting cellophane. This deduction follows logically from the generalization. The validity of the deduction must then be tested by further observations (testing of hypothesis). The observations must be carefully planned so that they constitute convincing evidence regarding the validity of the conclusion. The final step involves an inference on the basis of the observation concerning the soundness of the hypothesis.

Suppose the vice president's sales test had failed to indicate that the adoption of the cellophane packages would result in increasing sales. What could account for the failure of the evidence to confirm the soundness of the conclusion? In the first place, the hypothesis could have been wrong. Maybe the vice president had mistakenly assumed that cellophane packages were the cause of the increased sales that he had observed in other companies. The association between cellophane packages and increased sales could have been fortuitous. Or perhaps the use of cellophane packages was necessary for an increase in sales but not in itself sufficient for an increase. In the second place, the vice president might have been mistaken in thinking that his company's product was of the same type or would encounter the same reactions as the products of the other companies which he had observed. In the third place the evidence that the vice president had obtained in order to test the validity of his conclusion may have been misinterpreted. That is, the vice president could have either accepted a false hypothesis or rejected a true one. These three possibilities should be examined in turn by the vice president in order to decide whether the change to cellophane packages would in fact be justified.

The example given is a simple one, but it contains the essential elements of scientific method and indicates how this method might be applied in marketing research. Undoubtedly in any actual example, as we have hinted in the earlier illustration, the investigator would encounter many difficulties and practical obstacles not included in this example, and these difficulties and obstacles would undoubtedly cause some deviation from the smooth and logical step-by-step process outlined above. Nevertheless, it is important to have such a process clearly in mind. Often fact gathering which proceeds without such a plan yields data which the competent investigator would have to judge worthless for appraising the validity of any relevant hypothesis.

CHAPTER 4

SCIENTIFIC METHOD IN MARKETING RESEARCH

4.1 Introduction

In Chapter 3 the discussion of scientific method was not specifically oriented toward marketing problems or marketing research. We were content to point out the general characteristics of scientific method and to comment on certain general procedures that have been found useful. In this chapter attention will be focused upon the application of scientific method in marketing research. First, however, it will be useful to reexamine the basic purpose of marketing research in order to see how the steps of most scientific investigations can, in general, be useful in achieving this purpose.

Marketing research is worth its cost to the extent that it reduces the risk of wrong decisions. All decisions are based implicitly or explicitly on forecasts or predictions of the consequences of the alternatives between which choices or decisions must be made. Marketing research is useful, therefore, to the extent that it facilitates such forecasts.

It is possible to distinguish two types of forecasts which are made in marketing research. First, there are forecasts of the general economic environment in which the individual firm will be operating, regardless of any action of the firm. Forecasts of general business conditions or of changes in the size or geographic distribution of populations are of this type. Second, marketing research is concerned with forecasting the reactions of this general economic environment to changes in the policies of the firm. Forecasts of the rates of sale of products under different advertising appropriations are examples of this second type.

While forecasts of both types are obviously necessary, marketing research has been focused primarily on forecasts of the second type. Marketing research, thus, is largely devoted to the forecasting of the consequences for individual firms of alternative marketing policies or practices. The consequences of most concern in marketing research are variations in sales, and most marketing research considers explicitly or implicitly the probable effects on sales of alternative marketing decisions.

The forecast of the effects on sales of alternative marketing policies is done in order to choose that alternative which will be associated with the largest sales (assuming equal costs and the desire to maximize sales). Such

a forecast stems from a hypothesis that the selected alternative will result in larger sales than the rejected alternatives. Research devoted to selecting a new product design, for example, can be thought of as research devoted to testing the hypothesis that design A will result in sales that are larger than those resulting from the adoption of the alternative designs. The selection of an advertising appeal can be thought of as involving the hypothesis that appeal A will result in larger sales than appeals B, C, or D. A decision to modify a system of paying salesmen involves the hypothesis that the modified system of remuneration will result in greater efficiency in the use of dollars for sales effort than did the old system.

Conceived in this way, marketing research is useful at two stages. In the first place, marketing research can be useful in suggesting hypotheses. Many marketing research projects are exploratory in the sense that they are undertaken in order to provide descriptive or informative data without any explicit prior formulation of a hypothesis. An advertising agency, for example, might undertake the systematic investigation of the reactions of the audience of a television program. This investigation could provide descriptive data of great value to the agency or to the executives of the company sponsoring the program. The data could be valuable in suggesting desirable changes in the program. Ideas might be suggested which were new to the agency and to the company, and these ideas could suggest the hypothesis that the program would be more effective if modified in certain designated ways. Many companies systematically investigate the ways in which their product is used and the reactions of the users in order to get ideas for changes in the product or in promotion policy. The suggested changes can be thought of as hypotheses that the change will result in greater benefit to the company than the continuation of the old policies.

In the second place, marketing research is useful not only in suggesting hypotheses but also in testing their validity. Of course many changes (probably most) in marketing policies are made without any test of the validity of the hypothesis upon which the change is implicitly or explicitly based. The use of research offers some special advantages in testing hypotheses, however, which deserve our attention. In the absence of any research prior to or after the adoption of the change, the major evidence regarding the wisdom of the change or the validity of the hypothesis upon which it was based is the success of the company after the change as compared with its success prior to the change or with the success of other similar companies which have not made a change of the same sort. There are two major deficiencies in this sort of evidence.

First, this type of evidence is frequently extremely expensive to obtain. Although there are no expenditures devoted to research in getting this sort of evidence, there are frequently the relatively great expenses involved in

making a major or minor change in policy which turns out to be unwise and which therefore requires revocation or modification. This point is obvious: wrong decisions are expensive.

Second, this type of evidence is unsatisfactory because it is almost inevitably ambiguous. Consider, for example, an executive who had to appraise the wisdom of an adopted change in the package of his product on the basis of changes in the rate of sale after the change in package. If sales increased markedly after the change, the executive might be inclined to think that the change had been wise and that the prediction (hypothesis) underlying the change had been correct. The hypothesis just referred to is certainly consistent with the evidence, but it is not the only hypothesis which is. General business conditions may have improved at about the same time that the change in package took place, and it would be as plausible to attribute the increase in sales to the improvement in general business as to the change in package. Perhaps advertising expenditures also increased at the same time (perhaps in order to promote the new package). This increased advertising might have been the cause of the increased sales. Or the new package might have been adopted at a time of year when sales usually showed a seasonal increase. By considering evidence other than the change in sales alone, it will usually be possible to indicate that some of these hypotheses are more plausible than others. By considering the performance of competing companies which had not made the same sort of change, some of the possible hypotheses can be eliminated with reasonable confidence. Even under the best circumstances, however, it will be impossible to have great confidence that the complex network of causation has been sufficiently untangled to enable the investigator to impute to the change in package its correct amount of influence on sales. And it is almost always impossible to get any satisfactory evidence that the accepted change in package had a better effect on sales than other proposed changes which were not adopted.

By contrast, marketing research can often provide evidence which is less expensive and ambiguous. Even though research involves expenditures which would have been unnecessary if the result of the research indicates the desirability of the same course of action favored prior to the research, research frequently can forestall unwise changes in policy or can indicate the need for changes that would not otherwise have been made. In the example given above of the change in package, it would have been possible to conduct sales tests of proposed packages in order to test on a small scale the effects of the different package designs on sales. Or it would have been possible to get consumers to state their preferences concerning the proposed packages. These types of research could have provided evidence concerning the relative success of the packages at substantially less cost than would

be involved in the adoption of a package having a significantly adverse effect on sales. Also, this evidence, if properly acquired, would be much less ambiguous than the evidence consisting of changes in sales after the actual adoption of the new package. The sales test, if properly conducted, would enable the investigator to isolate the effect of the package from the effects of other factors such as seasonal fluctuations, changes in advertising, etc.¹ Or, if the statements of preference had been secured, it would have been possible to interpret precisely their significance, even though the relation between stated preference and behavior is not always predictable.

While research can often provide relatively unambiguous evidence, there are some disadvantages connected with research. The main disadvantage is that the hypotheses which can be tested by research are more limited than the hypotheses of ultimate interest. A sales test, for example, may evaluate the sales response to a new package design under a given set of conditions. But the sales test is by its very nature likely to be a relatively small-scale study, and the given set of conditions cannot typically include full national promotion nor can it reflect the reactions of competitors which might be important if the change in package design was made on a national scale rather than in a limited number of local markets. Similarly, it is possible for people to act in ways that would not have been predicted on the basis of their stated preferences. Because the hypotheses which can be tested by research are inevitably more limited than the hypotheses of ultimate interest, it is almost always true that there is room for the use of subjective evaluation in conjunction with the use of the objective evidence which research provides. In other words, business executives are almost always left with the task of considering the relationship between the hypothesis which has been tested (*e.g.*, the hypothesis that consumers will, upon inspection, indicate a significantly greater preference for package A than for package B) and the hypothesis that would have been tested if such a test were possible (*i.e.*, the hypothesis that consumers will buy more of the specified product in package A than in package B). Therefore scientific research, though useful, can almost never be used alone in reaching business decisions. Marketing research can be viewed as a means for reducing the risk of making wrong decisions. The risk can never be eliminated entirely. The cost of reducing this risk is the cost of performing the research.

Thus formal scientific method can make two distinct contributions to the solution of marketing problems: it is helpful in suggesting alternative courses of action and it is useful in evaluating these alternatives. Sometimes, of course, this distinction becomes blurred. In practice the neat

¹ For a discussion of the principles of experimentation which underlie properly devised sales tests, see section 4.2 in this chapter.

compartmentalization suggested by the enumeration of steps in scientific method is not so clear-cut as a literary discussion might suggest. Sometimes, for example, hypotheses may not be explicitly formulated, yet the information collected by research is used as a test of an implicit hypothesis. For example, a sales manager might have subconsciously in mind some norm of sales performance which can be expected from different territories or different individual salesmen. This norm is really a hypothesis, although the sales manager would almost certainly not think of it in this way. Yet when he looks at evidence, such as the routine reports on sales by territories, he is really looking for deviations from the norm implied by the hypothesis. If there are no such deviations, he accepts his hypothesis and does not look further. If there are important deviations, he is forced to reject his hypothesis of satisfactory performance and to look for explanations for the deviations and for possible action that may correct them. This process of comparing performance with some subconsciously held concept of what is normal or acceptable may be so automatic that it is not recognized as involving the use of objective information to test some hypothesis. In the remainder of this chapter we consider in greater detail the problems involved in systematically and objectively testing hypotheses. First, we make some general remarks which serve as background for this discussion.

Hypotheses are tested by comparing what happens with what would be expected to happen if the hypotheses were actually true. It is therefore essential that there be some evidence as to what actually happens. When this evidence is in numerical form, it is called "data," and it can be objectively analyzed and interpreted. Other forms of evidence—nonnumerical evidence—are also useful. The interpretation, however, is largely subjective and for that reason is almost impossible to discuss. Data arise in two ways.¹

First, data arise from the recording of events as they are observed. The accounting data of most enterprises are of this sort, as are the data which result from surveys. The data are recorded in order to make known what has happened; the events that are recorded are neither ordered nor controlled in order to provide data that would be useful in reaching decisions or testing hypotheses. Such data are properly called "observational" since they represent the recording of what has happened in the past without any definite intervention by the person doing the recording in order to influence the events being recorded.

¹ In Chapter 2, which was descriptive, we distinguished three kinds of data: experimental, survey, and historical. For our present purposes, which are analytical, we distinguish only two kinds; experimental and observational. "Observational data" include data previously described as "survey" and "historical."

The second type of data—experimental data—also results from the observation of events, but of events which have been ordered or controlled in order to get data that will permit the unambiguous testing of specific hypotheses. This second type of data results from experiments, which can be thought of as deliberately designed processes for the purpose of creating data which can be meaningfully interpreted. Both types of data are useful in testing hypotheses, but the peculiar limitations and dangers in using each must be thoroughly understood.

4.2 Experimentation

Experimentation can be described in a number of ways. We have said that experiments are processes deliberately designed or arranged in order to create meaningful data. Experiments may also be thought of as processes contrived by an investigator in order to make the conditions under which data are created conform to the assumptions underlying the use of available analytical tools so that meaningful tests of hypotheses can be made. Thus, the investigator may design a sales test so that the assumptions underlying the statistical theory which must be used to analyze the resulting data are met. Such a sales test would permit a more unambiguous statement regarding the effect of the tested stimulus than would the recording of sales data which resulted from unplanned comparisons of rates of sales.

Experiments have become the major means of verifying hypotheses in the natural sciences. In these sciences experiments have been widely used, and spectacular advances in knowledge have resulted. Some of the natural sciences, astronomy, for example, are much less amenable to experimentation than others, physiology, for example. Many people have long believed and stated that the social sciences are less accessible to investigation by experimentation than are any of the natural sciences. Certainly experiments have not been so widely nor so successfully used in the social sciences as they have been in the natural sciences. Since marketing research is concerned with social behavior (namely, the expenditure of money for goods and services and the forces that affect such expenditure) and is therefore an applied social science, the reasons for the relatively restricted and unproductive use of experiments in the social sciences deserve scrutiny here.¹

There seem to be four main reasons for the slight use of experiments in the social sciences. First, the willing participation of human beings frequently must be obtained, and human beings have almost always shown a

¹ The following discussion of experimentation follows rather closely a discussion of the same subject in an article by the authors, "Some Comments on Experimentation in Business Research," *Journal of Business*, Vol. XXIII, No. 2, pp. 94-103, April, 1950.

great reluctance to modify or reveal their behavior in the interest of increasing knowledge by experimentation. Many investigators, because of this understandable unwillingness, have had to use other than experimental means of investigation.

The second reason for the restricted use of experiments in investigations of social phenomena is that the human beings under investigation, when willing to participate or to be observed, are frequently affected by their participation or accessibility to observation so as to make it difficult to generalize on the basis of the experimental results. Perhaps the best-known example of the affective power of experimentation per se is the work done under the direction of Elton Mayo at the Hawthorne plant of the Western Electric Company. In his work, Mayo tried to experiment in order to discover the relationship between certain variations in working conditions and in productivity. It was soon discovered that productivity increased rather steadily throughout changes of all sorts in working conditions. When lighting was increased, productivity rose; when lighting was decreased, productivity rose, etc. One of the hypotheses that was advanced to explain this curious result was that the workers participating in the experiment increased their productivity because they knew that they were participating in an experiment. Even if this hypothesis is not completely satisfactory, the implication for social experimentation is obvious—when ever the persons participating in an experiment know of their participation, it is difficult to know the extent to which the results observed are a consequence of this knowledge.

The third reason for the relatively restricted use of experimentation in the social sciences is the instability of many social relationships. Although the tensile strength of aluminum or steel of specified composition is relatively constant through changing political administrations, fluctuations in business activity, wars, famines, etc., the same degree of constancy is not observable in the tastes and reactions of human beings in social situations. A consequence of this difference in temporal stability is that the results of experiments on the tensile strength of steel or aluminum can be generalized through time with great confidence, while generalizations through time of experimental data on social behavior must always be cautious and tentative. The advertisement which was shown to be relatively most effective last year might be least effective this year or next. Similarly, it is well known that different styles of clothes, cars, furniture, etc., are accepted to different degrees in different years as fashions change. The relatively great temporal instability of social relationships reduces the value of social experiments and discourages their use.

The fourth reason for the relatively infrequent use of experimentation in social research is the failure of many social scientists (and marketing re-

searchers) to learn or understand thoroughly some of the principles of experimentation that have made it such a powerful investigative tool in the natural sciences. This apparent failure has caused these investigators to turn to other investigative techniques when experimentation was both feasible and probably more informative. One of the most prevalent failures, if one is to judge both from observation and published accounts of research, is the failure to apply the principle of "statistical control," or randomization. Evidence of this failure is easily found in the journals of social sciences and marketing, where too often there appear statements to the effect that experimentation was very costly or extremely difficult or impossible because of the need to hold all variables constant except the variable whose effect is being studied. Such control is, of course, almost never possible when social behavior is involved, but fortunately such control is completely unnecessary for experimentation to provide unambiguous tests of hypotheses. In short, it is fallacious to contend that all variables except one must be held constant.

The four factors accounting for the slight and inefficient use of experiments in social research affect the different social sciences in their varied applications in different ways. The sociologist studying various aspects of family life can almost always be confident of meeting overwhelming obstacles to experimentation. Likewise the political scientist faces impressive (though not insuperable) difficulties in studying voting behavior by experimental means. The businessman interested in investigating that form of social behavior which involves cooperative production effort or that form which involves exchanging money for goods and services is less handicapped. It is possible to study experimentally the effect of most alternative business policies on costs and on the willingness of people to expend money.

The reasons for the relative accessibility to experimentation of studies of cost and of demand are easy to see. The businessman, by controlling the stimulus whose effect is being studied, often can avoid the necessity of securing the cooperation of the persons under study and often need not reveal to the participants that they are part of an experiment. The participants in a well-designed sales test have not consented to participate nor do they know that they are participating. The same is true of the readers of a magazine who are the unwitting participants in experiments based on the split-run technique or other techniques.¹ Many other types of market-

¹ The split-run technique involves placing an advertisement in one form in one-half the copies of a given issue of a magazine and in another form in the other half of the copies in order to compare the rate of coupon returns from the two forms of the advertisement. This comparison in turn should provide some basis for judging the relative effectiveness of the two forms of the advertisement.

ing investigations can be conducted experimentally without the knowledge of the persons whose actions provide the experimental data, and thus the first two obstacles to social experimentation—reluctance to participate and modification of behavior when participation is achieved—apply with substantially less force and pervasiveness to business research than to many other types of social investigations. The third obstacle to social experimentation—the temporal instability of social relationships—applies with approximately equal force to business experiments and to other social experiments. This fact will undoubtedly continue to restrict business experiments, but it should be understood that this temporal instability affects equally types of social investigation other than experimentation. Any reasoned attempt to forecast the future depends to some extent upon the continuation of patterns and relationships which have been noted in the past. The fact of instability must be taken into account in any type of investigation; about all that can be done is to estimate the importance and probable degree of instability in each attempted forecast. The fourth obstacle to social experimentation is less troublesome than the third in that it can be diminished or eliminated by application of the basic principles of experimentation. The remainder of this section on experimentation will attempt to indicate what these principles are, as well as to indicate some of the guides to efficient experimentation.

The marketing researcher can often arrange for the application of the stimulus whose effect is being studied to identifiable groups of human beings and for withholding its application from other identifiable groups. For experiments to take place it is necessary that this possibility exist. Such differentiation of stimuli among identifiable groups is possible, for example, whenever advertisements can be released in clearly defined localities or through media reaching known groups, or whenever products of different designs can be made available in different stores, or whenever different packages are made available in different and nonoverlapping markets. The significance of this possibility of control of the stimulus being studied is great. When this possibility exists, it is possible, by choosing the groups to which the stimulus will be applied and from which it will be withheld in a particular way (which will be described shortly), to reach unambiguous conclusions about the effect of the stimulus. It has often been said that unambiguous interpretation is impossible because of the impossibility of controlling—or more absurd, holding constant—the many variables other than the stimulus under study which affect the variable (usually sales) whose variation is being studied. It has been argued, for example, that it is impossible to reach unambiguous conclusions about the effect of an advertisement on sales (even though the advertisement can be restricted to identifiable groups of persons) because of the impossibility of controlling

other affecting variables such as competitors' advertising, changes in the prices of competing products, changes in the weather, and changes in the many other things that are believed to affect sales. The changes in these other affecting factors, it has been alleged, make it impossible to tell whether the differences in the rates of sale of the advertised product to the group exposed to the advertisement and to the group not so exposed should be imputed to the difference in advertising exposure or to differences in the effect between the two groups of the "other" factors. It is important to understand why this feeling exists and why it is wrong.

The belief in the necessity for such complete control of all factors affecting the variable whose variation is being studied is derived from common sense, perhaps, and from a misconception of the principles underlying experimentation in the natural sciences. Uncontrolled variation in these "other" factors is not fatal to unambiguous experimental results, although such variations do diminish the accuracy with which the effect of the stimulus being investigated can be measured. It is because of this diminished accuracy that natural scientists try to lessen variation in these "other" factors.

In the laboratory, elaborate apparatus makes possible a tremendous diminution in variation of "other" factors, especially the influences of the physical environment. Even so, the control is not perfect. Variation in the "other" factors exists and should be taken into account in designing laboratory experiments and in the interpretation of their results. For example, one ubiquitous factor that must be taken into account is the inability of human beings to make perfectly precise measurements, even in the laboratory. This deficiency causes variation in the results of experiments performed under the same circumstances by different investigators or in experiments performed by the same investigator at different times. And this variation must be allowed for.

The variation in "other" factors is almost always relatively larger¹ in social than in laboratory experiments, but the difference is merely one of degree. The same theory which permits unambiguous interpretation of laboratory experiments with the relatively slight variation in "other" factors that attend laboratory conditions permits unambiguous interpretation of social (marketing) experiments with their relatively great variation in "other" factors. The effect of atmospheric pressure on the boiling point of a liquid can be determined with great accuracy because all relevant variables other than pressure can be held within very narrow limits; the effect of fertilizer on crop yields can be determined with less accuracy

¹ Throughout this discussion it is realized that comparisons of variation between different fields of scientific investigation would be difficult operationally. The idea of such comparisons is sufficiently clear for our present purpose.

because variations in soil composition, seeds, sunlight, etc., from plot to plot cannot be so narrowly controlled; the effect of advertising can probably be determined with even less accuracy because of the uncontrolled and probably relatively great fluctuations in "other" factors—the numerous influences that affect the willingness and ability of consumers to buy a product. Nevertheless, the effect of advertising—like the effect of pressure on the boiling point of a liquid or the effect of fertilizer on crop yield—can be determined with measurable risks of error by experimental means.

The theory which underlies and makes possible this unambiguous and rigorous interpretation of properly conducted experiments is the theory of statistical inference. By applying the stimulus whose effect is being investigated to a group of people (stores, markets, etc.) which has been selected by a random process, this lack of ambiguity can be achieved. The term "random" is of crucial importance and probably needs explanation. A random process of selection is one in which a fraction of the total number of units in a group (the group under investigation) is selected in such a way that all the units in the group have an equal and independent probability of being selected.¹ The achievement of randomness requires very careful planning, and this achievement is essential for experimentation.² If and only if the selection of the group to which the stimulus will be applied (hereafter to be termed the "experimental" group) and the selection of the group from which the stimulus is to be withheld (hereafter to be called the "control" group) are at random, can the experimental results be analyzed statistically. An example will illustrate what is meant by the statistical analysis of an experiment and the "measurement of the risks of error."

Consider the example of the marketing researcher interested in measuring by experimental means the effectiveness of an advertisement. The effectiveness is defined as the difference in the change of rate of sale of the advertised product between the group exposed to the advertisement and the group not exposed. That is, if, after the introduction of advertising, sales to the group exposed to the advertising showed a greater increase (or smaller decrease) than sales to the control group, the advertising effectiveness would be equal to the difference between the increases (decreases) for the two groups. Assume that the group exposed is distinguishable from the group not exposed and that rates of sale to the two groups can be measured. These assumptions would be justified if, for example, local newspaper, radio, or outdoor advertising were used. If these media were used,

¹ This definition applies to simple random sampling. A more general definition would state that each element in the group has a known, positive probability of selection. Further qualifications as to the presence or absence of independence are also needed for a more complete discussion.

² There is a much fuller discussion of the problems associated with the achievement of randomness in Chaps. 6 to 10.

it would be possible to advertise in some local markets and to refrain from advertising in others and to determine the volume of sales in each.

The first methodological problem is selecting the markets for the experimental and control groups. A random selection could be achieved by numbering all the markets in which the product is sold and selecting a sample, say 20, by reference to a table of random numbers.¹ The twenty markets so selected could then be divided into two subgroups by a random process (*e.g.*, the use of the table of random numbers), and one of the groups (the experimental group) would be exposed to the advertising for a specified period of time. The rates of sale to the experimental and control groups would be measured prior to, during, and after the exposure to the advertising. Some conclusion would then be reached regarding changes in the rates of sale to the two groups.

A conclusion could be reached regarding the effect of the advertising in spite of the fact that nothing was done to control the effect of variables *other than advertising* that affected the rate of sales to the two groups. The hypothesis under test in this example was that the advertising would cause an increase in sales (or would cause an increase of a specified percentage). Only two conclusions could be reached: (1) the advertising was effective—meaning that the change in the rate of sale to the experimental group was significantly greater (*i.e.*, either a greater increase or a smaller decrease) than the change in the rate of sale to the control group, or (2) the advertising was ineffective—meaning that the change in the rate of sale to the experimental group was not significantly greater than the change in the rate of sale to the control group. Reaching the conclusion that the evidence was consistent with the hypothesis implies that the data show that the difference in the change in the rate of sale was greater than would ordinarily be expected from the results of the variation in the “other” factors alone. Reaching the conclusion that the evidence was not consistent with the hypothesis implies that the difference in the change in the rate of sale to the two groups was not greater than would ordinarily be expected from the effects of the uncontrolled variation in the “other” factors. Two sorts of errors are possible in reaching these conclusions: (1) accepting the hypothesis that advertising is effective when it is not effective, and (2) rejecting the hypothesis that advertising is effective when it really is effective. That is, it is possible to accept the hypothesis as valid when in fact it is invalid. It is also possible to reject the hypothesis when in fact it is true. The risks of making these errors must be faced.

The risks exist in any attempt to generalize about a group on the basis

¹ A table of random numbers is a listing of groups of digits. The listing has been so determined that the probability of selecting any digit, 0 through 9, is the same in any selection of contiguous digits anywhere in the list. There is further discussion of the use of these tables in the section of the book which deals with sampling.

of observations of only a sample of that group. So long as the selection of the experimental and control groups was made at random, however, it is possible to state these risks in probability terms. For example, *if* the advertising had had no effect, there might be a 5 per cent risk of coming to the erroneous conclusion that the advertising had actually been effective. This would mean that about 5 per cent of the time, in repeated experiments of the same kind (in which 20 markets were selected at random, divided at random into a control and experimental group, and subjected to differential exposure to advertising) one would expect to come to the wrong answer if, in fact, the advertising had really been ineffective. By the same reasoning one might expect to come to the right answer in about 95 per cent of repeated experiments if the advertising were really ineffective. Similarly, *if* the advertising had really been effective, there would be a given risk of coming to the incorrect conclusion that it had been ineffective.

In practice, of course, only a single experiment is typically performed, but the confidence in the method of experimentation is expressed in terms like those just given. The reason such probability statements can be made concerning the risks of error in spite of the uncontrolled variation in the "other" factors is that the group to which the advertising was applied was selected at random, as was the group with which the comparison was made. This means that the "other" factors could be expected to affect both groups about equally in a long sequence of repetitions of the experiment. In any one experiment the effect usually would not be equal, but deviations from this equality would be attributable to random fluctuations rather than to any systematic variation in the "other" factors. Since the deviations would be attributable to random fluctuations, the probability of deviations of any given size can be measured by the techniques of statistical inference. Therefore the probability of obtaining deviations greater than any specified size can be determined. Some size of deviation for which the probability is relatively small, say 0.05, is chosen. It is agreed that a larger deviation will lead to the conclusion that something else besides random variation in the "other" variables is at work. There is, however, always a 5 per cent chance that, if only random variation were present, a deviation this large or larger would be obtained.

The random fluctuations will be larger when the "other" factors have a relatively large influence on sales and when the variation in the "other" factors from one local market to the next is large. When the "other" factors have a great effect on sales or when the variation in the "other" factors is relatively great, the probability of a given deviation between experimental and control groups will be larger, for any given sample size, than if the opposite were true. But regardless of the importance of the influence of the "other" factors on sales or the variability of "other" factors from mar-

ket to market, the conclusions of the experiment can be expressed unambiguously in the form exemplified in the preceding sentences. It is this fact that explains the potential superiority of experimental methods over ordinary trial and error in interpreting the effect of one variable on another. Later in the chapter we treat this point further when we discuss the interpretation of observational data.

1. Controlling and diminishing risks of error. There are two means available for diminishing the risks of error in experimentation: (1) increasing the number of observations and (2) modifying the random process by which the experimental and control groups are chosen. In statistical terminology these methods are known as increasing the size of the sample and improving the design of the experiment. Another example will show how these methods are applied.

A company which operates a chain of 200 retail grocery stores is considering the advisability of installing self-service meat departments. Ten stores are selected at random and are divided randomly into two groups of five. One of the groups, the experimental group, will sell meat through self-service departments; the other group, the control group, will sell in the orthodox manner. Measurements of the rates of sale of meat will be made both before and after the introduction of the self-service departments, and some conclusion will be reached regarding the significance of the difference (if any) in the change in the rates of sale between the two groups of stores.

Again there are risks of interpreting the observed differences incorrectly, but these risks could be diminished by increasing the number of stores in the experimental and control groups. As has been stated, the selection of the experimental and control groups by a random process ensures that *on the average* (i.e., with many repetitions of the experiment) the effect of "other" factors on sales to each group will tend to be the same. It was also stated that for any single experiment the effect of these factors on the two groups could not be expected to be exactly the same and that the probability of differences of any given magnitude could be determined. If it were possible to perform more than one experiment and average the results, the probability of large deviations would be smaller than for a single experiment. Similarly, the probability of large deviations can be diminished by increasing the number of stores in the experimental and control groups on any one experiment. The risks of reaching erroneous conclusions therefore could be diminished by having a larger number of stores than five in the experimental and control groups. In general, in making inferences about a large group on the basis of observations on a portion of the group, the risks of making incorrect inferences decrease as the number of observations increases. Or, in statistical language, in sampling at random, the sampling (or random) error tends to decrease as the sample size increases.

Moreover, when an experiment is designed to test a hypothesis about some variable whose effect on the variable being investigated (sales of meat, in this example) is relatively unimportant, the sample must be large relative to that required to test hypotheses concerning relatively important causes of variation. For example, a smaller number of stores probably would be needed to test a hypothesis regarding the effect of self-service meat departments than would be required to test a hypothesis about the effect of relatively slight changes in display for conventional types of meat departments. The effect on sales of minor changes in display might be so slight that a prohibitively costly number of stores might be required in order to detect the effect being studied. That is, with a relatively small number of stores, differences between the control and experimental stores which are attributable to sampling might tend to obscure the effect of the changes in display.

The risks of reaching erroneous conclusions can be lessened not only by increasing the size of sample (number of observations, number of stores in the experimental and control groups)—a costly process—but also by modifying the way in which the experimental and control groups are selected. For the experiment to yield measurable risks of error, it is essential that the selection be random, but it is not essential that the process be the simple one described above. (See footnote on page 50.) In the example, risks of error could be diminished if the random variation between the experimental and control groups in some of the “other” factors could be reduced below the level to be expected in the experimental design suggested above. The effect of such variations is often diminished in the laboratory by the physical control of the environment. That is, the effect of variations in temperature and pressure, for example, are controlled physically by some form of mechanical apparatus. Similar control of “other” factors is seldom possible in investigations of social behavior, and therefore other means of diminishing the effect of variation in these factors must be found.¹

The chief means that has been found involves adapting the method of selecting the units to be included in the experimental and control groups to the problem being studied without compromising the randomness of the selective process. The guiding principle in this process of adaptation is that the experimental and control groups should be made as like as possible with regard to the “other” factors so that differences in the effect of these

¹ Sometimes a similar form of control is possible in marketing research. In sales tests in which sales are measured in the retail store, a frequently important “other” factor is the tendency of stores to run out of the product whose sales are being measured. It is common, therefore, to make special efforts to ensure that both the control and experimental stores have a supply of the tested product on the shelves continuously during the experiment. Sometimes factors such as shelf position, special displays, etc., are also controlled.

factors on the variation being investigated can be diminished. This principle can be implemented by the process of breaking up, or stratifying, the larger group (the universe) into smaller groups on the basis of the "other" factors and then selecting experimental and control units *at random* within each of these subgroups. In other words, an attempt is made to create subgroups within the universe, each subgroup containing stores as similar as possible with regard to one or more of the "other" factors. The group of 200 stores might have been subdivided into smaller groups on the basis of sales volume and location so as to make these smaller groups more nearly similar than was the group of 200 with respect to these two factors. The group of 200 obviously could have been divided according to more or fewer factors, the basis of the decision as to the number of factors depending upon estimates of the cost of such subdivision and the value of the resulting gains in the accuracy with which the experimental data could be interpreted.

In order to simplify the exposition, an example will be given of stratification according to only one factor. In the example, the 200 stores will be subdivided into five groups on the basis of the sales volume of meat. Within each of the five groups, two stores will be selected at random and one of the two will be selected at random for the experimental group, leaving the other for the control group. Thus there will still be five stores in the experimental and five in the control group. But each group will have one store with very large sales volume, one with moderately large sales volume, etc. Hence, the experimental and control groups will be much more comparable with respect to sales volume, thereby diminishing the disturbing effect of this variable and rendering the experiment more sensitive to the effect of the form of service in the meat departments. It will be possible, therefore, to draw inferences about the effect of self-service with smaller risks of error, even though the number of stores used in the experiment was not altered.

2. Simultaneous evaluation of the effects of several variables. Before leaving this rudimentary discussion of experimentation, it should be pointed out that it is possible to design experiments which permit a single experimental process to cast light on the relationships between several variables. It is theoretically possible, for instance, to devise an experiment which would indicate not only the effect of self-service but also the effect of variations in type of merchandise, prices, etc. Experiments indicating the individual and collective effect of several variables are commonplace in agricultural experimentation, where most of the complex experimental designs have been developed. These relatively complex designs have had almost no application in marketing and other types of business research where their use can quite possibly produce more efficient research. Experiments yielding information about several variables can usually be con-

ducted for approximately the same cost as experiments yielding information about only one of the variables. The use of these complex (or, indeed, any) experiments in marketing research requires thorough administrative planning and therefore presupposes that experiments are not resorted to only in moments of crisis when answers must be gotten quickly, but that experimentation is a well-established and pervasively used means of getting objective information about the effects of proposed changes in marketing practices and policies.¹

3. Further comments on limitations of experimentation. Experimental results, whether from complex or simple experiments, are useful to the extent that they permit generalizations. The selection by a random process of a sample from the population being studied ensures the ability to generalize about that population at that point in time. In other words, generalization through space is assured by the nature of the experimental process. The temporal instability of social relationships and the changing nature of the effect of different factors on marketing behavior may mean that relationships established at the time of the experiment fail to continue into the future. This temporal instability is a handicap to all methods of forecasting, however, and seems to be no more severe for experimentation than for other methods. Experiments, if they are to provide information about the nature of temporal changes in relationships, must be either periodic or continuous. Individual experiments of short duration usually should be relied on to provide information about short-run effects only. The length of time for which the experimental results can be accepted as valid can be increased somewhat by lengthening the period of the experiment. There are practical limitations to this because of the expense involved and because a prolongation of the period of experimentation might make it increasingly difficult to keep from the persons being observed the information that they are part of an experiment.

For the most part it is possible only to speculate about the difference between short- and long-run effects, and not always will the experiment itself help in the speculation. In the example discussed here, it is possible that the experiment would indicate a greater effect on sales than would continue to exist in the long run because of the probability that a successful introduction of self-service would cause competitors to institute similar service, thus decreasing the diversion of customers from orthodox departments in other stores to self-service departments in the experimental stores. On the other hand, it is possible that the experiment would understate the positive effects of the new form of service because of people's reluctance to try something new. It is difficult to provide any general rules which will

¹ For a more complete discussion of this subject see George W. Snedecor, *Statistical Methods*, Chap. 11, Iowa State College Press, Ames, Iowa, 1946.

indicate which of these two tendencies would tend to predominate. Moreover, it is equally hard to ascertain other tendencies that might influence the future or to forecast their importance.

The discussion so far has emphasized the theoretical aspects of marketing experimentation. In experimentation as in any other kind of research there are many administrative problems to be considered. In fact, experimentation probably requires more competent and active administration than most other kinds of research, since experiments involve the intervention of the investigator into the process which yields the data. The general problems of administration in marketing research will have to be delayed until the final part of the book.

Just one thing deserves mention here. Experiments are designed to indicate the effect of controllable stimuli. Each agreed-upon change in marketing practice or policy involves modifying the stimuli which affect the sale of a good or service. The time when it is least costly to find out the effect of agreed-upon changes is when they are introduced. If experiments are an accepted part of research procedure and if research is typically used to provide information about expected reactions, then experiments used *at the time of the introduction of the change in policy or practice* provide a powerful and relatively cheap means of measuring the effect of the change. Experiments at a later date are frequently unfeasible because of the relatively greater costliness of securing experimental and control groups. At the time of the introduction of the change, such groups are automatically provided by selecting (by some random process) a part of the enterprise in which the change will originally be introduced and selecting another part as a control. In other words, the introduction of changes "piecemeal" provides the basis of an experiment if the "pieces" are chosen properly. Since many changes are made piecemeal in any event, the possibility of experimentation is very frequently present and no unusual administrative measures may be needed.

4.3 Observational Data

Experiments are a very powerful investigative tool, but unfortunately experiments are sometimes impossible, undesirably time-consuming, or too costly. The vast bulk of marketing research does not involve experimentation. Most data of relevance in marketing research may be called observational data. The essential distinction between observational data and experimental data is simple. Observational data arise without the deliberate intervention of the researcher; they are simply a recording of the normal course of events as they would have proceeded had they not been recorded. Experimental data, by contrast, result from the deliberate intervention of the experimenter in the course of events being recorded.

There are two main sources of observational data in marketing research. Some data are available from accounting records or other internal records or outside secondary sources; other data are obtained from special marketing surveys.

The routine operations of the business enterprise provide a large number of records which give valuable information when properly analyzed. As a matter of course, an interested sales executive can expect to have data on sales classified by salesmen, by territories, by products or departments, etc. He can also expect, as a matter of course, to get with negligible difficulty demographic data on territories or urban areas and indices indicating the level of education, purchasing power, etc.¹ The use of these types of data in formulating marketing policy involves analysis which frequently goes by the name of "sales analysis." The term "sales analysis" may be misleading, since the methods used in sales analysis are essentially the same as the methods used in analyzing primary data of the type collected in marketing surveys. "Sales analysis" refers to the particular kinds of data being analyzed (*i.e.*, accounting information in conjunction with readily available information from secondary sources) rather than to a method of analysis. It is probably fair to say that the majority of marketing decisions which rest in part at least on observational data involve sales analysis rather than the acquisition of primary data by surveys or other techniques. The distribution of space in this book could, however, quite easily lead to the conclusion that marketing surveys are the most frequently used sources of data. This probably is not true, if importance be measured by the time and dollars devoted to the various research techniques in current practice.

Regardless of the source—government statistics, accounting records, or marketing surveys—the analysis of observational data presents problems that were not present in the analysis of experimental data from properly planned experiments. It is well to study the basic difference, because a clear understanding of this difference is immensely helpful in making intelligent use of observational data. In the interpretation of experimental data it is possible to conclude that one variable has had an effect on another, and the conclusion, while fallible, is subject to risks of error that can be evaluated in the sense discussed earlier in this chapter. Assuming that basic conditions remain as they were at the time of the experiment, it can be forecast that in the future a change in one variable will have a given effect on the other. If in the example of the sales test of advertising effectiveness it was found that sales in the experimental markets had signifi-

¹ One example of this kind of information is the "Buying Power Index" published by *Sales Management*. For other examples of the many sorts of data which are available and useful in analyzing marketing activities see D. R. G. Cowan, *Sales Analysis from the Management Standpoint*, University of Chicago Press, Chicago, 1938.

cantly exceeded sales in the control markets, it could be forecast that advertising applied on a national scale would produce effects on sales similar to those observed during the test, provided that the basic conditions (*e.g.*, competitors' advertising) that existed during the test remained the same.

Suppose, however, that advertising effect is analyzed by the usual methods of sales analysis. The country would be divided into a number of local markets.¹ In each local market both sales and the input of advertising would be measured. No attempt would be made, however, to decide randomly which local markets would receive a large amount of advertising and which would receive little or none; the essence of the observational data is that no effort is made to control in the interest of research the things being observed. Assume that in the markets in which sales were high, advertising was also high and that in the markets in which sales were low, advertising was low. There was therefore an association or correlation between advertising and sales. The next question is, "What value does this association have in making forecasts of what would happen if advertising were changed?" What would happen, for example, if advertising were increased in the markets in which, at present, both advertising and sales are low?

Before attempting to answer this question, recall the example of experimentation. If sales in the experimental markets were significantly higher than in the control markets, one could forecast that the introduction of advertising into the control markets would result in an increase in sales. In such a forecast there would be risks of having made an error in allowing for random variability in drawing the original inference, but the risks of error are measurable. Moreover, in making such a forecast, there is the additional risk that the basic conditions prevailing at the time of the test would not continue into the future.

In attempting an analogous forecast from observational data, however, one is immediately confronted with two additional sources of error that were not encountered in the analysis of experimental data. It was found that a high level of advertising was usually found in conjunction with a high volume of sales and vice versa. The probability of such a strong association arising from random variation alone can be assumed to be negligible for the present discussion. On the basis of this evidence it is impossible, however, to be confident that an increase in advertising in markets that now have poor sales and a small amount of advertising would result in an increase in sales. In addition to the qualifications that would be present in the analogous experimental situation, there are the two additional sources of error just mentioned but not yet identified.

¹ The 608 Hearst "trading areas" are frequently used for this purpose.

The first source of error is that "other" factors cannot be expected to be the same within the limits of random error. There may be great differences in "other" factors between regions in which advertising and sales are high and regions in which they are both low. These "other" factors, whose probable influence can no longer be evaluated as in the experimental situation, may be wholly or partly responsible for the observed association between advertising and sales. It is possible that the same factors which led to a high level of sales also led to a high level of advertising. For example, relatively high levels of income could have been responsible for both the large purchases of the product being sold and the large purchases of the magazines that carry the advertising for the product. The "other" factors, such as income, could be equalized within the limits of random error when experiments were used. This equalization was accomplished by the use of random selection of markets. This random selection of markets for the application of differential advertising stimuli is no longer possible in observational data, since, by definition, the researcher records events as they arise without any attempt to influence or modify them in the interest of research. As a result the "other" factors have not been approximately equalized within the limits of measurable risks of error. The many "other" factors besides the one whose effects are being studied—namely, advertising—may have accounted for some or all of the observed association between advertising and sales. If these "other" factors are in fact partially or wholly responsible for the observed association, it is said that these "other" factors are "confounded" with the variable of primary interest. When confounding exists, it is no longer possible to isolate the desired relationships.

The second problem in the use of observational data for making predictions (*e.g.*, about the effect of changes in advertising on sales volume) would be present even if the problem of confounding did not exist. Even if the observed association between advertising and sales were not attributable to "other" factors, there would still be reason to hesitate in forecasting that an increase in advertising would result in an increase in sales. This additional reason for hesitancy is that one cannot know from observation alone which variable "caused"¹ which. That is, observation does not tell whether sales were high because of large expenditures on advertising or whether large expenditures on advertising existed because of high sales volume. In the experimental approach to the problem of evaluating the effectiveness of advertising, this problem did not exist. The introduction

¹ While the use of the term "cause" leads to many problems if its meaning is explored deeply enough, it is felt that the everyday understanding is sufficiently clear for our purposes and that a more qualified mode of expression would cause unnecessary obscurity and confusion.

of advertising was made at random, and no attempt was made to place advertising in markets where it was felt that the advertising would "do the most good." In the observational example, the advertising executives responsible for the allocation of advertising effort to different local markets did not make their decisions at random. They consciously tried to put advertising where it would "do the most good." There is a fairly common belief (whose validity would be difficult to test) among advertising men that advertising will do the most good if it is allocated to markets where sales are already high or where a strong sales force and dealer organization already exist. The media men responsible for the allocation of the advertising to local markets thus typically allocate advertising in such a way that little advertising goes to markets where sales are low and a lot of advertising to markets in which sales are high or potentially high because of a strong marketing organization. It is not surprising, then, that the research man finds a strong association between advertising and sales. Even if advertising had no influence whatever on sales, this association would exist because of the way in which advertising is allocated.

These two sources of error in the analysis of observational data are formidable. The basic problem is to find ways of evaluating or neutralizing the effects of "other" variables and to uncover evidence on the "direction" of causation. The best methods are likely to be disappointing and, by comparison with experimental methods, unsatisfactory. Nonetheless, there are many occasions on which observational data can be used in making marketing predictions. One must, in fact, use them whenever experimental data are unavailable. The only other recourse is to make no use of data at all and to rely instead on the intuitive appraisal of one's own experience and informally acquired knowledge. Since observational data are of great value in spite of their manifest limitations, the rest of this chapter will be devoted to a discussion of the ways in which these limitations can be diminished by the use of statistical techniques.

1. Cross-classification. One of the two sources of error inherent in the use of observational data which are usually not present in experimental data is the possibility of confounding. In the example of the study of advertising effect by the use of observational data, it was seen that it was possible to confound the effects of advertising and of income on sales. In order to reduce the possibility of this sort of confounding, it would be useful to get some indication of the association between income and sales. This could be done quite easily by dividing the total group of markets for which data are available into subgroups on the basis of per capita income in each and computing per capita sales for each subgroup. Assume that such a process showed a significant association between income and sales such that high per capita sales tended to be associated with high per capita

income. The next step in seeing whether advertising had any effect independent of the effect of income would be to subclassify again on the basis of advertising. Such a cross-classification by income and advertising would look something like that shown in the accompanying table.

In each "cell" of the table are two numbers. The first number specifies the average per capita sales for the markets in the cell. For example, in markets in which both income and advertising were low, the average per capita sales per market were \$10. Immediately underneath this number is a second number in parentheses, which tells how many markets were

AVERAGE PER CAPITA SALES IN MARKETS CLASSIFIED BY INCOME AND ADVERTISING*

	Average per capita sales			Average per market
	Low advertising	Medium advertising	High advertising	
Low income	\$10 (100)	\$10 (50)	\$10 (10)	\$10 (160)
Medium income	\$20 (50)	\$20 (200)	\$20 (50)	\$20 (300)
High income	\$30 (10)	\$30 (50)	\$30 (100)	\$30 (160)
Average per market	\$14.375 (160)	\$20 (300)	\$25.625 (160)	\$20 (620)

* Sales, income, and advertising are on a per capita basis.

included in that particular classification. For example, there were 100 markets in the category "Low income, low advertising."

Now, concentrate on the bottom row of the table, which is labeled "Average." For markets with low advertising, the average per capita sales were \$14.375; for markets with medium advertising, \$20.00; for markets with high advertising, \$25.625. It is apparent, therefore, that as more and more advertising is encountered, average per capita sales tend to be higher and higher.

Next, examine the extreme right-hand column of the table, which is also labeled "Average." From this column it is seen that average per capita sales increase from \$10 to \$20 to \$30 from low-income to medium-income to high-income markets. Sales tend to be higher as income is higher.

These first two facts illustrate what previously had been assumed: that there is an association between both advertising and sales and between

income and sales. The next task is to examine the body of the table. The first row gives information for low-income markets with low, medium, and high advertising expenditures respectively. It is seen immediately that for low-income markets, the average per capita sales are the same regardless of the level of advertising expenditures; in each case, average per capita sales are \$10. Similarly, per capita sales in medium-income markets, while higher than sales in low-income markets, are the same regardless of advertising; the average is \$20 throughout. In the high-income markets the average per capita sales are \$30, but there is no difference between markets exposed to different levels of advertising.

For markets with similar per capita incomes, therefore, the average per capita sales no longer vary with changes in advertising. This finding may seem curious and paradoxical. Initially, there was a strong association between advertising and sales, but after cross-classification, the association disappeared. The reason is not hard to find: there is also an association between income and advertising. There are 100 markets with low advertising and low income, 200 markets with medium advertising and medium income, and 100 markets with high advertising and high income. These three classifications alone account for 400 of the 620 markets. There are relatively few markets (10) in which income is low and advertising is high or in which advertising is low and income is high (10). The association between income and advertising is responsible for the observed association between advertising and sales in the bottom row of the table. For the markets in which advertising is low are markets in which income also is low. The markets in which advertising is high are markets in which income is high. This association between advertising and income, taken in conjunction with the association between income and sales, is sufficient to account for the observed association between advertising and sales.

This example is stripped of the many perplexing details that would accompany a typical problem. It would be unusual to find that the association between income and advertising plus the association between income and sales "explained" all the observed association between advertising and sales. In the more usual cases, it would be much more difficult to leap to obvious interpretations, and additional analytical problems would be raised. In this simple example, however, the obvious interpretation is worth examining. It is this: "The advertising isn't doing any good." For, according to the evidence, the apparent effect of the advertising disappears with the introduction of the additional variable, income. It would appear that income had an important effect on sales and that advertising had none. This would be an unduly hasty and possibly incorrect conclusion, however, for the effects of the "other" variables of importance have not been considered. Income was a plausible choice for a first attack, but many other

variables might also have been explored. If the method of cross-classification were used in the exploration, the approach would be the same, but the tables would get more and more complicated and more and more difficult to interpret. (Usually the procedure of cross-classification would be carried no further than one additional variable for a number of reasons, some obvious and others more subtle, that we need not consider here.) The technical complications need not be further pursued here. Rather it will be helpful to consider a situation in which the conclusion to cease advertising would be incorrect. One such situation would be this: Advertising is at present canceling out or offsetting the effect of competitive advertising, but no more. If advertising were stopped, this offsetting effect would be lost and sales would be lost to competitors.

The method of cross-classification is relatively simple and can be used by persons without extensive training in statistics. The idea is to cross-classify data according to the variable whose effects are to be measured (*e.g.*, advertising), the variable to be influenced (*e.g.*, sales), and one or two of the "other" factors whose effect should be taken into account (*e.g.*, income). Doing this makes possible comparisons between groups whose components are similar with respect to the "other" factors. In this case, for example, the effects of different levels of advertising were compared within groups of markets which were similar with respect to income. When the cross-classification is finished, there are serious problems of interpretation remaining, even ignoring the vexing idiosyncrasies of real data that have been avoided in the example. One is never sure that all the important "other" variables have been examined. Just as the picture changed surprisingly when income was added to the simple association between advertising and sales, so the addition of another variable, such as region of the country, might add still new information. And one can never be sure that all the relevant variables have been considered, even if technical obstacles did not restrict the consideration to a relatively small number.

But there is still another problem of interpretation. Consider the example presented here, in which income apparently explained the association between advertising and sales. The second problem that arises in the analysis of observational data—the direction of causation—has still not been considered. Does income increase sales or do sales increase income? This question cannot be answered solely on the basis of the data at hand, but fortunately a general knowledge of economics suggests that the causation, if any, proceeds largely from income to sales rather than in the opposite direction.

Suppose, however, that the example had been set up so that there was an association between advertising and sales even after allowing for the "other" variables believed to be important. Suppose further that this

association is considered to be evidence of causation. It is then necessary again to ask in which direction the causation runs. Again other knowledge must be used. In this example, the other knowledge does not provide so clear-cut an interpretation. It is possible, on the one hand, that some effort is made to allocate advertising according to the volume of sales already attained in a local market. On the other hand, various scraps of evidence (comments of salesmen and consumers, for example) may suggest strongly that the advertising is also influencing sales. The only recourse, then, is to combine the evidence from the cross-classification analysis with the other information, and to use judgment in reaching a conclusion.

2. Other statistical tools. So far one statistical method of studying observational data, namely, cross-classification, has been examined. This method is simple and easily applied.¹ Other statistical techniques may be used for the same kind of problems: analysis of variance and covariance, regression and correlation analysis, etc. These techniques have both advantages and disadvantages over the simple method of cross-classification. If used appropriately in circumstances in which the assumptions underlying them are reasonably well fulfilled, and if the interpretation is handled by a person skilled in statistical theory and thoroughly experienced in the practical problems of application of statistical theory, results superior (*i.e.*, more precise, unambiguous) to those of the simple methods can be expected. If these prerequisites are not fulfilled, however, the consequences may be unfortunate. Powerful statistical techniques in the hands of people who know little more than the directions for computation are as dangerous in marketing research as a high-powered racing car in the hands of an individual who does not know how to drive. The best thing to do in either case is to stay away from the dangerous gadget.

Analysis of variance and the chi-square test for contingency tables are two relatively simple statistical techniques that can be easily learned and soundly applied by nonstatisticians. In the bibliography at the end of this book sources are given from which individuals with relatively little statistical training can acquire some proficiency in these techniques. The effort is worth while if for no other reason than that it gives insight into the problems of cross-classification and into the types of associations that can prevail.

Regression analysis and correlation analysis should be left to mathematical statisticians with a flair for practical applications of statistics. The computations are deceptively easy to learn, but a real understanding of the techniques—such that one can understand what the computations mean

¹ The chief usefulness of statistical sophistication in analyses of this kind is to enable the investigator to perform tests of significance that protect the investigator against being misled by random or accidental features of the data.

when he has gone through them—comes only to the advanced statistician, if to him.

It seems appropriate here to add the comment that a great deal of sophisticated thought and empirical work has been applied to the problem of extracting "causal" relationships from observational data. Workers in two different fields—econometrics and factor analysis—have led the way in these developments. The essential characteristic of these approaches is the use of theoretical models which are felt to underlie the observational data. In the Thurstone variation of factor analysis, for example, it is postulated that a relatively small number of mental abilities will largely account for the performance of a group of individuals in a large number of mental tests. Work of this kind may ultimately be of great value to market researchers.¹

3. Other methods of using observational data. A supplementary means of attack has been suggested above when it was pointed out that a person trying to find out whether advertising caused sales or sales caused advertising must use supplementary information. One method, for example, would be to ask consumers if the advertising had caused them to buy the product. This ingenuous approach (as we point out in Part IV) is a relatively weak way of getting at the problem, but one cannot afford to ignore sources of information, however weak, when confronted with such perplexing questions.

Another attack has been suggested at the beginning of the chapter. Instead of attacking the problem directly, it is possible to consider a more limited hypothesis that seems to be relevant to the more important but more elusive one. This is frequently done in consumer and dealer surveys. Instead of trying to make a direct prediction as to what would happen if advertising were changed, one might make a dealer survey to find out the attitudes of dealers toward advertising. It could be reasoned, for example, that the morale of dealers and their willingness to "push" the product would have a lot to do with sales. If a dealer survey showed that dealers tended to keep the product under the counter because of a failure to provide as good advertising support as a competitor does, it would probably be reasonable to assume that an increase in advertising would increase sales even if it had no direct effect on consumers. Studies of this kind provide observational data which make possible the testing of some limited hypothesis, which casts light upon the basic hypothesis whose validity must be appraised.

¹ For a reference which discusses these matters in greater detail and, in addition, gives an excellent bibliography, see Koopmans and Reiersøl, "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, Vol. XXI, No. 2, June, 1950.

CHAPTER 5

THE USE OF MARKETING RESEARCH FOR MARKETING PROBLEMS

In this chapter, we discuss the kinds of decisions for which marketing research can profitably be used. Undoubtedly, the great majority of marketing decisions are made without recourse to formal marketing research. Sears, Roebuck and Company, for example, has achieved its present great size and success without ever having had a marketing research department and without ever having made great use of the marketing research of outside agencies. On the other hand, many large and successful companies use formal marketing research in making a great many marketing decisions. Before going on to the problems of getting information through marketing research, it might be well to see the circumstances in which information so collected can be expected to be useful.

This discussion of the usefulness of marketing research for marketing problems will be divided into two parts: (1) the nature of problems and the processes by which they are recognized, and (2) the general principles which should guide executives in their decisions to solve these problems by the use of research or by other means.

5.1 The Recognition of Problems

There are three different kinds of problems in marketing research. The first arises from the need to identify those policies and practices of the enterprise which should be changed in the interest of the enterprise. As a subhead to this kind of problem should be added the problems arising from the need to find out if new policies and practices should be established where none existed before. The second arises from the need to formulate specific alternatives to the practices and policies which have been identified as being in need of change. The third kind of problem arises from the need to choose between the designated alternatives. In this chapter we discuss problems of the first two kinds; in the rest of the book we discuss the third kind of problem— getting information to enable executives to choose between designated alternatives.

Enterprises do a vast number of different things, and the task of deciding which of this vast number need changing is impressively difficult. In practice, the need for change is usually indicated by trouble. Personal

friction or disagreement, some palpable breakdown of the corporate machinery, some spectacular and successful change in policy by a competitor, etc., are the sorts of stimuli which typically indicate the need for a revision in the company's policies. Usually there is no continuing procedure within the enterprise designed to indicate the need for considering changes in current policies; usually there is only a response to trouble when it arises. Such a procedure of waiting for trouble is unscientific. To deem a procedure unscientific is not, however, to condemn it; we have not yet become so enamored of scientific method that we will reject anything which is unscientific. There are, nevertheless, rather obvious reasons for seeking some positive, continuing procedure for locating problems of this first kind.

The failure to change policies until trouble arises frequently results in subjecting the enterprise to adverse consequences which could have been avoided if the trouble had been foreseen. An army commander who failed to consider alternatives to his current disposition of troops until the enemy had attacked would be subjecting his troops and his position to unnecessary peril. Similarly, business executives who react only to the aggressive and successful action of competitors or to internal frictions subject their enterprises to unnecessary disadvantages. Moreover, the passive procedure which recognizes the need for change only after trouble has arisen causes decisions to be made because of urgent necessity and in an atmosphere of crisis. Under these circumstances, executives must often make decisions before they can secure much evidence that would be helpful in deciding the nature of the changes that should be made. A more active and objective process which anticipated trouble might permit executives to consider alternatives before the need for adopting them would otherwise be apparent and might permit, therefore, a more complete and systematic examination of the probable consequences of the alternatives. Finally, desirable alternatives are sometimes never found unless they are actively sought.

For these reasons, then, we shall consider what sort of active, exploratory, and relatively objective procedure can be devised. In the interest of finding such a procedure we shall make what seems to be a useful classification of problems of the first kind.

Some of the activities of companies are frequently recurring and involve frequently recurring expenditures. The payment of salesmen's salaries and other types of so-called sales expenditures are examples. Often it is possible to obtain rough measures of the revenues that are associated with these expenditures and thus to obtain a rough measure of the usefulness of the expenditures in making possible profitable operations. Merely be-

cause a current practice or policy yields a revenue that exceeds the associated expenditure, however, is no reason to assume that no change is appropriate and that alternative practices or policies should not be considered. In order for these rough measures to be useful in indicating the need for change, it is necessary that there be some standard or norm with which the measurements can be compared. It is not enough to know the revenues attributable to a particular salesman; it is also necessary to know whether the revenues so attributed are satisfactory or adequate on the basis of some standard of performance for salesmen operating under the existing circumstances. The first category of problems of the first type, then, involves appraising the efficiency of frequently recurring expenditures.

The second kind of problem concerns major policies which do not involve habitual, frequently recurring expenditures. The major decisions concerning product policy, brand policy, product design, price policy, etc., do not yield so easily, or perhaps do not yield at all, to attempts to measure their effects on revenues. The recognition of the need to consider alternative policies to those currently in force requires different procedures from those required to make possible the early recognition of problems concerning recurrent expenditures. We shall consider first the problem of early recognition of problems concerning frequently recurring expenditures.

1. Recognizing problems concerning frequently recurring expenditures.

As has been said, the first step in recognizing problems of this sort is the measurement of the results of such expenditures. The measurement of the results of expenditures is not always simple and may even frequently involve difficult technical problems. Now, however, it is desirable to assume that the problem of measurement has been or can be solved and to concentrate on a more difficult task.

The more difficult task in recognizing problems is that of establishing standards which will give meaning to the measurements. Unfortunately, there is no completely satisfactory way to do this. That is, there is no way of determining with a high degree of certainty the optimum policy or practice, because it is patently impossible to test the relative profitability of the infinite number of alternative ways of doing the many things for which a sales executive is responsible. The most that can be achieved is the establishment of some standard that experienced executives think is both desirable and attainable.

Standards are used for evaluating many types of sales personnel from salesmen to sales managers and are sometimes used for evaluating advertising campaigns as well as for other purposes. The standard (often referred to as a "quota") is usually conceived of in the same way as "par"

for a golf course. That is, the standard performance is one that represents a satisfactory level of performance for competent, energetic personnel.¹ We shall not go into details on methods for establishing such standards for all types of sales personnel or sales expenditures but shall limit the discussion to an example. The example will be the establishment of standards (quotas) for salesmen of a manufacturer who assigns salesmen to nonoverlapping territories.

Probably the most common method of establishing standards is to compare the performance of the persons being judged (salesmen, in this case) at the current time with the performance of the same sort of personnel at some time in the past. It is quite natural and almost inevitable that current activities be evaluated by reference to some historical period. In its simplest form, this method of establishing standards for salesmen is implemented by comparing the salesmen's current volume of sales with the volume that the same salesmen achieved either in the previous year or in some group of previous years.

When it is realized that the purpose of establishing this standard is to enable various salesmen to be evaluated relative to each other and absolutely, a basic weakness of this method becomes clear. Salesman Jones who was half as energetic or competent as salesman Smith would appear to do as well as Smith if the men were rated in terms of their own previous performances. If the base year or years which served as the basis of comparison remained fixed for relatively long periods, it would be in the individual salesman's interest to sell a small volume in the base period (assuming, of course, he knew what the base period was). If the base were not stable—that is, if the standard were always, for example, the preceding year or the average of the preceding five years—a salesman could do as well in terms of his own standard by working half or quarter time as by working full time. To return to the analogy of the standards used in golf, this method of setting standards is the same as would exist in golf if a player were rated as highly for shooting 100 as another player for shooting 72, simply because the first player had already established a score of 100 and the second a score of 72. The weakness in this method of establishing standards is that it creates a substantial difference between the activity which yields the greatest reward to the salesmen and the activity which yields the greatest reward to the employer of the salesmen. An important characteristic of a desirable method of establishing standards is that it creates very nearly an identity of interests for salesmen and their employer.

Sometimes the basic weakness of this method may be partially concealed

¹ This definition is purely a pragmatic one; obviously, the adjectives "competent" and "energetic" are vague.

when its implementation becomes complex. Various complicated formulas are often used, which consider expenses as well as sales (perhaps classified by customer or product or both). Many variations on this fundamental theme can be found in practice, but so long as each salesman is rewarded because of his current performance relative to some previous performance, the method is to that extent questionable. It is fairly easy to recognize this flaw, but it is difficult to remove it. The task involves judging the salesmen on the basis of their success in overcoming the difficulties which each faces rather than on the basis of relative performance in two time periods. The "par" for a golf course is established on the basis of the characteristics of the course and the difficulties which it presents to players. This same principle can be applied in establishing the "par" or standard for the markets in which salesmen operate. The objective of the golf player is to get the ball in the cup in as few strokes as possible. The objective of the salesmen from the employer's point of view is usually the creation of revenues which exceed associated costs by as great a margin as possible. Therefore, the difficulties which must be defined for each salesman are difficulties in creating a surplus of returns over direct costs.

The factors which seem to account for territorial differences in the difficulty of selling a product profitably can be subsumed under four headings: (1) the demand for the type of product being sold, (2) the sales activities of competitors, (3) sales activities by the employer, other than those involving salesmen, and (4) differences in the costs of making sales that are not at the discretion of the salesmen.¹ It is easy to see that Smith, who sells in a territory with 5,000,000 families which have an average income of \$3500 per year will probably find it more difficult to sell a given number of units of a product (*e.g.*, shoes) than will Jones who has a territory with 6,000,000 families having an average annual income of \$3750, assuming that other things are equal in the two territories. Furthermore, it is reasonable to assume that Smith will have a harder time than Jones if Smith must compete with 50 rival salesmen who are well supported with advertising while Jones has only 10 competing salesmen who are unsupported by advertising. Similarly, if Smith operates in a territory where his company's advertising is relatively light while Jones operates where the impact of advertising by his company is relatively heavy, Jones can be expected to have less difficulty. If Smith sells in a territory where transportation costs are high and the geographical dispersion of prospective customers is great, and Jones sells where transportation is cheap and the potential customers are highly concentrated, we would probably expect Smith's costs

¹ This categorization of influences is not mutually exclusive unless one assumes that sales effort by any firm in the industry results merely in redistributing sales among competing firms without altering the total sales volume of the industry.

to be higher than Jones's for a given volume of sales. These four sorts of factors, then, are the ones affecting the difficulty for Smith and Jones in achieving their employer's objective—the creation of profits. It is necessary to quantify these factors and then relate them to contributions to overhead before a numerical "par" for each salesman can be established. Doing this is always difficult and sometimes impossible. We shall not discuss these difficulties here, since our emphasis is on principles of analysis rather than upon the practical difficulties of application. If these practical difficulties can be successfully overcome, the resulting standard would have the following characteristics which are present in any ideal standard:

- 1 The standard would take into account the difficulties of selling in a given territory
2. The standard would ensure that the activities that are to the best interest of the employee are also to the best interest of the employer
- 3 The standard would have an unambiguous meaning for both employees and employer
4. The standard would permit a determination of the extent to which the employees attained a satisfactory level of performance

Making the measurements necessary for the establishment of such an ideal standard is often costly, frequently difficult and time-consuming, and sometimes impossible. There is reason, therefore, for considering briefly an alternative which is more easily adopted. An alternative which would achieve a fairly satisfactory compromise is the judgment of current performance by reference to previous performance. The deficiencies of such a procedure have already been pointed out in the paragraphs above. If these deficiencies are kept in mind and allowed for, such a procedure provides a practicable and useful means for locating possible trouble spots. The procedure would involve keeping current data on sales and expenses by salesmen and devising some easy means for detecting sharp deviations from normal in these variables. This method would lead the executive to investigate these sharp changes in order to find their causes and to take remedial action if it was indicated. The procedure would be insensitive to situations which needed changing but which did not involve fluctuations in sales and expenses. Consistently bad salesmen would go undetected. Also, the system might sound false alarms. It might be that the volume of sales by salesman or sales territory tended to fluctuate erratically from one time period to the next even when nothing fundamental had really changed. Salesman calling on large industrial accounts, for example, might show a violent change in their sales by losing a single customer. An investigation in such an instance might show simply that the salesman had lost a customer through no fault of his own or no particular merit of competitive salesmen. If, by contrast, a salesman calling on a large number of

small accounts showed a sharp drop in sales, something much more pervasive probably occurred, since a sharp drop could be explained only by the loss of a large number of accounts or a sharp reduction in the business obtained from each. Some method, then, would be needed for drawing the line between changes readily attributable to ordinary erratic variations in sales volume and changes that appear too large to be accounted for as erratic variation.

The simple system suggested here could be elaborated to allow for seasonal variations or changes in the business cycle. In a period of seasonal increase in sales, for example, perhaps no salesmen would show a decrease but many might show increases that were relatively small, and these salesmen might be checked. Considerable skill in the interpretation of evidence would be needed in order to set up a simple system of standards that would serve the purpose of indicating the need to consider change.

There are other alarm systems that indicate the need for considering the advisability of change. Survey data provided by panel studies or local newspapers are frequently useful. Other means, too, are used. We shall not discuss all these systems but shall rest content with having pointed out an ideal system and the reasons for its frequent impracticability and a feasible system with its attendant flaws. The usual function of the research personnel is the implementation of these alarm systems rather than correcting the ills that they reveal.

2. Recognizing problems concerning major policies. The scrutiny of recurring expenditures in the ways suggested above is often helpful in suggesting the need for changes in major policies, but such scrutiny frequently is not enough. A sales executive can see by looking at routine data that Smith has shown a sharp drop in sales and may thereby be led to investigate the situation and find a basic difficulty, but the need for many major policy decisions is not indicated by local crises of this sort. Usually such major changes are made by men of experience whose judgment is not based on any measurement of results. We cannot suggest ways of getting experience, but we can suggest means for organizing the material which men of experience should consider in forming their judgments. The next few pages indicate and illustrate a method of thinking about the factors affecting a business enterprise—factors which may indicate the need for revision of major policies.

The factors which influence the profitability of an enterprise can be divided into three groups: (1) those that pervasively affect the national or regional economy; (2) those that pervasively affect the firms within a single industry but which do not similarly affect other industries, and (3) those that affect a particular firm or firms within an industry without similarly affecting the whole industry. Data are available which cast light on all

three types of factors It is possible at fairly brief intervals to discover the position of the firm within the industry, the position of the industry within the whole economy, and the position of the whole economy with reference to seasonal, cyclical, and secular variations. This information can be extremely useful in suggesting the need for changes in major policies.

First we shall consider the methods of detecting changes in general business conditions either for the nation or for a particular region and the sorts of policy changes which should be considered. Rather than speak in abstract terms completely, we shall consider these problems from the standpoint of a sales executive for a large department store in the Middle West.

The customers and potential customers of such a store almost certainly live in a rather restricted trading area in and about the city in which the store is located It seems appropriate, therefore, for the sales executive to consider fluctuations in regional and local business conditions. The prosperity of most types of retail stores is importantly affected by, or highly correlated with, fluctuations in employment, personal income, industrial production, the volume of bank loans, and other components and indicators of business conditions. Such regional data are available from a number of sources such as the research departments of the Federal Reserve Banks, the U.S. Department of Commerce, the U.S. Bureau of the Census, the various state departments of commerce, the local chamber of commerce, U.S. Employment Service, etc. These data, to be useful in suggesting changes in policy, must indicate future business activity, for it is in the future that the changed policies would be effective.

The methods of using current data for general business forecasts are many, and arguments concerning their relative merits have often been violent. Although some economists have remarkable records for accurate forecasting, the methods by which these forecasts were achieved were not objective. Not only were relevant factors (*e g*, business expectations) evaluated subjectively, but also the joint influences on business were estimated by subjective means. The econometricians, who have made forecasts by creating mathematical models of the economy, have been rather unsuccessful forecasters. Two extremely simple objective models have been as successful as, or more successful than, the more complicated ones. The first indicates that the variables being considered will be the same in the next period as in the current one. The second indicates that changes from current levels will be of the same magnitude as changes from levels prevailing in the immediately preceding period. Whether the relatively subjective methods can be used depends upon the availability of personnel whose judgment has been tested and found adequate. On the other hand, use of objective means should not be relied on blindly merely because of the fact

that they are reached by precise mathematical means and are stated with exactitude

It is not our purpose solely to discuss methods of business forecasting, however. We also wish to see how such forecasts can be used to indicate desirable alternatives to current policies. There are no rules that show what changes should be associated with upward or downward fluctuations in business activity; if there were, business executives would need ability of a lower order. There are, however, certain general principles which can be discussed. In general, it has been found that lower levels of business activity should be associated with increased emphasis (more floor space, more sales personnel, more advertising) on lower priced goods, lessening of the volume of services, less liberal credit policy, reductions in inventories, etc. Higher levels of business activity usually should be associated with contrary policies. Marketing research can make the necessary business forecasts and suggest appropriate changes, thus allowing for the preparation and institution of plans before the obvious impact of changed business conditions is felt. A research department should make forecasts as often as the availability of data permits. In this way, problems can be anticipated and plans can be prepared to meet them before they become urgent.

The sales executive must consider the influence not only of general business conditions but also of factors affecting his particular industry. A manufacturer of buggy whips would have avoided discerning his most important problems if, in 1910, he had considered only the state of general business conditions or his firm's position within the industry. The overwhelmingly important fact for a buggy-whip manufacturer was the incipient secular decline in the demand for buggy whips. Adaptation of the firm's policies to the state of general business conditions or the adoption of the most effective product designs and brand policy would have done little to maintain sales and profits in face of the growth of the automobile industry.

Although the executive in the department store will probably not encounter any such spectacular forces as were felt in the buggy-whip industry, he must be sensitive to forces affecting the demand for the type of retailing which his store (with other department stores) can offer and the type of merchandise offered. The automobile, which destroyed the buggy-whip industry, was one factor responsible for the increase in the relative importance of department stores as distributive outlets. Cheap and rapid transportation such as is provided by mass-produced automobiles extended the trading areas for the urban centers, increasing the prosperity of department stores and hastening the decline of the general store in the rural areas.

The same automobile is now so widely owned and used that traffic congestion in downtown areas has decreased the accessibility of many department stores, diminishing to some extent the demand for their merchandise. The impact of population shifts, secular changes in tastes, changes in the distribution of incomes, etc., all affect the demand for the types of service and merchandise that department stores can offer.

Since the forces that affect a whole industry (or type of economic institution, in this example) are usually of gradually changing importance, it is essential that these forces be appraised regularly, though not necessarily frequently. Spectacular and sudden changes in population movements, incomes, tastes, means of transportation, etc., are rare. Adaptation of a company's policies to such changes, if it is to be prompt, must be based on the findings of vigilant research personnel who actively and systematically consider current information to discover its relevance for management. Such procedures could have indicated to the management of the department store in the Middle Western city the desirability of making provision for customers' parking and perhaps for the opening of suburban branches. The regular analysis of secondary data of the kind that we have been discussing should suggest hypotheses regarding changes in major policies.

We have sketched suggestions for the type of research activity which can indicate the need for a reconsideration of major policies in response to changes in general business conditions and in factors pervasively affecting a whole industry (or all institutions of the same type—e.g., department stores). We must now consider how research can reveal changes in the prosperity of a firm relative to other firms in the industry and how the cause of these shifts in competitive position can be determined.

Frequent and regular determination of intraindustry competitive position can be extremely useful in revealing the need for changes in major policies. A continuing "intelligence service" for competitors' activities is useful. Much information may be obtained through regular printed sources such as newspapers, trade magazines, advertising papers, and trade associations. Other information can be obtained from more informal sources, such as salesmen's reports and direct observation. Marketing surveys and panel investigations may also turn up information of importance. The sales executive of the Middle Western department store might have people who specialized in "shopping" competitive department stores to note changes in price, promotion, and other policies. When information of this kind is collected and studied on a continuing basis, it is much easier to formulate hypotheses that account plausibly for changes in competitive position within an industry.

There are systematic ways of analyzing such information. For example, when the intraindustry position of a firm has worsened, it is desirable to

look for simultaneous variations in the practices and policies of competing firms in order to find possible causes for the variation in intraindustry position. If, for example, firm A had for several years been selling from 18 to 20 per cent of the sales of the industry and suddenly found that its sales were down to 15 per cent, this would probably indicate the need for or desirability of choosing alternative marketing policies. Let us see how a marketing researcher, confronted by the information of such a decline in the share of the industry's market, could proceed to isolate the particular policy which needed changing or which gave greatest promise of restoring the company to its previous position within the industry.

Probably the first step in locating the "trouble spot" would be the securing of information about the changes of marketing practices and policies by firm A and by the competitors whose intraindustry position had improved. The major policies which would usually be considered are price policy, product policy, brand policy, advertising and promotion policy, distribution policy, service policy, etc. In the simplest case it might be found that company A had not changed any of these policies while a competing firm had made an important change in one of them. Suppose for example that A's policies had remained static while company B had effected a substantial change in product design. This information would suggest the hypothesis that the change in product design was causally related to the change in intraindustry position. This hypothesis could be checked more or less rigorously by securing, through attitude measurements, sales tests, or by other means, consumers' reactions to the new product design or to one similar to it. If such research indicated increased preference for B's design or type of design over A's design, the executive could accept the hypothesis regarding the deterioration of the firm's competitive position with greater confidence.

Unfortunately, the nature of the changes in policies by competing firms is often not so clear-cut. In some instances there are simultaneous changes in many policies by many firms. In such cases the finding of the explanation of the shift in intraindustry position is much more complex. The basic problem of inference is the same, however. It involves finding associations between variation in policy and in relative rates of sale and subjecting the hypotheses suggested by these associations to empirical tests, if possible.

Consider a more complex pattern of policy change in order to see how such analysis would proceed. Suppose, for example, that two firms, A and B, both changed their products and their prices. It is possible that both the product changes and the price changes resulted in deterioration of the relative position of firm A. Or possibly the changes together produced a joint effect to the detriment of company A. Evidence would be needed

for testing these various hypotheses unless decisions had to be made in the absence of evidence. When more than two major policies have changed, the problem again becomes more complex, but again the principle guiding solution is the same. The explanation for the variation in intraindustry position of one firm is sought by finding associated variation in the marketing policies of the firm in question and its competitors and in changes in their intraindustry position.

5.2 The Problem of Devising Alternatives to Established Policy

Once the area in which changes should be made has been located, there remains the problem of finding alternatives to established policy or practice. Sometimes the process which identified the policy that needs changing also indicates in a general way the sort of change which should be made. For example, when the cause of a worsening in intraindustry position is indicated by finding the change in competitors' policy which is associated with the worsening, the sort of change which is suggested is obvious. Presumably, a change in the factor that caused the deterioration would cause an improvement if it were adopted by the firm which was losing ground.

There are many other sources of suggestions for the kind of change that would be most profitable. One source of obvious importance is the opinion of marketing personnel about the reactions and needs of buyers. Salesmen usually have some useful impressions regarding the likes and dislikes of consumers, the uses to which they put products, their preferences, etc. Although these impressions from different salesmen are seldom systematically secured, consolidated, or analyzed, they can suggest types of changes in policy that might be useful. A certain amount of information from personnel relatively close to consumers (having opportunities for direct observation and conversation) will almost inevitably reach the attention of marketing personnel relatively remote from consumers. Thus ideas about possible changes in policy reach executives having the responsibility of forming policy. A second source of suggestions is the personal experience of these executives. A third source is the letters and other communications voluntarily submitted by consumers. The usefulness of suggestions by employees has led many managements to provide suggestion boxes. The usefulness of suggestions from consumers has led many managements to seize every opportunity to elicit and take advantage of voluntary comments from consumers. A fourth source of suggestions for proposed changes is the experience of specialists familiar with the technical properties and deficiencies of the product being sold. Finally, formal marketing research is a valuable source of suggestions. Market studies of consumers, their attitudes, and their behavior frequently suggest changes in products that will gain for them greater consumer acceptance. Marketing research

is useful in testing the hypotheses provided by the five sources of suggestions concerning changes in marketing policy

5.3 The Use of Research to Solve Marketing Problems

So far in this chapter we have discussed the ways in which problems of two sorts can be recognized. Once problems are recognized, there must be some decision as to how the solution shall be reached. Resources devoted to research should be economized like any other costly factor of production. Clearly it would be uneconomical to use research to gather and interpret evidence relating to all decisions that must be made. The consequences of such hyperobjectivity are exemplified by the story of the cautious chemist who refused to proceed on the basis of reasonable assumption and required objective evidence at every step in his investigations. Whenever he used water, for example, he refused to proceed until he had tested the properties of the liquid alleged to be water in order to see if it in fact was H_2O . The effort required for the acquisition of this objective evidence was very costly, since it involved the use of time and resources which could have had more productive alternative uses.

A business executive who demanded numerical evidence to support every step in his reasoning in arriving at a decision would be in much the same position as the compulsive chemist. Certainly he would make few decisions, and at best he would use up a great deal of time of the research department in his company in relatively sterile investigations.

It is difficult to state in general abstract terms the principles that should be used in deciding whether to use research or other means in gathering information. Obviously, resources should not be devoted to research when the information that is sought probably cannot be obtained with accuracy sufficient to make it useful. In Part IV of this book, dealing with the problems of obtaining information from given individuals, we indicate in a general way the kinds of information which it is probably foolish to attempt to obtain. Money would probably be wasted if it were used to get people to report on their luncheon menus for the previous month or on their motivation for many types of behavior. Also, money is wasted in an attempt to get information by a specified deadline when the acquisition of accurate information would probably require a substantially longer period than that allowed. Thus many problems amenable to solution by marketing research under ideal circumstances are not amenable to solution under the conditions imposed by the business environment.

In the second place, marketing research, usually being more costly than others means of getting information, should be used only when there is reason to doubt the accuracy of information provided by these other means and when the gain in revenue associated with the information exceeds the

cost of obtaining it. The idea just stated cannot be quarreled with in principle, but there would be some justice in the assertion that it is almost useless as a guide to action. The difficulty is that one can almost never know when information available through other means is sufficiently accurate. Many assumptions about consumers' attitudes and attributes which have for many years been unquestioningly accepted have been shown by careful research to be widely in error. Skepticism about prevailing beliefs is frequently profitable and leads to the conception of new ideas whose validity can be tested by research. A too-universal skepticism, on the other hand, leads to the same sort of wastefulness that characterized the work of the compulsive chemist. And these two principles are all that we can say at this point on the subject of the choice between research and other means of getting information.

PART III

SAMPLING

CHAPTER 6

AN INTRODUCTION TO SAMPLING

6.1 The Place of Sampling in Marketing Research

Typically in marketing research the researcher is interested in a group of people such as consumers or dealers about whom he wishes to obtain information. Ultimately all information useful for the purposes of marketing research is obtained by observing these people or by communicating (talking, corresponding) with them. Each consumer, for example, can be considered to be a source of potentially valuable information which can be obtained either by observing what he does or by asking him about what he has done, what his attitudes, preferences, or intentions are, etc. When the number of consumers is small, as is the case for companies selling industrial machinery to large manufacturing firms, it is possible and often profitable to tap all the available sources of information, it is possible to have detailed information on *all* one's customers or potential customers. The same type of situation would prevail for a company that wished to obtain information about the distributors through which its product is sold. For most industrial products and for most problems connected with the sale of consumer goods, however, the number of these sources of information is so great that the utilization of all of them would be prohibitively costly and time-consuming.

In these latter cases, therefore, there arises for marketing researchers the very important problem of selecting the sources of information which will be used. The producer of tooth paste, for example, who wished to ascertain consumers' reactions to a proposed change in the flavor and consistency of his product would find it too costly to consult each of the thousands or hundreds of thousands of consumers of his product. The problem of deciding to whom to talk or whom to observe is the problem of specifying the sources of one's information—or more briefly and less pretentiously, the problem of sampling—and sampling is the subject of Part III of this book.

The word "sample" is probably already familiar to almost everyone, but just to make sure that the following discussion is understandable, a word about the nature of samples will be injected at this point. The investigator of any problem is inevitably concerned about some characteristic of some group (the group need not be a group of persons—it could be horses, telephone poles, stores, rivets, or shirts). The tooth-paste manufacturer

of the preceding paragraph, for example, was probably interested in the attitude toward the new flavor of his tooth paste of all persons in the United States who might conceivably buy the tooth paste. If these attitudes were completely known, the "whole truth" would be known as far as this particular project is concerned.

All potential buyers of tooth paste in the United States—from whom the information on attitudes might be obtained—can be considered to be a "universe" or "population" (These two terms are interchangeable.) More generally, any complete group in which an investigator is interested is called a universe or population.

Sometimes, as in the decennial count of the population of the United States, a complete enumeration or "census" is taken. Typically, however, it is impossible to examine the entire population, the investigator must be satisfied with something less. For the tooth-paste manufacturer this something less would consist of the knowledge of some fraction of the total number of persons whose preferences were of interest. This fraction of the total number of individuals would be called a "sample." In general, a sample is a part of a universe, regardless of how that part is chosen. Since something less than the "whole truth" is known when samples are the source of information, the possibility of error is introduced. The sampling problem consists in selecting the sample in such a way that this error is controlled to make it as small as possible for a given dollar expenditure.

Sampling can sometimes be avoided and a census taken. It is interesting to note, however, that the sampling method is extremely pervasive. The two most important and obvious reasons for its pervasiveness are the savings in cost and in time. It should also be realized, however, that sampling assists in improving the reliability of the process of communication or observation. It is much cheaper and easier to select, train, and supervise 50 field interviewers who conduct 3000 interviews than it is to select, train, and supervise 150,000 interviewers who conduct a complete census of a human population. Thus, although sampling introduces "sampling error" which is absent when a census is taken, the *total* error in the research project—that is, errors arising from failure to achieve perfect communication, errors in recording, errors arising in tabulating, etc., as well as error arising from sampling—may actually be less when samples are used. For example, the relatively careful training and supervision of field-workers which the process of sampling makes possible may so reduce errors in the information secured from each respondent—that is, "nonsampling error"—that the introduction of sampling error is more than offset. One very interesting example of the superior results which can be obtained from the sampling process has been reported by P. C. Mahalano-

bis.¹ In this example, a sample costing much less than an attempted census of the same population was much more accurate.² The reason was simply that the census was such a staggering job that its results were very inaccurate. These three reasons, then, savings of time, savings of money, and the possible reduction of nonsampling errors, provide the rational basis for the use of sampling

6.2 The Nature of Sampling Theory

Most people would probably accept the assertion that a sample can provide useful information about the population from which it is drawn. There have been apparent failures of sampling methods such as the 1948 election predictions, where rightly or wrongly some of the blame has been popularly fixed on sampling, but the results of sampling surveys have been surprisingly widely accepted. In part, this acceptance reflects a tendency to believe "figures" without realizing clearly that they have been obtained from samples, but it also stems from the respectability which has come to be associated with sampling. Nonetheless, sampling theory and common sense are not always in accord, and when one is personally affected adversely by the results of a sampling survey, it is easy to find strong intuitive reasons for disbelieving the survey. For example, the point system for establishing priority of discharge from the army after World War II was based (democratically) on a sampling of soldier opinion. Yet one of the favorite comments of those who felt themselves unfairly treated by the point system was, "I'd like to find one of the guys they talked to." Fred Allen has turned skepticism into humor in his comments on Hooperatings, a sampling measure of the audience of radio programs: "The 'Hooperating' is a so-called service that allegedly tells you approximately how many listeners the average radio program theoretically has. It's like taking a bite of one roll and telling you how many poppy seeds there are in the country."³ Allen's jibes came at a time when these ratings indicated that his audience was decreasing.

Suspicion, disbelief, or uneasiness about sampling is also encountered

¹ P. C. Mahalanobis, "A Sample Survey of the Acreage under Jute in Bengal," *Sankhya*, Vol. 4, pp. 511-530, 1939.

² The reader may wonder why, in view of the advantages of sampling, the entire population of the United States is enumerated completely every ten years by the U. S. Bureau of the Census. Aside from the fact that the Constitution requires that this be done, perhaps the most important reason is that information is required for very small groups of the population—such as small towns, small areas of cities, etc.—as well as the whole country. Even so, however, about half the questions on the 1950 census were asked only on a sample basis.

³ Quoted in *Newsweek*, Jan. 17, 1949, p. 48.

among business executives. In part this feeling arises from findings which contradict well-established beliefs. But this is not all: people responsible for important decisions based on the information obtained from a sampling survey are not always satisfied by glib reassurances that the sample is "broadly representative" or that it provides a "miniature cross section" of the population.

Even though he may seldom be called upon for technical justifications, the marketing research specialist should not accept sampling on faith. Without some understanding of the theory underlying sampling, the marketing researcher is helpless to use comfortably, defend honestly, or criticize intelligently the results of sampling studies. Real expertness in sampling requires much more statistical training than most marketing research people find it feasible to acquire, but fortunately a basic understanding of sampling is not difficult. This Part of the book provides a beginning.¹

Unless the reader is already familiar with the logical structure of sampling theory, it might be helpful at this point for him to ask himself what reasons he can advance to explain why samples can be expected to "work." Aside from a possible citation of instances in which sampling *has* worked, he will probably have trouble in thinking of convincing explanations. The explanations are relatively simple, yet they probably would not occur to anyone without some knowledge of statistical theory.

Consider an example that dramatizes this point. In the depression years of the 1930's many estimates were made of the extent of unemployment, and many different methods underlay these estimates. The differences were huge—in 1932 there was a difference of several million people between the highest and lowest estimate. Furthermore, there was no way of being reasonably sure which of these estimates came the closest to the truth. Today, by contrast, there is one widely accepted estimate of employment and unemployment. This estimate is based on a sample of approximately 25,000 to 30,000 families taken monthly by the U. S. Bureau of the Census, which gives results within tolerances for sampling error that can be (and have been) computed and published. Now, suppose in a particular month the unemployment figure is reported by the Census Bureau as 3,500,000. Someone claims this is wrong—that there are really 7,000,000 unemployed. What argument can be advanced—other than arguments of authority—for the contention that 3,500,000 is likely to be much closer than 7,000,000 to the correct figure?²

¹ The readings suggested in the bibliography take the reader much further into the theory and complexities of sampling.

² There have been criticisms of the census estimates. These criticisms center on the way in which one *defines* employment and unemployment. This is not strictly a statistical problem, the *sampling* methods used are apparently accepted by most critics. See *Fortune*, September, 1949, "Those Unemployment Figures," p. 76.

Fortunately, there are two distinct justifications for sampling and the results based on sampling. The first is purely empirical. People have tried sampling experiments with known populations or universes, such as red and white beads in a box, and have found that the results conform with what would be expected from the abstract considerations of sampling theory. Suppose there were 500 red beads and 500 white beads in the box. If the beads were well mixed and 100 beads taken from the top of the box, we usually would find—as we would expect from sampling theory—between 40 and 60 red beads. Perhaps a good way to acquire a “feel” for sampling is to play a relatively simple chance game, such as matching pennies, and to observe the connection between what actually happens and what one would expect from sampling theory. One of the most important lessons of such a game—or of drawing beads from a box—is that untypical results occasionally occur. These “aberrations” might be written off by those ignorant of statistics as good (or bad) luck which violates the law of averages. Statistical theory makes possible, however, not only the prediction of what usually will be found but also the anticipation of such unusual occurrences and the making of precise statements about their frequency (though not, unfortunately, their exact *time*) of occurrence.

The second justification of sampling is a logical one. While a rigorous explanation would necessarily be mathematical, the main steps can be described in intuitive terms with only occasional recourse to symbols. The following chapters are devoted to such a logical development and to some of the problems which are involved in the application of such statistical theory to marketing research. A few words about the basic organization of these chapters may be helpful.

As has been said, whenever samples are used, there is the possibility of error. If 100 beads were drawn from the box previously referred to (which contains 500 red beads and 500 white beads), it would be a relatively unusual occurrence to obtain *exactly* 50 red and 50 white beads. It is even possible, though exceedingly improbable, that all 100 beads would be red. Hence samples typically do not mirror *perfectly* the characteristics of the population from which they have been drawn. It was said earlier, however, that ordinarily between 40 and 60 red beads will be found in any sample of 100. This assertion was made on the basis of a knowledge of statistical theory. More completely and precisely, the following assertion would have been made: The number of red beads in a sample of 100 beads from this box will fall somewhere in the range of 40 to 60. While this prediction may turn out to be incorrect for any particular sample, the risk of error is calculated (by methods not shown here) to be approximately 5 per cent. That is, in 5 per cent of all samples for which predictions of this kind are made, the predictions will turn out to be wrong. The inter-

esting part of this assertion is that a numerical value (probability) has been attached to the risk of error—the risk that the method which enables the prediction to be made will lead to the wrong answer. (In subsequent chapters something of the reasoning that underlies assertions of this kind is shown.) Whenever it is possible to so evaluate the risk of error, the samples obtained are called “probability samples.” Roughly speaking, probability samples are samples for which the risks of error are measurable. Other types of samples (nonprobability samples) are of great importance and are, in fact, even more widely used than probability samples in marketing research. But samples of both types can probably best be understood if probability samples are explained first. We proceed on that basis; the next two chapters refer only to probability samples. The subsequent chapter deals with nonprobability samples. The final chapter of Part III compares the usefulness of these basic sample types for marketing research.

CHAPTER 7

SOME FUNDAMENTALS OF SAMPLING THEORY

7.1 Introduction

In this chapter we are concerned with sketching the outlines of sampling theory. Our objective in doing this is neither to treat sampling theory exhaustively nor even to plunge very deeply into it, but rather to indicate its main steps in a nontechnical way so that the practical applications to be considered in subsequent chapters can be better understood. Above all, we hope to indicate briefly how sampling theory can give confidence in inferences drawn from samples because of knowledge of the procedure by which the sample was selected rather than because the sample results conform to the preconceived views of the investigator. In the following pages, then, we consider the sequence of main definitions and ideas that are needed in order to accomplish this understanding of sampling theory.

7.2 Populations and Samples

The population or universe may be thought of as the "whole truth," or all possible measurements relevant to some particular question. In actual practice the whole truth is usually unknown. If it were known, neither samples nor marketing research would be needed. Two possible universes are (1) all housewives in the United States and (2) all Dodger fans in Brooklyn.

These examples, however, are slightly misleading. In these cases the universe or population has been "personalized" by thinking of it as consisting of people, families, institutions, or the like. Often it is preferable to think of *measurements* made on people, institutions, etc., as being the universe. Thus, for example, one universe would consist of the most recent purchases of flour by housewives in the United States; another, of the ages of these housewives; another of their attitudes toward radio commercials. In each case the same people—American housewives—would be involved. In each case, however, the population would be different. It may seem that this distinction between people and measurements made on them is artificial. For ordinary speech it is artificial and throughout our discussion we shall use such short cuts as "the universe of housewives." Nonetheless, the possibility of a distinction should be seen at this point. For by seeing the possibility of making such a distinction, it is easy to under-

stand a problem of great practical importance in sampling. That is, the sampling methods best suited for studying flour purchases of a certain group of people may be different from those best suited for studying attitudes of this same group toward radio commercials or for finding out about some other characteristic of the persons involved.

A mean, a proportion, a median, a standard deviation, or any other summary measure of a characteristic of members of a population is called a "parameter." (The analogous summary measure for a sample is a "statistic.") For example, 0.04, or 4 per cent, of the labor force might be unemployed; the median income of families in Chicago might be \$4500; or the mean age of people attending movies might be thirteen years. A useful convention in statistical writing is to use Greek letters to symbolize parameters. Two of the most frequently used parameters are the mean of a population, represented by the Greek letter mu or μ , and the standard deviation of a population,¹ designated by the Greek letter sigma or σ . Many statisticians use this apparently esoteric symbolism to avoid a serious confusion, which is explained in a moment.

A sample is any part of a population. If 100 beads are drawn from the box containing 500 red beads and 500 white beads, the 100 beads comprise a sample from the population of 1000 beads. If 100 housewives are interviewed as to their attitudes toward radio commercials, these 100 housewives would be a sample from the population of all housewives. The definition of "sample" does not specify that the sample be drawn in any particular manner. An understanding of sampling theory will, however, be facilitated by thinking of the many possible different ways of selecting samples as falling into one of two major categories, probability and non-probability samples. In the sketch of sampling theory in this chapter we consider only the most familiar type of probability sample, the simple random sample. In Chapter 8 we introduce more complicated kinds of probability samples, and in Chapter 10 we discuss nonprobability samples.

Just as a mean, a proportion, a median, or a standard deviation of a

¹ The following brief comments are added for those who need review on the slightly elusive concept of "standard deviation." The standard deviation is one of a large number of summary measures known as "measures of variation" or "measures of dispersion," since they measure the variability or dispersion encountered in a set of data. One other common measure of variation, for example, is the range, which is simply the difference between the largest and the smallest observations in a group of observations. The standard deviation is defined as follows: (1) Obtain the deviation of each observation in a group of observations from the mean of all the observations in the group. (2) Square each deviation, and add the results. (3) Divide by the number of observations. (4) Take the square root of what remains. (Actually, for computational purposes, there are convenient short cuts which can be found in any elementary statistics book.)

population is called a "parameter," so a mean, a proportion, a median, or a standard deviation computed from a sample is called a "statistic." Sample statistics are used in drawing inferences about population parameters. In drawing such inferences, of course, it is necessary to remember that a sample will almost never mirror perfectly the population from which it is drawn and that statistics of a sample will almost never equal the population parameters. To emphasize this fact, many statisticians follow the practice of using Roman letters to symbolize sample statistics. The sample mean, for example, is called $\bar{x}$ (pronounced "x bar") while the sample standard deviation is called s . It is essential to distinguish $\bar{x}$ and s from μ and σ , the symbols for the mean and standard deviation of a population. We use all these symbols in the discussion of sampling theory, and any confusion between them will lead to confusion in the understanding of the basic ideas of the theory.

7.3 Probability and Randomness: Simple Random Sampling

The two closely related concepts of randomness and probability supply the foundation of the theory of statistical inference and, in particular, of the theory of sampling. Probability has a nontechnical, although imprecise, meaning which is familiar to everyone. Everyone is familiar with sentences such as, "His chances of winning are pretty small," "It's pretty likely that we'll have rain before tomorrow," or "You're probably right." In each of these examples the general idea of uncertainty is involved, and by use of words such as "chances," "likely," or "probably" this fact is acknowledged. All these terms are closely related to the nontechnical idea of probability.

Probability has a much more precise and subtle meaning in statistical inference. Without delving into the subtleties of this meaning, we shall first make a statement that conveys a part of the meaning and then modify this statement in order to convey a fuller understanding. The simplest definition of statistical probability (which we amend shortly) states that the probability of a particular event occurring under given conditions is equal to the relative frequency of its occurrence under those conditions. If 950 out of each 1000 live-born babies reach the age of one year, the probability of survival to the age of one year would be 0.95 (*i.e.*, 950 divided by 1000). If a coin seems to come up heads in about two-fifths of all tosses, the probability of a head would be said to be 0.40. Relative frequency can be as small as 0 if a particular outcome never occurs and as high as 1 if it always occurs. In most discussions it is customary to speak of an outcome with a probability of 0 as being "impossible" (since it never happens) and an outcome with a probability of 1 as being "certain" (since it always happens).

This first approach to an understanding of probability, then, equates probability with the familiar term "relative frequency." If one stopped with this definition, however, there would be trouble. For example, half the people in the United States are men, and therefore the probability that the mayor of Chicago is a man would seem to be one-half. This is of course absurd. A person is or is not a man, and the probability is certainly not one-half. People, or beads, or opinions are what they are and cannot be "probably this or probably that." Probability applies only to predictions about these immutable facts. To obtain a more adequate idea of probability, the closely related idea of randomness must be explored.

A concise, clear, and correct definition of randomness is beyond our present needs. We shall, however, attempt to get an intuitive grasp of the idea by proceeding with an example. Imagine a perfectly balanced roulette wheel with just 10 numbers, the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. Someone spins this wheel and waits for it to stop. When it stops, a pointer points at one of the 10 possible numbers. The number is recorded and the wheel is spun again. Each time the wheel is spun, the number designated by the pointer is recorded. After a series of spins, the following sequence of outcomes is obtained:

8092 9572 9931 3826 9066 0003 2184 4003 3574 3408 7331

A person would probably feel intuitively that each of these 10 digits would tend to have a relative frequency of about $1/10$ if the wheel were spun repeatedly. If, before each spin of the wheel, one were to predict that the digit 7 would turn up, he would be correct about $1/10$ of the time. And there is nothing that can be done to make the proportion of correct predictions significantly greater than 10 per cent. The reason is that while 7 will turn up about $1/10$ of the time, it is impossible to know *when* it will turn up. For the sequence of observed digits has no apparent pattern or regularity. Or, if evidence of regularity is observed in a part of the observed sequence, there is no assurance that this pattern will be of any use in increasing the number of successful predictions.

It is possible to think of the sequence of digits produced in this manner as a *sequence of random observations*. We have not really defined how this sequence was produced; we have simply described a mechanical process (spinning a roulette wheel) and assumed that this process would in fact lead to a sequence that lacks any persistent pattern in the order of digits produced.

A random sequence, then, may be thought of as a sequence of observations like that produced by some mechanical device similar to one described here. The mechanical process is not defined precisely; it must be understood intuitively. A process (mechanical or otherwise) is random if

it gives rise to a sequence of observations having no regular discernible pattern.

In view of the vagueness of the definition of a random process, it is especially desirable to have some way of knowing whether any particular process can be regarded as random. The only way to decide whether a process can be regarded as random is to judge by the appearance of the sequence of observations produced by the process. There are tests by which one can arrive at a judgment as to whether the process is random. Essentially these tests can be thought of as tests for evidence of persistence or regularity of pattern in the sequence of observations.

Long sequences of digits, similar to the short one given above, have been produced (though not with roulette wheels) and tested for randomness. These sequences are known as "tables of random digits."¹ Tables of random digits, in turn, are immensely useful in practical sampling applications, as will be seen later.

We are now in a position to complete the definition of probability: "The probability of an event is approximated by its relative frequency in a very long sequence of random observations." In the above example, we might therefore state that the probability of the outcome 7 in spinning the roulette wheel is approximated by its relative frequency in a very long sequence of such spins. We have assumed that this relative frequency would be about $1/10$.

The reason for probing this far into the definition of probability must now be explained. If, in the selection of a sample, one is able to assign a known positive (nonzero) probability of selection to every member of a population, statistical theory can be used to make rigorous inferences about the population from the information in the sample. There is no other known way in which rigorous inferences can be drawn.² When the process of selection is such that a known positive probability of selection can be assigned to every element in the population, the resulting sample is a probability sample. The only way in which a known positive probability of selection can be assured, as we have seen from our more complete definition of probability, is by a process of *random* selection, since probability is neither defined nor measured in the absence of randomness. The simplest process of random selection is one which assigns an *equal* probability on

¹ One example of such a table is given in George W. Snedecor, *Statistical Methods*, Chap. 1, Iowa State College Press, Ames, Iowa, 1946.

² In the chapter on nonprobability sampling, it will be seen that pragmatically useful inferences are drawn from samples in which known probabilities of selection are not assured. Indeed, most of current scientific knowledge comes from the results of nonprobability sampling methods. Being able to draw rigorous inferences is, however, often a great advantage.

each draw to each member of the population. Such a process of selection is called "simple random sampling," and the samples that result from such a process, "simple random samples." In this chapter we discuss only simple random samples.

The next task is to see specifically how a simple random sample would be obtained. To obtain such a sample, a table of random digits may be used. To illustrate the use of such a table we shall repeat the sequence of digits produced by the roulette wheel (which, in principle, would be capable of producing a whole book of random digits).

8092 9572 9931 3826 9066 0003 2184 4003 3574 3408 7331

This sequence of digits resembles a single line of digits in a table of random digits which, in turn, consists of several pages filled with digits in this fashion. Suppose that the particular line of digits reproduced above is one line in such a table.

The problem is to draw a random sample of charge-account customers in a large department store. This department store has 10,000 such customers. A list of these customers is made and the customers are assigned consecutive numbers as follows: 0000, 0001, 0002, 0003, . . . , 9997, 9998, 9999. The next step is to refer to a table of random digits and proceed consecutively from a starting point just as in reading a book. If the starting point had been the beginning of the line of random digits shown above, the 8092d customer would have been selected, since the four digits 8, 0, 9, and 2 were the first four encountered. The next four digits are 9, 5, 7, and 2; hence the next customer selected in the sample would have been the 9572d. One proceeds in this manner until the desired number of customers, say 225, has been obtained.¹

Before going on it is worth while to refute two fallacies about random sampling that are rather widespread. First, the idea of randomness is not the same as its everyday connotation of "haphazardness." The reason that tables of random digits are accepted is that they have been subjected to tests for randomness. By contrast, haphazard sampling processes, in which the sampler tries to be "aimless," seldom succeed in approaching true randomness.

Second, a random sample is not necessarily a representative sample or a true cross section. It has already been seen that it would be possible to

¹ For most sampling purposes, if the same number were to turn up twice, the second appearance would be ignored. Technically, this method of sampling is described as "sampling without replacement." One other possible complication is also worth mentioning since it causes some confusion. If there had been (say) 9214 customers instead of 10,000, the procedure would have been exactly the same with one exception: all numbers greater than 9213 would have been ignored if they had been encountered in the table. Hence 9572 would have been skipped.

draw by a random process 100 red beads from a box that contained 500 red beads and 500 white ones. The probability of this happening in random sampling, fortunately, would be very small. But if such an occurrence did come about in random sampling, the 100 red beads would be a random, although a most unrepresentative, sample. Whenever the term "representative" is used in practice, it is necessary to examine carefully what is meant, since, unless the characteristics of the population are known exactly, it is impossible to assure the selection of a perfectly representative sample.

7.4 Sampling Distributions

In understanding the theory of sampling, one of the most essential steps is understanding the concept of a "sampling distribution." We shall illustrate this concept by an example before defining it. Return to the sampling of charge accounts in which a sample of 225 charge accounts was chosen at random from the 10,000 charge-account customers. Suppose that one of the things (parameters) to be estimated was the annual average (mean) expenditure on furniture by charge customers.¹ (In practice, one would probably also be interested in estimating, say, the average expenditures on jewelry, men's suits, etc.) Intuition suggests, correctly, that a sensible way of doing this would be to compute the mean expenditure for furniture by the 225 charge-account customers included in the sample. That is, one would add up the total furniture expenditures by these 225 families and divide by 225. Suppose this sample mean $\bar{x}$ turns out to be \$50. The question to be answered is this: How good is the sample mean ($\bar{x}$) as an estimate of the mean of the population (μ)?

An obvious way to answer this question without further equivocation would be to examine the remaining 9775 charge accounts and compute directly the mean (μ) of the whole population. To do so, however, would increase the cost and time of finding out about the charge customers. If the accuracy of the sample estimate (\$50) could be appraised without an examination of the rest of the universe, the saving could be substantial and the potential value of sampling could be realized.

¹ Throughout the discussion of sampling theory, we shall illustrate ideas by reference to the mean. Other statistics, particularly the percentage or proportion, are also of interest in applied research. The entire discussion, however, applies equally well to the proportion since the proportion can be considered a special case of the mean. That is, to compute the proportion of people possessing a particular characteristic, one could assign the number 1 to each person who possesses the characteristic and 0 to each person who does not. The 1's and 0's could then be added up and the sum divided by the number of elements, just as in the computation of the mean. The only added information (the formula for the standard error of the sample proportion) needed by the reader in order to make the discussion apply directly to the proportion is given in the footnote on p. 100, Chap. 7.

If the population is not examined to see if the sample mean is reasonably good as an estimate of the population mean, the next best thing is to ask whether this particular way of drawing samples of 225 from a population of 10,000 charge accounts usually gives samples with means reasonably close to the population mean. More concretely, "What is the pattern of sample results to be expected if, instead of one random sample of size 225, a whole sequence of such samples had been taken?" In the present example, this question could be answered by direct trial. One could draw repeated samples and for each such sample compute the mean expenditure on furniture. For the sample actually drawn the mean was \$50. For a second, the mean might be \$57.34; for a third, \$46.67; and so on. If a large number of such samples were taken, it would be possible to make a "frequency distribution of the means."¹ This frequency distribution of the means of repeated samples of the same size which are selected in the same way (simple random sampling in this case) is an example of a "sampling distribution," since it gives the frequency distribution of values of a certain statistic (in this case, the mean) of samples drawn from a given population by a given sampling process.

Appraising the accuracy of a sample estimate by a process of repeated sampling is about as costly, and therefore about as objectionable, as a process of checking on the sample estimate by examining the whole population. Fortunately, it is usually unnecessary to determine sampling distributions by a process of repeated sampling. Beginning with the basic ideas of randomness and probability and using certain assumptions and rules from the theory of probability, which have not been mentioned, it is possible to *deduce* certain important facts about the sampling distribution of the mean (or of other statistics). Before seeing what these facts are, it is helpful to see intuitively why information about the sampling distribution of the mean might be helpful in answering the original question as to how good an estimate of the population mean is provided by the sample mean.

If very different means tended to result from the different samples, one would be justified in having only a low order of confidence in the accuracy of an estimate of the population mean based on a single sample mean. A second sample mean taken from the same population by the same process might give a very much lower, or higher, sample mean. If, on the other hand, about the same means tended to turn up for a series of samples, one would be justified (except for possible reservations to be introduced shortly)

¹ It is assumed that the reader is reasonably familiar with the term "frequency distribution," since this is one of the most important topics covered in elementary courses in descriptive statistics. A brief comment, however, may be helpful. In forming a frequency distribution, it is necessary only to classify observations into different categories according to their magnitude. A list of these categories, together with the number of observations in each, is a frequency distribution.

in having a higher order of confidence in an estimate of the population mean based on a *single* sample. Knowledge about the variability in a sample statistic that can be expected for different samples from the same population, then, is useful in appraising the sample mean as a basis for inference about the population mean; knowledge of the sampling distribution of the mean tells, among other things, what variability to expect. This sampling variability is often called "sampling error," although such variability is not the only, or necessarily the most important, source of error in inferring about populations from the information provided by samples. (There is also, as the next paragraph will show, a possibility of bias in the sampling process. Moreover, as mentioned briefly in the last chapter, there may be errors that would be present even in a census. Errors in the charge-account records would be one example of these nonsampling errors.)

It is important to know not only whether the results from different samples are consistent¹ (possessed of small variability), but also whether these results are consistently "right" or consistently "wrong." Samples could lead to incorrect inferences even if, to make an extreme assumption, the statistics from repeated samples were all the same. This would obviously be true if all sample means were about one-half (or any other fraction) of the population mean. The difference between the population mean and the average value of the means of all samples is called "bias." This bias is equal to the difference between the population mean and the mean of the sampling distribution of the mean.

Knowledge of the presence or absence of bias and knowledge of the sampling variability, both of which can be determined in probability sampling, provide a sound basis for appraising the accuracy of a sampling process. In the following section we discuss the relation of the mean of the sampling distribution of the mean to the population mean to answer the question of the presence or absence of bias; and in the section after that we discuss the standard deviation of the sampling distribution of the mean as a measure of the sampling variability of the sample mean.

1. The mean of the sampling distribution of the mean: unbiased estimates. First, it is important to see whether the sample mean will, *in the long run*, be a good or a poor estimate, regardless of its variability or constancy from one sample to the next. That is, does the sampling process provide an unbiased estimate? To answer this question for the charge-account example it is necessary to know the average of the sample means of a very long sequence of random samples of size 225. If the average tends to

¹ Here the word "consistent" is not to be identified with the use of "consistent" in advanced discussions of desirable characteristics of statistics, such as in R. A. Fisher, *Statistical Methods for Research Workers*, Chap. 1, Hafner Publishing Co., New York, 1948.

equal the mean of the entire population (*i.e.*, if the average of a long sequence of $\bar{x}$'s tends to equal μ), the sampling process is unbiased, since the average of many sampling estimates tends to equal the thing being estimated.

It is possible to show deductively, using the notions of randomness and probability plus simple axioms of probability theory which have not been discussed here, that the mean of the sampling distribution of the mean equals the mean of the population from which the sample is drawn.

The meaning of this last sentence must be made clear. It is *not* possible to deduce an exact numerical value of the mean of the sampling distribution of the mean. Moreover, the mean of any one particular sample may be quite different from the mean of the population. What *is* stated is that the average of the means of a long sequence of random samples tends to equal the mean of the population, whatever that mean may be. In symbolic terms, the mean of the $\bar{x}$'s tends to equal μ . Thus with random sampling the sampling process is unbiased since there is no systematic tendency toward error: if the sampling process were repeated, the average of the sample estimates would tend to equal the true value for the population. If this happy state of affairs did not prevail, the mean of the means of the samples would tend to equal some value other than the mean of the population. The difference between the mean of the sample means and the mean of the population would be the bias inherent in the sampling process.

Means computed from simple random samples are unbiased estimates. For other probability sample designs, sample means are usually either unbiased or are known to possess a bias of unimportant size. For nonprobability samples, as will be seen in more detail later, estimates are typically biased but the bias is of unknown magnitude—it may be huge, small, or possibly zero. Since nonprobability sampling does not assign known probabilities of selection to each element of the population, nothing is known theoretically about the bias inherent in the sampling process.

It is possible to give an intuitive argument showing why the sample mean is an unbiased estimate in simple random sampling. First, taking the mean of the means of a series of samples of the same size is equivalent to considering the series of samples as one big sample and computing the mean for this grand sample. In either way the answer would be the same. Now, in this grand sample made up of all the individual samples, each element of the population would be represented with a relative frequency approximately equal to its probability of appearing in any one sample. Since the probability of selection is the same for every element of the population, the relative frequencies of elements will be approximately equal.

Then the result obtained from calculating the mean of the sampling distribution (or the mean of the one large sample) would be approximately the same as the result obtained from calculating the mean of the original population.

2. The standard deviation of the sampling distribution of the mean: the standard error of the mean. Attention will now be focused on the variability of the results of the sampling process. As has already been seen, confidence in any inference based on a sample estimate depends not only on the amount of bias present in the sampling process but also on the variability of the results of the process from one sample to another. If this variation is large, one could have only a low order of confidence in any inference, even though no bias were present. If, on the other hand, the variation from sample to sample is small, one would be justified in having a relatively higher order of confidence in an inference based on the result of a single sample.

Of course, it would ordinarily be impracticable to measure this variability from sample to sample by actually drawing repeated samples from a given population. Rather, the amount of variability to be expected from such a process may be obtained from a knowledge of the dispersion of the theoretical sampling distribution for the statistic of interest (the mean, in our discussion). Various measures of dispersion are discussed in all courses in elementary descriptive statistics. The measure that is by far the most useful for statistical inference and the sort of problems that are considered here is the standard deviation.¹

The standard deviation of the sampling distribution of the mean is obtained by treating the collection of sample means as individual observations and computing their standard deviation just as for any other group of observations. The standard deviation of the sampling distribution of the mean is usually called by a much shorter term, "the standard error of the mean." The standard error of the mean, then, is simply the standard deviation of the sample means computed for a long series of means of samples of the same size.

In the consideration of the mean of the sampling distribution of the mean, it was seen that there was a direct, simple relation linking the sampling distribution with the population from which the samples were drawn. The mean of the sampling distribution was equal to the mean of the population.

There is also a relation between the standard deviation of the sampling distribution (standard error of the mean) and the standard deviation of the population from which the samples making up the sampling distribution

¹ We shall soon show why the standard deviation is the most useful measure of dispersion for the problem here.

were obtained. This relation can be stated as follows:

$$\text{The standard error of the mean} = \frac{\sigma}{\sqrt{n}}$$

where n is the size of the sample (selected by simple random sampling), and σ is the standard deviation of the population from which the sample was drawn.¹

This relationship can also be deduced from the basic rules of probability theory. No attempt will be made to give any intuitive argument for its correctness: it is one of the many remarkable results for which the practicing statistician is obligated to the mathematician. It is instructive, however, to see how closely this result agrees with what would be expected solely on the basis of intuition.

In the first place, as the above equation shows, the standard error of the mean becomes smaller as the sample size increases.² This result agrees perfectly with intuition: if samples are any good at all, big ones should be better than little ones. The above equation shows us that for simple random sampling, at least, common sense has been a sound guide.

In the second place, the more variable the population (the larger σ is) to which sampling has been applied, the larger the standard error of the mean (holding sample size constant). A moment's reflection on extreme cases shows that this conclusion is also plausible. If all charge account customers spent exactly the same amount for furniture (say \$50 per year), exactly the same mean would be obtained in every sample drawn from this population. The standard error of the mean, then, would be 0. At the other extreme, if purchases of furniture ranged from 0 to \$10,000, a higher degree of variability from one sample mean to the next would be expected. The inclusion or exclusion of the one customer who bought \$10,000 worth of furniture would make a huge difference in the sample mean of each sample from which this customer was included or excluded. The equation and intuition coincide.

There is one respect, however, in which the equation stated above does not coincide with intuition. It includes nothing whatever about the effect on the standard error of the mean of the *size* (number of observations) of the population from which the samples are drawn. Intuitively one would

¹ The analogous formula for the sample proportion is $\sqrt{\frac{P(1-P)}{n}}$ where P is the proportion of elements in the universe possessing a given characteristic and n is the size of the sample. In the next chapter the discussion of the mean can be translated into a discussion of the proportion by simple modifications of this formula.

² More precisely, the standard error of the mean is inversely proportional to the square root of the sample size.

probably feel that the bigger the population (even holding constant the variability, σ , of the population), the larger the standard error of the mean. If the formula for the standard error of the mean is right and intuition is wrong, the size of the population (10,000 in the charge-account example) has nothing whatever to do with the standard error of the mean. Actually, intuition is correct and the equation is not quite correct, except in the case of sampling from indefinitely large populations.

In practice, however, the adjustment that should be made to allow for the size of the population is usually of such insignificance that it can be safely ignored.¹ Only when the sample is a large fraction of the population is the failure to use the correction for the size of the population of much importance. In most applications, statisticians ignore the correction unless the sample is at least 10 per cent of the population from which it is drawn.* In the example (of the sample of charge-account customers), the sample was 225 from a population of 10,000. In this example it would, therefore, be unnecessary to take account of the correction factor.² In some instances, however, particularly when sampling from relatively small populations, it turns out to be desirable in practice to use this correction.

One fact about the standard error of the mean needs to be stressed strongly at this point because of its importance later. The formula $\sigma/\sqrt{n}$ applies *only to simple random sampling*. Analogous formulas have been worked out for other probability sampling designs. Usually these formulas are much more complicated than the formula for simple random sampling. Of immediate practical importance to researchers is the fact that in sampling with a given sample size from a given population, the standard error of the mean for many sample designs (or arrangements) is larger than for simple random samples, while for other designs it may be smaller.

When nonprobability sampling methods are used, no formulas whatever are available. Just as there is no way to ascertain whether or not nonprobability sampling processes are biased, so it is impossible to state any formula which represents the amount of variability to be anticipated from one sample to the next. The only way to judge this variability would be to perform repeated sampling, which, of course, is highly impracticable.

1. A Note on "Efficiency." The term "efficient" is frequently used in

¹ The adjustment, or correction term, is $\sqrt{\frac{N-n}{N-1}}$ where N is the number of elements in the population and n is the number of elements in the sample. This term is often called the "finite population multiplier." Its use is correct for random sampling without replacement from a finite population.

* The small error introduced by omission of the correction factor is a conservative error; that is, it causes an overstatement rather than an understatement of the standard error of the mean.

discussions of sampling. It is now possible to see what is meant by this expression. One sample design¹ is said to be statistically more efficient than another if, for a given sample size, its standard error of the mean is smaller. Different sample designs yield estimates with different standard errors. Simple random samples of size 225 from the population of 10,000 charge-account customers would have one standard error; other probability sampling designs of the same size (225) would have differing standard errors and would, therefore, be more or less efficient than simple random sampling.

The term "efficiency" is also used in another sense. When the size of standard errors of the mean of different sample designs is compared for samples of the same size, the comparison yields a measure of relative *statistical* efficiency. When, however, the size of standard errors of the mean of different sample designs requiring the same dollar expenditure is compared, the comparison yields a measure of relative *economic* efficiency.² In applied research in the business world it is, of course, economic efficiency which should be maximized, and the word "efficiency" is often used by sampling statisticians to mean only economic efficiency. That is, the efficiency of a sample design is evaluated by statisticians in terms of reliability of results per dollar of cost.

3. The "form" of the sampling distribution of the mean. There is still one more thing that can be deduced about the sampling distribution of the mean: its form. So far it has been seen that the mean and the standard deviation of the sampling distribution can be deduced from a knowledge of the mean and the standard deviation of the population from which the samples are drawn. The mean and the standard deviation are in themselves sufficient to describe one type of frequency distribution—the "normal" distribution.³ That is to say, if the mean and the standard deviation of a normal distribution are known, the distribution is known completely; it can be graphed, or a table can be prepared showing the proportion of observations falling between any particular set of values. Thus if one knew that the sampling distribution of the mean was normal, one would then be quite completely informed about its general characteristics (although not, of course, about the particular numerical values of the mean

¹ More completely, "one sampling design *and* method of estimation." There are different methods of making estimates, even for the same sampling design, and these estimation procedures do not all possess the same efficiency. This point is discussed briefly in section 9.2. "Statistical efficiency" is now very frequently used in technical discussions in reference to the efficiency of different procedures of estimation.

² For a further discussion, see *How Big Should the Cluster Be?* under Section 8.2.

³ A more detailed explanation of this fact and the general description of the normal distribution may be found in any textbook on elementary descriptive statistics.

and the standard deviation), since the mean and the standard deviation of the sampling distribution have definite relations with the mean and the standard deviation of the population from which the samples have been drawn. In the preceding pages these relations have been shown.

The mean and the standard deviation of the sampling distribution of the sample mean are closely related to the mean and the standard deviation of the population from which the samples are drawn. It might therefore be expected intuitively that the form of the sampling distribution of the mean would be closely related to the form of the distribution of the population from which the sample was drawn. If the distribution of the population were extremely skewed,¹ for example, it might be expected that the sampling distribution of the mean would also be skewed. Fortunately, however, such a surmise would be largely wrong.

The central limit theorem (which is one of the most important deductive steps in the development of sampling theory) states that the sampling distribution of the mean tends to approximate the normal distribution more and more closely as the sample size is increased, almost regardless of the form of the population from which the samples are drawn.² Investigations have shown that for a large number of populations, the approximation becomes very close for surprisingly small sample sizes, often on the order of 10 or less.

The full importance of the central-limit theorem—the fact that the sampling distribution of the mean tends to normality regardless of the form of the population—cannot be fully appreciated without a more leisurely exploration into the theory of sampling. Its importance, however, is shown by the results of the next chapters of the book. For the knowledge of the form of the sampling distribution plus the knowledge of the mean and the standard deviation of the sampling distribution are all that is necessary for the discussion of confidence-interval theory for estimates of the mean.

7.5 Confidence Intervals

Now the threads of the last few pages can be pulled together. The inquiry began by making the suggestion that the mean of a sample be used as an estimate of the mean of the population. In particular, the objective was to find out how good the mean of a random sample of 225 would be as an estimate of the mean expenditure on furniture of 10,000 charge-account customers. For a particular sample the estimate is either good or bad, and

¹ The term "skewed" means that the frequency curve, or graphical representation of the frequency distribution, is lopsided rather than symmetrical about the mean. The normal distribution is an example of a symmetrical distribution; the normal distribution, then, is *not* skewed.

² The central-limit theorem holds true for all finite populations, that is, all populations in which the elements can be counted.

one can't find out which without knowing the very thing for which the sample is being used to provide an estimate, namely, the mean of the population. Thus the original question was changed so as to ask if this way of selecting a sample gives good results: How good is the sample *process*? This question led to a consideration of the results the sampling process would give if it were repeatedly performed, that is, if a very large number of samples were taken rather than only one. The inquiry turned, therefore, to a consideration of the sampling distribution. It was recognized, of course, that in practice one almost always has only a single sample, but it was seen that certain characteristics of the sampling distribution would tell, indirectly, how much confidence could justifiably be placed in the sample estimate.

It was then pointed out that the sample mean is an unbiased estimate of the mean of the population. This did not signify that the mean of any particular sample was equal to the mean of the population, but it did signify that the *sampling process*, if repeated many times, would yield samples whose average means approximated the population mean. Even though this information does not tell whether a *particular* sample is close or far from the population mean, it gives us knowledge of the presence or absence of any *systematic* tendency toward error. If one were to say that the "best estimate" of the mean of the population (in the example of the charge accounts) is \$50, the statement would be correct.¹ More could be done, however, in appraising the accuracy of the sample estimate. It would be a very unusual sample that produced a mean exactly equal to the mean of the population from which it was drawn. It is therefore important to appraise the error that must be allowed for in interpreting this best estimate.

The concept that makes this appraisal possible is the "confidence interval." The argument by which the confidence interval is derived is not difficult, and we shall present its outlines here. First we define a new expression which is a simple combination of terms which are already familiar. We shall call this term k and assign it the following symbolic definition:

$$k = [\bar{x} - \mu]/[\sigma/\sqrt{n}].$$

In words, this definition can be stated as follows: "For each sample, subtract the mean of the population (that is, μ) from the mean of the sample (that is, $\bar{x}$) and divide this difference by the standard error of the mean (that is, $\sigma/\sqrt{n}$)." The definition of k is arbitrary but useful.²

¹ The characteristic of being unbiased is not the only requirement for a "best estimate," as we show in the next chapter, but it is an important one.

² The symbol k appears in many different forms in elementary statistic books. The terms "normal deviate" or "unit normal deviate" are sometimes used. Other letters used to designate the same concept are z , u and x/σ .

In practice, of course, the only quantities to which a numerical value could be assigned would be $\bar{x}$ and n . But for the moment the objective is to see how k would behave in repeated sampling if the remaining unknowns, μ and σ , were actually known. That is, the sampling distribution of the statistic k must be determined. The only quantity in the right-hand side of the above definition that would vary from one sample to the next is $\bar{x}$, for, as has been seen, the sample mean is typically different from one sample to the next. The quantities μ and σ , being population parameters, do not change from one sample to the next. The remaining quantity, n , is also constant from one sample to the next since the sample size is assumed to remain the same.

Because $\bar{x}$ is the only variable on the right-hand side of the equation, and since the sampling distribution of $\bar{x}$ can be regarded as normal, the sampling distribution of k is also normal. If it were possible to obtain expressions for the mean and the standard deviation of the sampling distribution of k , it would be possible to describe its sampling distribution completely.

In this case, it turns out that numerical values for the mean and the standard deviation of the sampling distribution are actually known from theoretical considerations. In repeated samples, values of $\bar{x}$ which are larger than μ will tend to offset values of $\bar{x}$ which are smaller: the average of the numerator of the equation for k will therefore tend to 0. Therefore the mean of the sampling distribution of k is 0.

Moreover, it is well known that if all the observations in a distribution are divided by their common standard deviation, the standard deviation of these new observations will be equal to 1. For example, if the sample means ($\bar{x}$'s) are divided by their standard deviation ($\sigma/\sqrt{n}$), the result is 1. The standard deviation of $\bar{x} - \mu$ is the same as the standard deviation of $\bar{x}$, or $\sigma/\sqrt{n}$. Hence the standard deviation of $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, or k , is simply 1.

Therefore, in repeated sampling, k is normally distributed with a mean of 0 and a standard deviation of 1. Since all the facts about the sampling distribution of k are known, it is possible to make probability predictions as to the range of values assumed by k in repeated sampling. These predictions can be made by the use of any standard table of the normal distribution such as is given in elementary statistics textbooks. For example, it is a familiar fact in elementary statistics that about 95 per cent of the observations in a normal distribution will fall within a range of plus or minus two standard deviations from the mean. Therefore, k will fall within the range 0 ± 2 , or -2 to $+2$, in about 95 per cent of repeated samples.¹

Now we come to the challenging part of the reasoning. First, in order to simplify the exposition, it will be assumed that σ is actually known.

¹ More precisely, 95 per cent of the observations will fall between $+1.96$ and -1.96 . For simplicity we use the number 2 throughout.

This assumption will later be relaxed by the device of replacing σ by its sample estimate, s . Second, even though everything about the sampling distribution of k is known, it is impossible to compute k in practice! The reason for the inability to compute k is that μ , the mean of the population, is unknown. In fact, the whole purpose of the sampling process is to estimate this unknown parameter. But even though one cannot compute k , one is able to say that in about 95 per cent of all samples, k will fall in the range from -2 to $+2$. Next comes the crucial step: "If k falls between -2 and $+2$, as it must do 95 per cent of the time, what will μ have to be?" The answer to this question is obtained by substituting values of k into the formula above and solving the equation for μ . We shall not show the algebraic manipulation. The final result, however, indicates that if k falls in the range of -2 to $+2$, μ must be included within the range given by $\bar{x} \pm 2\sigma/\sqrt{n}$. This result is the "95 per cent confidence interval." The reason for this name will be seen in a minute.

At this point the essential thing is to see that this result is dependent on the three basic facts (the mean, the standard deviation, and the "form") of the sampling distribution of $\bar{x}$ that were discussed earlier. The steps in the argument given above are interesting and will become more so when the verbal interpretation of the result is given. Nonetheless, the details of the argument should not obscure the fact that everything has been deduced from the three basic facts about the sampling distribution.

Before these theoretical points are pursued further, a specific numerical example will be given. In the charge account example, the mean ($\bar{x}$) was \$50.00. The size of the sample, n , was 225. Assume that the standard deviation of the population (σ) is known to be \$30.00. Then substitute all the specific numbers into the formula just given:

$$\begin{aligned}\bar{x} \pm 2\sigma/\sqrt{n} &= \$50.00 \pm (2) (\$30.00/\sqrt{225}) = \$50.00 \pm (2) (\$30.00/15) \\ &= \$50.00 \pm \$4.00\end{aligned}$$

The result, then, is $\$50.00 \pm \4.00 . Therefore, the 95 per cent confidence interval of this sample extends from \$46.00 to \$54.00.

We must now return to abstraction to explain the meaning of this result. In particular, it is necessary to concentrate on the expression "95 per cent confident." A very crude explanation would be that 95 per cent was a measure of the strength of the one's conviction that the confidence interval really included the mean of the population. A more precise explanation is also needed. It will be remembered that the entire discussion of the theory leading up to confidence intervals—namely, the investigation of the sampling distribution of the mean and the argument for the formula of the confidence interval—was a discussion of the characteristic of the sampling process. Therefore, it is not surprising that confidence is to be attached

not to the particular sample but rather to the sampling process which produced the particular sample. The confidence is in the *method* and not in the results of the particular sample. But confidence in the method makes one willing to use the results of the particular sample. We shall now show precisely why.

The rigorous meaning of the expression "95 per cent confident" is that in about 95 per cent of all samples produced by the sampling process, the confidence intervals constructed in this way¹ would, in fact, include the mean of the population. Since the sampling process would lead to the right answer in about 95 samples out of 100, on the average, it is reasonable to believe that it has done so in *this particular sample*. The reason for these two assertions should be clear from the logical development of the formula for the confidence interval. In effect, if k falls in the range of values in which it would be expected to fall in about 95 per cent of repeated samples, the value of μ needed to produce such a sample value of k would have to be somewhere in the range $\bar{x} \pm 2\sigma/\sqrt{n}$. Therefore, one really says, when looking at the results of a particular sample, "I'll bet that k has fallen between -2 and $+2$, because I know it will do so 95 per cent of the time. If my bet is right, then the formula for the confidence interval will include the true value of the unknown mean, μ . If my bet is wrong, as I would expect it to be about 5 per cent of the time if I were to keep on betting for a long time, k would not fall between -2 and $+2$. The value of μ which would bring about this unfortunate outcome is outside the range of values included in the confidence interval, and the confidence interval would not include the mean of the population. But since I know that my bet will be wrong in only about 5 per cent of the times that I make it, I am not going to withdraw it. I believe k has fallen between -2 and $+2$, and I'm willing to stick with this bet."

There is just one loose end in the above discussion. It was assumed that the standard deviation (σ) of the population was actually known, when in fact this parameter was not known. An approximation used in practice is to substitute the sample standard deviation (s) into the above formula in place of σ and to make some other very minor adjustments in the formula that can be ignored for the purposes of this discussion.² Hence, in practice, three facts are obtained from the actual sample, its mean, its standard deviation, and the sample size (that is, $\bar{x}$, s , and n), in order to compute the confidence interval.

¹ In general, one would not expect to get the same numerical values for the confidence interval from one sample to the next, since the mean ($\bar{x}$) varies from one sample to the next.

² Where, in computing the standard deviation, the formula customarily used in elementary statistics is modified by the replacement of n by $n - 1$ in the denominator. This method of estimating the standard deviation is slightly better than the other.

It is possible to use other levels of confidence in computing confidence intervals. If 68 per cent confidence were desired, the 2 in the formula above would be replaced by 1. (This follows from the fact that in the normal distribution, about 68 per cent of the observations are within one standard deviation of the mean; hence the prediction that k will fall in the range -1 to $+1$ about 68 per cent of the time. Similarly, to be 99.7 per cent confident, it would be necessary to use a 3 instead of the 2 in the formula above. Still other levels of confidence can, of course, be used. In general, for any given sample, the higher the level of confidence, the wider the range included in the confidence interval; the lower the level of confidence, the narrower the range included in the confidence interval.

7.6 Additional Comments

We have now sketched the outlines of sampling theory for a very simple problem: the estimation of the mean with a simple random sampling. The details of the development are not particularly important to the general practitioner, but its main outlines are vital. Above all it is essential to understand that sampling theory is focused on the method by which samples are selected rather than on the results of particular samples. Confidence in the results of particular samples really depends on a knowledge and understanding of the sampling method used. With simple random sampling (and other kinds of probability samples) it is possible to make precise confidence-interval inferences of the kind illustrated here. The details of the calculations vary from one probability sampling design to the next, but the basic interpretation of the end result is the same.

Again, the calculations involved in estimates of other parameters than the mean will be somewhat different in detail, but the idea is the same. It would be necessary to explore the nature of the sampling distribution of the statistic used to estimate some other parameter, just as the sampling distribution of $\bar{x}$ has been examined in this chapter. For simple random sampling, for example, the 95 per cent confidence interval for the proportion of a population having a certain characteristic (*e.g.*, purchasers of Wheaties) is given by

$$p \pm 2\sqrt{\frac{p(1-p)}{n}}$$

where p is the proportion of observations in the sample having the given characteristic.¹

One important caution, however, is needed. In practice, the probability sampling designs usually employed are more complicated than simple

¹ This formula is an approximation. As a rule of thumb in determining its applicability, the following is helpful: np and $n(1-p)$ should both exceed 5.

random sampling and the confidence-interval formulas are not the same as the ones given in this chapter for simple random sampling. Application of the formulas for simple random sampling to these cases may lead to seriously wrong conclusions. For this reason, the computation of confidence intervals should be delegated to a statistician or undertaken only with the help of one of the advanced treatises on probability sampling mentioned in the bibliography.¹ Specifically the reader should be warned against the indiscriminate use of the tables which purport to give confidence intervals. These tables are usually based on the formulas of simple random sampling and cannot be safely used for other probability sampling designs, except in rare circumstances.

Another caution, of course, is the fact that the theory of sampling discussed in this chapter does not apply at all directly to *nonprobability* sampling. We focus our attention solely on probability sampling for the next two chapters before turning to the subject of nonprobability sampling.

¹ In the discussion of probability sampling designs given in the next chapter, we occasionally indicate the nature of the necessary calculations without going into detailed directions.

CHAPTER 8

APPLICATION OF SAMPLING THEORY TO THE SAMPLING OF HUMAN POPULATIONS

8.1 Introduction

In Chapter 7 the broad outlines of elementary sampling theory were discussed. It was seen that the simplest kind of probability sampling, simple random sampling, can be used to draw inferences for which the degree of uncertainty can be stated as a numerical probability. In this chapter we discuss other more complex types of probability sampling. We do this because simple random sampling is seldom feasible for studying most of the human populations in which marketing researchers are interested. These more complex designs involve different formulas and computational procedures, but the theory underlying them is based on the theory of simple random sampling which has been discussed. We shall not go into much detail in discussing these more complex designs but shall, instead, aim at furnishing a discussion which will enable the reader to understand the principles of these designs and the kinds of circumstances in which they are useful.

The theory of simple random sampling is relatively well known. Frequently it is the only kind of probability sampling that is discussed in elementary statistics courses. The fact that investigators who are familiar with only this type of probability sampling have frequently given up probability sampling in favor of nonprobability sampling is not surprising, since simple random sampling is almost always impossibly costly in studies of human populations. A handful of investigators, however, in widely varied fields of social and marketing research have chosen other alternatives to simple random sampling. They have developed sampling designs which share the theoretical advantages of simple random sampling but which are at the same time feasible for practical work. There were isolated applications of very simple variants of simple random sampling well before World War I, but most of the developments in probability sampling of human populations have occurred within the last fifteen years, and the sampling designs in use today have, for the most part, been developed since World War II.¹ At the present time, probability sampling designs are being used

¹ The newness of this application of statistical theory is attested by the fact that the first two textbooks on the subject have appeared since January, 1949.

more and more frequently in applied social research, and theoretical developments in the details of these designs are keeping up with the demands for practical applications.

Before going on to a discussion of these more complex designs, it is vital to see why simple random sampling is so frequently unfeasible for the kinds of problems encountered in marketing research. In the example given in the previous chapters of the study of charge-account customers, simple random sampling was feasible. The reason is that the complete list of the universe was available. When such a list is available, it is relatively easy to devise a process of selection such that each element of the universe has a known positive probability of being selected. Further, in this example, the required information could be easily secured at relatively low cost, since the information was available from the accounting records of the store. There are other instances in which simple random sampling is also feasible. If the telephone company wished to secure certain information about its subscribers, it could use simple random sampling. Again the usefulness of simple random sampling rests upon the facts that a complete list of the universe is available and that relevant information can be secured at a low cost. The telephone company obviously has a list of its subscribers and can communicate with them cheaply by telephone (assuming that the information can be obtained by phone).¹ In other problems of marketing research in which complete lists are available and in which the cost of getting information about the elements on the list is small, simple random sampling can appropriately be used.

Usually in marketing research, complete lists are not available and the cost of getting information about the elements of the universe is substantial. Consider, for example, the problem of determining the average annual expenditure on furniture by the families in the United States. There is no list of the families in the United States. (The census lists are made every ten years, but even at those times the lists are not available for private use.) The lists which are available, such as telephone directories and mailing lists, do not coincide with the universe being investigated. The telephone-owning universe, for example, has a higher average income than the nontelephone-owning universe and can be expected to behave differently from nontelephone owners with regard to purchases of furniture.²

Even if a list of all American families were available, simple random sampling probably would be too costly unless a mail study was used. Probably

¹ Even when simple random sampling is feasible, more complex designs are, as we shall see, often economically more efficient.

² In England, because of frequent compulsory registration, current lists are available and are used in commercial samples. (C. A. Moser, "The Use of Sampling in Great Britain," *Journal of the American Statistical Association*, June, 1949, pp. 231-259.)

personal interviews would be necessary to determine family expenditures on furniture, and if this were so, interviewers would have to visit each of the families selected in the sample. With a simple random sample the families which were chosen would probably be widely scattered. It would be unusual to find two families in the sample who lived close to each other. After an interviewer had talked to one family, she¹ might have to take a 75-mile bus trip before she could interview the next family in the sample. The expense and time required to implement a sampling design in which such travel was involved would be immense. Thus, even if lists were available, simple random sampling would be unfeasible, since the time and expense of reaching the families in the sample would be unnecessarily great. In other words, simple random sampling would not be used because of its economic inefficiency.

These two obstacles, unavailability of lists and the cost of reaching people selected by simple random sampling in order to interview them, probably account for the infrequent use of simple random sampling in marketing research. Simple random sampling is simple in theory but is usually exceedingly difficult and costly to implement. The application of methods of random selection to human populations requires ingenuity.

"Cluster sampling" is a device which has very widespread applicability in marketing research in making possible the implementation of random sampling when it is to be applied to human populations. Cluster sampling requires that, instead of regarding each individual element of the population as a sampling unit, groups or clusters of elements be considered as sampling units. Later we shall examine the theory of cluster sampling; now, we shall show how it is applied.

The most important form of cluster sampling in marketing research is known as "area sampling." In area sampling of human populations, the clusters usually consist of individuals or families living in small, compact geographic areas (*e.g.*, city blocks). When these areas, rather than the elements of the universe which they contain (*e.g.*, individuals or families), are considered as the sampling units, the two major objections to simple random sampling can be largely overcome.

In the first place, it is easy to obtain a list of areas even when it is impossible to obtain a list of all the individuals living in the areas. All that is needed to get a list of areas is a map, and the map does not have to be either current with regard to political changes or precisely accurate in its representation of physical features. The only requirement is that the map permit the identification of mutually exclusive areas in such a way that each individual element of the population can be assigned to one and only

¹ Throughout this book interviewers are usually referred to by use of the pronoun "she". This does not mean that male interviewers are not often used.

one of these areas.¹ For example, a city map showing blocks should make it possible to find a block on the map and to locate the same block "on the ground." After blocks have been located in this way, it is then easy, of course, to recognize all the dwelling units that belong to this particular small area. Thus one can get a list of areas which include the relevant population. And the list does not get out of date. People may move around and social relationships may change, but fortunately—for sampling, at least—areas do not change enough to create difficulties in selecting a sample of them. With a list of areas, it is easy to select a simple random sample of areas. Each area can be assigned a known positive probability of coming into the sample.

In the simplest example of an area sample, the sampler selects for interview all the families (or individuals) living in the areas which have been selected at random from the list of areas. Suppose, for example, that 50 city blocks were selected from a list of the 1000 blocks in a city, and that each family living in these blocks was interviewed in order to find out about its expenditures on furniture (and any other information that would also be of interest). Since the 50 blocks were chosen at random from the list of 1000, each block in the city would have 50 chances in 1000 or 1 in 20 of being included in the sample. Once a particular block was chosen in the sample, any family in the block would certainly be chosen since the sampling method requires interviewing every family in any block that turns up in the sample. Therefore every family in the city, regardless of which of the 1000 blocks it lives in, has a probability of $1/20$ of being included in the sample. Since every member of the universe (in this case all families living in the city) has a known positive chance of being included in the sample, the result is, by definition, a probability sample of families living in that city.

It would not be necessary to interview all the families within the blocks chosen in the sample in order to maintain a probability sample. One might, for example, decide to take a simple random sample of $1/10$ of the families in each block selected in the sample. Since the block is a small compact area with a relatively small number of families, the random selection of $1/10$ of the families living in the block is a manageable task. It usually is not difficult to make a list of the families living in each of the sample blocks in order to facilitate the random selection.²

¹ It is possible to assign human beings to areas by associating them with their dwelling places. Some individuals have more than one dwelling place, and some have none. Simple rules can be devised to handle these relatively unusual cases.

² There are sometimes practical difficulties in listing. For example, one of the problems in listing families is that for many houses it is difficult to tell without knocking whether or not there is only one family there. One solution is to use addresses rather than families as sampling units.

With this sampling method, each family in the city would still have a known positive probability of selection. The magnitude of this probability can be obtained by applying one of the simple rules for the calculation of probabilities. The probability that any particular family would be selected is contingent upon the selection of the block in which the family is located. The selection of each family is in this sense conditional. The rule for calculating a conditional probability requires that the probability of the first event be multiplied by the probability of the second event which is contingent upon it. In the example, $1/20$ (the probability of selecting any particular block) would be multiplied by $1/10$ (the probability of selecting a particular family within a block, once the block has been chosen). Thus, under this sampling method, the probability of selecting any particular family in the city is $1/200$.

By using clusters, then, it is possible to overcome the first obstacle to probability sampling of human populations, the unavailability of listings. The second obstacle, the expense of making each observation, is also importantly diminished by the use of cluster sampling. By clustering the interviews into a relatively small number of areas, it is possible to reduce tremendously the average expense of traveling from one interview to another. The number of interviews which an interviewer can complete in a single day can sometimes be multiplied several times by the use of cluster sampling.

There are many variations in the ways in which clusters can be formed and in which clusters and individuals within them can be randomly selected. All these variations make it possible to get lists of the relevant population—by securing a list of the areas in which the population is found—and thus to secure a probability sample. And each variant is most efficient economically under particular sets of circumstances. Although it is not usually possible to select the most efficient design for particular purposes, a reasonably efficient design can be chosen for each problem.

Before considering the theoretical side of cluster sampling more closely, it should be noted that area sampling is not the only form of cluster sampling that is useful in the sampling of human populations. Suppose the owner of a chain of restaurants wished to obtain a probability sample of the opinions of his employees. Instead of drawing up a list of all his employees and then drawing a simple random sample, he might regard each individual restaurant as a cluster of employees and first draw a random sample of restaurants, after which he would select a random sample from the employees of the restaurants chosen in the sample. Or consider the problem of the telephone company attempting to draw a sample of its subscribers. Instead of numbering all the subscribers in the phone book in order to take a simple random sample, it might alternatively select

pages at random and then select names at random from the pages that were thus chosen. Still another example of cluster sampling illustrates the use of clusters made up of elements which are contiguous *in time*. A group of marketing students were studying the retail sales of different grocery stores in a shopping area of a large city. One store refused to provide information on dollar sales. The students were undaunted by this and reasoned that they could obtain the amount of sales by watching the amounts rung up in the cash register, which was visible from the window of the store. They knew that constant observation during the entire day would be fatiguing and moreover that the suspicions of the store manager would be aroused. They therefore selected several ten-minute intervals at random and observed the sales rung up during these randomly chosen intervals. From this information it was possible to estimate average sales per ten-minute intervals, from which, in turn, it was easy to estimate total sales. In this example the universe consisted of sales to individual customers. Each ten minutes of the day was a cluster consisting of individual sales, and these clusters were selected at random. The same sort of process could be used to estimate the size of television audiences in bars. A modification of this idea has been found useful as an alternative to conventional "store audit" measures of retail sales.

These examples have been cited in order to show that many of the principles to be discussed in the following pages are applicable not only to area sampling (which we use for most of our illustrations) but to other methods of cluster sampling as well. With this general introduction, then, we must next examine cluster sampling in more detail.

8.2 Cluster Sampling

We have introduced the general ideas of cluster sampling and have shown that cluster sampling is really a form of probability sampling in that it assures each element of a population a known positive probability of appearing in a sample. Now we shall consider the theory of cluster sampling. This theory can be quite complicated in its details. There are many ways in which a cluster sample can be selected, for there are a great many ways in which one can group members of the population into clusters and select clusters and individuals for the sample. Even if only geographical areas are considered, as in area sampling, there are many possibilities. The geographical areas can be large, like counties, or small, like city blocks or even sections of blocks. The areas can follow certain natural or political boundaries, or they can cut across such boundaries. They can be any shape: approximately square, as are most city blocks, or long and narrow, or irregular. The number of elements of the population in each cluster may be equal or unequal. Moreover, it is possible to select

all the elements of the universe falling within a given cluster or only some of them. One can interview all the families living in a given city block or only a sample (*i.e.*, a subsample) of them. If subsampling within blocks is used, one may select an equal number in every block, an equal proportion in every block, or other more complex arrangements.

These complexities are somewhat bewildering for the beginner in practical sampling. Rather than tackle all the complications at once, we shall first consider a very simple case in which sampling is from a population which has been grouped into clusters in such a way that the number of elements of the population in each cluster is exactly the same. In this case it is possible to see the close relationship between cluster sampling and simple random sampling without being overwhelmed by a host of new difficulties. The basic ideas discussed in the preceding chapter apply quite well in this new context. It should be said at the start, of course, that it is ordinarily impracticable to group the members of a population into equal-sized clusters and that natural groupings of this kind are unusual. The assumption is made only to aid in understanding the basic ideas. It is possible, however, to conceive a practical situation in which equal-sized clusters (or very nearly equal-sized clusters), might be expected to exist. One such example would be a sample of the population of all United States infantrymen in rifle platoons. Each such platoon would (theoretically, at least) consist of an equal number of infantrymen. Another example would be the sampling of tomato cans from cases of 24 cans. In our further discussion, however, we shall talk of area sampling rather than examples like these.

1. Cluster sampling without subsampling: equal-sized clusters. Visualize a hypothetical housing project in which there are 1000 blocks with 10 families in each block.¹ The problem is to draw a cluster sample in order to estimate average income per family. Whenever a block is selected at random, *all* the families within the block are interviewed in order to ascertain the income of each.

By a simple transformation of concepts, it can be seen that this new problem is no different theoretically from simple random sampling as discussed in the last chapter. For there are two ways of thinking about the population. In the first place it is possible to think of the population as consisting of the incomes of all the 10,000 families in the project. In the second place, the population can be thought of as the mean incomes of the people living in each block. This second population would therefore consist of 1000 elements—*i.e.*, one mean income for each of the 1000 blocks.

¹ It would be said, in this case, that the *size* of the clusters is the same for all clusters. By size of cluster is meant the number of elements of the population in the cluster and *not* the geographical area of the cluster.

The mean income for both these populations would be the same. For, since the number of families per block is the same, the mean for all 10,000 families can be obtained either by direct computation for all families or by first finding the block means and then averaging them. In the present problem, it is necessary to draw a random sample of blocks, find the mean income for each block, and average the mean incomes for the blocks in the sample.

Recall, now, the main objective in this sampling process, to make a "best estimate" and to construct a confidence interval. To do this it is necessary first to examine the sampling distribution of the mean, just as was done for simple random sampling.

1. The mean income per sample family (*i.e.*, the mean of the block means for blocks chosen in the sample) is an unbiased estimate of the mean income per block in the entire housing project. The sampling process is therefore unbiased.

2. The standard error of the mean is given by the same formula used in simple random sampling since, in effect, the present sample is a random sample of block means rather than of individual families. That is, the

standard error of the mean is given by $\frac{\sigma_c}{\sqrt{m}}$, where σ_c is the standard deviation of the 1000 block means and m is the number of blocks in the sample.¹ We have changed the symbol for sample size from n to m in order to emphasize that we are now dealing really with clusters (blocks) rather than elements (individual families).

It will be observed, as before, that in sampling from a given population the standard error of the mean decreases as the sample size (number of blocks) increases. As the number of blocks in the sample is increased, the standard error of the mean will decrease. The symbol σ_c is the standard deviation of the mean incomes for the 1000 blocks. If this standard deviation is relatively large and the sample size is fixed, the standard error of the mean would be large. If the standard deviation is small, the standard deviation of the mean would be relatively small.

This may be illustrated by extreme cases. If the mean income per block were identical for all blocks, σ_c would be 0. For any sample size, therefore, the standard error of the mean would be 0. This fact, in conjunction with the fact that the sampling method is unbiased, ensures that there would be no error at all in the sample estimates. Even if the sample consisted of only a single block, the mean income for the block would exactly equal

¹ We have ignored, for simplicity, the finite population correction. The statistician accustomed to neglecting this factor in other sampling applications should be warned that quite frequently in cluster sampling he may not be warranted in doing so, since the number of clusters in the population is sometimes quite small.

the mean income for all the blocks of the population, since all the blocks have the same mean income. Consider another possibility. If sales managers and other executives lived in some blocks, the mean income for these blocks would be relatively high; if file clerks and marketing professors were located in other blocks, the mean income for these blocks would be relatively low. The standard deviation between blocks would then be large, and, as was just stated, the standard error of the mean would be large.

3. Finally, the sampling distribution of the mean (the frequency distribution of the means of repeated samples of the same size from this population) would be approximately normal, provided that the number of blocks in the sample was reasonably large—say, as a rule of thumb, 20.

The theory of confidence intervals fits in perfectly at this point. The 95 per cent confidence interval, for example, is obtained as follows:

$$\bar{x} \pm 2\sigma_c / \sqrt{m}$$

where $\bar{x}$ is the average of the mean incomes of the blocks chosen in the sample. Again, as in the case of a simple random sample, there is one loose end to tie up. The standard deviation of the universe (σ_c) is unknown. The approximation used is identical with that in the earlier case: compute the standard deviation of the means of the blocks in the sample (s_c) and use this number in place of σ_c (the population standard deviation).¹ As in the earlier example, the minor modifications in the formula for the confidence interval necessitated by this step need not be considered for most practical purposes.

With this brief outline of theory, we can attack some of the important practical problems encountered in cluster sampling.

1. How Big Should the Cluster Be? In the above example, blocks were used as clusters. There was an equal number (10) of families per block. Suppose someone suggested a modification of this design. Suppose it were proposed that each block be divided in half so that there would be 2000 clusters with only 5 families in each. Suppose that someone else suggested a doubling up of blocks so that there would be only 500 clusters (each consisting of two blocks) with 20 families in each cluster. Clearly, there are many other possible proposals. We cannot consider all of them but shall focus on the problem of deciding whether either of the two alternatives indicated above is better than the original method of forming clusters of individual blocks. A consideration of this problem provides some insight into the kind of reasoning necessary to answer the practical question: "How big should the clusters be?"

¹ In computing s_c , $m - 1$ is used rather than m since this procedure tends to give a slightly better estimate of σ_c .

In order to answer the question, it is necessary to have some standard of judgment. When is one sampling design better than another? We shall use as a standard the concept of economic efficiency which was mentioned briefly in the last chapter. That is, for a given amount of money, the best sampling design for estimating the mean is the one that gives the smallest standard error of the mean. The design which will accomplish this objective is said to be the design with the highest economic efficiency.

There is a given amount of money. It is to be spent in such a way that the smallest standard error of the mean will be obtained. The formula for the standard error of the mean has just been seen to be $\sigma_c/\sqrt{m}$. It is necessary to figure out what m and σ_c will be for each of the three sampling designs in order to see which design yields the smallest standard error for the money.

First, one figures out m , the number of clusters which can be afforded in the sample, if the individual clusters are one-half block, one block, and two blocks, respectively. To do this some knowledge is needed of (1) the costs of interviewing and traveling from one family to the next in the same cluster and (2) the costs of traveling from one cluster to the next.

Second, σ_c must be approximated for each of the three designs, for the standard deviation of the cluster means depends on the way in which the clusters are formed. There would be a different value of σ_c for each of the three methods of forming clusters which are now being considered. There is no way to know *a priori* just what σ_c would be for each of the three designs. The only way to find out would be to examine the particular population being studied. Suppose for the moment that this has been done.

Then both m and σ_c would be known for each of the three proposed methods of forming clusters, and these values would be substituted into the formula for the standard error of the mean to find which of the three methods yielded the smallest standard error for the money. In this way it is possible to reach a decision as to the size of cluster to be chosen from the three proposed alternatives.

This procedure is simple enough and serves nicely to illustrate the kind of reasoning used to choose among alternative cluster sampling designs. There are two things in particular to be emphasized. First, it is necessary to know the costs—costs of interviewing and costs of traveling. Second, one must have knowledge of certain characteristics of the population, in this example the standard deviation of cluster means for different ways of forming the population into clusters. Organizations which periodically study the same population are able, in any particular investigation, to make the necessary estimates regarding the population on the basis of information secured from previous surveys. The U.S. Bureau of the

Census, for example, in its monthly surveys of national unemployment can make very good estimates of those characteristics of the population which must be known in order to design efficient samples.

In marketing research and other fields making frequent use of nonrepetitive surveys, it is often difficult to obtain the kind of information needed for efficient sampling. Marketing researchers can take comfort in the fact that even though it is seldom, if ever, possible to make such investigations in marketing research, it is frequently possible to use generalizations which have emerged from the investigations of other samplers, particularly those of the U.S. Bureau of the Census. It has been found that such generalizations apply to a surprisingly wide range of applications and are helpful as guides in many of the problems encountered in marketing research.

The general answer to the present problem is interesting. If all the families within a cluster are interviewed, it has been found that small clusters come close to the optimum size. That is, for a given amount of money the smallest standard error of the mean is obtained by the use of small clusters. In the extreme case the cluster would coincide with the individual family, and cluster sampling would coincide with simple random sampling. This extreme, however, is too extreme. Whenever economic efficiency is considered, the optimum size of cluster is larger than a single family. On the other hand, the optimum size is perhaps seldom larger than a cluster of the order of magnitude of a city block. The ordinary practice is not to interview all the families in each cluster, as we have been assuming, but rather to subsample within clusters—take only a sample of the families living in a block. (This is discussed further under Cluster Sampling with Subsampling, 3 and 4 of this section.)

Implicit in the discussion above is a very important idea that has not yet been brought out clearly. A preoccupation with *economic* efficiency has prevented adequate consideration of the effect of cluster sampling on *statistical* efficiency. For the types of problems with which marketing research is concerned, the use of cluster sampling almost inevitably results in a decrease in the statistical efficiency of the sample design. That is, for a given number of interviews from a given population, the standard error of the mean is higher for cluster sampling than for simple random sampling. Thus the abandonment of simple random sampling results in interviews which have a much lower average cost but which contribute less to the reduction of the standard error of the sample mean. The problem of defining the optimum size of clusters is the problem of finding out at what point the decreased cost of the interviews is just offset by the loss in statistical efficiency. This is simply another way of looking at the problem previously discussed.

One important consequence of the frequently large difference in the statistical efficiency of simple random sampling and cluster sampling deserves emphasis. *Simple random samples frequently have standard errors which are 50 per cent or less of the standard errors of cluster samples of the same size.* For example, the standard error of a simple random sample of 60 elements (that is, $\sigma/\sqrt{n}$, or $\sigma/\sqrt{60}$) might be 50 per cent or less of the standard error of a random sample of six clusters with 10 elements in each cluster (that is, $\sigma_c/\sqrt{m}$, or $\sigma_c/\sqrt{6}$) from the same population. Hence if one were to assume that the formula for the standard error of simple random samples applied to cluster samples, one would seriously underestimate the magnitude of the standard error and would have an unwarranted degree of confidence in the sample estimate.

2. How Should the Clusters Be Formed? We have discussed ideas useful in deciding the best size of clusters. We have suggested three alternative sizes: 20 (two blocks), 10 (one block), and 5 (one-half block). It is not necessary, of course, to group the elements of the universe by blocks or multiples or fractions of blocks. Ideally, for statistical efficiency at least, elements of the universe should be grouped into clusters whose means are as much alike as possible with regard to the characteristic being studied. As has already been seen, if cluster means were exactly alike with regard to the characteristic being studied, it would be possible to secure an errorless estimate of that characteristic for the population with the use of only a single cluster.

The desirability of similarity in cluster means can also be seen by considering the consequences of the opposite extreme—great dissimilarity of cluster means. If cluster means differed from each other as much as possible, all elements in a single cluster would be alike, making the variation between cluster means the same as between the elements of which the clusters were formed. This can easily be illustrated by attempting to divide a deck of cards into 13 clusters of four cards each in such a way that the means of the 13 clusters will be as unlike as possible. But if all elements in a cluster were perfectly alike, a sample of 50 clusters would be no better than a random sample of 50 individual elements. No more could be learned by selecting 50 blocks at random and interviewing every family in each block (500 families in all) than by selecting 50 families at random and interviewing them. The standard error of the estimate for both samples would be the same.¹ This is true because the first interview in each block would indicate the average income of the block with as great precision as would all ten interviews.

Hence, in the example of the study of average income, the securing of ideal clusters would require that all clusters have the same average income.

¹ Except for the possible effect of the finite population multiplier.

Such an ideal would seldom be obtained in practice for two reasons. In the first place, forming clusters so as to achieve similarity between cluster means implies a prior knowledge of the universe which would in practice seldom be possessed by the investigator. In the second place, forming clusters in this way, even if it were possible, would almost certainly entail additional administrative costs. There would be increased mapping costs in the planning of the sample. The increase in costs would come about also because the creation of clusters with similar means can ordinarily be achieved only by grouping together elements (*e.g.*, families) which are not geographically contiguous. The cost savings of clustering would then be partly lost. Costs per interview would go up because of the increased time required per interview. In most cases the costs would probably not be offset by gains in statistical efficiency.

Getting clusters which are very much alike implies that the elements *within* any given cluster will be relatively dissimilar. If it were desired to have the clusters alike with regard to average income, it would clearly be undesirable to have any of the clusters consist of poor people only or of rich people only. More desirable would be a definition of clusters so that each contained both rich and poor. More generally, in order to form clusters in such a way that the cluster means would be equal, the families within each cluster would have to be dissimilar.

There are ways of increasing this within-cluster dissimilarity. In the income illustration, the magnitude of the incomes of individual families would not be known in advance. However, other things which are rather highly correlated with income might be known. Average rentals for each block, for example, are usually available from census data. These data could be used in the formation of clusters consisting of more than one block, if the increased dissimilarity within the clusters seems to warrant the decline in statistical efficiency which would result from increasing the cluster size. By using objective evidence such as census data along with subjective impressions of neighborhoods and dwellings, it might sometimes be possible to increase the efficiency (both statistical and economic) of sample designs by modifying the way in which clusters are formed.

The attainment of the theoretically ideal clusters is limited by the availability of data for various areas and by the difficulties of administering sampling designs of different types. In the above example, city blocks were used as the clusters. It is easy to see why this is the natural way to approach the problem. A city block can be designated simply and unambiguously by naming the streets which surround it or by reference to a city map. Moreover, 1940 census data on individual city blocks are available for all cities which had more than 50,000 people in 1930.¹

¹ Similar data from the 1950 census will be available shortly after this book is published.

Departure from block boundaries in the definition of clusters causes administrative inconvenience. While some increase in economic efficiency might seem possible by subdividing blocks, the increase would probably be illusory. The reason is that there are other costs than the costs of interviewing and travel; administrative costs are important and can vary widely with the sample design. The administrative costs would certainly be larger if blocks were subdivided, and this increase in administrative costs would probably make such a design relatively inefficient economically. Besides, as briefly mentioned before and as will be seen later in detail, cluster sampling with subsampling within clusters usually dictates a larger size of cluster than cluster sampling without subsampling, which is the only design considered thus far. One would not, then, generally want to subdivide blocks.

It is possible, when clusters larger than city blocks are used (as would often be feasible if subsampling within clusters is employed), that the blocks can be combined in forming the clusters in order to achieve a greater dissimilarity within clusters than would be possible otherwise. There are striking examples of such possibilities in most large cities. A single one of these larger clusters in Chicago could extend from the Gold Coast to the slums. Similarly spectacular increases in intracluster dissimilarity could be achieved in New York City, where the very rich and very poor frequently live in contiguous blocks. Although it is sometimes possible to improve sample designs by combining groups of blocks in this way when clusters are formed, the administrative difficulties and expenses are most certainly increased. If repeated sampling is to be done in the same city, however, the initial administrative expense might be worth while.

2. Cluster sampling without subsampling: unequal-sized clusters. In order to simplify the discussion, we have so far been discussing equal-sized clusters. At the outset, however, we cautioned that in practice it is unusual to find clusters with natural boundaries such that the number of families (or individuals, depending on the problem) within each cluster is the same. Moreover, it is difficult and costly to form clusters so that equal size can be attained. As a result, the researcher must typically face a situation in which the sizes of the clusters are not equal. Fortunately, however, the general conclusions of the discussion of equal-sized clusters apply also to unequal-sized clusters. Now, instead of trying to determine the most efficient size of cluster, it is desirable to determine the most efficient *average size* of cluster, but the method of attack is similar. The principle of trying to achieve internal dissimilarity in the composition of clusters applies to clusters of unequal as well as to clusters of equal size.

There are, however, some minor amendments and additions that must be mentioned. Again it is necessary to ask the basic question about the

sampling distribution of the estimated mean in order to see the characteristics of the sampling process. In this case, however, only the mean and the standard deviation of the sampling distribution will be considered, since it will be assumed in this and all subsequent examples that the sample mean will be distributed approximately normally. The following comments are relevant:

1. The mean of the sample is no longer an unbiased estimate of the mean of the population. (The reason is not obvious and not worth exploring here.) The amount of the bias is known to be unimportant for practical purposes and to become smaller as the sample size is increased. One is still warranted in thinking of the estimate as being "essentially unbiased," in contrast with nonprobability sampling, in which the bias is of unknown dimensions and can be enormous.

2. The formula for the standard error of the mean is more complicated than before, since the variability of the sampling process is affected (increased) by the variation in the size of clusters. In effect, the change in the calculations consists of modifying the method of calculating the standard deviation of the cluster means. Instead of weighting each squared deviation in the calculation equally, as is ordinarily done in the computation of the standard deviation, each squared deviation (*i.e.*, deviation of cluster mean from population mean) is weighted by the square of the ratio of the size of the particular cluster to the average size of all clusters. In estimating the standard error of the mean from a sample, which must be done in this case as in the earlier ones in order to construct a confidence interval, an analogous computation is carried through with the standard deviation of the *sample* cluster means.

It should be remembered that the finite population multiplier is being neglected in this discussion.

3. Cluster sampling with subsampling; equal-sized clusters. The next complication to be considered is the case in which elements within selected clusters are sampled rather than completely enumerated. In the earlier discussion, if a given city block came into the sample, every family living in the block was interviewed. But rather than interview every family, it would also be possible to take two families at random in each block, or every fifth family in the block from a random starting point, or any other group of families so long as the probability of inclusion for each family is known and positive. Including only a fraction of the elements in each cluster is called "subsampling within a cluster," and a sampling design which uses this approach is called "cluster sampling with subsampling," or, alternatively, "two-stage sampling."

Suppose that in the housing project of 1000 blocks the following sampling design were used: Select 100 blocks at random from the 1000 blocks

and within each sample block select at random (*i.e.*, by simple random sampling) 5 families from the 10 living within the block. (We shall reintroduce for now the assumption that the number of families in each block is 10, *i.e.*, that the clusters are of equal size.) In a similar example on page 113 of this chapter, we showed that a sampling design of this kind results in each family in the entire population having the same probability of being included in the sample. In the present case each block would have 100 chances in 1000 (1 in 10) of selection, while each family in a block chosen in the sample would have 5 chances in 10 (1 in 2) of selection. The probability of selection for each family, as in the earlier example, would then be $1/10 \times 1/2$, or $1/20$.

We pointed out also in the introduction to this chapter that the selection of a simple random sample within each block presents no serious problem. In the present example it would be necessary only to visit each of the 100 blocks chosen in the sample, list the names of the 10 families living in the block, and select a simple random sample of 5 families in 10.

This sampling procedure can be thought of as a simple random sample within a simple random sample. That is, a simple random sample of families is chosen from the families living in each of the blocks chosen in the first sample. This illustration of two-stage sampling is a very simple example of multistage random sampling, which will be discussed later in this chapter. In this case the first stage is the selection of clusters; the second, the selection of families within clusters.

Once again we shall follow through our basic procedure in discussing sampling designs. That is, we shall examine the salient features of the sampling distribution. So long as it is assumed (as in the present example) that the clusters are of equal size and that an equal number (or what comes to the same thing in this case, an equal proportion) of families is selected in every cluster, the mean of the sample will be an unbiased estimate of the population mean. The standard error of the mean becomes a more complicated expression than the one given in the earlier example. We shall not write down the formula, but we can easily indicate the nature of the complications.

The only source of sampling variability in cluster sampling without subsampling was the variability of cluster means. In that case the basic procedure could be regarded as a simple random sample of cluster means; once a cluster (block) was selected in the sample, all families in the cluster were interviewed so that the mean of the cluster was determined without sampling error. In the present example, by contrast, only five families of the ten in each cluster are interviewed. Thus, in each cluster selected, only an estimate can be made of the cluster mean. Therefore, not only must one take account of the fact that the 100 clusters are only a sample

of the 1000 clusters in the entire housing project, but it is also necessary to allow for the fact that in each of the 100 clusters only a sample of the families living in that cluster is actually interviewed. Hence the formula for the standard error of the mean must make allowance for both these sources of error.¹

In spite of the fact that the formula for the standard error of the mean is relatively complex, a *sample estimate* of the standard error of the mean, which is needed in order to compute a confidence interval, is relatively simple. The estimate of the standard error of the mean is given by the following simple (though approximate) formula:

$$\frac{s_c}{\sqrt{m}}$$

where s_c is simply the standard deviation of the means of the subsamples from each cluster in the sample.² As before, m is the number of clusters which have been selected in the sample.

Cluster sampling with subsampling is usually more efficient, when cost is considered, than cluster sampling without subsampling. That is, with a given amount of money to spend, in most problems cluster sampling with subsampling would yield designs better than the best attainable by cluster sampling without subsampling.

The analysis of the relative efficiency of these two types of sampling is similar to that described earlier when we discussed the process of deciding which size of cluster was best. First, figure out what is possible with various sampling designs, given a fixed amount of money to spend. Then calculate the standard error of these various possible sampling designs, and select the sampling design with the lowest standard error. In this calculation, of course, it is necessary to have information about the variability of the cluster means and, since we are now discussing the possibility of subsampling within clusters, also about the variability of family expenditures within clusters.

It was pointed out earlier that seldom will the information necessary for such a careful consideration of alternative sampling designs be available in marketing research and that the marketing researcher must, therefore,

¹ From what has been said here, a statistician unfamiliar with survey work would probably have little trouble writing down the formula. He should be warned, however, that the finite population multiplier becomes important because of the fact that a large fraction (that is, one-half) of each of the 100 clusters is now being sampled.

² Again, division by $m - 1$ rather than m is appropriate in the calculation of s_c since the objective is the estimation of the standard deviation of the population of clusters from which the sample of m clusters was drawn.

accept the generalizations that have emerged from the detailed study of sampling problems in the applied work of investigators in other fields.

Evidence of this sort indicates that cluster sampling designs with subsampling are usually economically more efficient than cluster sampling designs without subsampling. It is easy to see intuitively why this is true. Suppose that the problem is to estimate the proportion of Negro families in a large city. Often in large cities Negroes are concentrated in certain areas. In any given block, therefore, it would not be unusual to find either all Negroes or no Negroes. Suppose that this all-or-none pattern were pervasively true and that 100 blocks were selected in a sample. Then only one interview from each block would determine precisely the characteristic of that block. The best sampling design under these conditions would be cluster sampling with subsampling within clusters, and the rate of subsampling would be one per block. For most problems the elements within clusters do not tend to be so similar, and therefore the optimum rate would usually be greater than one but less than the total number of families in the block. This example, of course, is identical to the earlier example in which it was shown that each interview after the first would be wasted if all families in a cluster had identically the same income. Both examples show intuitively that, instead of concentrating all interviews in a relatively small number of clusters, it is better to take a relatively small number of interviews per cluster in a relatively large number of clusters.

If it were practicable to design clusters so small that each contained, say, three families, it is possible that cluster sampling without subsampling would be as efficient as cluster sampling with subsampling using larger and more convenient clusters (such as city blocks). The greater convenience of the relatively large clusters almost invariably suggests relatively large clusters combined with the use of subsampling.

Indeed, it is possible to think of subsampling as merely a device for obtaining clusters. That is, the larger cluster (*e.g.*, the block) is divided into several subgroups which can themselves be thought of as clusters. Viewed in this way, the separation of the sampling process into two stages is simply for administrative convenience. While this way of looking at things may at first seem somewhat difficult, it leads to additional insight into the basic ideas of cluster sampling. For it is not at all necessary for clusters to be composed of contiguous elements, even though in area sampling clusters are usually of contiguous elements.

We have shown that if the cluster size is fixed and the elements within clusters are relatively alike, the sampling design could be improved by subsampling within clusters and sampling more clusters. Another reason

for subsampling may be seen by a relaxation of the assumption that the cluster size has already been determined. That is, it is possible to obtain the advantages of dissimilarity within clusters associated with larger cluster sizes by the expedient of subsampling within clusters.¹

In the discussion of the standard error of the mean for cluster sampling with subsampling it was seen that there are two sources of error: (1) error due to the fact that only a sample of clusters is selected and (2) error due to the fact that only a sample of elements within a cluster is examined. Ordinarily in the sampling of human populations the formation of large clusters tends to make the cluster means more alike and therefore to make the error arising from the first source relatively small. The error from the second source is increased but the net result is a smaller total error, assuming the number of elements in the sample to be fixed.

However, the increase of cluster sizes results eventually in substantial increases in the cost of interviewing, because much more travel is necessary in going from interview to interview within clusters. In addition, the cost of listing the dwelling units within clusters increases.

A balancing of these competing considerations results typically in clusters on the order of size of a city block or small groups of city blocks.

4. Cluster sampling with subsampling: unequal-sized clusters. In this section we consider briefly the combined effects of the two complications explained in the two preceding discussions: unequal-sized clusters and subsampling. If the proportion of families subsampled is the same for all clusters chosen in the sample, the present problem is simply the sum of the complications in the preceding two discussions without any new difficulty.² We shall therefore make this assumption and ignore, for the moment, another type of subsampling with unequal-sized clusters in which an equal, or approximately equal, number of families is selected in every cluster regardless of cluster size.

As usual, we shall briefly discuss the salient features of the sampling distribution for the sampling design being considered. Although, in this design, every family in the population has an equal probability of selection, the sample mean is no longer an unbiased estimate. The bias, however,

¹ In an earlier example, it was suggested that the formation of clusters containing widely divergent elements would be desirable. For example, in the city of Chicago it is possible to combine blocks in the Gold Coast with adjoining blocks in the slums. To the extent that this can be accomplished, the differences between cluster means are relatively small and the standard error of the sample mean would likewise be small.

² By subsampling the same proportion of families in each cluster chosen for the sample, we mean, of course, that if 2 families were selected from a cluster containing 10 families, then 10 families would be selected from a cluster containing 50 families, etc.

is of no practical importance, and in any event it becomes smaller (approaches zero), as the sample size increases.

The standard error of the mean, in addition to the simple expression already considered for equal-sized clusters without subsampling, must take into account the contribution to error of subsampling within clusters (discussed under number 3 of this section) and the error arising from the fact that the clusters are of varying size (discussed under number 2 of this section). Again, the computation of an *estimate* of the standard error, which is all that is needed in practice to compute a confidence interval, is not difficult. The procedure first described under number 2 of this section for clusters of unequal size without subsampling is applicable, except that one now must think of the subsamples within each cluster as being clusters.¹

5. Multistage sampling with more than two stages. For most probability samples of human populations, subsampling takes place at more than one stage in the sampling process. We have already given a simple example of subsampling in which clusters were selected at random and families subsampled within clusters. This idea is extended frequently to designs which include first a sample of large-sized clusters, then a subsample of small-sized clusters from the sample of large-sized clusters, and finally a sample of families within the small clusters. Sometimes even more steps are added. There are no new basic ideas involved, but the extension of the old ideas involves further complexities. Therefore we shall present a descriptive discussion so that the reader will have a concrete idea of the way the principles are applied in practice.

For sampling of cities or other relatively small areas, cluster sampling with subsampling would be sufficient (although it would, in practice, be improved by stratification and other devices to be discussed later in this chapter). Consider now the problem of estimating income for the entire country rather than for one particular housing project. If an investigator were to break the country up into areas roughly the size (in terms of the number of families in each) of a city block and to select a sample of those areas at random, he would be slightly better off than if he had attempted a simple random sample of families. The amount of traveling *between clusters* (or the expense of obtaining local interviewers) would be immense, because the "blocks" would be widely scattered throughout the country and the average distance between blocks would be great.

The economic efficiency of such a sampling design would usually be greater than that of a simple random sampling design because of the fact

¹ In place of "cluster mean" in the earlier discussion (see p. 124) for example, supply the words "mean of subsample within cluster." In place of "size of cluster" read "size of subsample within cluster," and in place of "average size of sample clusters" read "average size of subsamples within sample clusters."

that several families within each cluster would be grouped together within easy walking or (in rural areas) driving distance. It is usually possible, however, to obtain still greater efficiency by grouping clusters (blocks) within still larger geographical areas such as counties. Hence the usual procedure in national area samples is to divide the country into counties or groups of counties which are selected at random and to subdivide the counties or groups of counties chosen in the first drawing into smaller clusters such as blocks within cities or square-mile sections in rural areas. These smaller areas are randomly chosen in the second drawing. Finally, within blocks or square-mile sections, families are chosen at random.¹ It is customary to refer to the largest clusters (such as counties) as "primary sampling units" and the smaller clusters (such as blocks) as "secondary sampling units." A brief discussion of the problems of forming primary and secondary sampling units will contribute to an understanding of practical sampling designs.

In the formation of primary sampling units, it is sometimes desirable to depart from the use of individual counties as the primary sampling units. In particular, it is frequently desirable to merge adjoining counties into a single primary sampling unit (or large cluster). The reason for this is that it is frequently possible to combine contiguous counties in such a way that the internal dissimilarity of the resulting cluster is greatly increased. Adjoining counties, more than adjoining blocks within cities, frequently differ greatly with regard to characteristics commonly investigated in marketing research. For example, a county consisting of a large city will often be contiguous to a county consisting of predominately rural or suburban areas. The combining of such dissimilar counties usually results in increased efficiency in the sample design for the same reasons we have discussed earlier in this section under *How Should the Clusters Be Formed?* The limit of the extent to which counties can be combined is set primarily by the areas of the counties and transportation facilities within them. A resident interviewer, for example, would often be able to travel to different interviewing locations in two adjacent counties without spending an excessive amount of time in traveling. Ten counties, on the other hand, would almost always cover too great an area, and in some western areas, single counties are so large in area that it may be advisable to subdivide them into smaller primary sampling units. Again, if two counties are

¹ Many variations of multistage sampling are possible. For example, it would be possible to use large groups of counties (such as Hearst trading areas) as the largest clusters, subsamples of townships (or similar units) within these group of counties, and blocks or square-mile sections within townships. Still another example is provided by the complication added frequently to the process described above. Instead of directly selecting blocks within counties, small towns are first selected and then blocks are selected within the towns.

separated by a mountain range or by a river with few bridges, it is inadvisable to combine the two counties into a single primary sampling unit.

Fortunately, there is a great deal of information about counties, which can be used in making decisions on combining counties into larger sampling units. A good map will indicate roughly the probable effects of possible combinations of sampling units on traveling costs for resident interviewers who might be required to conduct interviews at any spot in the proposed primary sampling units. Moreover, data on the characteristics of counties can be used to guide the combination of dissimilar counties. The best source, perhaps, is the *County Data Book* prepared by the Department of Commerce. This book summarizes the most important facts about the characteristics of each county as revealed by the censuses of population, agriculture, business, and manufacturing. The government provides other sources of information, and there are many private agencies such as trade associations and publishing companies which provide similar information.¹

The problem of forming secondary sampling units within primary sampling units is somewhat more difficult in practice. Fortunately, however, time is saved by worrying about this problem only for those primary sampling units (counties) that have actually been included in the sample. One can completely forget about the remaining counties. The city block is the natural boundary of greatest usefulness within the corporate limits of towns. For large or medium-sized cities, maps are available which clearly show the blocks. For cities which had greater than 50,000 population in 1930, the U.S. Bureau of the Census has prepared books (*Supplements to the First Series Housing Bulletin* for the state in which the city is located) that give detailed block maps with the blocks already numbered (which aids in the mechanics of sample selection).²

For medium and small towns the procurement of maps or aerial photographs is often difficult, always time-consuming, and sometimes impossible. When maps are unavailable, various expedients can be used, including the preparation of rough sketch maps by local interviewers and devices of the kind that are discussed in the last section of Chapter 11. Often aerial photographs are available, and for sampling towns with populations less than 2500 these photos may be used as maps.³

¹ The Department of Commerce has published a pamphlet, *Sources of County Data*, which is extremely useful as a reference not only for sampling problems but for problems of sales analysis as well.

² These books also provide important block statistics for individual blocks: number and characteristics of dwelling units, including the average monthly rentals. Similar volumes are to be issued for the 1950 census.

³ It is also possible to save effort in procuring a map by first selecting the sample roughly in such a way that it can be ascertained whether or not any blocks will be chosen in a given town. If no blocks will be chosen, obviously no map is needed.

The problem of forming secondary sampling units outside town limits is for the most part quite simple. Almost all these rural districts have already been divided into small areas for sampling use by the Master Sample of Agriculture. These small areas have, for any particular region of the country, approximately the same number of families, which, as we have mentioned, is an advantage from the viewpoint of sampling efficiency. Rural districts are divided into open-country segments or "count units" which average about two or three square miles—more in lightly populated areas and less in heavily populated areas. The boundaries of these small areas usually are easily recognizable demarcations such as roads and streams in order to make the identification of the areas easier. All this information is recorded on maps of individual counties and can be procured at a modest cost from the U.S. Bureau of the Census in Washington, D.C. One of the first steps in a national area sample is to obtain the relevant county maps from the Bureau. Certain material supplementary to the maps, namely, lists of available aerial photos for the county and a record of the estimated number of farm and nonfarm housing units in each open-country segment, can also be obtained from the Bureau.

Unfortunately, the maps of the Master Sample do not mark off small areas of the open country outside the limits of cities in metropolitan counties. However, maps of such counties can be obtained and the sampler can either mark off the area into segments or use some of the short-cut devices discussed in Chapter 10.

1. Numbers of Clusters at Different Stages. Previously we discussed the kind of reasoning and investigation required to determine roughly the best size of clusters for cluster sampling without subsampling. Later we pointed out that the same kind of analysis could be applied to cluster sampling with subsampling in order to find the best rate of subsampling within clusters. We were able to generalize that the optimum rate of subsampling within clusters is usually relatively small; that is, relatively few families per block should ordinarily be interviewed.

It would have been possible to make a qualification to this generalization. If the variability of cluster means is relatively large by comparison with the variability within clusters, the best subsampling rate within clusters would be lower and the total number of clusters within the sample larger than if the opposite situation prevailed. That is, if the differences between clusters are relatively large, it would be desirable to have a relatively large number of clusters within the sample (with relatively few interviews within clusters). If the differences between clusters are relatively small, a relatively small number of clusters and a relatively large number of interviews in each cluster would be best.

In multistage sampling with more than two stages, the same kind of

analysis is used in determining the number of clusters at each stage. The general nature of the conclusions is roughly the same. For example, if the variability *between* county means was relatively large, one would want to include a relatively large number of counties in the sample. If the variability of block means *within* counties was relatively large, a relatively large number of blocks within counties would be sought. Finally, if the variability of families within blocks was relatively large, it would be desirable to include a relatively large number of families within blocks. In each case, of course, "relatively large" variability at one stage is to be interpreted as meaning "relatively large" with respect to that prevailing at the other stages.¹

As a practical problem in marketing research, the determination of the number of clusters at different stages of the sampling process is either relatively simple or quite complicated. It is relatively simple if the marketing research man elects to follow closely in the pattern set by other samplers who have been investigating the same problem. If a statistician skilled in probability sampling is not available, this is probably the wisest course to take. More or less intuitively, of course, it is possible to make minor modifications in these designs. For example, the U.S. Bureau of the Census uses about 70 counties or small groups of counties as primary sampling units in its monthly survey of the labor force. Within these counties, subsampling is quite intense. The purpose of this survey is to determine certain characteristics of the labor force, particularly employment and unemployment. The variations in the proportion unemployed in different counties are probably relatively small compared to variations in factors of interest in marketing research when, say, sales are under scrutiny. For in one county a manufacturer may have no distribution at all and in another he may have such strong distribution that he dominates the local area. Intuitively, then, if a marketing research man had the same amount of money for sampling as the U.S. Bureau of the Census, he would want more than 70 counties with less intensive interviewing within counties. In most marketing research applications, the amount of money for sampling would, of course, be much less than the amount available for the Sample of the Labor Force. In trimming down the sampling design to marketing research budgets, then, one would almost certainly cut down relatively more on subsampling within counties than on the total number of counties selected in the sample.

The problem of determining the number of clusters at different stages of the sampling design can also be approached by direct study and analy-

¹ For completeness, it should also be mentioned that costs are an important influence in determining the best size of cluster. If the total budget and the size of clusters are fixed, the number of clusters is automatically determined.

sis rather than by modifying the methods of other surveys, if an expert (i.e., well trained both theoretically and practically) sampling statistician is available or if it is possible to consult with such a statistician. This is the second method mentioned in the preceding paragraph.

8.3 Stratification

Stratification is another basic idea of probability sampling. In practice stratification is often used in combination with cluster sampling, but it involves a different idea entirely. The term "stratified sampling" appears frequently in popular accounts of sampling and is more or less understood by the layman. Unfortunately, however, the term is also very frequently misunderstood. For that reason it is necessary to proceed with caution in clarifying the basic issues before becoming involved in the application of stratification to the sampling of human populations in general and to area sampling in particular.

Stratification can be defined as the process of breaking up a larger group into smaller groups called "strata." The forming of clusters also resulted in breaking up a large group into smaller groups, but aside from this formal similarity, the two processes, stratification and clustering, are really quite different. One important difference lies in the fact that *all* the strata of the population are automatically included in a stratified sample, while only a *sample* of the clusters in the population is selected in a cluster sample.

We shall again proceed by considering an example. In trying to estimate the average income in a city by selecting a probability sample by the process of cluster sampling, it is possible that the sample would include too many blocks from high-income areas or too many blocks from low-income areas. Such a result is possible with probability sampling even though it is obviously in itself undesirable. By stratification it is possible to prevent this undesirable side effect of probability sampling without compromising probability sampling itself.

Suppose, for example, that the large group of blocks which comprise the city is broken up into smaller groups consisting of high-income blocks, medium-income blocks, and low-income blocks. Each of these three groups of blocks would be considered to be a stratum. This sort of stratification can be carried through with the aid of census data which provide information on the average rentals in each block for all large cities as of a previous census. Since it is true that rentals (even past rentals) are highly correlated with (present) incomes, these census data on rents can be used as a basis for stratification. After the process of stratification, blocks would be selected within strata by some random process, say, simple random sampling. (Subsampling within selected blocks, if desired, would be done

in the customary way.) Although blocks are selected within strata at random, the number of blocks selected within strata is fixed in such a way that the number of blocks within each stratum in the sample is proportional to the number of blocks within that particular stratum in the universe.¹ Thus the sample would contain the same proportions of high-, medium-, and low-income blocks as there were in the universe from which the sample was selected.

How has the sample been improved by stratification? Stratification automatically assures that the sample will include the appropriate number of blocks in each of the rental classifications. At the same time the sample is still a probability sample. In effect, the stratification of the city into three groups of blocks according to rental levels has created three separate populations (which, of course, ordinarily are called strata). Within each of these subpopulations a simple random sample of blocks (that is to say, a cluster sample) has been selected. Therefore in each of the subpopulations it is possible, in the manner already described, to calculate the probability that any family would be included in the sample. It makes no difference, then, whether one views the process as three separate probability samples or one large probability sample. The essential point is that by introducing randomness within each of the strata or subpopulations it is possible to calculate the probability of including any family in the sample, just as in simple random sampling or in the various cluster sampling designs.

An intuitive explanation of the desirability of stratification is possible. It seems desirable, intuitively, that the various groups in the universe be weighted in the sample according to their relative numerical importance in the universe. One way of achieving this is to see that the numbers of the various groups in the sample are proportional to the numbers of the corresponding groups in the universe.² By numerical proportionality in selection, it is possible to achieve the correspondence between sample and universe which intuitively seems to be desirable in achieving accurate sample estimates.

The next step in the discussion of stratified sampling is to examine the sampling distribution of the mean for stratified samples. For simplicity, think of a situation in which there is no clustering at all. The universe

¹ This proportionality is not always desirable, but for the purpose of simplified exposition at this point we shall assume that it is. A more complete exposition will be given later in this chapter.

² Another way of achieving proportionality is by weighting. Thus, although the sample may not have the various groups represented numerically in proportion to their numerical importance in the universe, the sample groups can be weighted into proportionality. This weighting is a computational procedure applied after the sample has been chosen.

might be the 10,000 charge-account customers considered in the preceding chapter. These accounts might already be kept separately by the section of the city or the suburbs in which the customers live. Therefore, instead of taking a simple random sample of these accounts, it would be possible to take a stratified sample. Assume, as before, that the proportion of accounts selected is the same in every stratum of the population.

The sample mean of the stratified sample is an unbiased estimate of the mean of the population. This is easy to see intuitively. The stratified sample can be thought of as a collection of simple random samples. In each of these subsamples, the mean is an unbiased estimate. The mean of the stratified sample is a combination of the means of the subsamples and is therefore itself unbiased.

The standard deviation of the sample mean is relatively simple to figure out. In each stratum (ignoring the finite population multiplier), the standard error of the s th stratum mean is simply

$$\frac{\sigma_s}{\sqrt{n_s}}$$

where σ_s is the standard deviation of furniture purchases of charge-account customers in the s th stratum and n_s is the sample size in the s th stratum. Next, let w_s represent the ratio of the number of charge-account customers in the s th stratum to the total number in the population (10,000, in this example). If there were 1000 customers in the s th stratum, w_s would equal 1000/10,000, or 1/10. One simply computes $\left(w_s^2 \cdot \frac{\sigma_s^2}{n_s}\right)$ for each stratum, adds up these individual terms for all strata, and extracts the square root. The result is the standard error of the mean. In order to obtain the *sample estimate* of the standard error of the mean which would be required to make an (approximate) confidence interval from such a sample, it is necessary simply to replace σ_s in the above formula by s_s , the standard deviation computed from the sample observations in the s th stratum.¹

In the computation of the standard error of the mean, the only standard deviations involved were the σ_s 's, that is, the standard deviations of observations *within strata*. In simple random sampling, by contrast, the standard deviation of relevance is the standard deviation of all observations in the population, which we have called σ . The variability of all observations in the population, in turn, reflects two components: (1) the variability of observations within strata, and (2) the variability of the strata means, which is often called "between strata" variability. But only

¹ It should perhaps be reemphasized here that whenever a sample standard deviation is used as an estimate of the population standard deviation, it is appropriate to divide by $n - 1$ (in this case, $n_s - 1$) rather than by n .

the first source of variability contributes to the standard error of the mean of a stratified sample. The second source, variability between strata means, is removed by the process of stratification. To see this, we abstract from *within strata* variability. Suppose that each of the estimates of the stratum means, computed from the observations in a simple random sample, happened to equal the mean of the particular stratum in the universe. The estimate of the mean of the whole universe would still be in error so long as (1) the proportion of sample observations coming from each stratum in the population was, owing to sampling error, different from the true proportion in the universe, and (2) the means of the different strata differed among themselves. The estimate would be more in error the greater the differences among the strata means. The greater the *between strata* variation, then, the greater the error. In stratified random sampling, the situation just described is completely reversed. The proportions are known but the strata means are not. Hence the *between strata* variation has no influence on the error of a stratified sample. Only the *within strata* variation has an effect. Hence, *at worst*, stratified sampling is as efficient statistically as simple random sampling (for a given number of observations) and, *at best*, stratified sampling can be much more efficient than simple random sampling.

This fact permits an important simplification of practical computations for confidence intervals. If it is assumed that no gain results from stratification, the confidence interval can be computed just as if stratification had not been introduced. The correct confidence interval can be no wider than the confidence interval so calculated, but may be smaller. In other words, an estimate of the upper limit on the sampling error is ascertained. Frequently, in practice, one may be willing to use such a confidence interval, which would be calculated by the methods already outlined for simple random sampling and cluster sampling, because of the computational simplicity.

It is now possible to see the criterion for forming the members of a population into strata. If σ_s were zero for all strata, the sample mean would be subject to no sampling error at all. To make σ_s zero for all strata, it would be necessary to form the strata in such a way that all the members of any given stratum were exactly alike with respect to the characteristic being studied—*e.g.*, purchases of furniture. In practice this is usually impossible. However, it is often possible to form strata in such a way that the elements of each stratum are relatively similar with respect to the characteristics under study. An alternative and equivalent statement of the same principle is that the variability of strata means should be very high; *i.e.*, the differences *between* strata should be relatively large.

This brings out clearly another very vital difference between stratified

sampling and cluster sampling. In stratified sampling the population is broken down into groups called strata so that the elements within strata will be as *alike* as possible with respect to the characteristic being studied and (equivalently) so that the differences between strata will be as large as possible. In cluster sampling the population is broken down into groups called clusters so that the elements within clusters will be as *dissimilar* as possible with respect to the characteristic being studied and (equivalently) so that the differences between clusters will be as small as possible. In practice, as we show later, it is possible successively to break down a population into both clusters and strata so as to fulfill both these objectives.

Internally similar strata can, to some extent, be obtained by splitting the universe into strata on the basis of some characteristic or characteristics which are believed to be highly correlated with the characteristic under investigation. In the example, the large group of blocks was broken up into three smaller groups of blocks on the basis of rentals because it was believed that rentals were highly correlated with incomes. The correlation would not be perfect, but in general, the blocks where average rentals are high would be blocks where the average incomes are high, and the same correspondence can be expected for medium- and low-income blocks. This subdivision results in creating smaller groups, within each of which the families are relatively similar with regard to income. In general, by stratifying according to a characteristic which is correlated with the characteristic being investigated, strata are obtained which are internally more similar with respect to the characteristic being studied than is the universe as a whole. Hence information about the universe can be used to improve the efficiency of the sample.

1. Combination of cluster sampling and stratified sampling. One of the easiest ways to visualize the combination of cluster sampling and stratified sampling in practical area sampling designs is the following. Think of the sample design as being basically a multistage cluster sample. That is, the population is formed into big clusters (*e.g.*, counties), and a sample of these clusters is drawn. Then each of the big clusters which is drawn in this sample is subdivided into small clusters (*e.g.*, blocks and open-country segments), and a sample of these small clusters is selected. Finally, a sample of families or individuals is selected within the sample of small clusters.

Thus at three stages of the process samples are drawn. These samples can be simple random samples, and during the discussion of cluster sampling we assumed that they were. However, there is no reason for not selecting stratified samples rather than simple random samples at each of these stages. It has been seen already that a stratified random sample is at worst as efficient (statistically) as a simple random sample and that it

is often more efficient. Therefore there is nothing whatever to lose by stratification except the time and expense needed to effect stratification.¹

We shall now examine ways in which stratification is introduced in practical sampling designs. Stratification is almost always introduced in the selection of primary sampling units on the order of size of counties. The idea, of course, is to form counties into groups of counties which are relatively similar with respect to what are believed to be close correlates of the characteristic under study. The degree of urbanization might be closely correlated with many characteristics of interest in marketing research, and the researcher might therefore wish to form groups of counties that were relatively alike with respect to urbanization. Another factor frequently correlated with characteristics of interest in marketing research is the region of the country within which the county is located. It might be desirable, for example, to further subdivide groups of counties of similar degrees of urbanization into smaller groups that were from the same part of the country. This process of subdivision can be carried as far as one likes, using, for example, such factors as type of industry, type of farming, and extent of home refrigeration.² It has been found in practice, however, that an extremely simple method of subdivision tends to yield good results for wide ranges of problems. This simple method of stratification is known as geographic stratification and involves nothing more than the formation of strata consisting of geographically contiguous counties. It is frequently used in combination with stratification by degree of urbanization.

The degree of fineness of stratification is ultimately limited only by sample size. In selecting 70 counties (out of approximately 3070 counties in the whole country), stratification would be carried to the point at which there were 70 strata. Then one county would be selected at random within each stratum.³ For there is nothing to lose except the time required to stratify, and added statistical and economic efficiency may possibly be obtained by each further subdivision of the population.

Statements of the kind just made—that there is nothing to lose by stratification—could be denied with justification under certain circumstances. If, for example, the data which formed the basis for the stratification failed to indicate correctly the relative size of the strata in the universe,

¹ We are assuming that the proportion sampled in every stratum is the same, and also that the exact sizes of the strata are known, as is always true in area sampling, since strata consist of areas which can be counted on maps.

² Information about factors such as these is to be found in the *County Data Book* and other sources mentioned earlier.

³ As we mention in the discussion of sampling with varying probabilities, it might be desirable to assign varying, rather than equal, probabilities of selection to the counties within a stratum.

the resulting sample mean would be a biased estimate. Suppose, for example, that the problem was the estimation of per capita cigarette consumption. It is assumed that the population consists of 50 per cent males and 50 per cent females. If the true proportion were 60 and 40, the resulting sample mean would be a biased estimate if the sample included 50 per cent men and 50 per cent women.

The magnitude of this bias would depend on the difference in per capita cigarette consumption for men and women. It would be zero if and only if there were no difference in average per capita cigarette consumption between men and women. This possibility of bias exists, and it has, unfortunately, led occasionally to the erroneous belief that stratified area sampling estimates are also biased. Stratified area sampling never need be biased, regardless of the antiquity of the data by which areas are assigned to strata. So long as the size of the strata (*i.e.*, number of areas) can be correctly ascertained, no bias need exist.

Fortunately for area sampling one can always know the total number of areas in the universe and the number of areas in each stratum. Even if counties (or blocks) were assigned to strata by some random process, bias would not be injected into the sample. The only unfortunate consequence of such a process of stratification—or of a process based upon the use of inaccurate data—would be a loss in the statistical efficiency of the sample. The use of out-of-date data can increase the sampling variability, but for area sampling, the use of such data need not cause bias. So long as proportional sampling is assumed within strata, this increase in sampling variability is limited by the fact that the sampling variability of a stratified random sampling process cannot exceed the variability of a simple random sampling process if the sample sizes are the same.

The second stage of random selection is the choice of secondary sampling units (*e.g.*, blocks) within those primary sampling units selected at the first stage (*e.g.*, counties). Here again stratified sampling rather than simple random sampling can be used. A mode of stratification often used is the division of counties in the sample into open-country areas and town areas. Still finer stratification can be used within each of these two strata. City blocks, as in the earlier example, might be stratified according to average rental values. Alternatively, it would be possible to stratify geographically. That is, contiguous city blocks could be grouped together into strata, while in the rural part of the county, contiguous open-country segments would be grouped together into strata. It has been found that, for many problems, geographic stratification of secondary sampling units is statistically as satisfactory as, and much simpler and cheaper administratively than, other more elaborate modes of stratification.

Finally, when blocks and open-country segments have been selected,

it is possible to stratify the individual families living within those areas in order to improve the final stage of the sampling process. Usually the expense of elaborate methods of stratification at this stage is large and the statistical gains are small. The practical solution usually is to use neither simple random sampling nor stratified sampling, but still another probability sampling design of great administrative simplicity which has not yet been discussed. This design, called "systematic sampling," is treated in the next chapter. It achieves approximately the same effect as simple geographic stratification in stratified sampling.

CHAPTER 9

FURTHER COMMENTS ON PROBABILITY SAMPLING FROM HUMAN POPULATIONS

9.1 Additional Refinements of Probability Sampling

In the discussion of probability sampling up to this point we have shown how two important ideas—clustering and stratification—are used in obtaining a probability sample of human populations. We have followed some of the ramifications of these ideas into actual sampling designs. In doing so it has been possible to obtain a fairly realistic picture of how probability samples are selected and what the finished product looks like. We have been able to show the close relationships between more elaborate probability sampling designs and the theory of simple random sampling which was discussed in Chapter 7.

But the demands of practical problems and the ingenuity of statisticians who have worked on these problems have combined to produce a large number of additional refinements of the basic methods that we have sketched up to this point. It would be nearly impossible, of course, to list all these refinements, and if we did, the list would be out of date within a month. It is worth while, however, to examine briefly some of these refinements in order to enrich one's understanding of the important features of modern sampling design.

1. Variable probabilities of selection. Up to now we have assumed the probabilities of selection, whether of clusters or families, to be equal at all stages of sampling. While this simplifies the theoretical exposition, it is frequently desirable to depart from this practice in order to obtain sampling designs of greater economic efficiency. That is, the probabilities of selection of different clusters might be made unequal rather than equal. This in no way compromises probability sampling since probability sampling requires known but not necessarily equal probabilities of selection.

Unequal probabilities of selection are usually applied in multistage cluster sampling in such a way that the final probability of selection for all families is the same. This general statement may be illustrated with a concrete example of the simplest kind of multistage sampling, cluster sampling with subsampling. Assume that the sizes of the clusters are unequal. For simplicity consider a universe made up of only three blocks, one of which contains 40 families, the second, 10 families, and the third,

200 families. These blocks could be tabulated as follows:

<i>Block number</i>	<i>Number of families</i>	<i>Cumulative number of families</i>
1	40	40
2	10	50
3	200	250

where the third column, "Cumulative number of families," will be explained shortly.

First consider the method of selecting clusters with equal probabilities. One block of the three is selected at random and then $1/5$ of the families are subsampled in whichever block is chosen. The probability of selection for each individual family in the universe is therefore $1/3 \times 1/5$, or $1/15$. If block 1 happened to be selected, 8 of the 40 families in this block would be sampled and the final sample would consist of 8 families. If block 2 were chosen, however, the sample would consist of only 2 families ($1/5 \times 10$). Similarly if block 3 were chosen initially, the final sample size would be 40!

These results may be quite surprising. For one thing, they illustrate the point that it is sometimes impossible to tell in advance how many families will be selected in the sample, even though it is known that all families in the population have an equal probability of selection. If the final number were as unpredictable as in this simple example—where the sample could consist of 8, 2, or 40 families, depending on which block was chosen—the administrative problem would be great.

It can readily be seen, moreover, that the selection of blocks with equal probabilities and the subsampling of a constant proportion of families in blocks chosen in the sample will lead to considerable variability in the amount of interviewing from one block to the next. This variability is undesirable for administrative reasons. It is frequently undesirable for statistical reasons as well: in the original discussion of cluster sampling with equal-sized clusters it was pointed out that the standard error of the mean is increased by variability in the size of the clusters.

There are three reasons, then, for departure from equal probabilities in the selection of clusters: (1) to secure better predictability of the total sample size, (2) to obtain easier administration of fieldwork, and (3) to obtain a more efficient sampling design.

The next step is to consider a way of improving on the design which has equal probabilities of cluster selection. Administrative procedure can be improved simply by specifying that, regardless of what block happened to fall in the sample, a sample of five families would be drawn. Surprisingly enough, it is possible to combine this requirement with the specification

that each family in the population have an equal chance of selection in the sample. Assume that each family is assigned a probability of selection equal to 5 chances in 250, or simply 1 in 50. This probability must be equal to the product of the probability of selecting the block in which the family lives and the probability of selecting the family, once its block has been selected. These two probabilities must be adjusted in such a way that the probability of selection will in fact be $1/50$ and so that exactly five families will be selected in the sample.

This can be done if, instead of assigning equal probabilities of selection to each block, one assigns probabilities proportional to the *size* of each block. In other words, block 1, which has 40 of the 250 families in the population, would have a probability of selection of $40/250$. The probability of selection for a family within block 1, once block 1 has been selected, would be $5/40$, since exactly five families must be selected, regardless of the block which has been chosen. Multiplication of these two probabilities, $40/250$ and $5/40$, gives the desired probability of $5/250$ or $1/50$. A similar result is obtained by letting the remaining two blocks have probabilities of selection proportional to their size. That is, the second block would have a probability of $10/250$, and the third block a probability of $200/250$.¹

In actual practice, of course, the current size of each cluster is usually not known before the survey has been made; and after the survey, the size of only those clusters which had been included in the sample would be known. An approximation can be achieved by using information such as the preceding census. For example, for cities whose populations exceeded 50,000 in 1930, the number of dwelling units per block (for 1940) is listed. The number of dwelling units is of course highly correlated with the number of families. There have been many changes in the number of families since the 1940 census, but by and large, blocks which were heavily populated in 1940 were heavily populated in 1950, and conversely. In other

¹ The mechanics of selection with probability proportional to size are relatively simple. In the original table there are 250 families in the three blocks. Each block is given a probability of selection proportional to the number of families it contains. Block 1, for example, has 40 chances in 250 and is assigned the numbers 001 to 040. Similarly, block 2 has 10 chances in 250 and is assigned the numbers 041 to 050. Block 3 would be assigned the numbers 051 to 250. The right-hand column in the table, "Cumulative number of families," indicates how this process can be applied quickly. The cumulative totals correspond with the highest number assigned to each block. A random number is selected in the range from 001 to 250, using a table of random numbers, precisely as in the simple illustration in Chapter 7. Suppose that the random number turned out to be 038. This number would be one of the 40 numbers assigned to block 1. Therefore block 1 would be selected for the sample.

words, there is a high correlation between the number of dwelling units per block in 1940 and the number of families per block in, say, 1950.

Imagine that the figures on size of block in the above illustration are not current figures but data from the 1940 census, and that the over-all probability of selection in 1950 is to be $1/50$. Probabilities of selection for each block would be exactly as before. Block 1, for example, would have 40 chances in 250, and if it were chosen, each family in the block would have a probability of 5 in 40 of being selected (instead of saying flatly that 5 families were to be chosen). Then the over-all probability of selection for every family in block 1 would be $40/250 \times 5/40$, or $1/50$, just as before. Similar provisions would be made for the remaining blocks.

But suppose that between 1940 and 1950 the number of families in block 1 had increased from 40 to 48. In order that the probabilities of selection for families remain equal, it would be necessary to have some rule established to enable the fieldworker to make the necessary adjustments in the field on the basis of the information regarding the change in the number of families in the block. The rule is easy. All that is necessary is that the ratio of interviews secured to the number of families in the block equal the ratio that was originally planned. In the example, it was originally planned to have, in that block, five interviews for the 40 families—a ratio of $1/8$. It would therefore be necessary, after it was discovered that there were really 48 families, to take six interviews in that block in order to preserve the ratio of $1/8$. Since population changes are usually gradual, the use of such a rule would retain the main advantages of the process of selection with probabilities proportional to *known* (rather than estimated) size. In practice, of course, it is almost always necessary to use estimated size.

Thus, again, in employing probability sampling it is possible to use information about the population being sampled in such a way that inaccuracies in the information need not destroy the merits of the probability sample. As changes in the population occur, a probability sample will tend to reveal them within the limits of sampling error. In fact, one of the most interesting uses of probability sampling in government work is in the estimation of total population.¹ Estimates of population size from probability samples are frequently referred to as "sample censuses."

The advantages of selection with probabilities proportional to size are especially great in selecting counties as the primary sampling units, and therefore counties are almost always selected in this way. This means

¹ We have chiefly discussed the use of probability samples in the estimation of means, but samples are useful in estimating other characteristics, such as the size of the population, as well.

that the most populous counties have a much greater chance of coming into the sample. The probabilities of selection within the counties are adjusted in such a way that each *family*, regardless of the size of the county in which it lives, has an equal chance of inclusion in the sample.

It might be helpful to carry through one more example, more complete than the ones already discussed, to illustrate why equal probabilities in the selection of counties is not desirable. Assume that the problem is to select a probability sample from the families living in the continental United States. Suppose that the latest census estimate indicates that about 42,000,000 families are in the United States, and that a sample of about 4200 is to be drawn. To obtain a sample of roughly this size it is necessary to pick about 1 family in 10,000 (since 4200 divided by 42,000,000 is $1/10,000$). The number $1/10,000$ is called the "sampling ratio."¹

The procedure is to select counties, blocks within counties, and families within blocks. The required probability of including each family could be achieved by selecting 1 county in 100, 1 block in 10 in the county selected, and 1 family in 10 in each block selected ($0.01 \times 0.1 \times 0.1 = 0.0001 = 1/10,000$). With this arrangement the probabilities of selection for the various clusters would be equal. But suppose that Cook County, Illinois, was one of the counties selected, that in it there were 30,000 blocks (and open-country segments), and that the average number of families per block was 40. According to the sample plan just discussed, 1 block in 10 would be chosen, or 3000 blocks in all. One family in ten within selected blocks would give roughly 12,000 families in this one county alone. This number would exceed the total number originally planned for the whole national sample (4200) and would place a very heavy proportion of the total field-work on the interviewers in the Cook County area. In short, the design would be economically and statistically inefficient.

The use of equal rates of subsampling (with associated equal probabilities of selection) thus makes it possible for the actual sample size to differ widely from the sample size that had been desired. This deviation exists because it is quite possible for the clusters selected for the sample to be clusters having a much higher than average population. The selection of such clusters, when equal rates of subsampling are used, results in the acquisition of a much larger sample than was desired. Conversely, it is possible for the sample to be much smaller than was desired. Further, the work of interviewing may be very unequally distributed when equal rates of subsampling are used.

These two problems—almost complete unpredictability of sample size and inequality in the work for the field staff (which is even more serious

¹ The reciprocal of the sampling ratio is called the "sampling interval." In this example the sampling interval would be 10,000.

or counties than for blocks within counties)—are mitigated when approximately equal numbers of observations are selected within clusters which are selected with probabilities proportional to the estimated size of the cluster.

Finally, it is not too efficient to select clusters with equal probabilities, at least when the clusters considered are counties. In the example just given, it is quite possible that more than half the interviews in what purported to be a national sample would come from Cook County, Illinois! The sampling process would be unbiased, but the sampling variability would probably be immense.

The use of unequal probabilities for the selection of counties would usually result in a large county like Cook County being assigned a probability of 1—certainty—of inclusion. The economic efficiency of the sampling method is usually relatively high when this is done. One block in one thousand might be sampled within the county, and 1 family in 10 chosen within blocks which were selected. This would give the desired sampling ratio of $1/10,000$, and every family in Cook County would have an equal chance of inclusion ($1 \times 0.001 \times 0.1 = 1/10,000$).

Or, it would be possible to assign variable probabilities of selection to the individual blocks and adjust the probability of selection within blocks so that the desired probability of $1/10,000$ was maintained. The three probabilities (selection of counties, blocks within counties, families within blocks) can be juggled in any way so long as their product is $1/10,000$. The idea is to juggle in such a way as to obtain an efficient and easily administered sampling design.

As a matter of practice, however, information on estimated numbers of families in small units like blocks is much less complete than for large units like counties. For blocks in large cities, census data are available, while in some small- and medium-sized types of towns, detailed maps (*e.g.*, Sanborn Insurance Maps) can be secured and the size of blocks can thereby be approximated. Personal inspection ("eye estimates") is also useful. In the open-country areas, the number of farm and nonfarm dwellings is approximated on the maps of the Master Sample of Agriculture; and the supplemental sheets to these maps, which were mentioned earlier, actually list these in cumulative form. Frequently, however, for small clusters such as blocks and open-country segments, the time and effort of selecting with probabilities proportional to an estimate of size are great; frequently, also, the necessary data are unavailable or difficult and expensive to obtain. Often, then, the smaller clusters may be selected with equal probabilities. Larger clusters, such as counties, should seldom be so selected.

2. Disproportionate stratified sampling. Throughout the earlier discussion of stratified sampling it was assumed that the proportion of ele-

ments sampled in each stratum—*i.e.*, the sampling ratio—was the same for all strata. The method of stratified sampling, discussed earlier, reproduces the proportionality of the universe in the sample. If, for instance, there were 100 high-income blocks, 600 medium-income blocks, and 300 low-income blocks in a city being investigated, a sample designed in this way could have 10 high-income blocks, 60 medium-income blocks, and 30 low-income blocks. If the sample were designed in this way, the “sampling ratio” would be said to be 1/10 in all three strata. The proportionality of the sample could be preserved by any other sampling ratio, such as 1/20, so long as the ratio was the same for all strata. But this is only one of two ways of reaching the desired objective of assigning weights to the various groups in the sample in proportion to their relative frequency in the universe.

For the sampling ratio does not have to be, and sometimes should not be, the same for all strata. One reason for unequal sampling ratios is the existence of differences in the cost of obtaining interviews in the various strata. For example, the cost of obtaining an interview in high-rental blocks might be more than in medium- or low-rental blocks. In such a case it would be economical to select relatively fewer interviews in the most expensive strata—the high-rental blocks. The reasoning behind this procedure might go like this: “By cutting out some of the most expensive interviews and spending the money thus saved on less expensive interviews, more total interviews would be obtained. If this process isn’t carried too far, and if the results are weighted correctly by the number of clusters in each stratum,¹ it seems reasonable that the ‘disproportionate’ sample might be more efficient economically than the proportionate one.”

Formal mathematical analysis supports this reasoning. The standard error of the weighted mean depends on the standard error of the means of the individual strata. Nonproportionate sampling increases the standard error of the means of the “expensive” strata (where relatively fewer observations are taken) and decreases the standard error of the means in the “cheap” strata (where relatively more observations are drawn). Up to a certain point, as the sample is made more and more disproportionate, the reduction of error in the cheap strata more than offsets the increases in error in the expensive strata, and the standard error of the weighted mean is reduced. Hence the standard error of the (weighted) mean in disproportionate sampling can be less than the standard error of the mean in

¹ Thus, if the proportion of high-income blocks selected for the sample was only half as great as the proportion of low- and medium-income blocks selected, the high-income blocks would have to be given a weight twice as great (*i.e.*, in effect, counted twice) as that given to the low- and medium-income blocks. Hence it is necessary to weight the sample results rather than use the (unweighted) mean of the sample observations.

proportionate stratified sampling, assuming always that the total cost is fixed. But beyond a certain point (which is called the "optimal allocation"), further departures from proportionality cause the error of the weighted mean to increase. If the point of optimal allocation is badly missed, the standard error of the weighted mean may be greater than the standard error of the mean for proportionate stratified sampling or even for simple random sampling.

In practice, differential costs of obtaining interviews are frequently encountered. For example, interviewing costs might be higher in very large cities because of the difficulty in finding people at home. It might be wise, therefore, to select a smaller proportion of clusters in the very large cities. It is necessary, of course, to weight the sample results in order to make an estimate of the mean of the population.

The existence of cost differentials between strata is not the only reason for the use of disproportionate sampling. A second reason can be illustrated by the problem of estimating the incidence of pregnancy among the adult population of the United States. One mode of stratification would be by sex, since sex is highly correlated with the incidence of pregnancy. Suppose, further, that half the relevant universe were male and half female. It would obviously be unwise (inefficient) to select a sample composed of 50 per cent males and 50 per cent females. The reason for the abandonment of proportionate sampling is not the difference in the cost of making observations of males and females, but rather the difference in *variability* in the pregnancy of males and females. An errorless estimate of the incidence of pregnancy in the male stratum can be obtained by a single observation.

Another slightly more probable example illustrates the same point. The problem is to estimate the proportion of automobile licenses in this country having slogans on them (*e.g.*, "America's Dairyland"). It would not be necessary to make any more observations in the stratum consisting of New York licenses than in the stratum consisting of Nevada licenses, even though the former stratum is much the larger. A perfect estimate of the number of slogans in each stratum could be made by taking only a single observation in each stratum, since the elements within each stratum are perfectly alike with respect to the presence of slogans. Naturally, in computing the proportion for the whole population having slogans, the observation from the New York state stratum would have to be weighted relative to the Nevada stratum in the same ratio as the ratio of the size of the New York to the Nevada stratum.

A generalization of these examples would be that the sampling ratio should be relatively high in strata containing elements which are relatively dissimilar with respect to the characteristic being studied and relatively

low in the strata containing elements which are relatively similar. Differences in the variability within strata are, then, the second reason for the use of disproportionate sampling.¹ Regard for this second reason in the study of pregnancy would have resulted in great savings. By using only one man in the sample and using the money that would otherwise be wasted on men for the selection of a larger sample of women, one could make a more precise estimate of the incidence of pregnancy in adults. A greater over-all accuracy could thus be achieved for the same expenditure of money; the economic efficiency of disproportionate stratified sampling would, in this example, be greater than for proportionate stratified sampling.

The same principle works, of course, with less extreme examples. Families living in high-rental blocks might have incomes with great variation, perhaps ranging from \$2000 to \$200,000 a year. While the average income is high, there is also a great dissimilarity in incomes. Families in the low-rental blocks, by contrast, might have much more similar incomes. Perhaps the range would be from \$500 to \$4000. In order to estimate income for the whole group, a relatively small sampling ratio in the low-rental stratum and a relatively high sampling ratio in the high-rental stratum would be called for.

In actual practice, disproportionate stratified sampling is used somewhat less often than might be expected on the grounds of its theoretical advantages. One reason is that the calculation of weighted means is cumbersome. There is some danger of numerical errors in calculations, and these errors might more than offset the statistical gains, especially when a great many different items are studied in the same survey and the number of calculations required is large. A more important reason is that knowledge of costs and, more important, of the similarity or dissimilarity of elements in different strata is quite imperfect. Nevertheless, considerable inaccuracy is allowable; one can come close to the optimum allocation of a sample with relatively imprecise knowledge. There are, of course, limits to the inaccuracy of knowledge which can be permitted, however. For with a disproportionate stratified sample it is possible to obtain a statistically less efficient sample than with a simple random sample of the same size. With proportionate stratified sampling, by contrast, at the very worst the efficiency of the sampling design will equal that to simple random sampling.

To illustrate the possible dangers of disproportionate sampling, suppose that one had become confused in the pregnancy study and had decided to

¹ This conclusion is analogous to that reached earlier in the discussion of the optimum rate of subsampling within clusters. It was seen then that the rate of subsampling should be relatively small in relatively homogeneous clusters.

sample only one woman while devoting the remainder of the sample to men. The information from such a sample would be almost valueless. Either a simple random sample or a proportionate stratified sample would have been better. Hence disproportionate sampling resembles speculation; there are risks of loss and chances for big gains, and the gains usually accrue to those with some accurate knowledge.

There are a few administrative tricks worth knowing if one is executing disproportionate sampling. Suppose, in a study of income, that the variability of incomes in high-rental blocks was relatively great, making it seem desirable to take twice as many interviews in them as would be taken in a proportionate sample. Even though 2.4 times as many might be closer to the optimum, a researcher would probably choose the integral number 2 for administrative simplicity. The procedure in computing the weighted average would then be to count each interview in the high-rental block *once* and to count each interview in the remaining blocks *twice*. When tabulation is done by means of mechanical tabulating equipment the efficient computational procedure often would be to duplicate punch cards corresponding to the interviews that are to be counted twice.¹

3. Systematic sampling. It is sometimes inconvenient to select a simple random sample of families within a block or of blocks within a city, since the use of a table of random numbers requires office time or very elaborate and awkward field procedures. To select 100 blocks out of 1000 in a city, the use of simple random sampling would require the drawing of 100 random numbers and the identification of the corresponding blocks for inclusion in the sample. Stratified sampling, in addition to possessing theoretical advantages, would be much simpler administratively if some simple mode of stratification, such as geographic stratification, was used. An even simpler method, that frequently has the statistical advantages of stratified random sampling with geographic stratification, is systematic sampling.

In systematic sampling, it would be necessary to select but a single number from 01 to 10. Suppose the number chosen was 07. Then the seventh block would be included in the sample as would every tenth block from 17 through 997. The total number of blocks included would be 100 just as in simple random sampling. The interval between blocks (10 in this case) is determined by the proportion of blocks within the city that are to be

¹ There is some discussion of mechanical tabulation in Chapter 22. We should point out here that the duplication of punch cards may, in practice, have some undesirable side effects. People often tend to think that the accuracy of the sample is increased by the simple act of reproducing punch cards. Moreover, the analysis of other subgroups of the population (such as educational groups) is complicated by the fact that we do not know how many cards are originals and how many are duplicates.

included in the sample. In this case $1/10$ of the blocks were to be included, so every tenth block beginning with the randomly chosen seventh block was included.

A systematic sample is a probability sample, because every element in the population has an equal probability of being chosen. In this example, the blocks from 1 to 10 each had an equal chance of being chosen. Each of the other blocks from 11 to 1000 is "linked" by this systematic procedure to one and only one of the blocks from 1 through 10. For example, the blocks 17, 27, . . . , 997 are linked to the seventh block and therefore have the same probability of selection as the seventh block. Hence all blocks in the population have the same probability of selection.

Systematic sampling can also be thought of as a type of cluster sampling, where the clusters are not formed by grouping contiguous elements but by skipping systematically through a list of the elements in the population. In the example above, the population was divided into ten clusters. The sample actually selected—consisting of the 7th, 17th, 27th, etc.—was just one of these clusters. Hence the systematic sampling design really consists of the selection of a single cluster and the sampling of all elements within the cluster.

Just as in ordinary cluster sampling, then, the standard error of the mean of a systematic sample can be either greater or smaller than that of a simple random sample of the same size, depending on the nature of the population being sampled. In the kind of cluster sampling discussed earlier the sampling error was usually greater, because contiguous elements in a population usually tend to be similar, making the internal similarity of clusters relatively large. With systematic sampling, however, the elements within clusters tend in most applications to be dissimilar. Each cluster contains elements spread across the entire population.

The statistical advantages of systematic sampling have resulted in its widespread use in all kinds of sampling designs and at all stages of the sampling process. It is very useful for the subsampling of families within blocks. In the 1950 census, for example, certain questions were asked of every fifth individual from a random start among the first five individuals in each enumeration district.

It is important not to overlook the fact that it is theoretically possible for systematic sampling to be extremely inefficient. This theoretical possibility is seldom, if ever, important in area sampling, but in other applications it may be. For example, estimation of the annual sales of a department store during a year by selecting every seventh day's sales would almost certainly be a bad sampling design. There would be one chance in seven of obtaining nothing but Sundays, for example. Whenever there is any fear that such a situation might prevail, the best alter-

native is to take a stratified random sample, which cannot possibly be a worse approach than simple random sampling. In the department-store example, one might select a single day's sale at random from each week.

While systematic sampling is often more efficient than alternative methods, it has one serious disadvantage: the standard error of the mean cannot be estimated from the sample.¹ One way of getting around this is to make the assumption that the error is the same as that of a simple random sample of the same size. In most cases this will be a conservative estimate of the standard error; that is, the estimate tends to overstate the true error.²

Systematic sampling can be adapted for use when clusters are selected with varying probabilities. An illustration will make this clear:

<i>Block number</i>	<i>Number of families</i>	<i>Cumulative number of families</i>
1	40	40
2	10	50
3	200	250
4	150	400
5	50	450
6	50	500

Suppose two blocks are to be selected with probabilities proportionate to size. The total number of families, 500, is divided by 2, the number of blocks to be included. The result, 250, gives the range or interval between sampling numbers. It thus becomes necessary to select a single random number from 1 to 250. Suppose that the number 36 was chosen. Block 1 would be included. The interval to the next block is 250. Adding 36 and 250 gives 286, which falls between block 3 and block 4, and thus block 4 would be selected.

4. Double sampling, or two-phase sampling.³ Double, or two-phase, sampling is a process in which a small sample is selected within a larger

¹ Statisticians will readily see why it is impossible to estimate the standard error from the sample, if they will recall the fact that a systematic sample is really a cluster sample with a single cluster drawn into the sample. The estimate of the standard error of the mean in cluster sampling is based on the variation of the means of clusters drawn into the sample. If only one cluster mean is available, no such estimate is possible.

² Another approach of the same kind is to pretend that each three, say, successive observations represent drawings from a single stratum and to compute the variance as though the sample were a stratified random sample. Such a procedure, of course, will give an estimate of the standard error that would probably be conservative in most examples.

³ The terminology at this point is perplexing, since the term "double sampling" is used also for sampling methods other than the one discussed here and since the term "two-phase sampling" resembles the term "two-stage sampling," as used earlier. It is hoped that the discussion will make the difference clear.

one. A fairly large sample might first be selected from a human population and a variety of general information secured for each person in the sample. This general information could be used in order to stratify the large sample for the purpose of selecting a smaller sample of relatively great statistical efficiency. One large marketing research firm uses double sampling for many of its investigations. It has selected a probability sample of about 30,000 families, for each of which it has obtained information regarding family size, economic status, education, occupation, etc. This basic information is used to stratify the large sample for the purpose of specific investigations among subsamples of the large sample.

Double sampling is frequently useful in studying relatively small, hard-to-locate subgroups of a population. Owners of FM radios, for example, would be one such subgroup. With information for a larger sample which revealed the simple fact of ownership of FM radios, it would be possible to obtain more detailed information about attitudes toward FM programs, listening habits, etc., from the owners of FM radios at little extra cost. This could be accomplished in two ways. Interviewers could carry a special questionnaire (concerning attitudes toward FM programs, for example), which would be asked only of owners of FM radios as they were uncovered in the larger study. A second method could be applied if the large survey had been conducted in the past. This method would consider the families in the old sample as being divided into two strata, FM owners and those who do not own FM sets, and a new sample would be drawn from the old. The sampling ratio could be unity in the FM stratum—everyone would be interviewed. In the other stratum a relatively small subsample could be taken to estimate the proportion of people who had acquired FM sets since the original sample and to investigate their characteristics. In this second example, double sampling formally resembles disproportionate stratified sampling, since the little sample within the big sample is a disproportionate stratified sample.

Problems of the kind just discussed occur very frequently in marketing research. Instead of being interested in the entire population, the researcher is interested in some subgroup of the population, such as television owners, freezer owners, drinkers of Jamaican rum, etc. If special lists of these people are not available, the problem may be attacked either by double sampling or by nonprobability sampling methods.

Disproportionate sampling is exceedingly helpful if double sampling is used. For example, in a study of dealers in radio supplies one might design an area sample in which the sampling ratio would be extremely high in known important shopping centers and low elsewhere. Similarly, to obtain a sample of people living in homes valued at more than \$30,000 a

relatively high sampling ratio for blocks of known high rentals would be used.

One important advantage of double sampling is that it is possible to obtain comparable information for both the over-all sample and the subsample. For example, a subsample of owners of FM radios from a larger sample would permit a comparison of FM owners and FM nonowners with respect to characteristics thought to be important.

5. Sequential sampling. Sequential sampling is a process originally developed to increase the economic efficiency of the sampling processes involved in industrial quality control and sampling inspection. The basic idea behind sequential sampling is quite simple. Sequential sampling is probability sampling in which computations are made after each observation in order to find out whether the information already collected is sufficient to reach a conclusion according to some preconceived plan and with designated risks of error, or whether further sampling is needed. (On the average (that is, the average for repeated sampling) this sampling method requires a smaller sample size to obtain a given objective than would be required by ordinary simple random sampling. Occasionally, however, it is possible for sequential sampling to result in a larger sample than simple random sampling with equal accuracy.

Sequential sampling frequently makes savings possible in industrial quality control, but its potential use in marketing research seems limited. In general there are two conditions which must prevail if sequential sampling is to be superior economically to nonsequential sampling. Sequential is superior to nonsequential analysis whenever (1) the data become available serially and (2) the cost of the data (in terms of time, labor, material, etc.) is approximately proportional to the amount of the data.¹ The first condition can seldom be realized economically in marketing research except for telephone interviews within a single city or for sampling from files, list of customers, etc. Even when data can be obtained serially, the time between successive observations may be very great. For example, if a mail-order house wished to apply sequential sampling to a study of its customers, it would have to wait for complete replies to the first mailing before deciding whether to send out the second, etc. The second condition is almost never realized in marketing research, because sampling from human populations involves fixed costs that are of relatively great importance.

Sequential sampling also involves costs that are not present in ordinary simple random sampling. For example, computations must sometimes be

¹ Statistical Research Group, Columbia University, *Techniques of Statistical Analysis*, p. 419, McGraw-Hill Book Company, Inc., New York, 1947.

made after every observation or group of observations¹ and a decision made whether or not to continue sampling. In industrial applications these computations can frequently be streamlined to the point where they are made quite cheaply or even done in advance. In the sampling of human populations, however, the necessity to make computations after each observation or group of observations would involve practical complications. While industrial sampling may frequently be done by a single inspector, in marketing research respondents are usually called on by staffs of field-workers in widely separated areas. Problems of communication would typically be very difficult and would result in heavy administrative costs.

Another reason why sequential sampling is not at present useful for marketing research is that the theory of sequential sampling (and multiple sampling) has been worked out only for simple random sampling and is not yet available for the more complex and efficient sampling designs which we have discussed in this chapter. We have shown that simple random sampling is frequently unfeasible for the sampling of human populations. It would be even more unfeasible if attempted sequentially rather than with a fixed sample size.

One misunderstanding has resulted in confusion about the possible application of sequential sampling. There has been some tendency to regard the returns from mail surveys as being successive observations in a sequential sampling process. There is unfortunately no reason to believe that these observations arise by any random process, and a random process is essential for using the theory of sequential sampling to interpret the results. The early returns from a mail survey are usually from a segment of the universe consisting of people more interested in the survey than the rest of those who have received questionnaires. Or the early returns may differ with regard to other characteristics. It would be misleading, therefore—and a misapplication of statistical theory—to consider these early returns as successive observations of a random sample.²

So far we have tried chiefly to indicate drawbacks to the use of sequential sampling in marketing research. We shall now consider briefly a few instances in which this sampling procedure is applicable. The telephone survey has already been mentioned. In this kind of study the observations come in serially to a central office where simplified computational procedures can be developed and applied (although nonresponse, even here, would be a problem not easily solvable with the sequential technique).

¹ When groups are used, the process is called "multiple sampling."

² An analogous example may clinch the point. It would be disastrous to consider the earlier election returns to be a random sample of the universe of election returns and therefore the basis for a statistical forecast of election results. It is well known that the early returns are usually from the large cities and that these returns usually differ from the returns from the rural areas.

The cost of such a survey is roughly proportional to the number of observations. We have also suggested that, where the time lag required in successive mailings is not an important consideration, the process of multiple sampling might be applied to mail surveys. Finally, there are possible applications in internal office procedures. As has been said, sequential analysis was originally developed in response to problems of sampling the output of industrial processes in order to evaluate the quality of this output. Office output in marketing research is in many respects similar to factory output. Editing, tabulating, mailing, etc., are repetitive tasks whose quality may be measured. There is no reason why sequential sampling (and other methods of industrial quality control), should not be useful in evaluating these office operations.

9.2 A Note on Methods of Estimation

Throughout this chapter in the discussion of sampling designs it has been assumed that the mean of the sample observations would be used in estimating the mean of the universe. The only design in which this was not true was disproportionate stratified sampling, in which a weighted mean was used. While it may be surprising at first to find that there is more than one (good) way to estimate the mean (or other parameter) of a population from a given sample, this statement is nonetheless true. In Chapter 7 it was seen that the degree of bias was one important criterion in deciding on a "best estimate." All the common methods of estimation used in probability sampling have either zero bias or a bias which is known to be reasonably small for fairly large sample sizes. However, there is another criterion of excellence for an estimate: its standard error should be as small as possible.

Fortunately, the mean of the sample observations usually has a relatively small standard error for the sampling designs we have talked about here (except for disproportionate stratified sampling). This fact and the ease of computing the mean warrant its use in most practical applications.¹

There is one method of estimating a population mean in cluster sampling that is worth brief comment. This method is the so-called "regression estimate." It is complicated computationally, but sometimes its use is warranted. Past data for the characteristic being measured must be available for all the clusters in the population if the regression estimate is

¹ Solely as an aid to reading fuller treatments of sampling, the following comments are appended. The mean of the sample is a special case of the "ratio estimate" in cluster sampling with unequal cluster sizes and with a constant ratio of subsampling in every cluster. If the average size of cluster for the whole population is known (which would be rare) it is possible to make another kind of estimate known as an "unbiased estimate" (the ratio estimate has a slight bias), but the ratio estimate is much to be preferred because its standard error is usually much smaller.

to be used. For example, one might know average rentals for each block in Detroit at the time of the most recent census and desire to estimate average rentals per block in the entire city today.

Suppose 100 blocks are selected by simple random sampling. Current average rentals for these blocks are obtained by interview and averaged for the 100 blocks. This grand average would be the ordinary basis for estimating mean rentals per block. But information from the past census can also be used to derive an estimate that might be even more satisfactory. The information from the past census tells whether the particular 100 blocks selected in the sample were above, at, or below the city average at the time of the last census. Since it is almost always true that there is a very high correlation between average incomes per block for two periods not too far removed in time, it is reasonable to assume that if the average for the 100 blocks had been above the average rental per block at the time of the last census, the sample average for current rentals would also be above the average for the entire city (which, of course, is unknown). It is possible to make use of the correlation between past and present average rentals in order to modify the original estimate. Since the actual computation involves the computation of a regression relationship, which is beyond the statistical prerequisites of this book, we shall not describe it here.¹

9.3 Practical Problems in Area Sampling

We have sketched in enough of the important theoretical ideas in area sampling to enable the reader to form a good, though not complete, picture of area sampling designs in current use. Necessarily we have omitted many important developments and features. We have not intended to give a technical presentation, nor an exhaustive description of the subject. There is, however, another important omission which must now be remedied. We have paid only incidental attention to the practical problems of making an area sampling design work in actual investigations. We have wished to postpone this discussion until the reader has had a chance to form a good picture of the theoretical ideal.

Every practicing area sampler knows of the practical requirements which modify the actual working of the theoretical design, and the need for meeting those practical requirements wisely is great. If an area sampling design is not well executed, it will obviously not yield the full theoretical advantages of this kind of sampling. The ability to meet these problems is developed best by experience. In the absence of experience, some under-

¹ For an interesting example of a regression estimate, see Jesson *et al.*, "On a Population Sample for Greece," *Journal of the American Statistical Association*, Vol. 42, pp. 357-383, 1947.

standing of and appreciation for the practical requirements may be acquired from reading about them. Therefore the following sections discuss some of these practical problems.

1. Identification of people with areas. The basic idea of area sampling is the possibility of connecting people or families unambiguously with areas. That is, each member of the population must be identified with one and only one area. In practice this is not always easy. At any given moment of time, every individual in the world can be identified as being on or over a particular area of the earth's surface. A difficulty arises because usually researchers are interested in and prepared to interview persons in only a part of the earth's surface—such as the continental United States. In attempting to obtain a sample of all families who live in the United States, a problem would arise because some of these families would be vacationing in Bermuda or Europe.

Moreover, interviewing will inevitably extend over many moments of time. There is a need to decide what to do about the people who move about and occupy several areas during the period of interviewing. The possible practical importance of this can be seen by considering the difference in results that would be obtained from a sample of persons in the central business district of a large city at noon and a sample of persons in the same area at midnight.

Although the problem of unambiguous identification of people with areas is of fundamental importance, overcoming it is relatively easy because almost everybody establishes a residence or has some place which is considered to be "home." These residences and homes don't wander around nearly as much as the people to whom they belong. The problem is therefore largely overcome by the selection of a probability sample of dwelling places within the probability sample of areas.

Other minor problems still remain. Some families maintain more than one residence and thus may be associated with more than one area. This is important (in principle) because of the fact that such families would have a greater probability of being included in the sample than persons with only one residence, unless some offsetting step is taken by the person designing the sample. One theoretical possibility involves considering each person to be associated with the residence in which he is found by the interviewer, provided that he has been living in the residence for some specified period of time—say one-half of the month preceding the interview. This method often excludes transients, people who never stay in one place for long.

The problem of transients, occasional travelers, and other people with no established dwelling units is so vexing that a common administrative decision is to exclude transient hotels, hospitals, resorts, and other in-

stitutions commonly inhabited relatively briefly by the persons who use them. This restriction is ordinarily of small importance, but it can occasionally cause serious loss of information, as it would if a study of gasoline purchases were made in mid-July when vacation traveling was at its peak. If no interviewing were done in summer resorts, too few of these vacationists would be represented in the sample. This problem is closely related to the problem arising when people are not at home when the interviewer calls. Further techniques will be discussed when this problem of "not-at-homes" is taken up.

Still another problem involved in associating people with areas arises when the person's property lies partly in and partly outside of the area which has been selected in the sample. This occurs rather frequently in samples of farmers. The usual solution is to include the farmer if his residence is within the sample area, even though the larger part of the farm lies outside.

2. People who aren't at home at the time of interview. Probably at least one-third of all attempted interviews result in "not-at-homes."¹ One might rationalize a decision to forget about these people by saying that the universe is restricted to the universe of people who "would be found at home." If this narrower universe were no different from the broader universe that had been the original object of study, nothing would be lost. Unfortunately, however, the restricted universe is usually different. For example, a smaller than average proportion of persons who are at home during the daytime consists of persons with jobs. Occupational status might be highly correlated with characteristics of interest in marketing research. There could be other important differences in the group of persons found at home during the daytime and those not at home, and even if there were no differences between the "at-homes" and the "not-at-homes" with regard to the characteristics being studied, the investigator would have no way of knowing this. The assumption that no difference exists might lead to a serious bias in the sampling process. Because of this possibility, investigators usually make strenuous efforts to reduce the number of "not-at-homes."

One way to reduce "not-at-homes" is to call back until the persons originally missed are found at home. Another is to call back on a *sample* of the people who weren't at home on the first call, and from this sample to estimate the characteristics of all the people who weren't at home. With either of these approaches, there will almost always be some people who cannot be reached with any number of call-backs which can practically

¹ Usually relatively fewer people are found at home in the city than in the country, and the smallest proportion of all is in the largest cities. The percentage of not-at-homes varies also by time of day, day of week, and season of year.

be made during the period of interviewing. Some people might be maintaining residences somewhere else—as noted above. There will be other people who are never at home during the hours at which an interviewer calls, even when call-backs are made in the evening or on week ends. Mailed questionnaires can be used to get information from some of these people, but some will not reply. To the extent that people who have been selected by the sampling process are not interviewed or successfully reached by mail or other means, the universe from which the area sample is drawn does not coincide exactly with the universe of interest in the study. This difficulty, however, is present in any kind of sample (area or list, probability or nonprobability) and would indeed be just as important in a census as in a sample. Almost all censuses fail to enumerate a small proportion of the population.

A third solution does not involve call-backs.¹ Such a solution is especially attractive since call-backs are costly and tend to make the cost per interview (though not necessarily per unit of accuracy) in area sampling relatively high by comparison with nonprobability sampling. In order for a sample to be a probability sample without the use of call-backs, it is essential that those persons who are not at home when the interviewer calls be represented in some way in the sample results. The probability of finding any person at home at a moment of time selected by a random process is equal to the proportion of the time that the person is at home. The probability of finding a person at home at a randomly chosen time would be 0.5 if the person were at home on the average half the time. Thus on the average about 50 per cent of the people who are habitually at home about half the time would be found at home on the first call. Similarly one would expect to find about 40 per cent of the people who are at home about 40 per cent of the time; 10 per cent of the people who are home 10 per cent of the time; etc.

In order to allow for this it is necessary to weight the responses of those who are at home so as to represent those who are not at home. The weight that is used is the reciprocal of the proportion of the time that each person is at home. Thus the persons who are at home all the time would be given a weight of 1. Those who are at home one-half the time would be given a weight of 2, since each person's response will have to represent not only himself but some other person, also at home about half the time, who was not found at home when the interviewer called. Similarly a person who is at home one-fourth the time would be given a weight of 4, etc. All that is necessary in order to put this plan into effect is to find out from each

¹ This method was described first by Alfred Politz and Willard Simmons, "An Attempt to Get the 'Not at Homes' into the Sample without Callbacks," *Journal of the American Statistical Association*, Vol. 44, pp. 9-31, March, 1949.

person at the time he is interviewed the proportion of the time that he is at home. Politz and Simmons suggest that this information can be obtained (approximately) by asking each respondent whether he was at home at the hour in which the interview takes place on each of the five days preceding the interview. A person at home on all six days (the day of the interview and the five preceding days) would be considered to be at home all the time. A person at home on none of the preceding five days would be considered to be at home on the average one-sixth of the time, etc. It is impossible, of course, to make any allowance for the people who are at home none of the six days, but Politz and Simmons believe this problem to be of small practical importance.

With a given amount of money, more interviews could be obtained under the Politz-Simmons technique than with call-backs. This is true because the average amount of time and travel *per completed interview* will be smaller when call-backs are not made. However, the larger number of interviews per dollar doesn't assure that the Politz-Simmons method will be more accurate. In spite of the larger number of interviews, the standard error of the mean might be larger and actual call-backs might therefore be preferable.

Poltz and Simmons also suggest the use of their method in combination with a single call-back. It also seems reasonable to use the Politz-Simmons technique on initial interviews in order to see if the differences between people seldom at home and people frequently at home are small or large. Call-backs could be curtailed if the differences were small, and they might be increased if the differences were large. Some provision must always be made in practice for a "cutting-off point" for call-backs. Usually this point is not more than two or three call-backs, but there are examples in which greater persistence is called for.

Making a wise choice between these three alternatives of handling "not-at-homes"—calling back on everyone, calling back on a sample, and the Politz-Simmons technique—is difficult. The objective, of course, is to obtain the greatest economic efficiency. The determination of the optimum method of handling "not-at-homes," however, requires not only knowledge of costs of interviewing but also some knowledge of the population being sampled. In any case, every marketing research man should realize the dangers of having a large number of "not-at-homes" and should be familiar with the ways for reducing this number.

It should be said that some investigations of this problem have shown that unless the differential cost per interview between initial calls and call-backs is very great, the procedure of highest economic efficiency is to call back on *all* the "not-at-homes" rather than to subsample them. An instance in which this would not be so, because the differential in cost would

be high, is afforded by a mail survey in which the nonrepliers are called upon by interviewers. In an example of this kind, subsampling would almost always be advisable.

The Politz-Simmons technique offers one very important advantage in marketing research, regardless of its economic efficiency, in comparison with actual call-backs. Many users (buyers) of survey data cannot understand the need for call-backs and may not authorize them when an interview can be obtained more easily by going to the person next door to the person not at home, regardless of whether this person would have been chosen by the original sampling design. The Politz-Simmons method, then, possesses practical advantages that may often justify its use.

3. Refusals. Another reason why people who are selected by the sample are sometimes not included in the sample is that they refuse to be interviewed. Good interviewers can reduce the proportion of refusals. Call-backs can be made on refusers, although a different interviewer or a mailed questionnaire is usually advisable. Even with persistent attempts, however, some people will continue to refuse. Again, as with "not-at-homes," the result of refusals is that the universe actually sampled will be different from the intended universe, and some bias from this difference can be expected.

The proportion of the universe excluded from all causes usually can be made to be small, say from 2 to 20 per cent, by a combination of careful planning, call-backs, etc. As a result, the error attributable to failure to sample the larger universe can often be made of little practical significance. There are some exceptions. For example, if all people with incomes over \$10,000 should refuse to be interviewed on a survey of consumer incomes, the sample estimate of mean income would be much too low even though the proportion of people refusing would be low. (The estimate of *median* income, by contrast, might be very good.)

4. Other problems. It may be helpful to mention briefly some of the other practical problems which must be faced. Some of these problems and others as well are considered again in Part V, when we focus more explicitly on the administration of marketing research.

One of the most persistent of these problems is that of handling interviewers. It is difficult for an interviewer schooled in the methods of non-probability sampling to adapt to the more stringent requirements of probability sampling. Probability sampling permits no discretion in the choice of respondents. The basis of the whole process is that respondents be chosen randomly. As a result interviewers have to learn to find respondents and learn not to substitute alternative respondents when they can't find the ones that have been preselected.

A problem which effects interviewers' morale is the so-called "zero

block." Occasionally a cluster selected in the sample will contain no families at all. Baseball parks, garbage dumps, parks, and swamps have been known to be selected by the impartial mechanism of area sampling. This does no harm to the sample (such areas are ignored), but it may impair an interviewer's morale if she travels a long distance and finds no one to interview at the end of the trail.

A host of minor questions of interpretation accompany every survey using a probability sample, and these require the constant attention of the statistician and the vigilance of the interviewer. Calling on a random cross section of the American public can result in more unexpected situations than we could hope to catalog, even if we devoted the whole book to them.

CHAPTER 10

NONPROBABILITY SAMPLING

10.1 Introduction

The description of probability sampling has probably seemed complicated. Even though probability sampling is fairly simple in principle, it is complex in its applications. Many technical points have either been skipped completely or given only intuitive treatment. The development of technical proficiency requires a great deal more information than could possibly be compressed into two chapters of this book. One must have not only a thorough grounding in general statistical theory but also an understanding of the many technical developments which have been made in this particular application of sampling theory. As in many other scientific fields, "lab work" or practical experience is also required to become proficient.

Moreover, as has been seen, the problems in the applications of area sampling do not end with the choice of the method of estimation and the design of the sample. Once an efficient design has been chosen, it must be put into practice. Interviewers must be recruited for every county in the sample (except where interviewers travel from one county to another), blocks and open-country segments must be found by the interviewers, families must be correctly selected within blocks and segments, interviews must be completed, call-backs made, unforeseen troubles corrected, etc. These problems are enormous, and much experience in handling them is necessary before one may be deemed an expert. Every novice usually encounters many situations which were not anticipated in advance. Each initial venture into probability sampling seems to yield a new store of anecdotes.

One of these anecdotes will suffice for illustration. Alfred Watson of the Curtis Publishing Company, who was a pioneer in area sampling and possibly the first one to use area sampling in marketing research, reported the following:¹

I'll never forget the report from Minneapolis. In the dozen or so areas allocated to that city, I had hit the state institution for the feeble-minded as one area, and a

¹ *Measurement of Consumer Interest*, edited by C. West Churchman, Russel L. Ackoff, and Murray Wax, p. 165, University of Pennsylvania Press, Philadelphia, 1947.

nice grassy plot in the suburbs neatly placed between two lanes of a super highway as another one. From that point on, what we had been calling a pioneer venture in area sampling was thereafter referred to as "Watson's Folly."

It is not hard to see why area sampling hasn't been automatically accepted, without question, as a blessing for marketing research. Despite the rapidly growing acceptance of area sampling in the last few years, only a relatively small (but important) proportion of the survey work in marketing research is done with area samples. Watson's venture, described in the quotation above, took place in 1940. Many of the developments reported in the last two chapters have taken place since that time.

Prior to the development of area sampling there were only isolated examples of probability sampling in social research. Probability samples involving the simple random selection of names from lists or the systematic selection of households within an individual city had been used periodically, although "not-at-homes" and other problems of nonresponse were not always treated correctly. But the vast bulk of sampling activity in marketing research and other fields of social investigation has always been non-probability sampling.

Nonprobability sampling seems to have arisen when practical men, confronted with the need for information, selected samples by rough-and-ready standards which seemed feasible and plausible. No disparagement is intended by this statement. As a matter of fact, the resort to some form of sampling probably represented an advance over reliance upon intuitive judgment or upon the limited experience of one or two "experts." Many problems arise for which such "expert" opinion is an unreliable guide. The realization that the gathering of a sample of the relevant evidence would be very useful in reaching a wise decision was an advance in marketing management. It is not surprising that the actual use of sampling in marketing research should precede the development of the complicated sampling theory which has become the basis for the refinement and improvement of sampling procedures.

Nonprobability sampling has almost always contained a tinge of probability sampling. The words "random" and "stratified" have often been used to describe nonprobability samples. People have incorrectly interpreted nonprobability samples as if they actually were simple random samples or stratified random samples. This is not surprising. Many of the methods of sampling appropriate for human populations are of recent origin, and many of the elementary statistics books still discuss simple or stratified random sampling to the exclusion of cluster sampling without pointing out the difficulties of using the simpler designs to obtain a sample of people (in contrast to inanimate objects such as products made in a

factory). Hence some practitioners have been led to assume that any sample whatever must be either a simple random or stratified random sample. Others have assumed that nonprobability samples are as close to simple random or stratified samples as it is practical to come.

The introduction of area sampling has caused some fairly violent controversies between people who have used nonprobability methods and the proponents of the new methods. Controversy has arisen not only in marketing research but also in attitude research and perhaps most of all among the pollers of political opinion. Much of the controversy arises from misunderstanding. Nonprobability sampling at its best is usually compared with probability sampling at its worst, and conversely. Users of nonprobability samples have sometimes felt that advocates of area sampling turn away with alarm whenever they come too close to a nonprobability sample. People experienced in probability sampling have often felt that the opposite school has never really understood probability sampling and has seized on isolated cost figures as a pretense for loud "viewing with alarm."

Misunderstanding does not account for all the controversy. As we show in the next chapter, there are real grounds for differences of opinion. There is little doubt, for example, that the cost and time limitations encountered in marketing research will sometimes eliminate area sampling from consideration, and it is not always helpful or realistic to urge that nothing be done if an area sample is not feasible. An attempt to urge the use of an area sample in such circumstances can do more harm than good and may contribute little concrete help for a marketing research man who has to get some evidence regarding a pressing question within a very short time. Since nonprobability sampling is frequently the most useful sampling process available, it deserves serious discussion and will receive it in this chapter. In the next chapter an attempt will be made to compare probability and nonprobability sampling and to indicate the circumstances under which each type will be relatively efficient.

Before getting into the major discussion of nonprobability sampling, perhaps a word should be said about the sequence in which probability and nonprobability sampling have been taken up in this book and about the much greater length of the discussion of probability sampling. We certainly have not intended to deemphasize the importance or usefulness of nonprobability sampling. There are, however, three basic reasons for the procedure in this book.

1. Area samples, and probability samples in general, are important in themselves. As probability sampling is better understood, it is likely to find wider and wider applicability in marketing research. An increasing range of problems will probably be attacked with probability samples.

2. One cannot generalize at all about nonprobability sampling without introducing analogies from probability theory. Without some understanding of probability sampling, one would have to be content with an enumeration and classification of nonprobability samples which have been used in practice, together with an evaluation of the extent to which they did or didn't "seem" to work—and such an evaluation is often impossible. Probability theory gives some basis for discussion of the sources of error involved in nonprobability sampling. Comparison with probability sampling gives some rough (very rough) idea of the magnitude of error which is likely to be encountered in nonprobability sampling.

3. Without some conception of probability sampling, one would be unable to design improved nonprobability samples, except by the crudest kind of trial and error. An understanding of the principles of probability sampling is helpful for the design of nonprobability samples. An example of this is given by the quota sample, which was almost certainly evolved by people who had heard of stratified random sampling. The quota sample was clearly an improvement on the methods of nonprobability sampling which had been used before.

10.2 The Distinction between Probability and Nonprobability Sampling

The distinction between nonprobability and probability sampling is often confusing to the beginner, since the two methods may have many surface similarities. The distinction is that certain assumptions of sampling theory which enable one to calculate a confidence interval are not met in nonprobability sampling. The reason for this is that the ultimate selection of respondents in probability sampling is by some random process which makes it possible to use probability theory and the statistical techniques derived from it. When the selection is by a nonrandom process, the probability of including any particular element of the universe in the sample is not known and the use of statistical techniques based upon probability theory is, therefore, not justified.

Thus, in probability sampling, it is possible to know that sample estimates are unbiased, or essentially so, and that confidence intervals may be constructed by estimating the variability of the sampling process. In nonprobability sampling, these things can never be known merely from a knowledge of the method by which the sample is selected. It is, of course, possible to perform the same arithmetic operations on sample observations from nonprobability sampling as are performed in the calculation of point estimates and confidence intervals from probability sampling, but these arithmetical operations do not supply unbiased estimates or a valid estimate of sampling error.

Random selection must prevail in *every* stage in the selection process if

the resulting sample is to be a probability sample. A sampling design might look just like an area sample except that families within sample blocks were not selected randomly. Instead of, say, selecting every fifth family from a random starting point in each block, the interviewer might be told only to select two "typical" families in the block. The actual selection would be in the hands of the interviewer and would not be random. Even though the block itself and the county had been selected by a random process, the sample could not be called a probability sample. For another example, suppose that a group of counties is selected according to judgment as to what a "typical" group of counties should be like. Within counties, probability samples are drawn, so that for each individual county, the sample would indeed be a probability sample. But a collection of small probability samples doesn't necessarily make a big one. The group of counties cannot be considered a probability sample of all counties in the country. Even if each sample county were completely enumerated so that no sampling error whatever could arise within counties, the sample would still be a nonprobability sample.

In the two examples which have been given, random selection is introduced at some, but not all, stages. Perhaps a more typical example is one in which all decisions are made by judgment. The counties, cities, and towns are selected by the person planning the study, and the people to be interviewed within these broad areas are selected by interviewers. The interviewer may be relatively free in selection of families or individuals, or she may be restricted to certain areas within a city or told to get specific quotas of different kinds of people; but she almost always has some discretion as to whom she interviews.

1. The use of judgment in selection. Selection according to judgment has a strong appeal. Nearly everyone feels that he is able to select a typical sample. It would even seem possible that sample selection by judgment would be more, rather than less, accurate than selection by a random process. Random selection occasionally turns up very bad samples; personal selection would seem to be immune to this shortcoming. Is this really true?

To select a "typical" sample a person must know a great deal about the subject of his investigation. If he is studying the purchase of gasoline, for example, he must know a lot about motorists before the investigation even starts. Sometimes this condition is met; people working in a particular field often become very well informed about it. Generally, however, if a person knows enough about a subject to select a "typical" or "representative" sample, the sample itself is likely to be unnecessary. Sampling procedures are usually used when new information is to be obtained or when one wishes to see if changes have taken place. For example, an investigator might

wish to see if motorists were buying more premium gasoline than last year. The factors that might be associated with a *change* in purchasing habits might render obsolete the investigator's ability to select a "typical" sample.

Actually, "typical" samples are usually considered "typical" because they have produced correct (or plausible) results in the past. One example is particularly interesting. In his preelection polls in 1940 and 1944, Elmo Roper deliberately permitted certain biases in his sample because his political knowledge told him that these biases would largely offset each other. His sample gave too much weight to the South, which was predominantly Democratic, but this overweighting was offset by a tendency to underrepresent groups having the least income and education (groups in which support for the Democrats was very strong). The results were excellent in 1940 and 1944. In 1948 Roper's prediction would have been less erroneous if this assumption had not been made and if a correction had been made for the understatement of the Democratic vote in the Northern states. Hindsight is easy. It can now be seen that one adjustment would have been best in 1948 and another in 1944 and 1940.

Hence there are dangers involved in an expert judgment of "typicalness," no matter how good the expert may be.¹ It is difficult to devise any theory based upon the use of intuitive judgment concerning the "typicalness" of a sample. The subjective abilities which are involved are difficult to describe and impossible to measure. About all one can say is that there are varying degrees of expertness. The confidence that one can have in the samples which such experts provide depends upon one's evaluation of the experts' expertness. In probability sampling, one's confidence inheres in a mechanical process of selection, even though judgment may be introduced in the design of the process (*e.g.*, in stratification).

Sometimes judgment is used not to determine how a "typical" sample would be selected but rather to devise a *process of selection* that would supposedly yield the equivalent of a random sample. If the interviewer were to be given discretion, her personal biases and habits of thought might tend to increase the chances of including persons of certain types and excluding persons of other types. The altered probability of including certain elements of the population, which would be introduced by allowing discretion to interviewers, cannot be measured, and hence statistical theory is inadequate to indicate sampling variability. In probability samples the interviewer is left no discretion as to the persons to be interviewed, and it is thus possible to ensure that every element of the population has a known probability of being included in the sample.

¹ Political elections probably give one of the best examples in which skilled judgment is likely to select a good sample, since there is a periodic "checking up" on the results.

Some types of nonprobability sampling do not remove all the discretion which interviewers have but try to limit it so that its operation will not introduce biases into the sample. A rationale of nonprobability sampling, then, might be thought of as embodying the principles which indicate how the discretion of the interviewers should be limited in choosing respondents so as to try to secure the benefits of randomization without using the rigid procedures required of "strict" randomization.

10.3 A Rationale of Nonprobability Sampling

Randomization, with known probabilities of selection for all clusters and elements in the universe, is ordinarily assured by the use of a table of random numbers. This is the strict method of achieving randomization and the only one mentioned so far. More generally, however, any sampling process is random in which the probability of selecting each element in the population is independent of the characteristic under study. As an example, consider a study of income in which the probabilities of selection are not equal, but are no higher, on the average, for families of high income than for families of low income.

To see how randomization might be achieved by taking the usual precautions necessary for strict randomization, consider the use of a table of random numbers. Think of the table as being a universe from which a random sample of digits is to be drawn. It isn't necessary to go to another table of random numbers to draw this random sample! One simply takes digits in order from a random starting point in the table. The table of random numbers has *already* been randomized, and any consecutive path of selection through the table will assure not only that the probability of selecting a particular digit is unrelated to the size of that digit, but that the probability of selection is equal for all digits. The probability of choosing a 7 is neither more nor less than the probability of selecting a 3. Such a process of selection is said to be "effectively random" because the universe itself has been completely randomized to start with.

It is possible that this randomization may be characteristic of other universes besides tables of random numbers. Phenomena of interest in marketing research may be distributed in such a way that, almost regardless of the freedom allowed to interviewers, the resulting sample would be effectively random. An example may make this more concrete.

A universe can be said to be effectively randomized if the factors that influence the interviewers in their selection of respondents are independent of the phenomenon being investigated. Suppose, for example, that a manufacturer of golf clubs wanted to know the incidence of left- and right-handedness. It is quite possible that the factors which cause interviewers to select particular respondents are not related in any way to the fact that

the respondent is left- or right-handed. If this independence does in fact exist, the interviewers could be left completely free to select respondents as they wished. The resulting sampling process would be as effectively random as a process of selection in which respondents were preselected for interviewing by the use of random numbers.

The difficulty in applying this concept of effective randomization is that a probability sample or a census would be required as evidence of whether a particular characteristic was independent of the factors influencing the selection of respondents. Otherwise the assumption of effective randomization, upon which the sampling process is based, would be untested. There may be information which gives credibility to such an assumption, but this information is not usually complete enough to be thoroughly reassuring. On the surface, for example, it would appear plausible that right- and left-handedness should be effectively randomly distributed in the adult population. At first glance, there appears to be no reason why this should not be true. Unfortunately, in examples like these, one soon begins to think of reasons why the assumption might not hold. One example will suffice. Suppose that one neighborhood contains a progressive school in which no attempt is made to "correct" youthful tendencies toward left-handedness. Another neighborhood possesses a traditional school in which every effort is made to turn prospective southpaws into right-handers. Even though the two neighborhoods tend to have the same proportion of left-handed people from genetic considerations, the environmental difference might cause important differences between the two neighborhoods.

One problem in marketing research, "taste testing," seems to involve a randomized universe. It seems fairly plausible to assume that individuals' preferences for two masked or disguised brands of tooth paste should be effectively randomly distributed in the population. On the other hand, tastes might be closely related to nationality, economic status, and geographical location. These factors in turn are frequently related to the factors that influence interviewers in their selection of respondents. A salad-oil manufacturer interested in developing an oil with maximum palatability might get strange results if his interviewers were left complete freedom in their selection of respondents. The segments of the population preferring strongly flavored olive oils to the more bland corn and peanut oils might be either under- or over-represented in the sample because of the biases of the interviewer in selecting respondents.

More often there is good reason to believe that the factors that are known to affect the selection of respondents are related to the variable being investigated. It is known, for example, that the selection of respondents is often influenced by their nationality, race, economic status, and location. These factors are related to most of the variables of interest in marketing.

Reactions to advertisements, patterns of consumptions, tastes, etc., are all known to be related in most cases to the same factors that influence the selection of respondents.

So the assumption of an effectively randomized universe frequently seems untenable. Efforts must then be devoted to designing a sample that secures effective randomization without imposing the costs and complexities that attend strict randomization. The device used in the hope of securing effective randomization is the quota sample.

1. Quota sampling. The quota sample limits the discretion of the interviewers in their selection of respondents with respect to those variables known to be associated both with the interviewers' biases and the variable being studied. In practice this means that interviewers are assigned quotas of respondents having the various characteristics which are known to be affecting factors. For example, an interviewer might be told to get a certain number of men who were white, native born, and who received a medium income; a certain number of women having these same characteristics; a certain number of Negroes of designated economic status and sex; etc. This means that the universe being investigated is broken up into subgroups or strata defined in terms of color, nationality, economic status, and sex. The assumption is then made that within these subgroups there is effective randomization. In other words, if the interviewers' discretion is limited with regard to the factors defining the strata, it is assumed that the discretion remaining will not bias the sample result. This is the basic assumption underlying quota sampling, which is still the most widely used type of sampling in marketing research. We shall now consider in greater detail the development and current status of this sampling technique.

The best-known illustration of the utility of this technique was made more dramatic by the failure of another sampling technique in the 1936 presidential election. The famous *Literary Digest* poll was based on a mailed questionnaire sent to more than 10,000,000 names selected from available lists of telephone numbers, car owners, etc. More than 2,000,000 of these people expressed their voting intentions in that poll. Although in earlier elections the *Digest* poll had been correct, it failed badly in 1936. It forecast a substantial Landon victory in the face of a Roosevelt landslide that took 46 out of 48 states and roughly 60 per cent of the popular vote.

The *Digest's* method was a nonprobability sample. There were two stages at which judgment can be seen in retrospect to have been wrong: (1) in assuming that the lists of car and telephone owners, etc., could be considered in some sense a random drawing from the adult population, and (2) in assuming that the people who replied to the questionnaire were somehow a random selection from the entire list. The sample was biased

seriously toward the upper-income groups, which tended to support the Republican candidate. Some people recognized this defect at the time and said so. William L. Crum made some rough corrections for the biases by reweighting the sample results by known characteristics of the population, such as the proportion of the people who had voted Republican and Democratic in 1932, and announced that Roosevelt would win. At the same time, George Gallup not only predicted that the *Digest* poll would be wrong but predicted what the *Digest* poll would predict before actual publication!¹

Public-opinion samplers Gallup, Roper, and Crossley, in this same year, each used relatively small quota samples—on the order of thousands rather than millions—and all forecast that Roosevelt would win. Their margin of error in the percentage of popular vote was relatively large, but they named the winner.²

The 1936 election predictions thus had two important consequences. First, the layman was shown that a huge sample was not a guarantee of success and that samples comprising a very small part of the total population could come out with the right answer. The popular acceptance of sampling was undoubtedly accelerated by this demonstration. The second result, more technical, was that quota sampling was shown to be a way of making nonprobability sampling work.

On the surface, quota sampling resembles stratified random sampling; and it seems reasonable to assume that the idea of quota sampling was borrowed from stratified random sampling. Even if this surmise is not correct, it is certainly true that quota sampling has frequently been confused with stratified random sampling. The basic difference between the two methods is that with stratified random sampling the selection of elements within strata is by a strictly random process, while with quota sampling randomness is not assured by the method of selection. For that reason, the treatment of quota sampling as stratified random sampling and the application of the sampling error formula of stratified random sampling is probably usually unwarranted.³ A particular quota sampling method

¹ On p. 75 in *A Guide to Public Opinion Polls*, Princeton University Press, Princeton, N. J., 1948, George Gallup refers to the prediction of the *Digest's* final figures by the American Institute of Public Opinion in the July 14, 1936, editions of the newspapers carrying the Gallup Poll.

² In 1948 the chief defect of the polls was that they attempted to forecast the winner without qualification and did so incorrectly.

³ Even if the selection of respondents within quotas were effectively random, there is reason to believe that the sampling process achieved *within quotas* would behave more like cluster sampling (since in quota sampling respondents are typically concentrated in relatively small areas) than like simple random sampling. That is, even though the selection of respondents within quotas might be unbiased, it is

may be excellent or completely wrong for the problem to which it is applied, but there is no objective way of estimating in advance how excellent or wrong it is likely to be.

Despite the surface similarities between quota sampling and stratified random sampling, stratification serves a very important purpose in quota sampling which is different from the familiar function of serving to make the efficiency of a sampling design greater than it would be if simple random sampling were used. In probability sampling, stratification is introduced to prevent *random* fluctuations in the number of elements (or of clusters) drawn into the sample from each stratum of the population. Suppose there were a small population made up of 1000 individuals who were divided into two strata consisting of 200 and 800 respectively. In a simple random sample of 200 from this population, 47 came from the first stratum in one sample and 35, 39, 32, 43, 38, and 49 in successive samples. The average for a large number of samples would be about 40, so that the average sample result would give $40/200$, or $1/5$, of the sample in this stratum, the same proportion found in the universe. Stratification would prevent this variability from sample to sample. There would be 40 observations from the first stratum in *every* sample.

In a nonprobability sample of 200 in which stratification was not used, the number from the first stratum of the population falling in successive samples might be 98, 109, 73, 131, and 84 (where these numbers are arbitrarily chosen for illustration of the principle that the nonprobability sampling process may be biased). The consistent difference between the proportion from the first stratum included in the sample and the proportion in the first stratum of the population would be a source of bias. This source of bias would be removed by fixing a quota of 40 for the first stratum.

The fixing of this quota would have two consequences. First, the possible bias caused by the (assumed) tendency to get about 100 rather than 40 from the first stratum into the sample would be eliminated. Second, the random fluctuations inherent in even a random sample would be eliminated. The quota sample would still not be a probability sample, for there might be additional biases because of the method of selection within each stratum. Nevertheless, one important source of bias would be removed. The removal of this source of bias is probably the main reason for imposition of quotas.

Next we shall indicate how quota sampling is used in practice. The procedure given does not refer to a particular quota sample and is not necessarily typical of all quota samples, but it suffices to indicate a not

plausible to expect that the sampling variability within each quota would be larger than for a simple random sampling process in which the same number of respondents were interviewed within each quota.

unusual approach. The country is divided into geographic areas, each of which includes several neighboring states. The New England states, for example, might comprise one such area. Each area is stratified according to the size of towns and cities in the areas; for example, all cities in New England between 100,000 and 500,000 as of the most recent census would comprise one of these strata. The population unaccounted for in each region by incorporated cities and towns larger than 500 people is grouped in a stratum called "farm," or "rural."

The whole country has now been divided according to region and city size. The number of interviews within each such stratum is proportional to the total number of people in that stratum according to the most recent census. The next step is the determination of actual cities and towns for interviewing. This is usually, though not necessarily, done by judgment. If interviewers are already available in some cities, these cities are especially likely to be included in the sample. "Farm" interviews may be assigned in open-country areas located near the towns and cities in the sample, and the same interviewers may be used for both urban and rural interviewing.

Every city and town would then have a quota of urban and farm interviews. These quotas might be filled in two steps. First, interviewers may be sent to scattered areas of the city—often, but not necessarily, selected by judgment. Secondly, each interviewer is given detailed subquotas of the kinds of people she is to interview. She may be required to obtain 50 per cent men and 50 per cent women, so many old people and so many young ones, certain proportions of persons having high, medium, and low incomes, so many of each occupational group, etc. Usually only one or two of these controls are imposed, since a larger number is unwieldy. The proportions in subquotas are assigned on the basis of available information about the population, such as past censuses or surveys. In the absence of more accurate information, arbitrary proportions may be set. For example, an interviewer may be told to get 15 per cent "upper-income people," 55 per cent "mediums," and 30 per cent "lows."

Within any subquota the interviewer has complete freedom of selection, a freedom which is justified by the assumption of effective randomization within subquotas. It is necessary to assume that the freedom left to the interviewers will not bias the sample estimate. This will be true if one of two conditions exists. First, it is obviously true that this freedom will not bias the results if the elements in each stratum are the same with regard to the characteristic being studied. If all the persons in any subgroup are exactly alike with regard to the variable being studied, quite obviously the same results will be obtained regardless of which of these persons are included in the sample. In practice, of course, complete uniformity is never achieved and possibilities for bias usually remain.

The second condition which would render the sample safe from the danger of interviewer bias would be the achievement of effective randomization within these subquotas. Effective randomization is achieved if the factors affecting selection by interviewers are independent of the variable being studied. Complete uniformity within strata need not be achieved in order to have effective randomization. Hence if the mode of stratification excludes from interviewer choice all the variables that would otherwise lead to bias in the sampling process, effective randomization may be secured within strata. Thus if it is known that the selection of respondents is influenced by their incomes and that their incomes are related to the variable being investigated, this particular source of bias can be eliminated by using income as a mode of stratification.

The following illustration may make these ideas clearer:¹

If someone takes a boat, goes out on the Atlantic and picks up a gallon of ocean water, he may call it a sample of ocean water. If he measures the salt content of his sample, he will find 3.6 per cent salt. We know from other sources that 3.6 per cent is right for the whole ocean; consequently, our sampler is entitled to claim that his sample is dependable as to the salt content. But the fallacy starts if he claims that since his sample is right on salt, it must therefore be right on temperature. (It is obvious that surface water cannot have the same temperature as the ocean as a whole.)

The population being studied in the example can be thought of as the salt content of all the buckets of water in the ocean. The person taking the sample was given complete freedom of selection. He did not pick a sample at random. The probability that a particular bucketful would be chosen was not known. It seems certain that buckets of water from two miles down did not have any chance of being included. This fact means that strict randomization did not exist. Effective randomization did exist, however, because the location of the buckets of water was independent of their salt content. Thus the freedom left to the person taking the sample did not bias the result. If the objective of the study had been the determination of the average temperature of buckets of water in the ocean, the freedom left to the person taking the sample would have biased the results. For the person left to his own initiative would have taken buckets of water at the surface of the ocean in a latitude convenient to the continental United States. Since temperature varies both with depth and with latitude, the results of the sampling process would be biased. One could imagine the investigator being given quotas by latitude and depth of water. This would remove this bias, at least in part.

¹ Alfred Politz, "Evaluation of Sampling Technique," talk on Apr. 8, 1947, before the Chicago Chapter of the American Marketing Association.

In sampling human populations it is often risky to assume that effective randomization within subgroups has been achieved. It is perhaps seldom possible to limit the discretion of the interviewers sufficiently to remove all bias from the sample results. By stratification (establishment of quotas) it is possible to ensure that the sample will not be biased with regard to the characteristics used as modes of stratification.

It is possible to receive further assurance regarding absence of bias by a process known as "validation." Validation is a process whereby a comparison is made to see whether characteristics of the universe for which information is available and which have not been used as modes of stratification are present in approximately the same proportions in the sample as in the population. Thus if education and economic status were modes of stratification, one could partially validate the sample by seeing whether the sample contained approximately the same proportion of males, Negroes, and home owners as did the population being studied. Validation provides some evidence for believing that effective randomization within strata has been achieved. The evidence is never complete, however, and must always, unfortunately, be regarded with skepticism. The following example¹ illustrates the danger of assuming that validation is sufficient evidence for believing that effective randomization has been achieved for other variables than those considered in the validation process:

There is an article that I think many people don't know about, published in 1926 by two excellent Italian statisticians, Gini and Galvani. The situation was this. The census of 1929 or 1930 was going to be taken, and they still had on hand the cards for the previous census. However, they had to make room for more cards because they didn't have a great deal of space. But though they were forced to get rid of the cards for the last census, they did want to save a sample. Professors Gini and Galvani found twenty-nine counties that in fifteen respects had exactly (or closely) the same averages, percentages, proportions, and so on, as the entire country, and therefore some people would have supposed that all they had to do was to save the cards for these twenty-nine counties, throw away the rest, and tabulate these at any time later to get a picture of the country. Whereupon Gini and Galvani proceeded to make some tabulations to see how closely they would turn out. They discovered—I won't say to their horror, because a scientific man always takes what nature gives him—but they discovered anyhow that in many characteristics other than the original fifteen the sample was no good. It gave terrible results.

A second example may further emphasize this point. The McCall Corporation periodically makes studies of readership of its magazines and the

¹ Reported by W. Edwards Deming in *Measurement of Consumer Interest*, edited by C. West Churchman, Russel L. Ackoff, and Murray Wax, pp. 167-168, University of Pennsylvania Press, Philadelphia, 1947.

magazines of competing corporations in order to determine the attributes of their readers and their reactions to what is read. McCall's has undertaken some studies of the relative accuracy obtainable by quota samples as compared with probability samples. One of these studies involved the readers in Milwaukee County, Wisconsin. Large groups of interviewers were selected and were given the same training and the same questionnaires. Half the group was then instructed to select respondents in certain proportions according to their socioeconomic level. In other words, this group of interviewers was to select a quota sample based upon this economic variable. The other half of the interviewers was instructed in the ways of selecting a probability sample of approximately the same size as that to be selected by the quota samplers. The results are interesting. Whereas two independent sources indicated that the average monthly contract rent for Milwaukee County was about \$38.60 the probability sample indicated an average rental of \$38.82 and the quota sample an average rental of \$35.63. The Audit Bureau of Circulations indicated that about 46 copies of leading women's magazines should be found in every 100 homes. The probability sample indicated that 47.3 copies would be found on the average in every 100 homes; the quota sample, 66.1 copies.¹ In other respects, however, the quota sample gave results as good as the probability sample.

It should not be concluded from these examples that quota samples are never useful or even that they are never as good as probability samples. We have stressed the potential weaknesses in quota sampling in order to encourage proper caution in its use rather than to minimize the usefulness of what is a valuable marketing research technique. There is reason to believe that quota samples are seldom as efficient statistically as are probability samples but that quota samples are often more efficient economically. That is, when costs are considered, quota samples might be the best way to get required information. The only difficulty is that the researcher can never know the reliability of the information provided by quota samples. It is certainly possible, however, that quota sampling provides much more reliable information than is provided by people who profess to have sufficient knowledge to be able to select "typical" samples.

Three relatively minor comments about the use of quota sampling remain to be made. First, the reduction of sampling variability which is achieved by the stratification of quota sampling is relatively slight compared to the reductions that are possible through the use of stratification of area probability samples. It has often been found that the major gains from stratification come when clusters rather than individuals are stratified.

¹ Findings quoted above come from *A Qualitative Study of Magazines—Who Reads Them and Why*, sponsored by the McCall Corporation, September, 1946.

Usually the larger the clusters the larger the gains from stratification. Substantial reductions in sampling error usually accompany the stratification of counties; relatively small reductions in sampling error are achieved by stratifying individuals. And in quota sampling it is almost always individuals or families who are stratified. The justification for quota sampling, compared with other types of nonprobability sampling, as we have already suggested, probably comes from its potential reduction in bias (which may be substantial) rather than the reduction in sampling variability. (The justification for stratified random sampling, compared with simple random sampling, comes from the reduction in sampling variability.)

Second, some of the reductions in sampling variability which are achieved by establishing quotas may be offset by inaccuracy in fixing the dimensions of the strata. In area sampling, the size of strata (number of clusters) is usually known exactly, because the clusters are typically geographic areas which can be counted on the map. With area sampling, there is no danger of inaccuracies in establishing the size of strata, not only because of the possibility of precise counting but because the areas do not move, multiply, or die with the passage of time. People, unfortunately, do not stay put, do multiply, and do die with the passage of time. Thus the data which are available on persons become less and less precise as time passes, and the inaccuracies thus introduced may be substantial. For example, in the nine years following the 1940 census, North Dakota had a net loss of population while California gained about 50 per cent. However, rough adjustments usually can be made. A similar minor source of inaccuracy is the incorrect assignment of individuals to quotas by the interviewer.

Third, in addition to being out of date, the basic data needed to establish quotas are sometimes completely lacking. Even when such data are available for large political divisions, such as the entire nation, a state, or a city, they may not be available for smaller political divisions such as neighborhoods of an individual city. For the purpose of quota sampling, one may divide a city up into neighborhoods and within these neighborhoods wish to assign interviewers quotas as to sex. The sex ratio may be known to be 50/50 for the city as a whole, but the sex ratio for individual neighborhoods may be unknown. Or again, the proportion of individuals or families in different income groups may be known for the nation as a whole but not for individual cities or states.

Before any final evaluation can be made of the relative efficiency of nonprobability sampling it will be necessary to inquire more closely into the costs and the sampling variability which are usually associated with the various approaches to marketing studies. It is the purpose of the next chapter to make such an inquiry and a comparative evaluation.

CHAPTER 11

SOME SAMPLING PROBLEMS IN MARKETING RESEARCH

11.1 Introduction

It is time to pause briefly to reconsider the sequence of sampling topics taken up in preceding chapters. In Chapter 6 we introduced the general subject of sampling and tried to indicate reasons why samples can be expected to yield useful information. In Chapter 7 we sketched briefly the theory of simple random sampling. The eighth and ninth chapters described some of the many techniques that can be used in obtaining probability samples of human populations. With this background we proceeded to a discussion of nonprobability sampling in Chapter 10. All these topics are important to the practitioner of marketing research. An understanding of these *individual* topics, however, is not in itself a sufficient theoretical background for facing certain persistent problems which arise repeatedly in marketing research. The marketing researcher, whether he be a statistical specialist or an all-round analyst, is repeatedly forced to answer two basic questions as a part of his day-by-day routine: "What kind of sample do we need?" "How big should the sample be?"

In this chapter we focus our attention on these two questions and try to use the relevant results from the earlier discussion of sampling theory in attacking them. While the two questions are closely related, it is easier to discuss them one at a time. We first take up the question of sample size and later return to the question of sample design.

11.2 "How Big Should the Sample Be?"

A very simple but incomplete answer to the question of sample size may be illustrated by an extension of the theory of simple random sampling. The standard error of the mean was given by the expression $\sigma/\sqrt{n}$.¹ Under the assumption that the standard deviation of the population is known, the 95 per cent confidence interval for μ —the unknown population mean—can be written as follows:

$$\bar{x} \pm 2 \sigma / \sqrt{n}$$

The smaller the standard error of the mean, the narrower the confidence interval and the more precise the inference. That is, given the desired

¹ The finite population multiplier is ignored for simplicity.

degree of confidence—in this case, 95 per cent—one can make more precise inferences by reducing the standard error of the mean. The standard error of the mean can be reduced by increasing the size of the sample. To determine sample size, then, one may specify the desired standard error of the mean and solve for n .

In the example of estimation of average purchases of furniture, the standard deviation of the population was assumed to be \$30. Suppose the desired standard deviation of the mean ($\sigma/\sqrt{n}$) is \$3. (The 95 per cent confidence interval for the mean, then, would be the sample mean, $\bar{x}$, plus or minus \$6.) Then simply solve the following equation for n :

$$\begin{aligned} \$3 &= \frac{\$30}{\sqrt{n}} \\ \sqrt{n} &= 10 \quad \text{or} \quad n = 100 \end{aligned}$$

The sample size required, then, would be 100.

If it is no longer assumed that the standard deviation of the population is known, an estimate of it must be made. General knowledge or previous surveys of a similar type could be used to refine the guess. Moreover, even if the guess were a very good one, the confidence interval obtained would be somewhat different from the confidence interval sought, since the *sample* standard deviation would be used in the final calculation of the confidence interval. The basic idea, however, is the same. Nor is the main thread of the reasoning essentially different for any probability sampling design other than simple random sampling. The complications concern details which need not be considered in a general exposition.

For nonprobability sampling design there is no available formula for the standard error of the mean, and the method just outlined is no longer applicable. However, it is probably safe to assert that the determination of sample size in nonprobability sampling, in so far as it is done on any objective basis at all, usually proceeds by analogy with the computations outlined above for simple random sampling.

But even for probability sampling, the approach we have been outlining is incomplete as an answer to the practical question of how large the sample should be. The outline of its logic is to proceed from a specification of sampling accuracy (expressed in terms of the standard error of the mean) to a determination of the size of sample needed to attain that accuracy. In the first place it assumes a satisfactory answer to the question, "How much sampling accuracy do we need?" In the second place it does not consider the cost of obtaining any specified sampling accuracy.

A complete theoretical answer to the question, "How much accuracy?" involves reasoning which, briefly, goes like this. One is willing to pay for

sampling accuracy because sample results have value in reducing the risk surrounding marketing decisions by clarifying the basic facts upon which decisions must be made. One is willing to pay differing amounts for differing degrees of sampling accuracy, and usually, to pay more for higher degrees of accuracy. If sampling accuracy is measured in terms of the standard error of the mean, one could set up a table or graph showing the amounts of money (the total "worth") one would be willing and able to pay for differing sizes of the standard error of the mean. Another table or graph could be set up to show the cost of attaining different values of the standard error of the mean. The optimum solution would be the standard error at which the total cost and the total "worth" were equal.¹

Earlier, in the numerical example of the charge accounts, it was assumed that the standard deviation of the population was known to be \$30. Given this assumption, a simple random sample of size 100 was found to be adequate to attain a standard error of \$3. The cost of obtaining a sample of size 100, in the sampling of the records of charge-account customers, might be a relatively small figure such as \$25. Hence a standard error of \$3 could be obtained at a cost of \$25. A standard error of \$2.50 would require a simple random sample of 144 and might cost \$36; a standard error of \$2 would require a simple random sample of 225 and might cost \$60; etc.

Similarly, it might be believed that an estimate with a standard error of \$3 was worth \$30, one with a standard error of \$2.50 worth \$36, and one with a standard error of \$2 worth \$55, etc. In this case one would choose to take a simple random sample of 144 charge accounts, with standard error of \$2.50, since this is the highest accuracy which can be attained if the total cost is not to exceed the total "worth."

This theoretical solution to the problem still has one serious practical shortcoming.² In practice, one does not know how to figure out the dollar values of given degrees of accuracy. A subjective procedure, however, may be helpful as an approximation. Begin by specifying a given degree of accuracy which is believed to be needed, and figure out the cost of obtaining this much accuracy. Perhaps this cost figure seems to be completely out of line—obviously more than the information is worth. If so, scale down the requirements for accuracy until the cost seems really "reasonable." This "reasonable" cost sets the sample size. By such a subjective process, it is possible to accomplish what could be accomplished better if

¹ It is possible that one might not be willing to pay the necessary cost of *any* survey.

² We have previously mentioned another difficulty: it might not be possible to compute the exact cost of attaining a given degree of accuracy, since we often do not know the standard deviation for the population.

there were objective information upon which to base a table or chart of total "worth." Unfortunately, such objective information is ordinarily unobtainable.

The reasoning outlined in the preceding paragraph might proceed in a slightly different way. Instead of specifying the requirements for accuracy, figuring the cost needed to attain this accuracy, and then juggling the requirements for accuracy until the associated cost seemed reasonable, one might specify a certain cost and then figure out the accuracy attainable for such an expenditure. If the accuracy seemed more than enough, the cost might be trimmed a bit; if the accuracy did not seem sufficient, a little might be added to the cost.

These two approaches will theoretically lead to the same answer since they are logically the same. Sometimes it is more natural to use one and sometimes the other. The more typical procedure in marketing research, perhaps, is for the total cost figure to be specified. Usually this total cost will be unspoken, but the research man can guess pretty closely what it is. The job of the research man, then, is to figure out the sample size which can be obtained for such an expenditure. Ideally, if the accuracy is more than is felt to be needed, the cost should be trimmed, and conversely.

The other situation is also encountered in marketing research. The research man is told to figure out how big a sample—and how much money—is needed. The research man in these cases must begin with some requirement of accuracy he feels to be reasonable, figure the cost, and then make the needed modifications if the cost is out of line.

In practice the whole problem is sometimes decided by someone's fiat. The vice president may feel that so much money should be spent and that no higgling should be allowed even if the research man feels that the accuracy obtainable is either too high or too low. Usually, however, the researcher is free to go through the kind of reasoning described above. If his judgment as to the worth of different degrees of accuracy is sound and his calculations of the cost of these specifications are accurate, then his answer should be satisfactory. The calculation of the cost of different degrees of accuracy is a technically soluble problem, but nothing much can be said about the judgment involved.

There are additional complexities to be considered in practical problems. First, most marketing surveys actually concern several items of information. This means that a sample of any given size will yield estimates of different degrees of accuracy for the different characteristics being studied, since the variability of these characteristics will undoubtedly be different. Which characteristic is to serve as the standard of accuracy? One solution is to use the item felt to be most important, and this seems frequently to be satisfactory. Another solution is to work out some compromise.

A second complexity arises when interest is focused upon estimates of subgroups of the population rather than upon an over-all estimate. It is possible, for example, that a sampling design which would yield satisfactorily accurate information for the entire United States would fail to yield satisfactory information for some important sales territory, such as the southwestern states.

One way to handle this problem is to figure out where, in the normal course of events, the sample would be "spread the thinnest," that is, where the sampling variability would be expected to be greatest. In some national studies, for example, variability might be expected to be greatest in the southwestern states—largely because of the anticipated small number of observations. The sample could be designed, therefore, so that the sampling variability in the southwestern states was reduced to the desired level by specifying that the sample within these states be sufficiently large. The rest of the sample could be designed proportionately in the knowledge that the sample would be satisfactory for other regions and for the over-all national estimates if it were satisfactory for the southwestern states.

The second practical solution is to consider the sample as a series of small samples—one for each subgroup of the population for which separate estimates are desired—and determine sample size separately for each. In effect, this would amount to a disproportionate stratified sample. In arriving at the over-all estimate, then, it would be necessary to use a weighted mean (or other weighted statistic).

11.3 "What Kind of Sample Do We Need?"

As we stated at the outset of this chapter, the questions of sample size and of sample design are closely interrelated. We were able to separate out the question of sample size only because we assumed that only one sampling design was being considered. In the numerical example, for instance, we assumed that this sampling design was simple random sampling.

The discussion of sample size, however, fits nicely, with only slight modification, into the treatment of a choice of sampling design. The starting point is a given cost specification. If the problem of sample design is assumed to be solved, one would figure out the accuracy obtainable with the sampling design already specified and then proceed as before. But if it is necessary also to decide the sampling design to be used, one would proceed (in theory) by considering all possible sample designs and selecting the one which yielded the smallest sampling error for the given cost.

In practice, however, it is impossible to consider all possible sampling designs. There is no mathematical procedure by which the best of all possible designs for a given problem and a given cost can be determined.

If only probability sampling is considered, there are two main steps by which the optimum solution may be roughly approximated.

First, the general sampling approach, such as cluster sampling, is specified. The decision to fix upon a particular sampling approach is made on a combination of judgment, past experience with different sampling designs in similar situations, administrative requirements (such as the feasibility of recruiting local interviewers, costs of traveling of supervisors, etc.), and knowledge of the population being studied.

Second, once the general type of sampling design is fixed, it is necessary to decide the best way to allocate the sample. That is, how many clusters should there be, how should they be formed, and how many interviews should be taken within each cluster? Questions like these are technical and must usually be tackled by a sampling expert. The nature of this assault was indicated in Chapter 8 when we illustrated how the optimum size of cluster might be approximated in cluster sampling without subsampling.

If sampling experts are not at hand, and probability sampling is decided upon, one approach is to get a sampling expert on a consulting basis. An alternative is to copy the sampling designs created by sampling experts for problems that seem similar to the one at hand.

Hence for probability sampling the selection of sampling designs, though technical, is simple in principle. The answers obtained by sampling experts can be accepted with confidence. If, however, nonprobability sampling designs are admitted as possible alternatives (as is usually wise in marketing research), the selection is no longer simple even in principle.

Although it is impossible to *measure* the accuracy of nonprobability samples of different sizes and designs, these different designs and sizes do have different sampling variability. In choosing between different types of nonprobability samples, the sampling errors of probability samples of similar general designs can be used as benchmarks. For most quota samples, for example, one could use the sampling error of stratified cluster samples of similar size and shape as benchmarks. The sampling variability of the nonprobability sample will almost always be greater, but some idea of the order of magnitude can be obtained by such a comparison.

Unfortunately, there is no practicable way to get a measure of the bias in nonprobability samples. By judgment and experience, however, one can perhaps make a plausible guess as to the extent of the bias. For example, it may be helpful to have knowledge of the habits of interviewers when they are given freedom to select respondents within the general restrictions imposed by quotas. It might be that interviewers usually tend to interview the lower segment of the upper economic group and the upper segment of the lower economic group. Moreover, validation (see Quota

Sampling, under section 10.3) provides some evidence as to possible bias, although this evidence must be regarded as relatively weak. A comparison of some aspects of probability and nonprobability sampling seems justified at this point.

While probability sampling has come into increasing use, most work is still conducted by nonprobability sampling. One of the most difficult decisions to be faced in practical marketing research is the choice between the two methods. The difficulty of developing satisfactory criteria for guidance has in part explained persistent and pervasive differences in points of view. Supporters of nonprobability sampling have granted that probability sampling has a proper place—in research by organizations with “unlimited” funds at their disposal. A clear statement of this point of view is given in a recent statistics textbook: “Since the method [area sampling] is so expensive as to preclude its use except by governmental or the largest private polling agencies, we will not give details here.”¹ Similarly, all but the most extreme advocates of probability sampling have admitted a place for nonprobability sampling. It has often been conceded that nonprobability sampling is more attractive the less vital the decisions to be made on the basis of the study. The reason, of course, is that with nonprobability sampling the bias and sampling variability can only be guessed, while probability sampling gives an objective evaluation of sampling error. Therefore in collecting data for very important decisions one does not like to risk nonprobability samples if probability samples are available. Philip Hauser, former acting director of the U.S. Bureau of the Census, has expressed this position:²

In surveys in which the stakes are high, where results must have great precision, where important policy determination may be involved, the risk is usually too great to use judgment nonprobability samples. Consider, for example, the consequences of serious bias and large errors for policy determination by the Congress! In this type of situation you should use the probability sample for unbiased results and measurable accuracy.

Other probability samplers have recognized that occasionally the assumption of effective randomization is sufficiently well fulfilled to permit nonprobability sampling. All these viewpoints are at least partially correct and can in part be integrated into a comprehensive scheme for looking at the problem of deciding whether to use probability or nonprobability sampling.

¹ Quinn McNemar, *Psychological Statistics*, p. 336, John Wiley & Sons, Inc., New York, 1949. (Reprinted by permission.)

² “Factors in the Choice of Alternative Sampling Designs,” Philip M. Hauser, talk on Feb. 3, 1948, given before the Chicago Chapter of the American Marketing Association.

One point must be brought out clearly as a preliminary to the remainder of this chapter. A probability sample is never (hardly ever) obtained in the most rigorous sense. For example, there are almost always some people who will not reply to questions and others who are never found at home. Even though these people are a small proportion of the total universe, it is necessary to make some arbitrary guess as to what would have been found out from them had it been possible to reach them. Otherwise it is impossible to make an estimate which applies to the entire population.¹ Ordinarily, of course, no one worries very much if the excluded portion of the population is a small proportion of the total. But when the excluded proportion is large—as would be the case if no call-backs were made on the families who are not found at home on the first call—the sample is usually termed a nonprobability sample. Obviously no sharp logical distinction can be made, but it is usually possible as a practical matter to reach agreement on what is and what is not a probability sample. Nonetheless, it is also important to keep in mind that the distinction between probability and nonprobability sampling cannot be made a sharp one in practice. It has been argued sometimes that the lack of a sharp boundary between probability and nonprobability sampling means that there is really not much difference in results between the two approaches. As a consequence, the argument goes, there is not much point in going to the trouble of attempting a probability sampling design. This argument seems untenable; it could be used to discourage any undertaking in which a plan could not be perfectly executed.

1. Requirements for accuracy. In considering which of the two types of sampling to use, it seems clear that if the decision which is to be made on the basis of the sample estimate is extremely important, the estimate should be extremely accurate and the accuracy of the estimate should be measurable in the sense of permitting the construction of a confidence interval. One would want a very narrow confidence interval and a high level of confidence for an estimate of unemployment which was to be the basis of public policy. Only probability samples permit the construction of confidence intervals and fairly accurate advance planning as to the width of the confidence interval that will be obtained. Probability sampling, therefore, seems clearly justified because the cost of an erroneous decision based upon a seriously inaccurate estimate would almost inevitably be many times the cost of a survey using probability sampling.

¹ An alternative procedure would be to restrict the definition of the population to people who "would have been" included in the sample and regard the sample which has actually been obtained as a probability sample from this restricted population.

When the potential consequences of a bad decision are not so great, the need for an objective evaluation of the accuracy of the sampling process diminishes. It would be foolish to insist that all decisions be based on information whose accuracy could be measured in terms of confidence intervals. Often relatively inaccurate information is satisfactory. Seldom in marketing research would accuracy be needed of the degree necessary for a correct election forecast in a close race. Whenever the accuracy specifications are sufficiently low, nonprobability sampling is likely to be preferable.

For example, suppose that an advertising copywriter insists that "everybody" reads the advertisements he writes. An account executive in the advertising agency feels that the copywriter is wrong and should be corrected. The account executive talks to 10 employees of the client (outside the advertising department) to find out if they have read the advertisements in question. Only three were able to recall even having seen them. Since the employees of the client would be expected to be more conscious of the advertising than the public at large, this small nonprobability sample would probably suffice to show the copywriter the error of his ways. If the copywriter was still not convinced, he might go out on the street and ask 25 people if they had read his copy in the advertisement. Assuming no bias in his questioning procedure, it is safe to say (on the basis of general knowledge unearthed by marketing research about the readership of advertisements) that he would find that the majority had not read his copy. Questions of this kind arise surprisingly frequently in business, and nonprobability samples may be completely satisfactory for answering them.

The demarcation of the exact boundary between accuracy specifications that would indicate the use of nonprobability sampling and those which would not is much more difficult and must be resolved by judgment in particular problems.

2. Cost of obtaining information. Since the degree of accuracy can be measured with probability sampling designs, the cost of obtaining information of a given degree of accuracy can be estimated in advance or measured after the study has been made. Once a given degree of accuracy is specified or the given total cost has been set up, a skilled probability sampler can usually design a sample that approximates the optimum fairly closely. That is, if a given degree of accuracy is required, he can design a sample that will achieve this objective with relatively low cost; if a given cost is specified, he can design a sample that will yield relatively low error. Thus with probability sampling it is possible to know approximately the accuracy attainable for a given cost and come reasonably close to maximizing accuracy for the given cost. It is probably true, however, that

under some conditions nonprobability samples may, for a given cost, yield more accurate estimates than probability samples; but, in general, this is true only if the given cost is low.

The reason for the possibility of economic superiority of nonprobability sampling lies in the difference in the structure of costs for the two types of sampling. Probability sampling involves much larger fixed costs—costs which do not vary with the size of the sample. In the first place, probability sampling involves the costs of obtaining expert statistical guidance. As we showed in Chapters 8 and 9, probability sampling of human populations can be very complicated and can require statistical competence of the highest order. In the second place, a more elaborate apparatus is required for probability samples in order to select the sample. Maps must be assembled, and a great deal of clerical work is required. The corresponding routine tasks connected with the selection of a nonprobability sample are relatively simple and inexpensive. Third, there is some additional clerical work in making the statistical computations for confidence intervals when probability sampling is used. Fourth, the designing of efficient probability samples is usually a relatively time-consuming process. The work required for probability samples is certainly greater than for nonprobability samples of corresponding scope. Interviewing costs are not completely fixed costs—they certainly vary with the sample size—but the average cost per interview of actual interviewing for samples of all sizes is usually greater for probability samples. This is true because of the greater amount of training required (which will probably diminish as probability samples are more widely used) and the fact that probability samples more often than nonprobability samples include remote and inaccessible places as well as hard-to-interview people who require relatively large amounts of time per interview.

This enumeration of the reasons for relatively high fixed costs connected with probability sampling may perhaps convey the idea that these costs are greater than they actually are. In fact, there seems frequently to be a tendency in discussions of this subject to overstress these costs. It is particularly relevant here to observe that for local studies, particularly studies involving medium- and small-sized towns, it is easy to select an area sample, and the fixed costs of area sampling may be relatively small. Moreover, even in local studies involving large cities, the fixed costs are often relatively small once the sampler acquires the requisite knowledge.

There are other causes for the relatively high average cost of interviews in probability samples. The cost of call-backs in order to include the pre-selected respondents who were not at home is one of the major costs of probability samples. A way has been suggested by Simmons and Politz (see *People Who Aren't at Home*, under section 9.3) to avoid call-backs on

not-at-homes in probability samples, and this may remove or diminish the cost differential.

These differences in the cost structure of the two types of samples do not necessarily indicate that nonprobability sampling is to be preferred when costs are considered. The cost per interview may be higher for probability sampling, yet the "value" of each interview may be considerably higher also. However, the relatively heavy costs—particularly fixed costs—of probability sampling are important considerations when the funds available are small and when the surveys are not of a repetitive nature. When these conditions are fulfilled, the case for nonprobability sampling is often strong.

An organization which is continually surveying a given population will find the burden of fixed costs less oppressive than would the research department of an enterprise that conducts consumer surveys at infrequent intervals. It is often possible, for example, to use the same sampling design repeatedly when repeated surveys are being made. Quite obviously the costliness of a particular study would be much diminished if the expert statistical guidance, the trained field force, the sample designs, etc., were available not only for that study but had been available for other studies which preceded it and would be for others which would follow. It is possible, too, that much the same result ("spreading the overhead") can be achieved by the use of research agencies specializing in probability sampling and working for many clients. But for studies of very small scale, nonprobability sampling is very often indicated.

3. Recapitulation: probability vs. nonprobability sampling. The following very general conclusions have emerged from the comparison of probability and nonprobability sampling: (1) Decisions of great economic importance require the use of information of known accuracy even though the total cost of obtaining it may appear to be high by comparison with the cost of securing a comparable number of interviews with nonprobability sampling, where the accuracy cannot be evaluated. Probability samples should be used in these circumstances whenever possible. (2) The relatively high fixed cost of probability sampling is a disadvantage of the method if the budget available is extremely limited or if sampling is resorted to infrequently. The use of probability sampling may sometimes be completely out of the question and nonprobability samples may be clearly indicated. (3) When the required information is of relatively minor importance and great accuracy is not required, nonprobability sampling is often perfectly satisfactory.¹

¹ Odell and Peterson mention a special advantage of probability sampling that may sometimes be of importance in selecting the sampling design. ("Selecting Sampling Methods in Commercial Research," *Journal of Marketing*, Vol. 15, No. 2.

Other rather obvious comments are also in order. If there is good evidence that nonprobability sampling will work well, say, because of success in the past, one would be more willing to use nonprobability sampling. The danger here, of course, is that nonprobability sampling may not continue to work so well in the future. The *Literary Digest* failure of 1936 was preceded by successes in earlier elections.

Again, there may be strong reasons to believe that an effectively random sample can be obtained by the use of nonprobability sampling. Taste tests have been cited earlier as an instance in which effective randomization might exist.

Finally, there are some investigations that probably cannot be conducted by other than nonprobability sampling. It is hard, at least, to imagine Dr. Kinsey's interviewing a probability sample of the population in which he was interested. Some subjects—other than sex—of interest in marketing research are almost equally inaccessible to probability sampling.

The rough guides suggested above may be helpful in making the sorts of practical decisions which we have discussed. Before leaving sampling, we shall *briefly* mention some other approaches to sampling problems which seem useful in overcoming the sorts of practical difficulties encountered in marketing research. These other approaches are, of course, variations in probability and nonprobability sampling, since these classifications are all-inclusive, but the variations did not seem to fit well into the basic discussion of these types of sampling.

11.4 Adaptations of Sampling Designs for Marketing Research

There are two ways in which traditional sampling methods can be made more useful for marketing research. First, existing probability sampling designs can be made more widely useful by reducing the magnitude of fixed and variable costs which such designs entail. Second, existing nonprobability designs can be modified by introducing into them certain features of probability designs so that accuracy is improved without incurring much greater costs. Both lines of attack should help produce sampling designs which are relatively cheap and easy to apply to the sort of research

pp. 182-189, October, 1950.) Probability sampling designs permit an estimate of population *totals* solely from the information contained in the sample itself, while nonprobability sampling methods do not. Odell and Peterson mention an illustration in which it was desired to estimate *total* shipments of livestock from a certain farming area. Probability sampling was therefore used in this sample.

In a private communication, Roe Goodman of the Survey Research Center points out, however, that probability sampling designs permit an estimate of population *totals* from the information contained in the sample only if "extreme care has been used to see that the sample areas are thoroughly canvassed and that all cases of non-response have been taken into account."

which is often called for in marketing. In this book we shall consider only the second of these two methods of adapting sampling designs for marketing research, since the problem of developing new probability designs is a more technical statistical problem.

The underlying theory in the attempt to improve nonprobability sampling is already familiar from the discussion in Chapter 10. Bias may enter nonprobability samples because of the discretion left to fieldworkers in selecting respondents. This bias can be lessened if the discretion is lessened or controlled. There are various ways of doing this. The way most widely used in nonprobability sampling is stratification, which limits discretion by assigning quotas of certain types of respondents to each interviewer. Samples probably could be improved by a more imaginative conception of modes of stratification. Typically, strata are used because of the belief that the variables used to define the strata are highly correlated with the characteristic being studied. For to the extent that this correlation exists, the elements within strata will be similar.

This consideration of stratification is the only relevant one in devising means for reducing sampling variability, but stratification has, in nonprobability sampling, the additional task of reducing bias. In order to reduce bias, it is essential that the variables defining the strata be related not only to the characteristic being studied but also to the factors that influence the interviewers in their selection of respondents. Since interviewers often avoid the hard-to-get interviews, an obvious suggestion for modes of stratification is that strata be defined in terms of the factors that make interviews hard to obtain. For example, in addition to the conventional strata, the researcher could design samples in which strata were established for first-, second-, and third-floor residents, for the proportion of people not at home on the first call, etc. The major limitation on the use of such operational controls is that it is seldom possible (or, at least, easy) to obtain information about the size of these strata in the population under investigation. The idea is promising enough, however, to warrant further exploration.

In the interest of lessening bias, another way in which the discretion of the interviewer (researcher) can be lessened is by specifying procedures by which the interviewer (researcher) replaces respondents who were originally approached but who for some reason did not get included in the sample. The problem of replacing respondents who were originally selected but who failed to get in the sample is most obvious and pressing in the use of panels, though it exists for all sampling problems. Even though a consumer panel was originally selected by probability sampling,¹ the subsequent "mor-

¹ Experiments have been made with probability samples for panel work, but these experiments have not been very successful.

tality" because of migration, negligence, refusals to continue cooperation, etc., would soon make the panel a nonprobability sample. Whether the original panel was a probability or a nonprobability sample, the problem of replacement is important so that the estimates which the panel provide are unbiased or only slightly biased.

One possible solution is to make random replacements, but this might not be satisfactory. If, for example, single individuals and two-person families show a greater tendency to drop out of the panel, a random replacement of these individuals would result in a higher and higher proportion of larger families. On the other hand, the problem cannot be ignored by simply selecting a big sample to begin with and then not replacing families who cease to cooperate. The proportion of small families in the panel would drop slightly more rapidly if this policy were followed.

A better solution is to "match" the family which does not cooperate with another family which will. A two-person family dropping out of the panel would be replaced by another two-person family, possibly selected to conform also in other characteristics such as income and location of dwelling area. In effect, then, even though one were to begin with a probability sample, the result would soon be a quota sample in which panel members were selected to conform with certain quotas which were established from the original sample. The matching procedure is applicable, of course, not only for panels but for any sampling situation in which call-backs or further call-backs are not possible.¹ The procedure of matching may contribute to sampling bias, but the alternative is probably an even larger bias.

An approach closely related to matching is that of weighting. This method has already been illustrated by the *Literary Digest* poll, in which a statistician used weighting to diminish the bias in an inherently poor sample. Weighting involves establishing the same relative importance for groups in the sample that the same groups have in the population.

Other variations in sampling design permit marrying some of the statistical advantages of probability sampling to the cheapness and convenience of nonprobability sampling. Suppose, for example, that an organization has a field staff in various counties throughout the United States but that these counties have been initially selected by judgment and for convenience. The organization might not wish to recruit new interviewers, because of the expense, but it might decide to institute probability sampling methods *within* the counties. Still another possibility would be to select

¹ Other approaches are possible. Many area samplers prefer to ignore families from whom interviews are not obtained rather than to attempt to replace these families by similar families.

blocks within a city by random methods and then allow interviewers to pick out the families to be interviewed within these blocks. These two suggestions should suggest other similar adaptations.

Finally, and perhaps most intriguing from the standpoint of practical marketing research, is the possibility of removing bias attributable to discretionary selection without assuring known probabilities of selection to elements in the population. We shall consider two methods of doing this. The first was apparently introduced by Alfred Politz under the name of "the random walk." While the details of this method haven't been published, apparently the idea is to start an interviewer from a point chosen by some random method and then devise instructions so that the interviewer, as he proceeds from house to house, would have no choice as to the route or choice of respondents. If he comes to a street corner, for example, some random method, such as a coin toss, might be used to determine the direction he would turn. Systematic sampling could be combined with this approach so as to specify uniquely the respondents to be included.

The second method also involves diminishing the discretion left to interviewers in their selection of respondents. The method can be illustrated by the use of an example. Consider the dilemma of a marketing research firm engaged to study a characteristic of all the housewives in the United States on a total budget of \$500. The budgetary limitation precluded, of course, the use of a national sample or the use of a probability sample on even a regional scale. The study was, therefore, confined to a small part of the country, namely Philadelphia and all small towns within a 50-mile radius. A list (in geographical order) of all the shopping areas in Philadelphia was obtained, and a systematic sample of these areas was selected. (This required about four hours of clerical time.) Next an interviewer was stationed at a central point in each shopping area drawn into the sample. Most of the housewives in each area visited its shopping center, and most could be expected to pass the point at which the interviewer was stationed. The interviewer attempted to obtain an interview from the first woman passing by. As the interviewer finished the questionnaire, she looked up and attempted to interview the nearest woman, regardless of whether she appeared to be easy or hard to interview. The process continued in this way. A similar method was used in a systematic sample of the small towns around Philadelphia.

The opportunity of personal judgment in the selection of respondents can be almost completely eliminated by this approach if the interviewers make an honest attempt to follow directions. (There are instances in which judgment will be necessary, as when several women are walking side by side.) Even though the opportunity for bias in selection of respondents

has been lessened, the sample is not, of course, a probability sample. For example, too few of the women who shop at the busy times of day will be selected and too many of those who shop during off hours will be chosen. (During a busy period several women may pass by while a single interview is being completed; during other times, the interviewer may have to wait for several minutes between interviews.) Women shopping at the larger shopping areas will probably have a somewhat smaller chance of being included than will women in the smaller shopping areas. This is true because the areas are selected with equal probabilities, and a smaller proportion of the women in the larger areas will be interviewed.

There are other characteristics of this proposed technique which distinguish it from probability sampling. It is true, however, that the method does eliminate two major sources of bias: the selection of areas by judgment and the selection of respondents within areas by judgment.

Another illustration of this approach has been used by George Brown. Suppose the problem is one of selecting a sample from a city for which block statistics are available. It is then possible to select blocks with probabilities proportional to the number of inhabitants in each block at the time of the last census. Within any block selected in this fashion, some simple mechanical rule is devised to provide a starting point, that is, a first family to be approached by the interviewer. Then further mechanical instructions are given for the selection of families other than the first. For example, the interviewer might be told to proceed east from the starting point, keeping on the same side of the street, and possibly passing into additional blocks in the course of her interviewing. All families encountered would be interviewed.

The interviewer may be instructed to complete a certain number of interviews in this fashion. For example, the number might be roughly equivalent to a day's work. The problem of nonresponse might or might not be ignored. Brown suggests that in the early stages of the study, checks be made to determine the approximate magnitude of differences between those who are found at home and those for whom call-backs are necessary. Such checks are used to decide whether or not it is worth while to worry about the problem of nonresponse.

The method thus described is objective and indeed resembles in many respects cluster sampling without subsampling, in the case in which cluster sizes are equal. Actually there are unknown biases in the method, such as the slightly varying probabilities of initial selection of starting points occasioned by the use of earlier census data. For example, the use of out-of-date information in the design of the sample may possibly introduce bias, whereas in a strict probability sample this would not be so. None-

theless it seems reasonable that the various biases inherent in the method would be small and that an application of the simple formula for confidence intervals for the case of sampling with equal-sized clusters would be quite satisfactory for many marketing problems. This method, it should be observed, is similar in certain respects to both the random-walk method of Politz and an innovation known as the "equal-take" approach, apparently originated by Raymond Jesson of Iowa State College.

PART IV

COMMUNICATION AND OBSERVATION

CHAPTER 12

INTRODUCTION TO THE PROBLEM OF COMMUNICATION AND OBSERVATION

12.1 Introduction

A marketing research project is designed to provide information which will be useful in making marketing decisions. Typically, this information is about consumers and is secured either from consumers or from retail outlets, although occasionally attention is focused on the distributive system itself and information may then be secured from wholesalers, brokers, salesmen, dealers, etc. The first operational problem—sampling—concerned the specification of the particular consumers or outlets which could best provide the information required by the research project. This problem was considered in the preceding chapters in Part III. Now, in Part IV, we are ready to discuss the principles relating to the securing of information from the consumers, dealers, etc., who have been selected by the sampling process. In Part III we tried to answer the question, “*From whom* is the information to be obtained?” In Part IV we try to answer the question, “*How* is the information to be obtained?”

Both processes are essential to marketing research, and both present difficulties. The difficulties exist because errors in the information that is sought can arise in both processes. Since research is undertaken in order to get accurate information, a major objective of this book is an investigation of the possible sources of errors.

The basic objective in discussing these sources of error is the discovery of ways to measure and control them. Measurement and control are closely related. To see this, reconsider briefly a part of the discussion of Part III. There are two sources of error inherent in sampling: sampling error and bias. Bias can either be eliminated or kept within narrow bounds by the use of probability sampling, and sampling error can always be measured or approximated so long as probability sampling is used. Most of the theory of probability sampling is directed to methods of minimizing the sampling error, assuming a certain fixed amount of money to be spent on the sample. Therefore it is possible to exert some control over the error arising in sampling; and when such errors cannot be eliminated, it is possible to measure the error which remains by the calculation of confidence intervals. The main shortcoming of nonprobability sampling methods lies

in the inability to measure the error or to be sure that prescriptions for controlling error are really successful. The effectiveness of devices for control must remain uncertain in the absence of the possibility of measurement.

In Chapter 11 we indicated that nonprobability sampling could best be discussed against a background of probability sampling theory. The discussion of the second operational problem—obtaining information from the people included in a sample—often will be much like Part III would have been if probability sampling were unknown. If probability sampling had not been discovered, we would have been confined to an exposition of the rule-of-thumb procedures which have been evolved in practical applications of nonprobability sampling, and we might also have added *a priori* speculations and guesses as to which of these procedures were best. The present treatment of the problem of obtaining information from specified individuals must at times consist merely of such speculations because of the lack of measuring devices.

For example, consider the problem of asking an individual about his income in the previous year. One might get many possible answers or none at all. The respondent might say that he made \$3000, thinking merely in terms of round numbers; with more care he might say that he earned \$3182.72; if he wished to brag, he might say \$5000; and if he wished to evoke sympathy, he might say \$1200. The actual response could be affected by many things such as the wording of the question, the way the question was asked, the personality of the interviewer, the personality of the individual being interviewed, the circumstances of the interview, etc. Under certain conditions people will not or cannot give the correct answer to a question. In the absence of any objective test of the correctness of the answer, one can do no more than (1) attempt to classify the types of influences that are operating, (2) indicate, on the basis of reasonable surmise and previously accumulated empirical evidence, the kinds of effects that can be expected, and (3) suggest how the affecting factors should be manipulated so as to minimize response errors.

As a consequence of the lack of measuring devices we often will reach conclusions that are meager and disappointing by comparison with the relatively positive statements we were able to make on problems of sampling. Yet we hope to furnish the reader a useful framework to guide his own thinking and practical work, and to present the generalizations which are warranted by present knowledge.

In this introductory chapter we attempt to devise a useful classification of (1) the kinds of response errors that can be expected, (2) the sources of these errors, (3) the basic techniques for getting information, (4) the criteria for judging these techniques, and (5) the kinds of information that are useful.

12.2 Kinds of Response Errors

The concepts of statistical inference provide one classification of response errors. The answer to a particular question, for example, can be considered as a single observation from a universe made up of all the possible answers which might have been obtained. Unfortunately such a universe is different from the kinds of universes considered in Part III; one can't assign a number to every possible observation (response) and use a table of random numbers to obtain a random sample from it. It is necessary to take the answer that the respondent gives without checking the randomness of the process by which the answer was obtained. Thus the answer actually obtained is a nonprobability sample made up of one observation. This answer is subject to two sources of error—random error and bias.

In some instances the bias might be zero; for example, the respondent might not know his *exact* income and might be equally likely to guess too high by a certain amount or to guess too low by the same amount. Then only random error would be involved, and the response might be thought of as an unbiased estimate of his true income. That is, if the question could somehow be repeated many times under essentially the same conditions, the average of all the answers would be approximately equal to his true income. The repetition of the question under essentially the same conditions is only conceptual, however. In practice, the individual might be irritated at the first repetition of the question and refuse to give an answer; certainly his patience would soon be exhausted. Further, it is extremely doubtful that the successive answers would be independent of each other. Thus it seems highly improbable that any *measure* of this sampling error can be obtained, although attempts have been made. And, as we soon show, it is unlikely that this kind of error can in practice be distinguished from other kinds.

A term often used to describe these random fluctuations is "reliability." If the random variability is small—if one would get about the same answer every time a question was repeated to the same individual under essentially the same conditions—it would be said that the *reliability* of the answer to this question was high (with respect to this individual).

Suppose, now, that the respondent tends always either to understate or overstate his income. Then the average of all the answers obtained from repeated questioning, under essentially the same conditions, would not equal his true income; the difference between the average of his answers and the true income is the bias. The term "invalidity" is frequently used as being synonymous with "bias." If the bias is small, the validity is high, and vice versa. While the rules for eliminating (or controlling) bias in the sampling problems considered in Part III were relatively simple, no such straightforward rules are available to dispose of this second sort of bias.

1. Extension of reliability and validity to a group of individuals. So far we have considered the responses of a given individual. Some modification is needed in the conceptual framework for thinking of the responses for a group. For example, one person might tend to understate his income while another might tend to overstate his; although the individual answers would be invalid, the errors introduced would be partially or completely compensating. It is possible that the answers for the *group* would be free of the biases of the individual answers.

Even though the individual biases might tend to cancel for the groups, there could still be random response errors (which would include the biases in the individual answers). While the answers for the group might be free of invalidity, they would still be subject to unreliability. The unreliability for a *group* of answers, then, can be attributed to both the invalidity and the unreliability in the *individual* answers.

This classification of response errors into unreliability (sampling error) and invalidity (bias) is a common one in the literature of opinion and attitude research and marketing research. That is why we have mentioned it. We do not use it in this book, however, for two reasons. First, there seems to be no practicable way to measure separately these kinds of errors, although many attempts have been made and many claims of success have been reported. These attempts at measurement take one of two forms.

In one form, the information is sought in the same way at two or more times. If personal interviews are being used to get the information, the same question is either asked twice in different forms during the same interview or during two or more interviews. This approach has serious weaknesses. If the question is put in two different ways during the same interview, and if the change in wording is slight, the identity of the meaning of the questions often will be apparent to the respondent and he will tend to give a second response consistent with the first. If the wording of the questions is sufficiently different to disguise the intended identity of meaning, the meaning of the questions may actually be different *for the respondent*, weakening the possibility of measuring the reliability of the response to a single question. These same difficulties exist to a different degree when the same question is asked during successive interviews. Further, additional ambiguities are created when the questions concern attitudes or intentions, for these may have actually changed between interviews, making it impossible to distinguish such changes from unreliability. While the concept of reliability and the methods outlined above for testing it may have some value, the concept does not seem so useful for our purposes as the ones to be developed.

The second approach to measuring unreliability and bias involves comparing two independently arrived at measures of whatever is being investi-

gated. Thus, if aggregate income was being measured, the survey results could be compared with data of the Bureau of Internal Revenue. Such an approach permits an estimate of the combined importance of random variation and bias in the group responses. It is so seldom possible to make such a comparison, however, that the practical usefulness of the method is small. Further, the comparison does not make it possible to separate the errors attributable to random variation and to bias.

Another reason for rejecting this classification of errors into unreliability and invalidity is that this classification does not suggest practicable means of controlling these errors as well as another classification might. The justification for this second reason, it is hoped, is to be found in the alternative classification suggested in this chapter.

12.3 Sources of Response Error

Alternative methods of controlling errors can best be judged by measurement of the errors that are associated with each method. In the absence of such measures, methods of control must be judged by theoretical reasoning and some imprecise comparison of the results of using alternative methods. The classification of errors discussed so far is a kind of statistical description without reference to the causes of error. If errors can be classified according to these causes, there is greater hope of controlling and lessening them.

There are two basic causes of all response errors: the unwillingness and the inability of respondents and investigators to provide and record the correct information. The providing of information can be thought of as involving two states: formulation and communication. The two basic causes can operate at either stage. When the respondent is either unwilling or unable to *formulate* the correct information, we call the resulting response error "inaccuracy."¹ When the respondent is either unwilling or unable to *understand* what is requested or to *communicate* coherently the desired information, we designate the resulting error "ambiguity." Since in the following chapters we use this classification of errors ("inaccuracy" and "ambiguity"), it is important that these terms be understood completely at this point.

Ambiguity exists either because a respondent does not understand what is said (written) to him, or because the investigator does not understand a response, or both. After World War II there was a national survey which

¹ The investigator, too, can be unwilling or incapable of formulating or recording the correct information. The role of the investigator in formulating information is primarily of importance in the observational techniques of getting information. In the interest of brevity, most of the following discussions in this introductory statement will consider only the respondent.

asked the question, "Is the United States better off now than before the war?" An answer of "yes" to this question is ambiguous. It is ambiguous not because of the answer but because subsequent investigation showed that some people thought they were answering a question about military security, others thought they were answering a question about the government's financial status, and still others answered what they believed to be a question about the happiness and welfare of the people. In other words, the question that many people answered was not the same question that the research agency thought it was asking. In the field of attitude and opinion research such a question would be termed "multidimensional." Multidimensionality is one source of ambiguity. Many marketing questionnaires about preferences are multidimensional because not all the relevant circumstances that would attend the actual presentation of the alternatives in the market place are indicated when the request for the statement of preference is made. For example, some respondents might indicate a preference for A over B on the assumption that the prices of A and B would be equal, while other respondents might answer in the belief that there would be a significant difference in the prices.

Sometimes responses are ambiguous even though the question is clearly understood to have the same meaning by the research agency and by all the respondents. For example, a questionnaire designed to measure public opinion on disarmament included a question about the optimum size of the air force. Some persons, unfamiliar with military affairs and desirous of maintaining a preponderance of military power in the United States, answered, "One thousand planes." To the interviewer this answer was an indication of a desire for virtually complete disarmament, and the response was so recorded. The source of ambiguity in this case was a lack of agreement between the respondent and interviewer (or analyst) on the significance of the response itself. This is the second source of ambiguity. If both question and response have the same meaning for research personnel and for respondents, a lack of ambiguity will be achieved. In the following chapters there is a discussion of ways of detecting and diminishing ambiguity.

Inaccuracy, in contrast to ambiguity, arises not because of deficiencies in the communication between investigator and respondent but because of unwillingness or inability of the respondent to formulate or make available the correct information. If a respondent were asked to state the size of his income during the previous year, inaccuracy would arise from either unwillingness or inability. The respondent might know his income and be unwilling to divulge the information, or the respondent might not know his income and thus be unable to figure it out. On the other hand, it is possible that the respondent did not know his income and was incapable of figuring it out under the circumstances in which the request for the information

was made. Such inability might lead to a refusal to provide any answer or to make a guess differing significantly from the true income. Response errors of this type are inaccuracies. An error of this sort differs in its cause and control from an error attributable to ambiguity. If the respondent had thought that the term "income" had been meant to include only wages (and not dividends, interest, transfer payments, etc.), the resulting reported income would be in error because of a breakdown in communication and would, therefore, be attributed to ambiguity.

By dividing the process of providing information into two stages—formulation and communication—and by considering the basic causes of all response errors, the classification of these errors into inaccuracy and ambiguity suggested itself. This classification seems better suited to a discussion of ways of detecting and lessening response errors than does the conventional classification into unreliability and invalidity.

12.4 The Techniques for Securing Information

Some information can be secured by armchair speculation—usually involving an intuitive discovery of the implications of previous experience—and some, it has been alleged, has been secured through revelations or insights not based on previous experience. The method of intuition is an essential and powerful tool of scientific inquiry, but there is very little one can say about how to go about "intuiting" things. Perhaps even less could be said about the method of revelation. Therefore a formal discussion of marketing research can cast little light on these two methods.

1. Communication and observation. Ultimately, marketing research personnel rely for their information either on what is heard (read) or what is seen. Thus, the two general methods by which information can be secured from people are communication and observation.

The internal data of enterprises and outside secondary sources provide valuable information which at first thought seems to involve neither communication nor observation. It is difficult or impossible, however, to find secondary data which did not originate as a recording of either a communication or an observation. Even when secondary data are considered, then, it seems safe to say that all marketing information is obtained by communication or observation.

The number of ways in which consumers can communicate verbally with research personnel is great. The mail questionnaire, the consumer panel, and the personal interview, in its many forms, provide mechanisms of communication. Communication is capable of securing all types of marketing information. The influence of factors affecting the efficiency of each communicative mechanism differs from project to project, and the ways of deciding which mechanism is best adapted for a particular project will require discussion.

Observation, too, is various in its techniques and applications. Inventories of merchandise on pantry and bathroom shelves give information about what brands and colors of pudding, salad dressing, tooth paste, and cough medicine have been purchased and used. Inventories of bookshelves, magazine racks, waste baskets, and trash piles provide some information about the distribution of periodical and other literature. Observation of the actual sales of a product is one of the most useful devices for studying consumer behavior. When such observation is used in the "sales test," a kind of marketing experiment, information about the relative appeal of alternative packages or different methods of display, about the influence of price changes on rates of purchase, and about many other things can be gained. The usefulness of such information in forecasting future behavior is evident.

Observation, unlike communication, cannot yield *directly* information of certain kinds (*e.g.*, intentions or attitudes and opinions). One may, of course, draw inferences from observation which tell something about these "nonobservables," but such inferences are strictly valid only when one knows that intentions or attitudes are reflected in actual behavior in some definite way. It is always possible, unfortunately, that true attitudes and intentions may not agree with the inferences made solely from observation.

12.5 The Criteria for Judging Techniques

Techniques for getting information should obviously be judged in part by the extent to which the information that is secured is free from error—either inaccuracy or ambiguity. The same kinds of errors occur when observational techniques are used as when communicative techniques are used, and the same criteria are therefore appropriate.

In addition to accuracy and lack of ambiguity, a third criterion must also be used—the cost of alternative techniques. In the discussion of sampling it was seen that inferences can be more or less precise—*i.e.*, subject to greater or smaller errors—and that inferences of each degree of precision have associated costs. The costliness of the different kinds of information and of the different means of securing each type must be considered. Although a complete psychoanalysis of a sample of New York's population might possibly yield both less ambiguous and more accurate information about the motives of people with regard to their allocation of income than would alternative techniques, it is almost certain that the cost of such a procedure would make it unfeasible.

12.6 A Classification of the Kinds of Information Useful in Marketing Research

Different types of information can be used as a basis for marketing decisions. Because of the multiplicity of means for reaching the objective of

marketing research—enlightened marketing decisions—it is necessary to understand the difficulties in securing, and the potentialities of fruitfully using, each type of information. We shall provide in this chapter what it is hoped will be a useful classification of the basic types of information required in marketing. This particular classification, although not logically neat, permits a simpler and clearer exposition of the problems and techniques associated with each type of information. Each of the remaining chapters in Part IV deals in detail with one of the basic types of marketing information.

The following example, supplied by Market Facts, Inc., shows (among other things) that information of different types may not always point to the same conclusion. The distinction, then, between different types of information is of real importance in practical problems. A sound understanding of the strengths and weaknesses of each type of information is necessary in arriving at a practical decision as to what should be done. The example is presented here just as it was written up in a preliminary report, except that the product name has been changed.

The sales tests and the interviews were conducted in order to see how consumers could be expected to react to advertisements which emphasized different themes. The different research techniques produced different answers, and an understanding of the limitations of these techniques is essential for deciding which answer is probably correct. There are, of course, two interpretations of the data in the following tables, although these interpretations overlap somewhat and do not exhaust the possible interpretations. The first is that all the techniques measure essentially the same thing but that some are more sensitive than others. The second is that some techniques give measures which are really relevant to the problem while others do not. In deciding between these interpretations, one would not only have to understand the inherent limitations of the different types of information but would have to take into account the statistical significance or insignificance of the differences between reactions to the advertisements indicated by each of the techniques.

Ideally, almost all information sought in marketing research is used as a guide to future action. Information is useful as a guide in so far as it helps in predicting the consequences of marketing decisions. A manufacturer is interested in reasons why people buy his product because he can use this information to revise, or to refrain from revising, marketing policies in order to attain some objective such as increasing sales without increasing selling costs. Any decision he makes about these policies would then involve the assumption that consumers, on the whole, would react in a certain way after ("because of") the decision. This assumption is a prediction, whether the manufacturer consciously recognizes it as such or not. Any information about consumers might be useful in arriving at such a predic-

SUMMARY OF FINDINGS: TEST STUDY ON RESEARCH TECHNIQUES MEASURING RELATIVE EFFECTIVENESS OF TWO ADVERTISEMENTS

I. *Interviewing Approach.*—Customers were asked their preferences. (Method IA and IB were rotated.)

A. "Here are two advertisements. Which one do you like best?"

	Among former purchasers of canned hamburgers	Among nonpurchasers of canned hamburgers	All persons interviewed
Like "New" advertisement best	61.6%	59.3%	60.4%
Like "Fresh Flavor" advertisement best	37.7%	39.5%	38.7%
No preference	0.7%	1.2%	0.9%
Number interviewed	151	172	323

B. "Here are two advertisements. Which one do you think would be more likely to cause you to buy this product?"

	Among former purchasers of canned hamburgers	Among nonpurchasers of canned hamburgers	All persons interviewed
"New" advertisement more likely to cause to buy	60.0%	55.8%	57.9%
"Fresh Flavor" advertisement more likely to cause to buy	39.4%	39.7%	39.6%
Both would have about same effect	0.6%	4.5%	2.5%
Number interviewed	165	156	321

II. *Attention Approach.*—The ads were placed on a stand near the entrance in alternate 15-minute periods and the reactions of people were observed.

	Among people passing by the	
	"New" advertisement	"Fresh Flavor" advertisement
This per cent picked up, stopped, looked at the ad	21.1%	13.0%
Number passing by.	270	292

III. *Sales Test Approach*.—Handbills were passed out to customers (alternating between the two ads—first one, then the other). Then actual sales of the product associated with the advertisement being used were determined.

A. Results when 10¢ coupons were used.

	Among people receiving the	
	"New" advertisement	"Fresh Flavor" advertisement
This per cent redeemed the coupon	17.6%	17.4%
Number of people receiving coupon and advertisement	580	580

B. Results when only the advertisement, and no coupons, was used.

	Among people receiving the	
	"New" advertisement	"Fresh Flavor" advertisement
This per cent bought the product	3.2%	2.5%
Number of people receiving the advertisement.	285	285

tion. Any attempt at prediction (and all decisions involve such an attempt, either explicitly or implicitly) rests upon the belief that the relationship between what has happened and what will happen is not capricious and arbitrary but to some extent stable and understandable. Unless he has faith in occult arts, the marketing researcher must believe that consumers' behavior tomorrow depends in some way upon yesterday's experience.

1. Information about behavior. The simplest, and possibly most frequently sought, basic type of marketing information is information about the behavior of consumers, dealers, distributors, etc. By making certain assumptions, tacitly or explicitly, the research man or marketing executive can use such information as a guide to what will happen in the future. Under favorable circumstances the light of experience and logic may be focused on information about past behavior to reveal a reflection of the future.

Any marketing research project which indicates, for example, what brands of tooth paste have been used and at what rate, what advertisements have been read, how incomes have been spent, or how products have been used is an example of marketing research that provides information

about the behavior of consumers. A correct interpretation of such data can be useful in predicting future brand preferences and rates of consumption, future advertising readership, etc., and such predictions can be the basis for marketing policy. One may appraise the behavior of consumers and make some projection into the future. Sometimes people are reluctant to take this step. There is an alternative, though it has dangers, which we now consider.

2. Information about intentions. Each consumer might appraise his own previous experience and make some evaluation of it with regard to his own future behavior. This is what consumers did after World War II when they were interviewed concerning their buying intentions by the research staffs of many large publishers. The marketing man who uses data from a research project of this type interprets not past behavior but expressed intentions regarding future behavior. Thus information about prospective or future behavior is a second basic type of marketing information. Although the first type of information (behavior) also sheds light on the future, it does so less directly, though possibly more satisfactorily, than information about intentions. (Most people would probably consider last year's deeds a better guide to prediction than this year's New Year's resolutions.) Consumers providing information of the first type answer the question, "What did you do?" or "What are you doing?"; those providing information of the second type, "What will you do?"

3. Opinions and attitudes. Sometimes predictions can best be made by asking, "What do you think?" or "What is your attitude (or opinion)?" Although this third type of information is perhaps more important in the forecasting of political behavior than in the forecasting of marketing behavior, opinion and attitude research has important applications in marketing. After World War II numerous important industrial strikes afflicted the economy and received widespread publicity. Many industrial managements thought it prudent to keep informed about the public's attitude toward management during this period of strife. Attempts were made to measure the public's attitude, and the information was used to help formulate—among other things—advertising policy. More frequently marketing executives seek the consumers' expressions of preference regarding designated alternatives. Sometimes the alternatives are packages; sometimes advertisements; sometimes shapes, flavors, colors, gadgets, or trade names; and sometimes store locations, methods of delivery, etc. Such statements of preference are, of course, merely expressions of attitudes toward a series of stimuli (for example, advertisements) *relative to each other*.

The difficulties and techniques associated with the securing of information on attitudes or opinions are different from those associated with the

first two basic types of marketing information. For example, psychometric concepts and techniques may be needed. Moreover, while the accuracy of information of the first type can often be verified by direct observation of the evidences of past behavior (*e.g.*, home inventory of branded articles on pantry shelves) and that of the second type by observation of future behavior (*e.g.*, measurement of future sales), no reference to overt behavior is possible in checking the accuracy of statements concerning attitudes or opinions. As has often been pointed out, no one has ever reported seeing an attitude. One could, in principle, verify the statement, "I will buy a secondhand Buick next week." No such verification is possible for the statement, "I prefer this advertisement to the other." Sometimes attempts at verification are made by relating expressed attitudes to observed behavior, but there is a certain self-contradiction involved in these attempts. Consider, for example, the statement, "I was never a Nazi at heart." One couldn't refute this statement even if it had come from Adolph Hitler; for it is perfectly conceivable that people may act in one way and "believe" something which *seems* entirely inconsistent with their actions. At the same time, we can't push this too far; people can, and often do, make useful inferences about attitudes from observation of behavior.

The terms "attitude" and "opinion" are often differentiated in psychological discussions, but the distinctions made are not always the same. One such distinction is that an attitude is a propensity to act in a certain way while an opinion is a person's verbal expression of an attitude. In this book, however, we use the two terms interchangeably.

4. Knowledge and habits of thought. A fourth type of information that is useful concerns people's knowledge and habits of thought. One way of testing the effectiveness of advertisements is to see the extent to which people have learned the message conveyed by the advertisement. One way of selecting a new name for a product is to see what associations alternative names have for respondents and to measure the ability of respondents to remember these alternatives.

5. "Reasons why." There is a fifth basic type of marketing information, and it is more complex than any other. Although it is frequently possible to predict merely by observing the pattern of the past, by accepting expressed intentions, or by noting attitudes and opinions, attempts are often made to discover causal relationships before a prediction is made. When information of any one of the first four basic types is secured, knowledge of causal¹ relationships can be secured only by inference by the research man. The tools of inference, it may be added, are not particularly rigorous ex-

¹ "Cause" is an unfashionable word for many modern philosophers and scientists. In its place some phrase such as "invariant conjunction" or "association" is used. We shall be old-fashioned in this discussion.

cept when experiments are employed. This task of inference on the part of the marketing researcher can be transferred to the consumer. By introspection the consumer can *attempt* to discern his own motives. Thus he may be asked the reasons for his behavior or attitudes. Under these circumstances the consumer answers the questions, "Why did you do that?" "Why will you do that?" "Why do you feel that way?" Just as asking information about future behavior allows the consumer to interpret past behavior and project it into the future (and leaves the research man the task of interpreting expressed intentions), so securing information about motivation permits the consumer to infer causal relationships (and leaves the researcher the task of interpreting such inferences). A project designed to ask consumers *why* package C is preferred to A, B, or D or *why* advertisement 4 is preferred to 1, 2, or 3 is a project seeking information of this fifth basic type.

12.7 The Organization of Part IV

In Part IV there are five chapters after this introductory chapter. Each deals with one of the five types of information useful in marketing research. At the beginning of each of these chapters is a discussion of the kinds of marketing decisions for which the type of information is most useful. Following this is a discussion of the means of getting each type of information and of the degree of inaccuracy and ambiguity, and the cost, of information associated with each means.

CHAPTER 13

SECURING INFORMATION ON BEHAVIOR

13.1 Introduction

Before analyzing the problems peculiarly associated with the securing of information on behavior, we shall mention some business situations in which such information is useful.

One basic problem of manufacturers is knowing the current rate of retail sales of their product. Surprisingly enough, this information is often unavailable without the use of the special techniques of marketing research. Factory shipments to distributors, for example, may be known with great precision, yet the movement of the product from the retailer to the consumer may be known only through crude estimates. It is possible for factory shipments to be increasing while consumers are sharply curtailing their purchases of the product—resulting in accumulation of inventories by distributors. These inventories could become large enough to cause a reduction in orders at the factory and factory shipments. Between the time when consumer purchases were first curtailed and the time that factory shipments dropped off, several weeks or even months might have elapsed. Had it been possible to find out about the decline in consumer purchases at the time it occurred, adjustments in factory production schedules, promotion and advertising plans, etc., might sooner have been made.

In late 1948 and early 1949 there was a sharp decline in factory shipments for many durable consumer goods. It was later apparent that one of the main reasons for this decline was that dealers and distributors had become anxious to reduce the large inventories that had been accumulated. Consumer buying did not fall off as sharply as did factory shipments. Unless there had been some direct way of finding out about consumer sales, the decline in factory shipments could have been seriously misinterpreted. Of course, in this illustration, the manufacturer could have obtained information informally from salesmen, distributors, dealers, etc. who could have told him in a rough way what was actually happening. But such information is likely to be distorted by wishful thinking, waves of pessimism, etc., while direct information on consumer buying behavior would be free of these biases.

For manufacturers of many nonperishable consumer products, such as canned goods, soap, or sugar, information about the rate of purchase by

consumers may not be sufficient, because inventories of *consumers* can fluctuate. A rumor of the imminent government rationing of sugar might cause housewives to stock up on this product. Then, when the rumor proved to be false, housewives might curtail purchases of sugar until their stocks were down to a more nearly normal level. Retail sales would first rise sharply and then fall below normal. The desirability of obtaining knowledge about actual consumer behavior is obvious. This problem is particularly important with respect to special short-term price reductions, special promotions, premium offers, etc. In order to evaluate the effectiveness of these special "sales," actual consumption of the product before, during, and after the "sale" must be measured.

A basic consideration in planning advertising is the selection of media. The size of the audiences¹ of the competing media is of importance in this selection. For example, in deciding whether to put an advertisement in *Good Housekeeping* or *Better Homes and Gardens*, it would be desirable to know (among many other things) for each periodical the average number of readers per copy. This information cannot be obtained from sales figures alone, which are directly and easily available either from the publishers' own records or from the Audit Bureau of Circulation. These sources unfortunately do not have data from which it is possible to infer the average number of readers per copy. The further data must be obtained directly from consumers. The measurement of readership is an important example of research devoted to finding out about behavior.

The accounting records of the individual firm, particularly those which deal with sales, are a frequently used source of information on the behavior of aggregates of consumers. Analysis of these records is often called "sales analysis." A discussion of the many ways in which such data can be analyzed belongs more appropriately in a specialized discussion on statistical analysis rather than in a general study of marketing research.² The weaknesses of such data for a wide variety of problems are suggested in the examples given at the beginning of this chapter and in Chapter 4, and the usefulness of these data in suggesting hypotheses regarding consumer behavior are discussed in Chapter 5.

There are many other examples of the usefulness of information on behavior. A judgment as to the relative effectiveness of advertisements can often be made after determining either how many people have actually read the advertisements or how many people have been induced by the advertisements to buy the product. A firm appraising the value of a spe-

¹ "Audience" is a general term including the readers of periodicals, the people listening to radio programs, seeing television shows, etc.

² For a lengthy treatment of this subject see Donald R. G. Cowan, *Sales Analysis from the Management Standpoint*, University of Chicago Press, Chicago, 1938.

cial mechanical attachment on its product would want to know whether people who have already bought the product actually use it. A company analyzing its sales might want to measure variations in consumer buying in different parts of the country, among different economic classes, etc. The list of examples could be extended, but enough has already been said to indicate the importance of this type of information. We must now turn to a consideration of the techniques for securing information about behavior.

The two main techniques are communication and observation: one can ask people what they have done or observe current activity or evidence of past activity. In addition there is a third method, which shares some of the characteristics of each of the first two, in which people themselves make records of what they do. All the variants of these basic techniques (*e.g.* mail questionnaires, telephone interviews, personal interviews by field investigators who may conduct any of various kinds of interviews, store or home inventories, etc.) are limited in the quality of the information obtained by the ability and willingness of people to provide information.

The level of willingness and ability for most respondents is probably less than the inexperienced researcher would guess. Cantril, for example, found that respondents interviewed and reinterviewed within a three-week interval gave consistent answers regarding car ownership only 86 per cent of the time. Reinterviews concerning voting behavior in the preceding election yielded the same responses as the original interview only 87 per cent of the time.¹ It seems reasonable to believe that statements concerning less important events would yield even less satisfactory results.

The challenge to personnel in marketing research is a real one. Often problems which seem intrinsically very simple, such as finding out how many people have read an advertisement, require great skill and ingenuity. The various techniques will yield information of greater or lesser usefulness according to the extent to which they are compatible with the willingness and ability of respondents. The criteria for judging the information provided by these techniques are accuracy, freedom from ambiguity, and cost. These criteria are discussed in the preceding chapter and will now be used in discussing the techniques for getting information on behavior.

13.2 Observation

Observation is frequently used in obtaining information required in marketing research investigations. Instead of (or in addition to) *questioning* a respondent as to the brand of tires he uses, the fieldworker might *look* directly at the respondent's car. Similarly, the fieldworker might inspect

¹ Hadley Cantril, *Gauging Public Opinion*, pp. 102-103, Princeton University Press, Princeton, N. J., 1947.

kitchen and bathroom shelves, magazine racks, closets, etc., in order to find out about past purchases. Direct observation of kitchens in order to check on inventories of food products—particularly packaged and branded products—has been used so frequently that a special name, “pantry inventory,” has been applied to this technique. In each of these examples, the things being observed are the evidences of past behavior.

Purchasing behavior may also be observed as it takes place. This use of the observational technique seems to have attracted less attention in discussions of marketing research than observation of the results of past behavior, yet it is at least equally valuable. A marketing research man trying to find out the reasons for lagging sales can often find out a great deal by observing consumers in the act of purchasing. He may even go one step further and, while observing, participate in the purchasing process himself, either as a customer or as a salesman. In this way useful hypotheses concerning the reasons for lagging sales can often be discovered. Sometimes these hypotheses may be so plausible that policy decisions can be taken without further research. On other occasions the hypotheses may serve to guide further research.

Some further examples suggest the variety of ways in which observation may be used.

1. Up-to-date information on competitors' actions, particularly with respect to pricing, is needed by department stores. Hence a regular practice has evolved of “shopping” the competitors, that is, sending an observer to find out what the other stores are doing.

2. One marketing research man arranged to take moving pictures of consumers in supermarkets while they made up their minds on purchases. These pictures were shown to executives in the sales and advertising divisions of a large manufacturer. These executives found out some things about consumer behavior that they had never suspected before, in spite of their long experience in marketing. For example, some housewives took articles off the shelf and examined them and fingered them for several minutes before making up their minds on purchasing.

3. An important variable in determining the sales of a branded product is its display position in a store and how regularly it is kept in stock. A manufacturer can keep informed on the kind of display his product is getting, the frequency with which it is out of stock, etc., by sending observers into a sample of stores.

4. Periodic retail-store inventories are used to measure sales from stores to consumers. Perhaps the best known of these services is the Nielsen Food Index.

5. In order to gain more insight into reading habits, many attempts at direct observation have been made. Studies have been made of the news-

paper-reading habits of commuters by looking over shoulders. The so-called "one-way mirror" technique permits observation of people who are reading magazines, etc. without the reader being aware that he is being observed.

6. The relative popularity of brown and black shoes for street wear has been estimated by observation of samples of pedestrians.

7. One of the most ingenious of observational devices is the Nielsen Audimeter (and other similar developments) which are attached to radio sets and record automatically the stations to which the set is tuned whenever it is on.

The examples might easily be multiplied. There is a surprising range of opportunity for gathering accurate information by observational methods, and the ingenuity of the research man should be applied in devising methods of applying them to new situations.

1. Accuracy of observation. When there is a choice between obtaining information by direct observation or by communication, observation is usually more accurate. Errors in observation are, of course, possible. Psychological experiments have frequently shown that people in unexpected or startling situations do not observe accurately. In marketing research the errors are likely to be much slighter. While it is possible to make counting errors in a store inventory, and while observation may be extremely difficult in studying reading habits, by and large the errors of observation are relatively nominal or can be made relatively nominal by reasonable care and conscientiousness.¹ Indeed, in order to assess the accuracy of information obtained by communication we shall repeatedly suggest the comparison with results achieved by the observational method, which are assumed to be correct. Even in observation there is the possibility of forgetting between the time something is observed and the time it is recorded, but this danger is comparatively unimportant.

2. Unambiguity of observation. It must be pointed out that observation *alone* often cannot provide useful information that is unambiguous. Observation of a package of Pillsbury's cake mix on a pantry shelf does not indicate when or where it was purchased or even that it was purchased, the way in which it is used, the rate of consumption, the respondent's satisfaction with the product, etc.² Similarly, merely seeing *Fortune* in a magazine rack does not indicate that it was either purchased or read.

Observation, alone, can provide information about store distribution, display, etc. For most purposes, however, observation must be supple-

¹ Inaccuracy due to the dishonesty of the fieldworkers doing the observing is not considered here but is discussed in Chapter 20.

² The observation is ambiguous in the sense that it can have a different significance to the researcher than to the house owner.

mented by communication to obtain unambiguous answers to the questions to be answered. For example, the observation of a tube of tooth paste on a bathroom shelf might be followed (or preceded) by a questioning of the respondent concerning the place and frequency of purchase, the persons using it, etc. Questions about the purchase of tires, the mileage on them, etc. might be combined with direct observation of the brand of the tires.

In many instances observational and communicative techniques can be used to supplement each other in providing information that is more accurate and less ambiguous than the information which could be obtained by either technique alone. For example, a manufacturer of flashlight batteries wished to determine his market position by comparison with several leading competitors. The brand name of one of these competitors was so firmly established that people often stated that they were using this particular battery in their flashlights even when they were not. To overcome this difficulty the interviewers were asked to examine the flashlights owned by respondents to determine what kind of batteries were actually being used. At the same time, however, the communication technique was also used to obtain other information which could not be collected by observation alone.

3. Cost. Where observation is possible without the need for special permission, this technique is likely to be extremely cheap. Store displays, for example, may be observed without special permission from store managers. On the other hand, permission to take store inventories may be difficult and costly to obtain.

When the same information can be obtained by asking a question or by actual observation, observation may sometimes be more costly. A housewife can state very quickly whether or not she has Cream of Wheat on her pantry shelf. Checking this statement by direct observation requires time both in persuading the housewife to permit such inspection and in the actual observation. On the other hand, observing whether certain canned goods were in stock in a store might be quicker than asking the store manager.

4. Ability and willingness of respondents to permit observation. Often the ability and willingness of respondents are irrelevant for the use of observational techniques. One can observe customers in stores and one can observe evidence of competitors' display or price policies regardless of the ability or willingness of the persons observed. But for pantry or store inventories it is obviously necessary to have the cooperation of the housewife or store manager. In these cases willingness to permit observation is essential. (The methods of increasing willingness are similar to those suggested later in this chapter in the discussion of communication.) Securing the cooperation of dealers or store managers, as in a store inventory, is

more difficult. Remuneration in some form is often necessary. The dealer may be promised a copy of the report of the research findings. Even more important, however, the store manager must be shown that inventory taking will not be a nuisance to him. This promise must be kept or further cooperation may be withheld. When the store is a part of a chain, the permission of the central office as well as the store manager may be necessary.

13.3 Self-observation

Before going on to the discussion of the technique of verbal communication, we now consider a technique which has some of the characteristics of observation and some of communication. In this technique, consumers (dealers, etc.) are asked to keep a record of the things they do. For example, reports on radio listening, observation of television, purchase of magazines, exposure to special premiums and other promotions, purchase and use of various products, etc. have been obtained in this way. When records of this kind are kept over a long period of time, either intermittently or continuously, it is customary to refer to the technique as the "consumer panel."¹

The use of self-observation is, however, not limited to panels. The Market Research Company of America, for example, in some studies leaves a special questionnaire for a period of one week in selected households. On each day certain questions are answered by the respondents. For example, on one day the brands of breakfast cereal consumed by the family might be recorded; on another, the purchases of cigarettes. Another example is the "self-administered" questionnaire often used for special products tests. For instance, a housewife might be asked to bake two pies, one by her ordinary baking methods and the other by the use of a new pie-crust mix. Then she might conduct an "experiment" and record the reactions of members of her family.

In another sense these techniques are devices for providing information by verbal communication. In each instance the consumer considers a set of questions and records the answers. It can be seen that the consumer considers, and answers, questions in a way that is not fundamentally differ-

¹ The term "panel" refers to the repetitious or continuous nature of the process of information gathering rather than to the fact that consumers themselves record observations on their own behavior. For example, a series of interviews with the same group of people during the course of a political campaign has been termed a "panel" study. During the war a consumer panel was maintained by the government as an aid to the rationing program, and the members of this panel were studied by frequent personal interviews. In marketing research, the term "panel" has frequently been associated with continuous studies in which consumers record their own behavior (attitudes, etc.). Other types of panels in which, for example, consumers answer periodic questionnaires are also used frequently in marketing.

ent from that in which she considers and answers questions that might have been posed by a fieldworker. The main difference is that less time (it is hoped) has elapsed between the event and its recording in the consumer panel than in the interview by the fieldworker. In fact, the difference between the observational and the communicative techniques disappears when the elapsed time between an event and its recording is so brief as to render memory errors highly improbable.

Self-observing techniques are not limited to the obtaining of information about actual behavior. For example, in the so-called "radio diary," in which respondents are asked to make a record of radio listening, questions as to likes and dislikes of particular programs, reasons why certain programs are or are not listened to, etc., may be included.

1. Accuracy and unambiguity. It might seem, from the discussion so far, that self-observation is an ideal tool for obtaining information about behavior. In principle it would be possible for consumers to record behavior very soon after it happened and so avoid the errors of memory which, as we show later, are important in the communicative techniques. Moreover, behavior which might not be observed by any outside observer, such as a field investigator, is accessible to self-observation. Problems of ambiguity arising from failure to understand questions can be reduced with the consumer panel, since it is possible to instruct the panel members with great care at the start and to point out initial errors in filled-in records until the panel member becomes quite competent.

With self-observation, however, there are problems which don't arise in such serious form with other techniques. With the consumer panel, as noted in section 11.4, the usual sampling difficulties associated with non-response are intensified. There is an inevitable "mortality" in panel membership; people drop out because of loss of interest, illness, etc. It is impossible to replace these people in such a way that the panel can still be regarded as a probability sample, even though it was initially selected by probability sampling methods. In the best panels this "mortality" is reduced to a surprisingly low level, but it is always present. Moreover, the initial rate of refusal is likely to be greater for panel studies than for other investigations which are less demanding of the participant's time. In addition to these sampling problems there are two other difficulties.

The first is a more or less frequently encountered lack of diligence and care on the part of some consumers in recording their behavior. Without such diligence and care, inaccuracy may be greater than it would have been with skilled fieldworkers who relied explicitly on the memory of respondents. Various devices and incentives can be used to increase the diligence of consumers in recording their own behavior. These devices are of two sorts.

One sort aims at facilitating the recording of behavior. Various devices have been developed for making such recording easier, reducing the physical and mental effort, and making the opportunity for confusion as small as possible. Moreover, types of behavior which are difficult to record accurately—such as purchase of bubble gum by younger members of the family—are usually eliminated from the list.

The second sort of device aims not at facilitating the mechanical task of recording but at creating an attitude among the panel members conducive to the prompt, regular, and careful recording of the required information. Although a friendly feeling toward the administrators of the panel is of some importance in creating this attitude, it has usually been found necessary to supplement the effects of any possible feeling of brotherhood and friendship with some form of material remuneration. This is understandable, because the members of the panel are required to perform a repetitive chore which in spite of any mechanical ingenuity must soon, if not immediately, cease being pleasurable and which will inevitably consume in the aggregate a large amount of time if the family continues reporting information.

The form of payments typically used is that of merchandise premiums. That is, instead of being rewarded with money, panel members are given the opportunity to earn various types of merchandise. Some type of incentive system is ordinarily used. For prompt reporting, reporting special information, etc., panel members are rewarded with additional credit points beyond those which they could earn for a minimum performance of their duties.

Possibly the greatest hope of increasing accuracy through manipulation of rates and methods of remuneration lies in establishing a rate high enough to make panel members willing to submit to supervision and tests of accuracy and to care about the results of such tests. In other words, the fluctuations in remuneration—adjusted to the results of tests of accuracy—must be great enough to affect the behavior of the respondents in their participation in the panel. If the members of the panel were supervised and controlled to the same extent that fieldworkers are (or could be), it seems highly probable that the panel would be more effective than other techniques for getting information about behavior. The greater effectiveness would result, as has been said, from the greater range of behavior open to direct observation and the diminished reliance on memory with its attendant errors.

The second difficulty in the consumer panel which has received considerable attention is that the act of recording attitudes or behavior (and being rewarded for it by merchandise premiums) influences and modifies whatever is being recorded. Up to now we have been discussing prob-

lems centering around willingness to give information. The question now is one of ability to give information; for if the process of giving information changes the information being given, the ability to give the information *actually desired* has been diminished. This point has been much discussed by marketing research personnel, but so far as we know it has not been satisfactorily analyzed by objective research. Panel studies in other fields of investigation, such as political behavior, have been conducted in such a way that the "conditioning" effect of continuous study can be isolated and, if necessary, allowed for. In the absence of any objective information on marketing panels, we would hazard the guess that conditioning is not likely to be serious with panels which simply record purchases, so long as (1) the total amount of compensation is kept within reasonable bounds and (2) the types of merchandise premiums offered are not close substitutes for the products whose purchase is to be measured by the panel. For other types of panels which obtain information on attitudes, preferences, consumer knowledge, etc., the danger of conditioning is likely to be more serious.

For the other techniques of self-observation which do not rely on repeated observations, many of the disadvantages of the panel approach are not encountered. The sampling difficulties, for example, are less serious, and the act of recording is much less likely to influence the behavior reported, although even with radio diaries kept for short periods there is some danger of this. Lack of diligence in filling in accurate information is probably less serious, since the activity may have interest because of its novelty. However, the relative freedom from disadvantages does not necessarily mean that such techniques of self-observation are superior. The panel technique, by the very nature of continued observation, obtains information on long-term behavior that can be obtained in no other way. For example, data on patterns of brand loyalty and brand switching could probably be obtained in no other way than by the panel technique, for memory is not likely to be satisfactory for giving detailed information on behavior over a long period.

13.4 The Communicative Techniques

Communicative techniques differ from observational ones because the former seek to elicit information under circumstances in which memory errors may be important and which require verbal or written communication from respondent to researcher. As has been said, for those problems amenable to treatment by both techniques, the accuracy and unambiguity of the observational technique are usually greater. They are greater because observational techniques are relatively free from errors of memory. Many types of behavior, however, can neither be observed by fieldworkers nor inferred from observation of some evidence of the behavior. For these

types of information, and for those in which observation, while more accurate and less ambiguous, is much more costly, complete or partial reliance on communication is necessary. Our discussion is organized about the two main problems in obtaining information about behavior by communicative techniques: (1) memory errors (inaccuracy) and (2) imperfections in the process of communication (ambiguity).

1. Memory. "If anyone should be inclined to overrate the state of our present knowledge of mental life, all that would be needed to force him to assume a modest attitude would be to remind him of the function of memory. No psychologic theory has yet been able to account for the connection between the fundamental phenomena of remembering and forgetting: indeed, even the complete analysis of that which one can actually observe has as yet scarcely been grasped."¹ Freud wrote this several decades ago, but it is appropriate and only slightly less accurate today. No thorough understanding of the processes of learning and retention has been achieved in spite of the voluminous work of psychologists. Some modest hypotheses of very limited generality have been formulated, and a little fruitful empirical work has been done.

Unfortunately for marketing research, most of the work in psychology has not progressed beyond highly simplified laboratory experiments whose meaning for marketing research is not clear in some cases and absent in others. Further, most of this experimental work has been focused on the process of learning—for which memory is obviously essential—and is therefore of greater relevance for the problems of conveying information by advertising and personal selling than for the problems arising in obtaining information in marketing research by methods which depend on memory. This information has been of relatively limited use in indicating either how field interviews can evoke more accurate responses concerning behavior or in pointing out the sorts of behavior about which accurate information can be secured. We shall, however, try to indicate the usefulness for marketing research of what has been done.

Bergson, in *Matter and Memory*, distinguishes two kinds of memory—habit and independent recollection. The learning of a poem so that it can be recited without reference to a written text is an example of habit memory. The conscious awareness that the poem was read and reread is independent recollection.² Marketing research requires memories of both types. When information is required regarding consumers' familiarity with advertising themes or regarding shopping habits, habit memory can

¹ "Psychopathology of Everyday Life," in *The Basic Writings of Sigmund Freud*, edited by A. A. Brill, p. 95, Modern Library series, Random House, New York, 1938.

² Bertrand Russell discusses the subtleties of this distinction in *The Analysis of Mind*, pp. 166-167, The Macmillan Company, New York, 1921.

help. When the researcher wants to know whether a particular advertisement was seen or what radio program was listened to on a particular evening, the respondent usually must rely on independent recollection.

These memories can be considered to be affected by the following variables: (1) the event to be recalled; (2) the person attempting the recollection; (3) the length of time between the event and the attempted recollection; (4) the experience of the person during this period; and (5) the stimulus to recollection received by the person doing the recalling. Although all these variables are important and must be considered in evaluating the possible value of any research project relying on the memory of respondents, only two, the length of the period between the event and its recollection and the stimulus to recollection, can practically be controlled by the researcher. In addition, the type of event being recalled is important because the researcher must have some appreciation of the kinds of information for which memory does not provide satisfactory results. Hence, we consider first the two variables which the researcher can influence and then the types of problems for which the securing of information by memory seems to be appropriate.

1. *Seeking Information about Recent Events.* Probably the easiest and most obvious way to reduce memory errors is to seek information about events as soon after their occurrence as possible. This principle has been widely recognized and applied in commercial research. For example, the measurement of the rate of consumption of branded articles is often made by repeated surveys which are designed to secure information for the brief periods immediately prior to each survey rather than by a single survey seeking information about patterns of behavior extending farther into the past. If a survey were being conducted to find out the stores in which people purchase clothes, the application of this principle would indicate that respondents should be questioned about their most recent purchase rather than about more remote purchases. If the "recall" method were to be used to determine the audience of a radio program, the period over which past behavior would be studied probably would be a few hours or at most a day, and even then many people would be unable to remember their listening behavior accurately, except when habit memory was involved.

The meaning of this principle needs no further elaboration or illustration. It might be well, however, to comment on the work that has been done in attempting to quantify the relationship between the length of time between an event and the accuracy of its attempted recollection. The pioneering experimental work in this field was done by Hermann Ebbinghaus¹ in the 1880's, and almost all subsequent work has served merely to

¹ Hermann Ebbinghaus, *Memory, A Contribution to Experimental Psychology*, trans.

confirm his findings, to refine them, and to give them greater generality. Ebbinghaus found that the rate of forgetting¹ is relatively rapid in the few hours immediately after the event and that the rate diminishes sharply and steadily thereafter.

Subsequent work has, however, established that the longer the elapsed time between an experience and its recollection, the greater the loss in accuracy and completeness. And it has been established that for independent recollection the most rapid diminution in accuracy occurs in the first hours after the event. But this helps relatively little in practical marketing research applications. It is not enough to know that memory becomes less satisfactory with the passage of time; it is necessary to know, in addition, *how much* less satisfactory. For example, in an investigation of purchases of motor oil, should one ask about behavior in the past week, the past two weeks, the past month, the past year, or what? Unfortunately, only at the extremes of time—*immediately* after and *long* after the event—are any conclusions possible, and these are of little practical importance.

With regard to the period immediately after an event the following quotation is relevant:²

Between memory-image and sensation there is an intermediate experience concerning the immediate past. For example, a sound that we have just heard is present to us in a way which differs both from the sensation while we are hearing the sound and from the memory-image of something heard days or weeks ago. . . . Everyone knows the experience of noticing (say) that the clock *has been* striking, when we did not notice it while it was striking. And when we hear a remark spoken we are conscious of the earlier words while the later ones are being uttered, and this retention feels different from recollection of something definitely passed. A sensation fades gradually, passing by continuous gradation to the status of image.

This period of transition is very brief. For things like striking clocks and other ordinary sounds or movements this "immediate memory" probably does not last over a few seconds or minutes. So the absence of memory errors in such a period is of little practical importance for marketing research, except in such specialized techniques as "coincidental" radio measurements or interviewing consumers immediately upon leaving a grocery store to find out about purchases of groceries.

At the other extreme there is evidence and intuition to suggest that the ability of persons to recall events accurately is usually not importantly different in the fourth year following the event than in the third or fifth year, and for most events, that the loss of memory probably proceeds at

by Henry A. Ruger and Clara E. Bussenius, Teachers College, Columbia University, New York, 1913.

¹ "Oblivescence" is the technical word, or jargon.

² Russell, *op. cit.*, p. 174. Used with the permission of The Macmillan Company.

an insignificantly slow pace after a few days, weeks, or months. But such generalizations as this are misleadingly plausible. Because forgetting proceeds at a slow *rate* in, say, the fourth month after the event doesn't mean that the total "amount" of forgetting in the fourth month is not important.

In the great and important period between "immediate memory" and the time of an imperceptible rate of loss in the ability to remember, there are no precise guides at all for the marketing researcher. Introspection and conversation will probably indicate that the rate at which one forgets the substance of Thursday's breakfast is different from the rate at which one forgets the events surrounding a marriage or a broken leg. A man can probably recall his Thursday's breakfast menu with great accuracy on Thursday morning, but errors in recollection will probably begin to be significant by supper time and can be substantial by Friday morning. The recollection is probably little less accurate on the following Thursday than on the following Wednesday, however. A wedding or a broken leg, on the other hand, can frequently be recalled with distressing accuracy for considerably longer periods. Marketing research would benefit greatly from a methodical and careful measurement of the rates of forgetting of the sort of events important in marketing research.

What we have said so far suggests that the marketing researcher, confronted with a practical decision as to how far he should probe into the past on a particular study, would simply have to rely on introspection and the vague guides mentioned above. It is true that no neat solutions are available. On the other hand the problem is not nearly so imposing as the tone of the past paragraphs may have indicated. Fortunately it is not difficult to tell when people can't recall very well. For one thing, people will often state they don't remember, or that the answer they are about to give is "just a guess." Hence in the pretesting stage it is not hard, as a practical matter, to make defensible decisions about the distance into the past which is to be explored. ("Pretesting" is preliminary small-scale interviewing carried on to find out defects in the questioning procedure so that these defects can be corrected before the main study.)

In the pretesting, skillful interviewing and a good questionnaire are needed to find the maximum period in the past for which people can give accurate answers. A poor interviewer, or a good interviewer working with a poor questionnaire, might report that hardly any one was able to recall accurately the thing being studied. It is helpful for the research man himself to participate in such pretesting to supplement the reports he gets from his regular pretesters and any preconceived ideas that he may have had.

2. *Stimuli to Memory.* The accuracy of information can be controlled

not only by controlling the elapsed time between event and recollection but also by the way in which the stimulus to recollection is devised and presented. The wording of the question and the structure of the questionnaire as well as the time of their presentation affect the accuracy of respondents' memories. Nor are these the only important stimuli to memory. The interviewer and the reaction he creates are part of that stimulus, for the attitude of the respondent to the interview affects his ability to remember. For example, it has been found that in laboratory experiments the ability to recall is significantly affected by the degree of relaxation of the respondent—the greater the degree of relaxation, the greater the ability to recall.¹ While work on this relationship has been slight, it appears from the most casual observation that the ability of the interviewer to put respondents at ease is one of the most essential skills in conducting interviews. The only really debatable question is how to go about putting respondents at ease, and how to train inexperienced interviewers to do it. Suggestions are given in Chapter 20.

The scientific work on questions and questionnaires has been of greater value even though it, too, provides little detailed guidance in the formulation of questions and questionnaires. Two closely associated main guiding principles have emerged, and they probably would have occurred to the untrained armchair "psychologist" in the absence of any empirical work. The principles are (1) the questions should suggest to the respondent phenomena associated in the mind of the respondent with the event about which information is sought, and (2) the questionnaire should, in so far as possible, not suggest to the respondent phenomena not associated in his mind with the event about which information is sought.

These ideas are certainly not new. The importance of "association" for memory has long been recognized and was well stated by William James at the end of the last century:²

In mental terms, the *more other facts a fact is associated with in [sic?] the mind, the better possession of it our memory retains*. Each of its associates becomes a hook to which it hangs, a means to fish it up by when sunk beneath the surface. Together, they form a network of attachments by which it is woven into the entire tissue of our thought.

This whole network of attachments can often be recalled if a person only can be made to think of one element of the network.³ Getting the respond-

¹ Gerald R. Pascal, "The Effect of Relaxation upon Recall," *American Journal of Psychology*, Vol. 62, No. 1, pp. 32-47, January, 1949.

² William James, *Psychology*, p. 294, Henry Holt and Co., Inc., New York, 1915.

³ "Redintegration" is the psychological jargon used to refer to the process by which a network of memories is created by a fraction of the original stimuli.

ent to recall this element is the task of the researcher. We shall now consider several approaches.

The recognition method. An obvious, useful, and frequently used technique for the performance of this task is the presentation to a respondent of a facsimile of possible essential elements in the network. If, for example, information was required about which brand of cake mix was currently being used, the relevant memory might be evoked by presenting to the respondent packages (or facsimiles) of the various brands of mix that are currently sold on the market. This is an example of stimulating the respondent to the relevant recollection by presenting an element of the associated network. In this case the association is close and direct. The association is so close and direct that the process by which the stimulus evokes the memory is called by a special term, "recognition."

When the association is less close and direct—when, for example, the interviewer merely succeeds in getting the respondent to think of the most recent occasion on which a cake was baked—the process by which a memory is evoked is termed "recall." The process of recognition has been demonstrated to yield in most cases more accurate information about the past than have the less intimate associations underlying the process of recall. Whenever possible, then, the respondent should be given the opportunity to recognize rather than recall.

Some dangers attend the use of techniques relying on recognition. In the first place there is evidence of susceptibility to a so-called "positional" or "order-of-presentation" bias. If a series of objects (packages of different brands of cake mix, for example) are presented to a respondent in the hope that he will recognize the one which was actually purchased or used, the respondent may tend to choose a package which was presented first even though it is not the correct one. The possibility of such bias is widely recognized, and its effect can be nullified by a simple procedure. The net effect of bias of individuals can be offset for the sample as a whole by rotating the order of presentation of the packages so that each package appears first, last, and in each intermediate position as often as any other.

It is desirable also that the *order* of presentation be varied. For example, suppose that the readership of six advertisements, 1, 2, 3, 4, 5, and 6, was to be measured. The following six combinations would meet the first requirement that each advertisement occupy each position with equal frequency:

1, 2, 3, 4, 5, 6	3, 4, 5, 6, 1, 2	5, 6, 1, 2, 3, 4
2, 3, 4, 5, 6, 1	4, 5, 6, 1, 2, 3	6, 1, 2, 3, 4, 5

Even though this first requirement is met, positional bias is possible. If there were any advantage in an advertisement's being presented to respond-

ents immediately after, say, advertisement 1, then advertisement 2 would benefit from this advantage in five of the six combinations. Even though advertisement 2 is in each of the six positions an equal number of times, it has an upward bias because it follows advertisement 1 about 83 per cent of the time.

This difficulty can easily be overcome by a simple extension of the ideas of random sampling. Random drawings are made to fill each position. There are two restrictions on the random selection: (1) once an advertisement has been selected, it cannot be drawn again for the particular combination of six; (2) if an advertisement is drawn for the same position as it occupies in a combination already selected, this drawing is completely ignored.

A second possible distortion is the danger of a so-called "affirmative bias." Not only are the relative rates of purchase or use sometimes distorted because of the order in which the stimuli are presented, but also there is often distortion in the absolute rate reported for all stimuli. This latter affirmative bias occurs because people generally have indicated a tendency to say "yes" rather than "no" in response to questions, regardless of the question or its correct answer. When presented with a list of advertisements, for example, people may consciously or unconsciously tend to overstate their readership of all advertisements in the list. In so far as this is due to a desire to impress or please the interviewer, it may be combated by techniques suggested later in this chapter.

Whatever the cause of the bias, the errors introduced—if any—can sometimes be measured. In the cake-mix example, it would be possible to take an inventory of the pantry shelf to see whether the statements of respondents conformed to the direct observation of the fieldworker. Once the degree of distortion of the verbal statements had been established, it could be used to correct information received subsequently through verbal communication alone. This process has its danger, however. Few research projects resemble each other to such an extent that information about biases from earlier studies can be used to adjust verbal responses on a subsequent one. This procedure, therefore, is subject to the same kind of risks incurred by political pollers who assumed that because techniques worked in the past they would do so in the future.

The better procedure is to rely on observation completely, unless (1) observation is too expensive, or (2) other information not susceptible to observation must also be obtained. If observation is expensive, a simple application of the technique of double sampling (see Double Sampling, under section 9.1) can be used. That is, for a subsample of the sample of respondents, both techniques, recognition and observation, are used. Using the relation between the results of recognition and observation in the sub-

sample, adjust the findings obtained for the sample as a whole by the recognition method.

A simple illustration of the double sampling technique may clarify this procedure. Suppose 1000 people are approached to find out by the method of recognition what brand of coffee they are currently using. Two hundred people, or 20 per cent, state that they are using Chase and Sanborn. A subsample of 300 people is selected at random from the original 1000; and by means of a pantry inventory, the brand actually used by these 300 people is observed. Suppose the following results are obtained for these 300 people:

1. 90 people are actually using Chase and Sanborn (as determined by actual observation).

2. 45 people had stated (in the recognition test) that they were using Chase and Sanborn.

From these subsample results, the best estimate is to double ($90/45 = 2$) the reported use of Chase and Sanborn for the sample as a whole, as obtained by the recognition method. Hence 20 per cent is multiplied by 2 in order to obtain 40 per cent as the "adjusted" estimate of the incidence of use of Chase and Sanborn coffee.

Another potential bias in the recognition technique is the possible tendency to designate a particular object as having been used (purchased, read, etc.) because of its general familiarity rather than because it was actually used (purchased, read, etc.). For example, in the study of flashlight batteries mentioned earlier, it was found that even the presentation of facsimilies of different brands of batteries failed to overcome completely the tendency of people to name the best-known brand instead of the brand that they were actually using.

This problem is even more serious in surveys intended to estimate readership of particular advertisements rather than familiarity with advertising themes, brand names, or trade-marks. Many who have done such studies have observed the tendency for successive advertisements in a campaign made up of advertisements with similar themes, layouts, etc., to obtain successively higher reported readership. Indeed, experiments have been made which showed advertisements that had never been published to have obtained high "readership" figures.¹

Aided recall. So far the discussion of the ways in which researchers can evoke accurate memories has not proceeded beyond the process of recognition. It has been seen that recognition involves the use of the most

¹ For a method of measuring this bias, see D. B. Lucas, "A Rigid Technique for Measuring the Impression Values of Specific Magazine Advertisements," *Journal of Applied Psychology*, Vol. 24, pp. 778-790.

direct and intimate association—a facsimile of the original stimulus (or the original stimulus itself)—in order to evoke the relevant memory. It is also possible to use a verbal symbol, instead of a facsimile, to represent the original stimulus. It is, for example, possible to ask a person whether she is currently using Gold Medal, Pillsbury, Robin Hood, etc., brand of all-purpose flour. In this way, verbal symbols—brand names—are substituted for facsimiles—different flour sacks.

The use of verbal symbols in this way is termed “aided recall” and is subject to the same biases and limitations as the process of recognition. It has some peculiar disadvantages and advantages, however. Since the association between the verbal symbol and the original event is less direct—and in a sense, less vivid—than that between the facsimile and the original event, it would seem that aided recall would be less satisfactory. The facsimile has this greater power because it creates directly associations of form, package design, color, etc., in addition to providing the simple presentation of the brand name. A second advantage of the facsimile is the avoidance of a need for detailed verbal description. Magazine stories or articles, for example, are usually more easily identified by presenting a copy of the magazine than by an attempt to give a brief and meaningful verbal description. This brings to mind the familiar advertising proverb, “One picture is worth a thousand words.”

The main disadvantage in using the facsimile is its frequent impracticability. It may not always be practicable for fieldworkers—especially those who must depend on public transportation—to haul around trunks full of clothes, crates full of tires, reels of moving pictures with projecting equipment, or record players and records, in order to enable respondents to recognize the brands and styles of clothing purchased, tires used, moving pictures seen, or radio commercials heard.¹ Even in problems that seem relatively simple on casual inspection, the practical obstacles may be serious. Suppose the objective were to find out the magazines that people read. If the recognition method were used and the interviewer given one copy of each magazine whose readership was being investigated, there would be a difficult problem of transporting, say, 40 magazines. Hence compromises, such as a reproduction of the cover page from each of the

¹ The word “impracticable” and the phrase “too costly” are roughly interchangeable. The hauling of clothes, tires, moving pictures, etc., is termed “impracticable” because (1) even if willing to try, a fieldworker unequipped with trucks and strong men could conduct very few interviews per day, which would make the securing of information from the designated sample of respondents too costly, or (2) the equipping of a fieldworker with trucks and strong men in order to make possible more interviews per day would make too costly the securing of the required information from the designated sample.

magazines, might be employed. But reproductions of the cover are less satisfactory as aids to memory than the original magazine.

There is another drawback applying both to recognition and to aided recall, which comes up most frequently in attempts to measure radio audiences by the use of aided recall. Suppose a respondent is presented with a list of programs and stations and asked which of these stations and programs he listened to the preceding day. This procedure will cause no trouble so long as interviewing is done in an area—say, a large city—in which a relatively small number of stations can be heard. But if the sampling includes a larger area in which many stations can be heard, the list of stations which the respondent *might* have heard would become too long and unwieldy. On the other hand, the shortening of the list in order to make it more manageable might suppress valuable information.

A similar situation might arise in studying a product made by a large number of small manufacturers. It would not do to list only the eight or ten largest manufacturers, since the information on the others would be subject to serious understatement. On the other hand, a complete listing might be unwieldy. A partial solution is to use an "all other" category.

Unaided recall. Many researchers feel that the use of techniques based upon either recognition or aided recall involves disadvantages that more than offset the advantages of supplying stimuli to memory. These researchers are willing to abandon the use of the relatively direct associations which these techniques employ (and their possible biases) and rely upon the more difficult techniques of unaided recall. "Unaided recall" is misnamed. The researcher frequently does make serious efforts to aid the respondent in her attempted accurate recollection of past behavior. The term "unaided recall" is used because the research man foregoes the use of facsimiles or verbal symbols which directly represent the thing he is trying to find out about.

If information is sought about the brand of tooth paste which has been purchased, the user of unaided recall foregoes the use of either packages of tooth paste, facsimiles thereof, or lists of brand names in order to secure accurate recollection. He relies instead either upon the simple question such as "What brand of tooth paste did you last purchase?" or upon other forms of questions which avoid the direct suggestion of the brand name. Sometimes these other forms of questions attempt to suggest elements other than the brand name in the "network attachments" of which the brand name is a part.

The use of the simple question, "What did you buy?" "What did you read?" "What did you use?" etc., is free of both affirmative and positional bias, as discussed in our treatment of recognition. It has sometimes been

found, however, that questions of this type sometimes distort the true rates of purchase, readership, or use. There may be a "negative bias," analogous to the "affirmative bias" suggested earlier, caused by forgetting. There is danger of interviewer bias; for, in attempting to explain the question to a respondent who doesn't understand it, the interviewer may inadvertently give a "for example." The respondent might not know what a detergent is, so the interviewer might say, "You know—like Dreft." Further distortion may arise because some items are relatively well known even though they are not proportionately so frequently bought, read, or used. The flashlight-battery example is one of the best illustrations.

No way has been devised for eliminating this last bias in individual responses or for offsetting it in the sample as a whole. The greatest hope for securing accurate information lies in measuring the bias and allowing for it. The measurement of the bias must involve the establishment of the relationship between stated and actual behavior, and in most cases this means that observation must be used as a check against respondents' statements. Again a double-sampling technique, in which both questioning and observation are combined for a subsample of the over-all sample, would seem appropriate. It is also possible, of course, to assume that the bias discovered on one study will be encountered on subsequent studies, but any such assumption is subject to the serious risks we have already noted.

Sometimes it is possible to do more than merely ask the simple sort of question referred to in the preceding paragraphs. It may be possible to get the respondent thinking of links in the relevant "network of attachments" and thereby to succeed in recalling the essential link in the network. One device for doing this is to get the respondent thinking of the time of the action about which the interviewer seeks information. This device is widely used and is commonly referred to in textbooks on marketing research and elsewhere by phrases such as "working along the time line," "chronological suggestion," etc. In order to understand this device or to use it intelligently, it is essential to have some insight into the process of thinking about previous periods of time. William James offers some interesting comments on this point:¹

... and how can we think of the past? ... We have seen that our intuitive or immediate consciousness of pastness hardly carries us more than a few seconds backward of the present instant of time. Remoter dates are conceived, not perceived; known symbolically by names, such as 'last week,' '1850'; or thought of by events which happened in them, as the year in which we attended such a school, or met with such a loss. So that if we wish to think of a particular past epoch, we

¹ James, *op. cit.*, p. 288.

must think of a name or other symbol, or else of certain concrete events, associated therewithal. Both must be thought of, to think of the past epoch adequately. And to 'refer' any special fact to the past epoch is to think that fact *with* the names and events which characterize its date, to think it, in short, with a lot of contiguous associates.

Proceeding on the basis of James's description, it is possible to see that an interviewer may start a fruitful series of related thoughts if he can get the respondent to think about the "contiguous associates" of the event which is of interest in the research project. It is not always possible to know which events are contiguous to "the" event, but it is usually safe to assume that events which are contiguous in time are also contiguous—closely associated—in the minds of respondents. If, for example, information were required about the radio programs to which respondents listened the preceding evening, the questioning might cover everything done by the respondent from, say, 5 P.M. on. The questioning might, therefore, begin with dinner, leaving the office, or some other event which would probably be well remembered because it is either habitual or important. Just as people frequently try to recover misplaced keys, fountain pens, galoshes, by harking back to the last time they remember seeing the missing article and proceeding step by step from that period to the present in the hope that the present location will be recalled, so the researcher can try to get the respondent to proceed step by step from the well-remembered event (*e.g.*, having dinner) through the period about which the information is needed. Such a procedure can often produce more accurate information than the cross-examiner's question, "What did you do on the night of the 23d of January?"

On other occasions the researcher may not be able to start from a well-remembered event and proceed through the period of the event which is being investigated. There is another method of creating the desired associations. An extreme example of this method is the "free association" method employed in psychoanalytic treatment or in "nondirective" counseling. In this technique the patient or client is encouraged to express verbally whatever he is thinking about. The verbalization of the patient's thinking indicates to the analyst the network of association in the patient's mind. The individual sentences of the monologue reveal much information about the patient's past, and the associations between the thoughts indicate much of the implicit significance of the sentences for the patient.

Now it is clear that this technique of free association could seldom if ever be applicable to market research.¹ The respondent, unguided, prob-

¹ However, similar tools, such as projective tests, have been used in attitude studies in marketing research problems. See Inaccuracy Attributable to Unwillingness, under section 17.3.

ably would ramble for months before his comments would be relevant for the interviewer. Under such circumstances the cost of the information would almost inevitably exceed its value.

Under some circumstances, however, it is possible to guide the respondent into the general area of interest to the researcher. If, for instance, the researcher were interested in the places where the respondent purchased suits, the amount expended, the brands selected, etc., the respondent could be started in the right direction if the interviewer were to open a discussion about clothes and insert at the appropriate points in the conversation the specific questions of interest. This technique has been given the name "interview in depth," although this name has unfortunately also been applied at one time or another to almost any kind of relatively lengthy interview. Clearly the successful application of such a technique depends upon the skill and experience of the interviewer and involves very difficult problems of field supervision and control as well as of interpretation. The interviewer, for example, must be a catalyst; she must influence the direction of the conversation without injecting her own ideas or opinions too strongly. This may be extremely hard when the purpose is to obtain information on attitudes instead of factual information, although it is hard enough for factual information. The replies must be recorded as completely as possible, and the classification of verbal responses—coding—must be done very carefully. Ideally, however, interviewing in depth can be reserved for those particular problems in which the expense is warranted and in which the other techniques seem unrewarding.

2. Types of events and the ability to remember. It is important to know not only how to increase the accuracy of recollections but also how to decide in advance the sort of subject about which it is reasonable to hope to secure adequately accurate recollections. Probably no researcher would be foolish enough to try to get respondents to provide accurate information about their luncheon menus for the preceding month. Probably all researchers would recognize the good prospects for obtaining accurate information about the number of radios owned by respondents. But there are many other things which might not be so obviously accessible or inaccessible to successful investigation. Of course, no one can provide formulas or mechanical guides in this area of doubt. There are, however, some principles which, if not taken too literally, may be useful.

1. Events belonging to classes of events which are often repeated and nearly identical (except for the time of occurrence) are less quickly forgotten than events not having these characteristics: a person who almost always eats the same things for breakfast is better able to recall the previous day's breakfast menu than a person who frequently varies his breakfast. The former person is better able to recall the previous day's break-

fast if it was unusual *because it represents a deviation from established habit*. The absence of any deviation—that is, the continuation of a routine—can also be recalled, as can any repetitive act that becomes established as a habit memory. Similarly, people who habitually read certain magazines, listen to certain radio programs, shop at the same store, etc., can recall their behavior in these respects better than persons whose habits are more irregular. The person who makes strenuous efforts to arrange his schedule so as to avoid missing the Jack Benny or the Big Sister program will be able to recall the occasions when the program *had* to be missed.

2. Important events are better remembered than less important ones. This is obvious; the difficult thing is to define “important” so that the statement can be usefully applied to marketing research. In the first place, an event’s importance can be considered to be directly related to the magnitude of its effect on the size or allocation of one’s income. In general, the larger the change in income or the larger the expenditure, the more important is the event. In this sense the purchase of an automobile or a house is more important than the purchase of a pair of socks or a flashlight battery. In the second place, an event’s importance can be considered to be directly related to the magnitude of its effect on the allocation of one’s time. In this sense, reading a short story in *Collier’s* is less important than reading *Gone with the Wind*, even though the effects of either of these actions on the allocation of one’s income would be inconsequential. The reading of *Gone with the Wind* would probably be better recalled than the reading of the story in *Collier’s*. In the third place, the importance of any event is related to its emotional impact. Events which create strong emotional responses are better remembered than the more numerous day-to-day happenings. Events which anger, frighten, embarrass, move to pity, etc., are relatively well remembered. One’s first arrest will probably be better remembered than the place where cigarettes were purchased on the same day. A wife may remember well a small present from her husband in recognition of an anniversary.

3. Events which deviate strongly from the expected are longer remembered than the commonplace. A person would better remember seeing a green Rolls-Royce drawn by six white horses than a gray Chevrolet. Similarly, respondents could be expected to remember longer a radio program such as Orson Welles’s staging of the invasion from Mars¹ than an hour of song with Kate Smith or Frank Sinatra.

Since these three principles are general, their usefulness depends on the

¹ Before World War II, Orson Welles, directing a commercial radio show, announced the landing and progress of invaders from Mars. The program was so convincing that many people were seriously alarmed, sought weapons to repel the threat, etc.

judgment of the planners of marketing research projects. As we mentioned earlier in this chapter, general rules of guidance can be supplemented by actual investigation, preferably in the preliminary or pretest stage of the research investigation. When it is possible to check on accuracy by observation, in the way already suggested, there is little excuse for conducting a marketing research inquiry into some subject about which people cannot supply adequate information.

3. Selecting the person best able to provide the information. We have discussed ways of increasing people's ability to supply information and suggested criteria for deciding the kinds of information that can be obtained by questioning people on their behavior. One further issue requires brief mention: who can best provide the information we want? This question has nothing to do with the problem of sampling. The family about which we want to obtain information is selected by sampling techniques, but the decision as to the person in that family who can best provide information about the *whole family* is a problem that should be discussed here.

Often in market research the answer is extremely simple. It is necessary only to interview the "housewife" or "female head of family." Many of the consumer products about which information is desired are purchased almost exclusively by the housewife, and if they are not, she is aware of their purchase by other members of the family. Even for products like these, some practical problems will come up: for example, some families do not have a female head and others seem to have more than one. These practical problems are sometimes perplexing but are infrequent enough to present no major theoretical problem. There will ordinarily be little trouble in deciding whom to talk to within the family even if there isn't a single person who unambiguously fits the description of "female head of family."

But some products may be purchased by more than one member of the family, and it is quite possible that the husband, for example, may not know of his wife's purchases and vice versa. Cigarettes might be such a product. In instances like these, one satisfactory solution is to try to talk to all the members of the family. Moreover, if a complete picture of the interrelation of purchase patterns within the family is needed, the only method is to talk to all members of the family who buy the product rather than to use subsampling. To do this the interviewer may have to call back at least once and possibly oftener.

Sometimes the family unit may be inappropriate and it may be desirable to talk to individuals as such rather than to people as members of families. One way of doing this is to draw a sample of families and decide at random the member or members of the family about whom the information is to be

obtained.¹ For example, in many problems it will not be known in advance which member of the family is likely to be important in actual purchases, nor can the interviewer expect to know the extent to which members of a single family are likely to be informed on each other's purchasing behavior. This is often true when a particular product is being studied by marketing research for the first time. Everyone connected with the study may have an idea of who does the buying, but these ideas are usually based on limited experience and may be badly in error. Moreover, one of the important objectives of any such study is to find out who actually does buy the product.

For example, it is now well known, partly because of marketing research studies, that some deodorants are used by men as well as women. If this fact were unknown and interviewers were instructed to talk only to the female members of families, a very important fact might not be discovered.

The problem of whom to talk to is much more serious in dealer studies and in investigations of wholesalers or of problems in industrial marketing than it is likely to be in consumer studies. Everyone knows from experience that the first person encountered in a business organization may know little or nothing about the subject of interest. Yet it is often perplexing to find the proper person, for problems connected with *willingness* to supply information are extremely important and difficult to solve in this kind of work. For example, the store clerk or even the store manager in a chain store may be unable to give the desired information, and the job of finding the right person at the central office of the chain system may be difficult. All we attempt to do here is to say that the problem exists. Methods of solution are similar to those used by salesmen who have to find the right man to talk to in selling their product, and it seems inappropriate to talk about such methods here.

4. Ambiguity in communication. It has been assumed so far that the respondent has understood the meaning of questions put by interviewers and that interviewers (coders, analysts) have understood the responses to these questions. Unfortunately, such assumptions are often not justified. It is frequently difficult for both researchers and respondents to achieve the clarity of expression needed for mutual understanding (unambiguous communication). Such understanding, at least in reasonable degree, is essential if the results of communicative techniques are to be useful for marketing decisions.

The difficulties of communication are much less serious in projects requiring information about behavior than in projects investigating inten-

¹ A practical method of doing this is suggested by Leslie Kish, "A Procedure for Objective Respondent Selection within the Household," *Journal of the American Statistical Association*, Vol. 44, No. 247, pp. 380-387, September, 1949.

tions, attitudes, motivation, etc. It is much less difficult to communicate meaningfully concerning purchasing soap or hearing a Jack Benny program than about an attitude toward the Standard Oil Company of New Jersey or a reason for choosing a Westinghouse refrigerator. Since the difficulties involved in the simpler tasks of communication are also present in the more complex, it is well to begin in this chapter a discussion of some of the basic difficulties in communication, whatever its purpose. We therefore begin with some of the difficulties of conveying a question, or information necessary to explain a question, to a respondent.

Meaningful communication requires that the words used in a question have the same meaning for respondents as for the researcher. The use of words *unfamiliar* to respondents clearly violates this requirement. Until recently, at least, the question, "Do you use a detergent for your dishes?" would have been a good example of the use of an unfamiliar word. The unfortunate thing about using such words is that many respondents, unfamiliar with or uncertain about the language, have been found to be reluctant to seek help because of a fear of revealing their ignorance to the interviewers. Such a reluctance makes it especially important that the investigator take all precautions to avoid unfamiliar words.

This principle is simple enough, yet in spite of its self-evident nature, there has been a tendency for some researchers to be naively optimistic about the level of literacy of respondents. Questionnaires using such words as "fiction," "drama," "detergent," "whole-grain cereals," "dentifrice," "women's service magazine," "canned luncheon meat," are shaky tools of communication. One excellent example of trade jargon which tends to go counter to popular usage was the interviewer who asked a woman about her home freezer and didn't realize until half-way through the interview that the respondent was talking about her refrigerator. Not until the term "deep freeze" was introduced did the woman understand the question being asked.

The principle of simplicity can hardly be pushed too far. "Buy" should be used instead of "purchase"; "look at," in place of "observe"; "cooking," instead of "meal preparation"; etc. In cases of doubt the pretesting of questionnaires provides, under certain circumstances, an opportunity for checking on the familiarity of respondents with the words of the questionnaire. The pretesting is useful for this purpose if the following conditions hold:

1. The respondents in the pretest include members of the lowest levels of literacy expected to be encountered in the final survey. A pretest using the office force of the agency conducting the survey would, therefore, be inadequate because the literacy of such a group is presumably quite high. (Such a group, of course, might be typical in other respects.)

2. The interviewers in the pretest should employ some means to *check* on the respondents' familiarity with the language. The reluctance of people to reveal their ignorance exists in pretests as well as in final surveys. Sometimes it is possible to ask respondents directly the meaning of the questionable word. Another useful scheme is to ask the respondent to repeat the question in her own words.

It is not enough for the words of the questionnaire to be simple and familiar. Words such as "often," "seldom," "regular," "cheap," "expensive," "good," "bad," etc., are simple enough to be familiar to almost any English-speaking respondent, but they should be used in questionnaires only with great caution. The reason is simple. While these words are familiar to almost everyone, their meaning varies from person to person. One person might consider himself a "regular" listener to a daily radio program if he heard it once a week, while another might not consider himself a regular listener if he missed it once a week. Similarly, one respondent might consider a \$50 suit "cheap"; another respondent, "expensive." Such words are difficult to use in questionnaires because they are abstract. The degree of confusion would be lessened if concrete words were used. If the respondent had been asked either, "How much did you pay for that suit?" or "How often did you hear that program last week?" the verbal confusion, at least, would have been eliminated. (It is still possible, as we have seen, that respondents might be unable or unwilling to answer these questions.)

Still another difficulty arises with regional differences in words. For example, the term "jelly doughnut" in New York means the same thing as "Bismark" in Chicago. In Chicago a "show" is a movie and in New York it is the legitimate theater. Constant alertness is needed to avoid these sources of confusion.

Even with simple, familiar, and concrete words, trouble may arise if the words are not composed into sentences that are easily understood. The following anecdote illustrates the point:¹

I remember once during the war we were doing a survey of the WAC's and one of the problems which was to be covered was the extent to which there was a strong desire to be permitted to wear civilian clothes, and on what occasions and why. The person in charge of the study was interested in exploring the suggestion that the girls felt too conspicuous in public in the uniform, under too many constraints to have any fun. Well he gave considerable thought to the problem and how to word it and came up with this: "Are there any things which you can't do when you're wearing your uniform that you could do if you weren't wearing it?" Actually, the question almost got asked before its hilarious ambiguity was perceived.

* ¹ "Study Design in Opinion Research," Shirley A. Star, talk on Nov. 3, 1949, to a joint meeting of the Chicago chapters of the American Statistical Association and the American Marketing Association.

A rule of thumb which could usefully be heeded is that questions should be brief, simple, positive interrogations. Moreover, they should be as conversational as possible, since artificiality is conducive to ambiguity. Even when all these rules are met, however, difficulties remain. For example, two questions which have the same literal significance for the researcher may produce statistically significant and important practical differences in responses. This is rarely true for questionnaires seeking information about past behavior; it is often true for questionnaires regarding attitudes or motivation. The discussion of this vexing problem, therefore, is reserved for the chapters dealing with these subjects.

Also, there is little need to discuss at this point the effects on responses of the order or sequence of questions in the questionnaire, or of the form in which the response is asked to be given. These issues, too, are of far greater importance for the more complex problems of securing information regarding intentions, attitudes, or motivation. Realizing, then, the serious incompleteness of the discussion in this chapter of communication, we will postpone further discussion of the subject until later chapters.

13.5 The Willingness of Respondents to Provide Information

Up to now the discussion has emphasized the ability of respondents to give information. Marketing researchers are limited, however, not only by the ability but also by the willingness of respondents to give information. By understanding the causes of this unwillingness it can probably be diminished. Or at least, researchers can learn to forego seeking the sort of information that will not willingly be given or will not be given at all.¹

There are three basic (though not completely separate) causes of unwillingness to provide information:

1. Respondents may make incorrect statements because they feel that there will be some association between certain responses and either reward or loss for themselves.

2. Respondents may make incorrect statements because the emotional reaction to the incorrect response is pleasanter than the reaction to the correct response. For example, a person questioned about magazine readership might state that he was a regular reader of *Harper's*, which he had seen only once, rather than *Flirt*, which he read regularly (every issue). Pride might be associated with the incorrect statement; embarrassment with the correct one. A somewhat different instance of emotional reaction would be a desire to please the interviewer, to keep from hurting someone's feelings or getting someone in trouble, etc.

¹ We define "willingness" not as the willingness of respondents to answer questions but as the willingness to try to answer them correctly. "Passive resistance" to the questioning process is a possibility which must never be forgotten.

3. Respondents may make incorrect statements because of indifference as to the actual statement and a reluctance to take the time to determine and make the correct statement. Such respondents give the easiest response, the one which will most shorten the interview, regardless of accuracy.

We shall now consider each of these sources of unwillingness in detail and examine methods of overcoming them.

1. Unwillingness because of anticipation of reward or loss. Unwillingness that derives from the first cause is rare in marketing research projects seeking information about behavior. It is more frequent in surveys exploring attitudes or motivations, as is discussed in Chapters 15 and 17. Rarely will respondents believe that they can gain or lose anything because of their response to questions about magazine readership, purchases of branded articles, etc.—behavior which is the subject of such a large part of this type of research. Certainly any research man seeking valid information should try to eliminate or minimize any possible feeling that rewards or losses will be associated with particular responses.

One way to achieve this is to make the respondent feel that his response cannot be identified with him. This is particularly important in government surveys and less important in marketing research studies. Such a feeling of anonymity is easy to achieve in mail questionnaires, when it is seldom necessary to ask the respondent to sign his name. (If it is necessary to identify the respondent, a simple coding device will do this.) Another way to increase willingness may be to indicate the auspices of the inquiry, especially when the survey is being conducted by a research agency or any other organization whose interests do not apparently have any relation to the substantive findings of the study.

The problem of unwillingness is much more acute when an enterprise conducts a survey among its own employees. Accurate information can be secured under these circumstances only if the relationship between employee and employer has been such as to make the former have confidence in the professed motives of the latter. Such confidence might be established, say, by conducting a series of surveys which were followed in each case by neither dire nor beneficent consequences for the individual participants. The benefits received by a respondent *as an individual* should have no relation to the particular answer he gave.

There is another situation in which unwillingness results in a refusal to answer questions at all for fear that the interview is just a pretense for an entry into a sales talk. As surveys have become more and more widely used, the number of door-to-door salesmen who pose as survey takers has increased, particularly in the large cities. While there have been some attempts to stop this practice, it seems a difficult one to eradicate. Hence

legitimate interviewers are likely to find it progressively more difficult to gain cooperation. Special letters of introduction, identification cards, and above all, the ability to make the true purpose clear and win over the potential respondent quickly before she slams the door; all these are helpful. If the respondent is at all suspicious, the interviewer should immediately and explicitly attempt to put her at ease.

2. Unwillingness for emotional reasons. The second cause of unwillingness is a difference between the emotional response which will follow if the correct answer is given and the one which will result if an incorrect answer is given. It has long been known that people are motivated to some extent in both speech and actions by the hope of approval by (1) themselves and (2) others.¹ Even when respondents are questioned briefly about relatively impersonal subjects by fieldworkers whom they have never seen before, there may be a tendency to provide inaccurate data because of a need for giving socially approved responses. There may be something less than complete candor in responses to questions about reading habits, prices of objects purchased, ownership of automobiles, occupation, income, etc.

The problem of eliminating the bias injected by this unwillingness to give completely accurate information is difficult and permits no easy generalization. It is often impossible to estimate the magnitude of this bias and make adjustments for it. When verbal statements can be checked by observation, it is easy to use the double-sampling technique referred to earlier. But often the thing being measured—readership, for example—cannot be checked by observation. Often when observation is in principle possible, the very reason that makes the respondent reluctant to give the correct answer in the first place makes him equally reluctant to permit observation which would reveal that his answer was incorrect.

The measurement of the readership of periodicals is one of the basic problems in advertising research. Although the sales of magazines and newspapers are easily determined by consulting the figures of the Audit Bureau of Circulations or the publishing companies' records, the determination of readership (*i.e.*, the number of people who read the magazine in contrast to the number of people who buy it) is much more difficult. At the same time, however, readership is much more relevant information in the selection of a media schedule—a list of magazines, etc., in which advertising is to be placed. There is considerable danger, however, that people will tend to overstate their readership of "prestige" periodicals (*e.g.*, *The*

¹ This distinction is made for logical nicety rather than for its practical importance. For all except the seriously maladjusted, the values of the individual are derived from and largely identical with the values of the society with which the individual identifies himself. Therefore, in almost all cases, the response which will increase the self-esteem of the respondent is the response which he feels is most highly regarded by his society.

New York Times, *Harper's*, etc.) and to understate their readership of magazines devoted primarily to fantasies of romance and adventure, sex, or crime and of some tabloid newspapers. The ratio of *reported* readership to actual sales of so-called "prestige" periodicals, if taken seriously, would sometimes suggest that copies of these magazines or newspapers are passed hurriedly night and day from person to person. On the other hand, the ratio of reported readership to sales of magazines dealing with the sex lives of the inhabitants of Hollywood sometimes suggests that such publications must often be purchased for purposes other than reading.

Several methods have been suggested and tried in order to mitigate or overcome such biases. Examining fingerprints on used copies in order to determine the number of different persons who have handled each copy is one method. The sampling difficulties of this approach are, of course, staggering, and therefore probably no accurate inference could be made about readership, even if one granted the dubious implicit assumption of a high correlation between handling and readership.

Others have attempted to estimate the bias by questioning readers both directly about the periodicals they read and also about their knowledge of particular features unique to each periodical. Such an approach has several possible shortcomings. (1) People who do not read a periodical regularly or who did not read it "yesterday" may be familiar with the features unique to it. For example, many subscribers to the *New York Times* are familiar with Westbrook Pegler, or many people who seldom see *Esquire* may be familiar with its Varga or Petty girls. (2) People who do read a periodical regularly or who did read it "yesterday" may not be familiar with some feature unique to it. There may be some "canceling" effect for (1) and (2), but it is seldom prudent to rely on the hope that biases will offset one another.

A third approach seems more promising. A group of respondents could be asked, "What newspaper(s) did you *buy* yesterday?" The bias in the responses to this question could be measured by the figures on sales either of the Audit Bureau of Circulations or of the newspapers. The factors determining the bias of responses to such a question are probably not importantly different from the factors determining the bias in responses to the question, "What newspaper(s) did you *read* yesterday?"¹ The measured bias in the responses to the question on purchases could be used to adjust the figures on the responses to the question on readership. For example, suppose that the ratio of *reported* to *actual* sales of newspaper A

¹ This assumption of the identity or similarity in the nature of the biases of these two questions is necessary for the usefulness of this approach. It seems a more or less plausible assumption, although the validity of deductions based upon it has not been tested.

was 1.3/1. Then the reported readership figure could be divided by 1.3 to achieve the "correct" figure on readership.

This method makes use of a measurable bias (in reported sales) in order to estimate what is assumed to be a highly correlated but unmeasurable bias (in reported readership). The same principle is applicable to other types of marketing research projects. What is required in each case is a variable which is (or can be assumed to be) highly correlated with the variable of primary importance. The bias in the first variable must be measurable, so that one may use the measured bias in this variable to estimate the unmeasurable bias in the variable of primary interest.

This approach seems quite attractive at first, but the underlying assumption of similarity in the nature of the biases in responses to the two questions is only plausible and not of established validity. It might turn out that the biases were really quite different. Another approach to the basic problem is interesting. Suppose a person is shown a list of newspapers and asked which newspapers on the list he reads regularly. His answer to this question would be subject to the full bias of inflation for "prestige" papers and of deflation for "nonprestige" papers. Several more questions which gave him plenty of opportunity for displaying his urbanity, sophistication and patriotism might next be posed. Then one might ask quite casually the *particular* newspapers he *happened* to read the previous day. This question would be made to seem an unimportant anticlimax to the questions which went before, and the incentive for a face-saving response would be relatively small. For example, a person who claims to be a regular reader of the *New York Times* or *New York Herald Tribune* may without qualms acknowledge that yesterday he "happened" to read a less distinguished paper. Therefore one may hope that even though the biases may be very large to the "read regularly" questions, the bias may be quite small for the "read yesterday" question. A similar approach has been applied in studies of magazine readership.¹ Like the preceding method, of course, this method rests on an assumption that would be difficult to test adequately—the assumption that proper questioning procedure can, in fact, reduce the incentive for face-saving responses.

A classic example of this general approach has long circulated among marketing research personnel. The problem was to measure actual readership of *Gone with the Wind*. Many people were ashamed to admit that they had not read this book and therefore claimed to have done so even when they had not. An ingenious variation of the wording of the question seemed to overcome this difficulty quite satisfactorily. Instead of asking, "Did you read *Gone with the Wind*?" respondents were asked, "When do

¹ *Continuing Study of Magazine Audiences*, Report No. 9, Life Magazine, Nov. 15, 1947, Appendix.

you plan to read *Gone with the Wind*?" If the person had already read the book, he would naturally say so. If he had not, he could easily say that he intended to read it during his vacation, next week, etc. This would be a quite satisfactory face-saving response; yet at the same time, the fact that the book had not actually been read would have been established. The ability to formulate questions in such a way would undoubtedly be useful in many perplexing problems which arise in marketing research. The marketing research practitioner can seldom afford to be content with stereotyped procedures; there is usually a need for imagination and ingenuity.

There are many other variations of this general approach, which may be described as an attempt to reduce the emotional impact of certain key questions. Sometimes this is done by preceding the questions of crucial interest for the survey by a series of questions which probably have little emotional significance (e.g., questions concerning attendance at sports events, liking for picnics, reactions to the weather, etc.). A skillful interviewer can often use this method successfully by conveying to the respondent the feeling that the questions of primary interest are of the same order of importance as the initial questions and that they (the questions of interest) have as little power to reflect credit or discredit on the respondent.

In many problems the task of direct measurement of bias in response does not seem to be as overwhelming as it seems with measurements of readership. The task may, of course, be expensive. For example, a manufacturer of washing machines once thought it desirable to ascertain the amount of washing done by different-sized families, the place where such washing was done, etc., in order to determine the best size of a projected new model of his washing machine. Statements from respondents regarding the size of their weekly wash were believed to have an upward bias. The magnitude of this bias in responses could be overcome by *observing* clothes lines, laundry hampers, laundry bags, or laundry bundles of respondents. Such observation would be more costly than questioning, but the general idea of double sampling might again be used. When the refusal rate is high, the results of observation may themselves be subject to a sampling bias which may be of serious practical importance.

In this discussion of unwillingness to give the correct answers for emotional reasons, we have not discussed one particular class of emotional reasons—the desire to please the interviewer. People with practical experience in fieldwork and supervision of fieldwork seem to agree that this problem is potentially very serious, but there apparently is no evidence as to how serious it is. In suggesting ways to overcome this bias, all that we can do is mention the common-sense injunction to impress on the respondent that accurate information is sought rather than any particular "right answer."

3. Unwillingness to take time to provide thoughtful answers. The final general cause of unwillingness is the reluctance of respondents to expend the effort or take the time to provide thoughtful (or any) answers to questions or to provide (or make accessible) evidence for observation. In its most extreme form this reluctance results in a refusal by respondents to participate in any way in the survey. In a perhaps more insidious and costly form this reluctance results in completed interviews containing statements of great inaccuracy. The significance of a direct and complete refusal is clear. On the other hand, a questionnaire filled in with careless answers looks very much like a questionnaire filled in with careful answers, and sometimes the most sensitive methods won't distinguish one from the other. Hence there is a real danger of self-delusion if the researcher is not fully aware of this kind of unwillingness. Therefore we must examine further the problem of "semirefusal," or passive resistance.

Unwillingness to give careful answers has quite different consequences for the observational and the communicative techniques. For observational methods, unwillingness is much more apt to result in refusals to participate rather than in halfhearted participation yielding data of questionable accuracy. It would be rather unusual for respondents to provide or make accessible to fieldworkers for observation only *part* of the relevant evidence. Seldom will workers be permitted to view only one, two, or three automobile tires, or to look at only parts of pantry and bathroom shelves. And such halfhearted participation is relatively easily spotted by workers, who can state precisely the nature and magnitude of the problems thus created. The main danger of incomplete observation is that of not knowing all the places to look. If the interviewer asks to see only the pantry or kitchen, she may miss a 100-pound sack of sugar stored in the basement.

For communicative techniques, on the other hand, halfhearted participation is probably more pervasive and is almost certainly harder to recognize. Although careless answers can sometimes be identified, very often it is difficult or impossible to know whether a series of "no's" or "don't knows" are valid responses or the offhand responses of respondents trying to give merely formal rather than real cooperation.

Two general ways of diminishing unwillingness from this cause have been found practicable. First, it is possible to offset the effects of reluctance to respond by giving the respondent some material compensation for cooperation. This method has been almost exclusively used in mail questionnaires. For example, a quarter or some merchandise premium may either be sent with the questionnaire itself or offered if the respondent will return a complete questionnaire. Most experimentation with different types of rewards for mail questionnaires has been kept private rather than pub-

lished. It is clear, however, that the rate of return of questionnaires is substantially improved by such methods. The only question is which of the alternative rewards is most effective. But no one, to our knowledge, has investigated the extent to which those who are induced to return questionnaires by these methods give merely token cooperation.

No direct evidence exists as to the effect of paying respondents for their participation in personal interviews; indeed, any such scheme is conspicuously absent from commercial practice. A study by the National Opinion Research Center¹ indicates that a large proportion of persons who refuse to participate at all in personal interviews do so because they are genuinely busy and cannot spare the time. There is no evidence as to whether or not the same is true for those who participate only halfheartedly.

The second method of attempting to overcome reluctance to cooperate wholeheartedly is to design questionnaires and conduct interviews so as to create a sympathetic and helpful attitude on the part of respondents. It would be difficult to find a major meeting of marketing personnel in recent years which did not consider ways of reducing this source of error—as well as outright refusals—by establishing rapport between respondent and researcher. The interest in this problem is so general and intense that a large body of “home remedies” have been developed. Only few of the many generalizations have received either rigorous testing or widespread acceptance. The following generalizations are those for which the evidence seems most convincing:

1. The greater the interest of the respondent in the subject matter of the questionnaire, the lower is the rate of passive resistance. This generalization is of limited usefulness, since the primary subject of the questionnaire is determined by the project design and cannot be chosen or altered merely to reduce the rate of refusal. It is possible, however, to start a questionnaire with preliminary questions which are of great interest to the respondent although of no, or little, interest to the researcher and thus to increase the proportion of willing respondents. The transition from the interest-evoking questions to the essential questions must be done so as to disguise from the respondent the purpose of the first questions. But such transitions, or any transitions, are likely to create suspicious and unwilling respondents. If there is no connection between two sets of questions, the respondent may discern the purpose of the questionnaire design and thereby destroy its effectiveness in achieving real cooperation.

2. Questionnaires that are easily answered encounter less halfheartedness than more difficult questionnaires. It is well known, for example, that questionnaires requiring the computation of rates of use or purchase are

¹ See p. 455, Paul B. Sheatsley, “The Public Relations of the Polls,” *International Journal of Opinion and Attitude Research*, Vol. 2, No. 4, pp. 453-468, Winter, 1948-1949.

less well answered than questionnaires which secure the required information by other means. People probably less often take the trouble to answer carefully a question such as, "About how many pounds of bread do you eat each day, on the average?" than a question such as, "How many slices of bread did you eat yesterday?"

3. The reluctance to answer brief questionnaires carefully is thought to be less than that for longer questionnaires, particularly when the subjects studied are of relatively little interest to respondents as most marketing research studies, unfortunately, are. Answers to questions toward the end of a long questionnaire are often given with less care than answers to earlier questions.

Sometimes methods are available for offsetting biases of this last type. In readership studies of the type that require a respondent to go through a magazine or newspaper and tell whether or not he saw each page, one interview is started from the front of the newspaper or magazine while the alternate interview is begun at the back. Similarly, wherever alteration of sequences of questions will not seriously impair the questionnaire, and where relative rather than absolute measures are needed, some rotation of questions may be used.

4. The age, sex, appearance, method of approach, and personality of the fieldworker have important effects on willingness to give careful responses. Furthermore, it is fairly generally agreed that an interviewer who can get careful answers from one group of the population may not be so successful with other groups.

There is unfortunately no good objective way of detecting halfhearted answers other than by the costly method of checking by observation. Sometimes internal inconsistencies may be detected in the editing process,¹ but these are often difficult to correct from the evidence contained in the questionnaire itself without introducing arbitrariness which may itself lead to bias. For example, if a respondent indicates at one place that she has bought two pairs of shoes in the last year, while later she mentions three specific purchases, one would probably be safe in changing the original response from two pairs to three pairs. But in other instances the correct answer in two conflicting responses cannot be so easily ascertained. Of course, an alert and skilled interviewer can detect and reconcile such inconsistencies during the interview itself. But it is relatively rare that the error due to halfhearted response shows up in internal inconsistencies, so at best this approach gives only a hint of the full extent of the error due to careless response.

It has been suggested that such interviews can be detected by asking persons different questions with the same literal significances and different

¹ See Editing, under section 22.2.

wording. Theoretical objections to this procedure are mentioned in Chapter 12.

Moreover, as a practical matter it is hard to devise questions which are identical in literal significance yet different in wording without the respondent's being made fully aware of this fact. Two undesirable by-products may then follow. In the first place the respondent may be irritated and less inclined than ever to give wholehearted effort in his responses. In the second place, the length of the interview is necessarily prolonged, and this may lead, in turn, to greater inaccuracy than would have resulted if a shorter interview had been held. The use of repetitious questioning aimed at measuring inconsistencies in response offers few theoretical advantages and many actual or potential disadvantages. Moreover, it seems to be thoroughly unpopular with most interviewers, probably because of the unpleasant situations with respondents that are created.

Probably the best way to detect careless respondents is to employ intelligent, skilled interviewers who are permitted and encouraged to comment on interviews which they believe to be inaccurate because of careless responses. When this is done in the pretesting stage, questions which particularly tempt careless responses can be weeded out or modified. When it is necessary to ask difficult-to-answer questions in order to get information which can't be obtained in some other way, the same skill and intelligence which enables good interviewers to recognize careless or halfhearted answers is of great importance in curtailing this inaccuracy at its source.

13.6 Cost

So far almost nothing has been said about the relative costs of the techniques discussed. Nothing elaborate need be said about costs of securing information about behavior. The principal differences in cost exist because of differences in the time required for each interview or observation. Only in the case of techniques based on unaided recall are there additional costs in tabulating and interpreting the reports of fieldworkers. The differences in the time required for observational techniques as compared with communicative techniques will vary from project to project and can in each case be estimated on the basis of either pretests or previous experience. In general, however, it may be stated that taking inventories in the observational techniques requires more time per respondent than the communicative techniques require.

CHAPTER 14

SECURING INFORMATION ON INTENTIONS

14.1 Introduction

This chapter is concerned primarily with an appraisal of the usefulness of statements of intention and with problems of eliciting unambiguous and accurate statements of intention. Before proceeding to this task it seems desirable to inquire into the factors which determine the relative usefulness for marketing forecasts of information on behavior and information on intentions.

When the marketing researcher obtains information about the behavior of respondents, he is the one who must analyze the experience of respondents and who must discern its implications for their future activity. When the researcher secures stated intentions, each respondent analyzes his own previous experience and expresses his guess as to its implication for his own future activity. The researcher is left the task—frequently a difficult one—of appraising the meaning and accuracy of the respondent's statements. Intentions need interpreting, since the correspondence between intention and behavior is varying and not obvious. Robert Burns made this same point more charmingly in his poem, "To a Mouse, on Turning Her up in Her Nest with a Plow":

But, Mousie, thou art no thy lane,
In proving foresight may be vain;
The best-laid schemes o' mice an' men
Gang aft agley,
An' lea'e us nought but grief an' pain
For promised joy!

Of primary importance for the decision as to who—the researcher or the respondent—is best qualified to forecast future activity is the extent to which each has access to most of the most relevant evidence. It might seem that the individual respondent would inevitably have access to more relevant information concerning forecasts of his future behavior than would the researcher. Since an individual's future behavior is determined by his previous experience and his reflections on it, it is not immediately obvious why the researcher should ever undertake the task of forecasting the behavior of individuals when the task could be left to the individuals them-

selves. Nevertheless, in some cases the researcher is more competent for this task than the individual respondent.

An example may indicate the reasonableness of this assertion and suggest reasons for its validity. The Eureka Beverage Company (manufacturer of soft drinks) has had much greater success in forecasting its sales from year to year by extrapolating its own sales data and the data for the industry than by asking respondents how much soda pop they intend to drink in each of the forthcoming twelve months. Why is this so?

First, the sales of soda pop have shown a steady secular rise, a regular seasonal pattern, a general imperviousness to cyclical forces, and relatively small random or erratic fluctuations. The Eureka company predicts sales by saying that sales next month will be like sales this month, with an adjustment for the relatively stable pattern of seasonal variation and secular change. The difference between the predicted value of sales and actual sales is on the average quite small. The forces affecting the consumption of Eureka soda pop seem to change through the years and seasons systematically and consistently. The first reason for the researchers' relative competence in forecasting rests, then, upon the adequacy of the evidence available to him for this purpose. Assuming the accounting records to be accurate, he has a completely accurate record of total sales of Eureka soda pop.

The second reason concerns the inadequacy of the evidence available to the respondent. Whereas the researcher has available accurate data on the volume of sales by months for many years, the individual respondent has available only an indistinctly remembered personal history of soda-pop consumption. Probably few respondents could do more than say that more soda pop was consumed in hot weather than in cold and that consumption for each week varied within certain rather wide limits. The inaccuracy of respondents' memories is attributable largely to the characteristics of the events (purchases of soda pop) being remembered. The purchase of soda pop is for most people neither regular, nor important financially, nor of great emotional significance. Nor does it seem probable that the respondent would be capable of making an accurate forecast merely by considering his current tastes and resources. The purchase of soda pop is usually on the basis of whim and fugitive circumstances rather than upon the explicit, prolonged weighing of alternatives. It is not surprising that the researcher can in these circumstances do a better job of forecasting by considering the detailed information on the past behavior of respondents than by asking respondents to make individual forecasts on the basis of their imprecisely remembered personal histories.

It is quite possible, of course, that any *individual* consumer might be able to forecast his *own* behavior better than the Eureka Beverage Com-

pany. If the individual had never drunk soda pop, his forecast that he will drink no soda pop during the next year might be more accurate than a forecast made on the basis of the company's sales records. Nonetheless, the *aggregate* of individual predictions would probably be less accurate than the company's forecast based on its knowledge of past sales.

What generalization does the preceding example suggest? When the relevant forecasts concern the future consumption of a good which (1) has been widely consumed for a long period of time, (2) has exhibited stable temporal consumption patterns, and (3) is purchased because of impulse rather than because of conscious weighing of alternatives, information on behavior rather than expressed intentions will probably yield the more accurate forecast.

There are other factors which sometimes make information on behavior a more promising basis for forecasts. These other factors concern the relative difficulty of communicating about intentions as compared with communicating about behavior. For a statement by a respondent concerning his own future activity to be unambiguous, it is essential that the researcher know exactly the nature of the circumstances which the respondent expects to exist at the time of the future activity and the relevance of these circumstances for the respondent's decision regarding that activity.

The statement by a respondent, "I am going to buy a new car within the next year," probably carries with it unexpressed contingencies concerning the respondent's income, family status, health, the price of the car, the price of other things, etc. The respondent is probably meaning to say, "I am going to buy a car in the next year, if I get the raise that my boss has promised me, if my son comes home from college and gets a job, if the family continues well, if I can get the car at list price with no accessories, if the prices of food and clothes don't go very much higher, if I can rent a garage near the house, and so on."

Hence the first problem of communication peculiarly associated with statements of intention concerning future activity is the problem of unexpressed contingencies. The researcher must attempt to get the respondent to be explicit about the contingencies which importantly affect the future activity which is the subject of the respondent's statement. The problem of doing this is discussed later in this chapter, but it is important to indicate at this point that this problem exists in getting information about intentions and that there is no parallel problem connected with getting information about behavior, *i.e.*, past behavior.

There is a second and closely associated problem of communication for interviews concerning intentions. Some of the circumstances which may be relevant for a respondent's decision concerning future activities may be known to the researcher and not to the respondent. For example, a re-

spondent who considers the price of the clothing, the time that it becomes available on the market, its retail outlets, alternative brands of clothes, etc., to be relevant for his decision concerning the purchase of a designated brand of clothes may not be informed on these points at the time he is interviewed. For the respondent's statement of intention to be unambiguous, it would be necessary that the researcher know exhaustively what the relevant circumstances are and that he convey to the respondent that information which he has and which has not been available to the respondent. Unfortunately, getting fieldworkers to do this without seriously biasing the responses received is almost impossible. This again is not a problem in research seeking information on behavior. These difficulties of communication are of much greater importance for some types of future activity than for others, and in many cases they are important enough to make it advisable for the researcher to attempt the relatively easy task of getting information about behavior.

The discussion so far has indicated that in certain fairly frequently encountered situations more accurate forecasts can be made on the basis of information on behavior than on the basis of statements of intention. Most of the comments have indicated the great difficulties involved in getting useful statements of intention. There are circumstances, however, in which such statements are useful and in which unambiguous, accurate statements can be secured. In the next section of this chapter we discuss the usefulness of statements of intention in making certain types of unconditional and conditional forecasts—assuming that these statements are unambiguous and accurate. In the final section of this chapter we discuss the problems in getting unambiguous, accurate statements.

14.2 The Usefulness of Statements of Intention in Forecasting

We now consider the usefulness of statements of intention in forecasting the level of general business conditions or the volume of sales expected in any specified industry. Statements of intention are not sufficient, perhaps, to play the major role in making forecasts of either sort, but these statements can be helpful in refining the interpretation of data of other kinds.

Consider first the problem of forecasting general business conditions. Various techniques have been used with varying success. Probably the most successful forecasters have been men of experience and good judgment, who have systematically and subjectively appraised the relative importance of various factors conducive to expansion and contraction. Other forecasters have tried to objectify the process of forecasting and to this end have devised mathematical models of the economy which indicate various economic magnitudes and their relationships. Whatever method of forecasting is used, the plans and expectations of the members of the society

whose economic future is being forecast must be considered; for the plans of individuals not only are adapted to changes in the economic environment but also are important determinants of what that environment is. Statements of intention are, of course, one means by which those expectations and plans can be discovered. The problem of integrating information on intentions with information on other things, such as the physical volume of industrial production, price changes, changes in inventories, etc., is a problem in economics rather than marketing research. The problem of getting the information on intentions belongs to marketing research.

Stated intentions can be useful for forecasting the sales of particular industries in much the same way as for forecasting general business conditions. It is important in using such statements that there be some basis for knowing the relationship in the past between intentions and behavior. Consider, for example, an automobile company which had secured statements of intention from a sample of people, in which it was found that 5 per cent of the people intended to buy automobiles in the forthcoming year. Abstracting from the possibility of sampling error, if 10 per cent of the same population had purchased automobiles in the preceding year, would the automobile company be justified in expecting a large decline in sales? The answer is "no." Without some information on the intentions of the population in the period before the year in which 10 per cent bought automobiles, there is hardly any basis for appraising the significance of the stated intentions in the current year. If, in the period preceding the year in which 10 per cent bought automobiles, only 5 per cent had intentions of buying, there would be no reason to expect a decline in sales solely on the basis of information that current intentions to buy were indicated by only 5 per cent of the population. On the basis of this information alone, no decline in sales would be anticipated. If, on the other hand, the number of people intending to buy had declined since last year, this information in itself would lend some support to a pessimistic forecast of sales.

Intentions to buy many kinds of consumer durables are measured annually by the Federal Reserve Board, and these measurements are published in the *Federal Reserve Bulletin* as a part of the Survey of Consumer Finances. These data, if used with care, can contribute greatly to the forecasting of general business conditions and of the sales of particular industries. Special marketing research projects can fill in further details on this kind of information.

Marketing research more often concerns a different kind of forecasting than that discussed so far. Marketing research more often involves the attempt to predict the reactions of consumers to certain marketing practices or policies which might be followed. For example, the marketing

research department might be called upon to select the best of several proposed alternative packages for a product. This problem might not seem to involve forecasting at all; it might appear rather to be a technical problem which would be referred to experts on package design just as production problems are referred to industrial engineers. But the "best" package design is definable only in terms of its influence on sales and cost. In order to select the "best" package design, then, it is necessary to predict the sales volume that *would* result *if* each of the alternative packages were actually used. Perhaps one would be satisfied to predict only relative sales volume for each package design, but in any event prediction would certainly be involved.

For this problem it is not possible simply to look at sales this year and say that next year will be the same, nor to use the variations suggested earlier in this section. Such methods are useful for forecasting sales if the present package is retained, but they would not be helpful in indicating whether a new design, which at present is only an artist's sketch, would cause sales to rise or fall. Many other problems encountered in marketing are similar, since one of the most pervasive kinds of problem attacked by marketing research is the problem of forecasting consumers' responses to new stimuli. Information on past behavior is clearly of little use in such forecasts; other types of information are needed.

Probably the most valuable kind of data for these conditional forecasts of the effect of new stimuli is experimental data. For the reasons already discussed in Chapter 4, experimentation provides data which are less ambiguous and which permit a more precise definition of relationships than observational data—either historical or survey. A frequently used kind of experiment in marketing research is the sales test, in which the differential effect of alternative marketing practices and policies on sales is estimated. Whenever possible, experimentation should be used for forecasting the responses of consumers to new stimuli. Unfortunately, sales tests and other forms of experimentation are often too time-consuming, costly, or difficult to administer; and other types of evidence must therefore be used. It may be instructive to see in more detail why sales tests frequently cannot be used and why, therefore, other research procedures are called for.

First, under some rarely encountered circumstances, sales tests yield data which are of little value for forecasting. Experiments are of limited use for forecasting whenever the conditions prevailing in the experiment are not expected to continue during the period for which the forecasting is made. For example, in the years immediately following World War II, the demand for certain consumer goods was so great that almost anything which was produced could be sold. There would have been no point in trying to find out by a sales test whether a soap package should be pink or blue when

the demand for soap flakes (at prevailing prices) greatly exceeded the available supply.

Second, it is often too costly to use sales tests to determine consumers' reactions. The cost of the sales test is often prohibitive—as when new products are being tested, for example. It would probably be foolish to attempt to determine consumers' reactions to proposed designs for automobiles, washing machines, or other durable goods by sales tests. The necessity for having in service only standardized models, the further need for national publicity to accompany the introduction of new designs, the desirability of keeping new developments secret until their introduction to the public, and other reasons account for the relative costliness of sales tests under these circumstances. Further, sales tests necessarily involve the creation of the new product, the new package, the new advertisement, or the new anything else and its submission to the consumer.

Third, sales tests sometimes take too long. Sales tests involve time-consuming preliminary steps and must extend over periods of time frequently running into weeks or months. Sales tests require an amount of time below which the expenditure of money is powerless to effect a reduction, while there is almost no limit to the extent to which the time required for surveys can be diminished. By hiring more personnel, the time required for interviewing and clerical work can be reduced.

For reasons such as these, evidence in the form of stated intentions (or attitude measurements, etc.) can often more profitably be used to forecast consumers' reactions to new stimuli. Therefore we now inquire into the factors influencing the relationship between stated intentions and actual behavior in order to see how useful such intentions can be for forecasting. In order to simplify the subsequent discussion, we introduce here two bits of terminology suggested by John Dollard.¹ The situation in which the respondent makes a statement of intention will be called the "test situation"; that in which he makes the decision being investigated, the "criterion situation."

Now, consider what characteristics of statements of intention—assuming that they are accurate and unambiguous—affect their value for the prediction of behavior. First, the extent to which the respondent at the time of the interview knows the facts which he will think it desirable to know before reaching a decision importantly affects the usefulness of his expressed intentions. Persons typically respond to questions concerning buying intention with positive statements of intention even though they know only a small proportion of the facts which will be considered in making the purchase. When discrepancy between the knowledge of respond-

¹ John Dollard, "Under What Conditions Do Opinions Predict Behavior?" *Public Opinion Quarterly*, Winter, 1948-1949.

ents in the test and in the criterion situations is large, the usefulness of statements can be seriously impaired.

Such a discrepancy is apt to be greater for specialty and shopping goods than for impulse goods. The purchase of refrigerators, for example, is usually preceded by an examination of many brands in order to compare price, size, appearance, special features, etc. Not infrequently such an examination, with its consequent increase in the respondent's knowledge, will change the intention—tentatively held and based perhaps on hearsay or advertisements—that was the basis of the statement to the researcher. Several surveys of buying intentions have shown a closer association between statements of intention and advertising budgets than between purchases and advertising budgets. This fragmentary evidence is consistent with the thesis that intentions based upon advertising are significantly modified as new facts are obtained in preparation for making the purchase decision. Predictions based upon types of behavior previously experienced will obviously be less subject to error from this cause. For example, respondents who have been driving cars for years and who have bought many automobiles will probably make statements of greater usefulness than will prospective first-time purchasers.

In the second place, it seems reasonable to believe that the stated intentions of a respondent whose motives for the purchase or activity in question are partially subconscious will be less useful for forecasting than the statements of a respondent whose motives are relatively fully recognized. Subconscious motives can express themselves through action in the market place without ever being recognized by the persons whose motives they are. Such motives need not be and often are not discernible in statements of intention. Possible examples of activities determined in part by subconscious motivation are purchases designed to indicate the purchaser's affluence or to gratify his taste for lurid literature. Just as people have chronically exhibited a reluctance to admit *having read* lurid literature, so they are reluctant to admit *intending* to read it.

In the third place, the strength of the respondents' intentions affects the correspondence between statements of intention and behavior. If the survey requires that respondents state their intention regarding the purchase of several competing brands, the statements may more frequently be carried out if one brand is preferred very strongly to its competitors than if the respondents border on indifference.

A fourth factor (closely related to the first) affecting the usefulness for prediction of statements of intention concerns the interval between the test and criterion situations. Any statement concerning future plans is, of course, made on the basis of certain expectations or assumptions concerning events that will happen in the future and affect the person making the

statement. The accuracy of any statement quite clearly, therefore, depends upon the extent to which the expectations which form the basis of the statements are realized. So it is necessary to consider the factors affecting the fulfillment of these expectations. C. O. Hardy used to preface his business forecasts by saying that the forecast would be good only for the foreseeable future. He would then go on to say that the future was never foreseeable and that he would have to compromise with this rigorous standard of "foreseeableness" and make a forecast for three or four or six or eight months. The point of all this is, of course, that expectations are more often realized when they extend only a short time into the future than when they extend a long time into the future. It will follow, therefore, that statements of intention concerning the immediate future can be expected to be carried out more often than statements concerning the distant future. A person's actions can usually be expected to correspond closely to his statement regarding his intentions to buy an automobile within the ensuing week, but such statements are less often fulfilled as the period for which they are relevant becomes longer. Probably there would be gross discrepancies between statements covering several years and actions during that period.

Not only the length of the interval between the test and the criterion situations, but also the nature of the contingent circumstances, affects the accuracy of statements. Most intentions concerning the number of meals that will be eaten or the city in which one will be living can be expected to be fulfilled. Intentions concerning meals can be expected to be fulfilled because they concern an activity which is habitual and of a high order of importance. People have more often been observed to alter their daily routine so as to make possible the eating of a given number of meals than to alter the number of meals eaten so as to leave undisturbed some other daily activity. Further, there are physical limits on the extent to which food can be foregone for the sake of other activities. Because of the social and physical importance of meals, therefore, the circumstances which will be allowed to affect them are few and of relatively infrequent occurrence. Intentions concerning the city in which one will be living can be expected to be fulfilled because of the infrequent transfer of persons from city to city and because of the fewness and great importance of the circumstances which are usually necessary to cause people to move.

The circumstances which can alter people's plans concerning most other activities are usually neither so few nor of such necessarily great importance. A plan to buy a dress, for instance, can usually be postponed without creating either physical hardship or social distress. Such purchases will often willingly be deferred or foregone because of a great variety of circumstances which would cause people neither to change their eating

habits nor to move to distant cities. Plans to buy an automobile or some other durable consumer good can usually be postponed or given up with even less social or physical pain. The relative willingness with which respondents will allow circumstances to alter their plans concerning dresses and cars, for example, causes statements of intention regarding them to be less often fulfilled than statements concerning activities of more fundamental importance and of less vulnerability to shifting circumstance.

It is partly for the reason that people's plans to buy consumer durable goods can be altered with so little pain and are altered for such a variety of reasons that the volume of purchase of durable goods is difficult to predict and intentions to buy them are the subject of so many marketing research investigations. The success of these studies to date is not great enough to warrant optimism regarding the achievement of solutions to the difficult problems involved. Indeed, there is reason to believe that forecasts of behavior of this type will never be consistently accurate so long as the decisions being investigated are subject to influence by so many unpredictable circumstances.

Perhaps a more fruitful alternative to securing statements of intention is the measurement of consumer preferences which can indicate the distribution of sales among competing firms within an industry. It is probably much easier to secure accurate data on consumers' rankings of different brands of automobiles than it is to predict the total number of automobiles of each kind that will be purchased. Total purchases of automobiles can probably best be estimated by statistical analysis of historical data available to the automobile industry.¹

14.3 Accuracy and Ambiguity

In the preceding paragraphs we have discussed the factors which influence the predictive value of statements of intention, *assuming* them to be accurate and unambiguous. Now it is time to drop that assumption in order to discuss the probabilities and possibilities of getting accurate and unambiguous statements as to what consumers really intend.

As with statements concerning past behavior, so with statements concerning intentions, intentional deceit is often a source of serious inaccuracy when respondents feel that material gains or losses or unpleasant emotional reactions will be associated with particular responses. The discussion of this subject in the preceding chapter is relevant at this point as well. In briefest summary, we reiterate that such expectations should be minimized.

¹ Methods for measuring preferences are discussed in the next chapter. For an excellent example of a statistical analysis of the aggregate demand for automobiles see *The Dynamics of Automobile Demand*, General Motors Corporation, New York, 1939.

Aside from intentional deceit, it is meaningless to talk of inaccuracy in statements of intention. For a statement of intention is said to be accurate if the respondent making it intends it to reflect truly his state of mind. Therefore a desire to speak truly about one's own intentions necessarily creates an accurate statement. Unfortunately, however, the desire to reveal one's intentions and the ability to do so unambiguously are quite different things. Achieving unambiguity is a difficult problem which we now consider.

As has been seen earlier, the statement of intention, to be unambiguous, must include a statement of the circumstances upon which the decision concerning an activity is contingent. Since these circumstances are frequently numerous and occasionally have complicated interrelationships, a full statement regarding them frequently taxes respondents' capacity for organized expression. It is much easier to say, "I now own a Brand X refrigerator which has a freezing compartment for freezing and storing frozen foods" than it is to say "I expect to buy a home freezer within the next few months if I can get a 4-cubic-foot model for less than \$200, if its outside dimensions aren't too great for it to fit in the kitchen between the refrigerator and the sink—unless, of course, refrigerators with freezing compartments big enough for our needs become available for much less money."

When information is sought about past behavior, very often respondents can be relieved of the necessity for coherent self-expression by providing check lists which include almost all possible answers to the questions being asked. The number of possible contingent circumstances for statements of intention makes it highly impracticable in many cases to provide check lists which include a major proportion of the possible combinations of circumstances. The researcher is thus often left with the more complicated and less routine task of aiding the respondent in making long and "iffy" statements.

The researcher, in helping relatively inarticulate respondents—whether the incompetent or the lazy or the bored—to state the important contingent circumstances, not only overcomes ambiguity from this cause but also may diminish ambiguity from another source. Ambiguity exists not only because people cannot express the relevant contingent circumstances but also because in the test situation they may not think of them. The researcher theoretically can diminish ambiguity from both causes by suggesting in the test situation the circumstances which will probably be relevant for a decision in the criterion situation. Thus if respondents were being asked to indicate their intentions regarding their purchase of alternative dress styles, the researcher could convey to the respondent information concerning the price of the dress, the size range, the material, available colors, etc. Or if persons were being questioned concerning their intentions with regard to

the purchase of an automobile, ambiguity could probably be diminished if the researcher suggested the probable prices for new cars, the terms upon which credit could be secured, a range of trade-in values of old cars, expected changes in the design of automobiles in the forthcoming annual models, etc. The decision as to what circumstances should be suggested by the researcher might be determined by a careful examination of his own motives in purchasing, by experience in surveys of buying intentions, or by pretesting a sample of respondents in order to determine what factors were currently considered of importance.

Although ambiguity can theoretically be diminished by the remedies just suggested, the same activity which would serve to diminish ambiguity would make the interviews very difficult to administer, relatively costly, and possibly subject to influences that would diminish their predictive value. By trying to suggest to the respondent all the circumstances that might be important in making the decision in question, the interviewer might cause the respondent in the test situation to weigh factors which neither had been previously considered nor would be considered in the same way in the criterion situation. In addition to the difficulties in communication created by the existence of contingent circumstances, all the other difficulties which were discussed in the previous chapter also exist. This discussion will not be repeated even though it is relevant in this chapter as well.

CHAPTER 15

SECURING INFORMATION ON ATTITUDES

15.1 Introduction

Sometimes psychologists and students of public opinion distinguish the words "attitude" and "opinion" by defining the former as a tendency to act in a specified way and the latter as the verbal expression of an attitude. The words are often used synonymously, however, and we use them synonymously in this book to refer either to a tendency to act or the verbal expression of such a tendency. The specific meaning should be clear from the context. Information on attitudes, like other sorts of information, is useful in marketing research to the extent that it facilitates the accurate forecasting of sales.

Information on attitudes is of two sorts. The first involves securing the judgment of individuals regarding competing alternatives. This sort of judgment is secured by asking respondents to state which of several packages, product designs, flavors, advertisements, etc., is preferred. Such judgments are sought in order to be able to rank the alternatives according to the degree to which they are approved by respondents.

The second sort of information involves securing statements of approval or disapproval toward a single psychological stimulus. This type is useful in marketing research when the effect of a marketing policy or practice upon consumers must be determined. The effectiveness of an advertising campaign might be judged by such information, which permits ranking individual respondents according to the degree of approval which each expresses toward the psychological stimulus being investigated.

Whether information on attitudes is used to rank stimuli relative to each other (as when respondents are asked to express their attitudes toward competing package or product designs) or to rank individuals with regard to their reactions to a single stimulus (as when the attitudes of a population toward a product are determined prior to and after an advertising campaign), the usefulness of the information is generally restricted to indicating the direction of change in sales that can be expected or that has been achieved. It is usually impossible to use information on attitudes to forecast the absolute volume of sales. Seymour Banks attempted to predict absolute sales on the basis of attitude measurements and found the practical difficulties which a priori speculation would have led one to ex-

pect.¹ On the other hand, Bank's attempt to forecast relative rates of sale was rather successful.²

15.2 Getting Accurate and Unambiguous Expressions of Attitude

Expressions of attitude are not useful in even the limited way indicated above unless they are accurate and unambiguous. The problems of securing accurate and unambiguous expressions are discussed in this section.

1. Techniques for obtaining attitudes. The problem of devising question and response forms which yield accurate and unambiguous information concerning brands of soap purchased, radio programs heard, and other types of past behavior is relatively simple although not at all so easy as the neophyte might guess. More complex is the problem of devising a mechanism of communication to secure information on the intentions of respondents. The additional complexity is introduced because ambiguity can be avoided only if respondents' statements indicate what contingencies are assumed. This is troublesome, to be sure, but understanding the nature of the problem is not difficult.

Understanding the problem of devising an adequate mechanism of communication for securing information on attitudes is more difficult. Different methods are used for getting information concerning attitudes toward a single stimulus and information on the judgment of individuals regarding competing alternatives, and we consider these techniques separately. We begin by considering the mechanisms of communication available for classifying and ranking individuals according to their attitudes toward a single stimulus.

1. The Single-question Technique. The simplest and by far the most widely used method of determining attitudes of this type involves the "single-question" technique. This technique is one which requires but a single question to secure from respondents all the information that is considered necessary to judge their attitudes on any issue. Of course, almost never is a survey conducted which uses questionnaires having but a single question. Many questions are usually used in order to get attitudes on many different issues or, in some cases, on many aspects of the same issue and on biographical facts concerning the respondent.

Perhaps this technique and its limitations can best be understood by considering the problem of finding out attitudes toward the quality of radio programs. To simplify the example, assume that there is no problem of

¹ Seymour Banks, *Some Applications of Psychological Preference and Attitudes to Marketing Research*, doctoral dissertation, University of Chicago, 1949.

² Seymour Banks, "The Prediction of Dress Purchases for a Mail-order House," *The Journal of Business of the University of Chicago*, Vol. XXIII, No. 1, pp. 48-57, January, 1950.

unwillingness—that respondents are perfectly willing to tell us whatever we want to know, in so far as they are able. Assume also that there is no ambiguity arising in the process of communication—that all the words used are understood in the same way by all. Each respondent might express his attitude in many ways: "Radio programs bore me"; "The radio is very good company when I am alone in the evening"; "It's a cheap and altogether wonderful form of entertainment"; "Programs create bad habits in children"; etc. Respondents could thus provide a collection of words describing attitudes toward radio.

In such form, however, information on attitudes toward radio would be useless for almost any conceivable purpose, predictive or descriptive. It is true that one could read through all the attitudes and get an impression about attitudes toward radio, but such an impression would be vague and difficult to communicate precisely. Further, the lack of precision would make it difficult, if not impossible, to detect small but possibly important changes in attitudes. It is, therefore, necessary to devise some way to summarize briefly and precisely the data on hand. The separate bits of information must be placed into precisely defined categories or classifications.

The problem of classification is a critical and difficult one. Suppose that all attitudes toward radio are to be classified into one of five pigeonholes designated as follows: "very favorable," "favorable," "neutral," "unfavorable," and "very unfavorable." (The particular names for the pigeonholes are used only for illustration.) We shall assume—what is not self-evident—that people's attitudes can be classified and that it is possible to compare attitudes of different people. Now consider the pigeonholes used above for classification, to determine their usefulness for making such a comparison. (Not all schemes of classification are meaningful.) How would the attitudes toward radio, previously listed in this chapter, be classified by this scheme? The first attitude was "Radio programs bore me." This would apparently fit into the "unfavorable" category. The fourth comment, "Programs create bad habits in children," might also fit this category. Since these two comments have been placed in the same category, they must have something in common. They do, because in each case an individual indicates something about radio that he apparently does not like. In one case the attitude refers to the effect of radio on the respondent, while in the other it refers to the effect of radio on other people (children). The person who said that radio bored him might feel that radio was a wonderful thing for children, while the person who thought that radio creates bad habits in children might enjoy radio personally.

Hence the preliminary classification of both attitudes into the same pigeonhole—"unfavorable"—is in some ways not satisfactory. It indi-

cates as being the same, attitudes which differ importantly. In order to avoid this, one could adopt two criteria of classification, one referring to the effect of radio upon the respondents themselves, the other referring to its effect upon children. But even two criteria might not be sufficient. One person might think radio had a bad effect on children's habits; another might think that it had a bad effect on their moral character; a third, that it stultified intellectual development. All three attitudes would be classified "unfavorable, with respect to radio's effect on children," yet again one would be placing quite different attitudes—or rather attitudes about different things—in a single category or pigeonhole.

Hence categorization might have undesirable consequences. Suppose, for example, that someone's attitude were labeled "unfavorable." One would not know whether the attitude was unfavorable to communists, breakfast cereals, sin, radio, or chromium-plated mufflers for automobiles. It would help some to be told that the person was "unfavorable to radio," because one could at least rule out unfavorableness to Communists, breakfast cereals, sin, or chromium-plated mufflers for automobiles. Then those "unfavorable to radio" could be reclassified into finer categories. The process of reclassification could be continued almost indefinitely. The process of subdivision is halted at the point at which further specificity would make no difference in the practical action to be taken on the basis of the data. Hence if an investigator wanted only to obtain people's feelings about radio—whether on balance each person liked it or didn't—he might classify attitudes as first suggested—very favorable, favorable, neutral, etc. If he wanted to find out how people felt about radio advertising rather than radio as a whole, he would not be interested in what people thought of the dramatic qualities of soap operas. Hence it is necessary to define the subject about which attitudes are to be obtained. The more specific this subject, the easier it is to construct a meaningful classification.

Assume, as an illustration, that this rule has been successfully followed and that the subject has been narrowed down to the following issue: How do people feel about the length of the commercials on the Rhoda Lyfe soap opera? The classification is as follows: "much too long"; "too long"; "about right"; "too short"; "much too short." That is, people can be roughly ranked relative to each other according to the degree of favorableness of their attitudes by considering the category in which each person belongs. Such a ranking of attitudes would not, however, imply anything about "how much" one attitude differs from another. Although it is true that one can devise schemes which will assign numbers to each person's attitudes, these schemes are essentially arbitrary. Someone else could devise a different scheme which would assign different arbitrary numbers to each person's attitudes. Although both schemes might *rank* attitudes

correctly, the units employed might be completely different. In other words, one may succeed in ordinal measurement (the ordering or ranking of data), but one cannot (as yet, at least) succeed in cardinal measurement (definition of observations as meaningful absolute magnitudes).

An alternative form of response which *apparently* solves the classification problem involves the use of a line, or linear scale, ranging from one extreme of the attitude being measured to the other, upon which respondents are required to indicate the position corresponding to their own attitudes. Sometimes the line is subdivided by the use of a point of indifference in the middle. At other times the line may be divided into 7, 10, 12, or some other number of numbered segments. This response form is ambiguous to the extent that the units of the linear scale have different meaning for different respondents. It is almost certainly true that Smith, who locates himself halfway between indifference and approval, has a somewhat different attitude from Jones, who locates himself at the same point on the line. The need for a correct definition of the issue involved remains as before.

The difficulties of classification are not the only difficulties in the techniques considered thus far. Much of the empirical work in opinion and attitude research in recent years has been devoted to demonstrating that statistically significant and important practical differences in responses can be achieved by altering the wording of a question, its position within the questionnaire, and the method by which the question is presented and the responses recorded. A question seeking attitudes toward a package may achieve different results if, for example, the package is described as holding a "full quart" rather than "one quart." The degree of approval which consumers indicate in answer to affirmative questions is often different from that which is indicated in response to negative questions. Similar differences can be demonstrated when the questions are personal as compared with impersonal.

Although differences in results obtained with the variations in wording and in question form have been impressive, it is difficult to say which of the very large number of possible variants in question form is "the best." Some have contended that the best form is the one which indicates attitudes with the greatest *accuracy*. The term "accuracy" in this context refers to the correspondence between the meaning of the verbal statement of the respondent and the state of mind of the respondent to which the statement refers. It seems impossible to use such a definition. Perhaps each different question form gives an equally accurate indication of different and closely related attitudes. If this is true, the "best" question form is that which produces information having the best predictive value of the kind of behavior being investigated. No rigorous procedure

has been devised for selecting the "best" variation, but some comments seem worth making.

First, leading questions should not be used. Since a "leading" question is defined as a question which produces biased (inaccurate) answers, there can be no argument about the undesirability of such questions.¹ It is not hard to decide that leading questions are undesirable, but it is often difficult to recognize one.

Lyndon Brown has suggested some types of questions which experience has shown to be leading.² The first type is leading because answers to it are particularly subject to an affirmative bias.³ Most questions which are answered "yes" or "no" are subject to this bias. Examples of such leading questions are, "Does this advertisement appeal to you?" or "Do you like corn flakes?" (The possibility of a negative answer is not mentioned explicitly and is therefore relatively easy to overlook.)

A second suggestion made by Brown is that questions which include words such as "always," "ever," "never," "all," "none," etc. are leading in that uniformity of action or attitude is seldom great enough to permit an accurate affirmative answer to questions using such words. The bias resulting from such questions exists because respondents who regularly listen to Jack Benny frequently respond that they "always" listen to Jack Benny. People who very seldom attend the movies frequently answer "no" to a question, "Do you ever attend the movies?" This sort of bias is one reason for avoiding such questions, but a more important reason is that answers to such questions, even if accurate, seldom provide any useful information.

Brown's third suggestion concerns the use of questions involving check lists. The use of such lists frequently results in an affirmative bias, which is a special case of bias of the first type.

A fourth bias is associated with the order in which alternatives are presented. The sequence of preceding questions may condition responses to subsequent questions.

A fifth type of leading question is one which reveals the sponsor of the survey. Respondents have shown a tendency to give answers which they think will be pleasing to the sponsor of the survey, and this tendency has resulted in an overstatement of favorable attitudes toward the sponsor's product.

A sixth type is one which contains more words that are capable of

¹ Usually a leading question is defined as one which suggests an answer. This, of course, is but another way of saying that the question creates biased results.

² Lyndon O. Brown, *Market Research and Analysis*, pp. 232-235, The Ronald Press Company, New York, 1947.

³ For discussion of this bias see The Recognition Method, under section 13.4.

creating emotions or conveying information than are essential to the purpose of the question. An example of such a question would be "What is your opinion of Shiny Shampoo as advertised by Scott Elledge?" The person's response may reflect his attitude toward Elledge rather than toward Shiny Shampoo.

The sad thing is that even if one can avoid all questions which are identified as leading on the basis of Brown's suggestions, there are still many possible variations in question form which are capable of producing statistically significant and important practical difference in responses. What can be done about this? Two solutions exist if the researcher clings to the use of the single-question technique.

The first solution entails the elimination of all questions which are leading according to a few rules of thumb such as those suggested above. The remaining variations in question form are then used together in the actual survey, with the final results being some average of the results from using each of the variations. In some cases some interviewers are given one set of questions, other interviewers are given other sets. In other cases each interviewer uses all sets of questions, rotating them systematically from interview to interview. There can be no assurance that the average of results from the use of various questions will not be biased, because there is no assurance that the individual questions will not yield biased results. In any event, practical difficulties usually preclude the use of this remedy.

An alternative solution to this problem involves the use of highly trained fieldworkers and very simple direct questions which the respondent answers in his own words. People who rely upon such devices believe that intelligent, highly trained individuals can be sure that the meaning of their question is correctly understood and that the response which the original question and *subsequent probing* evoke can be also understood and evaluated.

As a consequence of the necessity of evaluating the nonstereotyped responses, a great burden is placed either upon the fieldworker or upon coders. In some cases it is the fieldworker who evaluates the responses of the persons being interviewed. In other cases the responses of the interviewed persons are recorded by the interviewer (either verbatim or in edited form) and are interpreted by coders so as to place the various respondents in the categories which have been established.

It should be pointed out that a method which requires that the fieldworker record an edited version of the respondent's statement and which leaves the interpretation of these edited statements to coders in reality splits the responsibility for the interpretation between the fieldworker and the coder. This is true because the editing process itself involves inter-

pretation. When the appraisal of the respondent's statement is done at the time of the interview by the fieldworkers, the accuracy of interpretation is aided because the fieldworker can consider, in addition to the actual words of the respondent, his mannerism, inflections, and other intangibles which may be valuable in discerning the respondents' attitude. On the other hand, those who favor the verbatim recording of respondent's statements point out that the interpretation of such statements by a single individual—*e.g.*, a coder—eliminates the differences in interpretation which would be caused if the interpretation were by interviewers with different backgrounds, intelligences, and prejudices.

Another defect of the single-question technique is that it renders responses particularly susceptible to influence by interviewers. Work done at the National Opinion Research Center has indicated that the prejudices or attitudes of interviewers have been significantly correlated with the responses secured in surveys attempting to determine public opinion on political and social issues.¹ For example, interviewers who were Republicans found a higher degree of approval for points of view advocated by Republicans than did interviewers who were Democrats. (The attitudes of the interviewer were determined prior to the beginning of the survey by having the interviewers fill out the questionnaires which were to be used in the field.) Interviewers can bias results by the way in which they ask questions, by their preparatory remarks, by their remarks of elaboration or explanation, or by their recording of answers. An attempt to control this sort of bias by requiring that interviewers adhere strictly to a printed form denies the interviewers the flexibility which is often required if unambiguous communication is to be established. Furthermore, even strict adherence to a printed form permits the injection of bias through mannerisms and inflections.

Professor Cantril suggests that such bias on the part of individual interviewers can be offset for the survey as a whole by selecting equal numbers of interviewers with opposing points of view. Others have sought the answer in more thorough interviewer training. Both points of view have merits and limitations. It is true, however, that no matter what is done, the single-question technique is more subject to this bias than alternative techniques. It is probably true that the danger of this sort of bias is much greater when attitudes are being sought upon important political and social issues than when attitudes are being sought concerning packages, advertising copy, or product attributes. Nevertheless, even in marketing research such bias can be important.

¹ Don Cahalan, Valerie Tamulonis, and Helen W. Verner, "Interviewer Bias Involved in Certain Types of Opinion Survey Questions," *International Journal of Attitude and Opinion Research*, Vol. 1, No. 1, pp. 63-77, March, 1947.

A summary of the preceding comments on the single-question technique would reveal that it is subject to important limitations. As it is most frequently used, it succeeds in placing individuals in only a relatively small number of categories and fails to reveal the often great differences in attitudes which exist within categories. Attempts to overcome this deficiency by use of self-rating, which is either verbal or involves the use of linear scales, is inefficient because of the ambiguity introduced by the fact that different individuals mean different things when they use abstract words and attribute different meanings to the units of the scale. Interviewer ratings by these devices are subject to the same basic criticism.

The second serious deficiency in the single-question technique lies in the fact that variations in wording, form, and order produce variations in results, and as yet no one has been able to determine which, if any, of the possible variants yield unbiased results. The conventional rules of thumb concerning leading questions do not solve this difficulty. The best solution which has been found involves the use of simple direct questions, highly trained interviewers, and nondirected response forms. But even this solution has its flaws, as has been seen.

Finally, there are some dangers that interviewers will interject bias into the survey results by the way in which they conduct the interview and ask the questions. Despite these drawbacks, the single-question technique has been of great practical usefulness. In some applications, however, the drawbacks are sufficiently serious to warrant the search for an alternative.

2. Ranking Individuals According to Attitudes toward a Single Stimulus. An alternative to the use of the single-question technique is the use of attitude scales. The construction of these scales typically involves the use of a battery of questions designed to determine each respondent's attitude, and in this respect attitude scales contrast with the single-question technique. In a sense, the single-question technique also produces a scale in that respondents are assigned to categories which can be ranked. The multiple-question techniques, however, make possible a more detailed ranking.

The details of the multiple-question scaling techniques are complicated and difficult, and anyone seriously interested in the subject can refer to the references that we cite. It is important in studying these methods not to lose sight of the basic assumption on which any attitude measurement must be based: the attitudes of different people can be compared. It is equally important to remember that the most that has been done in "quantifying" attitudes is to rank them, to say, for example, that one person is "more" favorable than another. One cannot hope to say "how much" more favorable.

The scaling techniques typically attach numbers to each person's atti-

tude, but actually, there is always an arbitrary stage in the process of obtaining these numbers in the techniques now used. The most one can hope is that the arbitrariness does not affect the ranking of attitudes; it always will affect the quantitative relationships among the numbers which are used. That is, one might attach the number 1 to one person's attitude, 2 to another's, and 3 to a third's. At most these numbers mean that the three attitudes can be ranked. The numbers do not mean that the second person's attitude is twice the first's, and they don't mean that the distance between the attitudes of the first and second and the second and third is equal. One might equally well have assigned the numbers 1, 3, and 7 to these attitudes, since these numbers would lead to the same ranking of attitudes.

It is theoretically possible that some degree of quantification of attitudes may be developed. One approach is based on the "theory of games" and has been expounded by Friedman and Savage.¹ Another has been suggested by Clyde Coombs of the University of Michigan.² These two methods are new and have been tried only in a very preliminary way, so we do not discuss them further.

The pioneer work on the construction of attitude scales was by Professor L. L. Thurstone in 1928.³ The standard methods of construction of attitude scales have much in common, and for our purposes it will not be necessary to consider exhaustively the procedures associated with each. We attempt only to outline the major steps associated with some of the most widely used methods; for a detailed picture the reader will have to refer to more specialized sources. The exposition of the Thurstone method will be more detailed than the others. This emphasis is not attributable to any deficiency of the other methods but rather to a desire to go into enough detail on one of the methods to indicate clearly the general nature of the scaling techniques and to prevent any possible confusion with approaches which involve multiple questions but which do not attempt to derive scales. We begin with the major steps in the Thurstone method.

1. It is extremely important that the attitude be rigorously defined so that all statements from respondents can be interpreted. The objective in measuring attitudes is to rank them—to be able to say, for example, "Jones approves of the present political administration's policy on arm prices more than Smith approves, and Smith approves more than White, who approves more than Black, etc." It is this ordering of individuals

¹ Milton Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, Vol. LXI, No. 4, pp. 279-304, August, 1948.

² Unpublished, privately circulated manuscript.

³ L. L. Thurstone, "Attitudes Can be Measured," *American Journal of Sociology*, Vol. 33, No. 4.

from one extreme of approval to another which is sought. But in order to rank individuals, one must be sure that their statements concern but a single attitude, and therefore the rigorous specification of the attitude is required.

2. The second step involves the securing of a large number of statements of different opinions concerning the stimulus being investigated. Statements of opinion can be secured by asking various groups to write them, by examining newspapers, by making them up, etc. An original list of between 100 and 200 statements should be secured in this way. The sources of these statements do not matter. The usefulness of an attitude scale is independent of the sources of the statements which provide the basis of that measurement.

3. The third step involves cutting and editing this initial list of statements. Thurstone suggests several criteria to be used in this process. First, brief statements are to be preferred here for the same reasons that they ordinarily are preferred in other kinds of investigations. Second, the statements should be capable of being understood by the respondents being surveyed. This requirement is already familiar and suggests the desirability of statements which are positive, brief, and employ familiar words. Thurstone has an objective criterion of ambiguity which cannot be applied at this stage but which can serve as an ultimate check. Third, each statement must be such that agreement or disagreement with it by a respondent will tell something about the respondent's attitude. Statements which do not meet this requirement are irrelevant; Thurstone has devised an objective criterion of irrelevancy. This criterion, however, cannot be applied at this stage of the process; it is therefore necessary to do some preliminary culling on the basis of subjective judgment.

4. The fourth step entails the sorting of the edited statements into various groups according to the degree of approval or disapproval which they represent. Each statement is typed on an individual card. A complete set of cards is given to each member of a group of two or three hundred individuals. (This number seems unfeasible for almost all marketing research problems, so in practice it would almost certainly be reduced.)

Each person is then requested to separate these cards into piles (11 were used originally by Thurstone, although other numbers could be and have been used). The statements placed in the first pile represent the most extreme degree of disapproval; those in the fifth pile, indifference and neutrality; those in the eleventh pile, extreme approval. The intermediate piles represent gradations by intervals which *seem* ("appear") to be equal to the persons doing the judging. That is, the least favorable statement in the ninth pile is supposed to be as much more favorable than the most favorable statement in the eighth pile as the least favorable statement in

the eighth pile is more favorable than the most favorable statement in the seventh pile.

The usefulness of an attitude scale does *not* depend upon the attitude of the persons who judge the degree of favorableness of these statements. These judges need not be carefully selected; their biases are unimportant. The validity of attitude scales rests upon this independence. This independence was originally assumed by Thurstone. It has been subject to experimental tests for more than twenty years, and there is no convincing evidence that his original assumption was not warranted.

5. The fifth step involves the calculation of the scale values of the sorted statements. The scale is represented by a line divided into eleven equal segments corresponding to the eleven piles into which the statements were sorted. The first segment thus represents the least favorable attitude, and the eleventh the most favorable. Beginning with pile number 1, the cumulative frequency with which each statement is judged in each pile is then plotted. The scale value for each statement is determined by dropping a vertical line to the scale from the intersection of the lines representing the cumulative percentages with the line representing the 50 per cent level (see Fig. 1). For example, it is seen from Fig. 1 that 50 per cent of the judges placed statement A in pile 6 or below, and 50 per cent placed it above pile 6. The scale value of A is therefore 6. The scale value is determined for each statement in this manner. In other words, the median judgment of the judges is used to determine the scale value of each statement. (The complications which may arise in this procedure are not discussed here. They are amply discussed in the references for this chapter in the bibliography.)

6. The next step—the sixth—involves the elimination of some statements on the basis of a criterion of ambiguity. Ambiguity is measured by uniformity of classification for each statement. If, for example, all 200 judges put a particular statement in category 6, the ambiguity would be very small. On the other hand, if a statement were placed in category 1 by 18 judges, in category 2 by 19, and so on until perhaps 18 judges put the statement in category 11, the statement would be very ambiguous. Arbitrary criteria have been developed to determine “excessive” ambiguity.

7. The next step, which Thurstone suggests but which is frequently omitted in practice, is the elimination of statements which are judged irrelevant by the use of an objective criterion. The omission of this step is dangerous, because it is the only means which permits the elimination of statements which do not refer to the single attitude of interest, and such an elimination is of fundamental importance in any scaling—ranking—of attitudes. A statement is termed irrelevant if persons who endorse it also endorse statements with very different scale values.

To illustrate, we give one example. Suppose a statement had a scale value of 5, but that the people who agreed with this statement tended to agree also with statements having a scale value of 1 or 11—that is, the extremes on the scale. The interpretation is that since this statement—which had been assigned a value of 5, or indifference—appealed to people of widely differing attitudes, it really is irrelevant to the attitude being studied. Again, quantitative criteria have been devised for drawing the line between acceptable and nonacceptable statements. The reason that

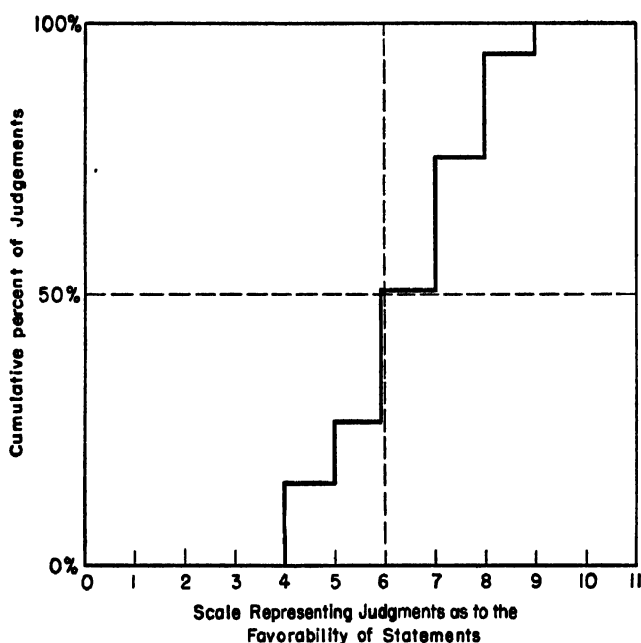


FIG. 1. Determination of the scale value of statement A (method of equal-appearing intervals).

this technique is often not applied in practice is that it adds another time-consuming and expensive step to the procedure.

8. The next step, after the elimination of irrelevant statements, involves the selection of about 20 of the remaining statements which have scale values covering the entire range of the scale and have approximately equal intervals between them.

9. Now, after eight preliminary steps, the final list of statements may be presented to the sample of respondents being surveyed. Each respondent is required to indicate which statements he endorses and which statements he rejects. The score for each respondent is an average (mean or median) of the scale values of all the statements he endorses. It may also

be of interest to determine the variability of the scale values of the statements endorsed by each respondent.

It will be recalled that responses to single questions could be altered radically by altering the wording of questions and their order in the questionnaire. It is interesting to inquire, therefore, as to the extent to which attitude measurements based upon Thurstone's attitude scales are affected by alterations in the wording of statements which do not alter their literal significance and by altering the sequence of statements. Fortunately there are ways of determining the degree of dependence of the results of attitude scales upon such variations in wording and order. Theoretically a person's score on an attitude scale is determined solely by the scale values of the statements which he is required to endorse or reject. Therefore if different sets of statements, each of which contained statements having the same scale values, were to be administered to the same group of respondents, it would be possible to test the extent to which individual scores were determined solely by the scale values of the statements used. Such testing of these scales has been carried on, and the correlation coefficients between the scores achieved by using the alternative sets of statements have typically been between .70 and .90 and most frequently have been around .80. It is probably a wise precaution to conduct such tests whenever a new scale is constructed. The evidence is fairly convincing that the results of attitude scales of this type are affected less by the wording and order of questions than are the results based upon single-question techniques.

It also seems probable that the measurements based upon the method of equal-appearing intervals are less subject to distortion because of the mannerisms or inflections of fieldworkers than are results based upon single-question techniques. No empirical testing of the susceptibility of these two techniques to this type of bias has been done and none seems possible, but it seems reasonable to believe that the presentation of the list of 20 statements to a respondent with the request that he endorse or reject each statement provides less opportunity for the injection of interviewer bias than does the asking of an individual question which alone determines the respondent's attitude.

A third advantage of the scaling technique is the fact that the placing of individuals along a scale provides much more complete and useful information than does the placing of individuals into a few categories. Consider the frequency distributions given below, each of which shows the relative frequency of occurrence of scores (Fig. 2) of a group of individuals. Each score represents the attitude of an individual toward a particular psychological stimulus, which in this case will be considered to be a purchasable object. The frequency distribution labeled A represents fre-

quency of various attitudes toward object A, and the other frequency distribution represents attitudes toward other purchasable objects (B and C). From a measurement of attitudes of this sort, it is possible under some circumstances (*i.e.*, when certain assumptions are met) to predict the relative frequency of purchase of objects A, B, and C. The way in which these attitude measurements can be used for this purpose was originally explained by Professor Thurstone in 1945.

The method of prediction of choice developed by Thurstone is not difficult to implement, but a thorough explanation of the method would involve statistical concepts that are not assumed as necessary background for this book. The interested reader is strongly urged to study Thurstone's original article.¹

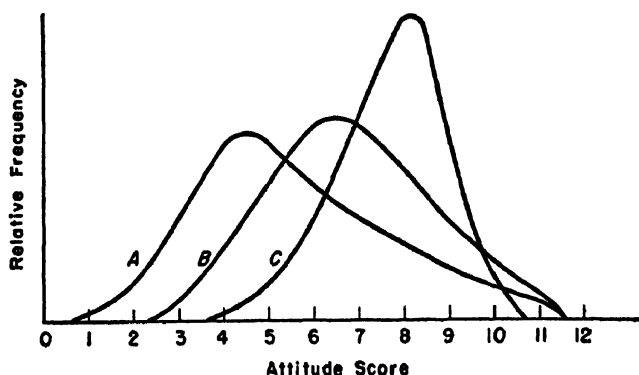


FIG. 2. A frequency distribution of attitudes.

The real usefulness in marketing of such elaborate methods of attitude measurement and analysis is their sensitivity in detecting shifts of attitudes. Suppose, for example, that one wanted to appraise the influence of an advertising campaign on attitudes toward a product. One method of doing this would be the single-question technique, but only three or five gradations of attitudes might be obtained with this method. An individual might be classified "unfavorable" toward a certain product at the start of an advertising campaign. At the end of the campaign his attitude may have softened somewhat, but not enough to move him out of the unfavorable category. The Thurstone method, obtaining, in effect, a much finer gradation of attitudes, would detect this shift.

In deciding whether or not to use this method in a practical problem, its advantages must be weighed against its great cost. For the same expenditure, for example, one might conceivably be able to interview a much

¹ L. L. Thurstone, "The Prediction of Choice," *Psychometrika*, Vol. 10, No. 4, pp. 237-253, December, 1945.

larger sample of consumers with the single-question method, since a relatively small part of the total cost would go into the preliminaries necessary under the Thurstone method. The reduction in sampling error with the single-question technique might, or might not, more than offset the advantages in accurate measurement of response with the Thurstone or other multiple-question techniques.

There are other methods of scaling, and we shall discuss them very briefly, again urging the interested reader to refer to the original articles. The reason for brevity in discussing these techniques, it should be repeated, is not a desire to deemphasize them, but rather a desire to discuss one scaling technique (the Thurstone method) in enough detail to give insight into scaling methods in general.

The second method to be discussed is the "Method of Summated Ratings" developed by Rensis Likert in 1932.¹ This method, like that of Thurstone, involves as an initial step the collection of a large number of statements (*e.g.*, 200) from which attitudes toward a single variable can be inferred and which must be edited in accordance with the criteria suggested above in the discussion of Thurstone's methods. These statements are presented to a sample of respondents who are required to indicate the extent to which they agree or disagree with each statement. Usually five categories of responses are provided, although more or fewer categories have been and can be used. Typically these categories are "strongly agree" (approve), "agree" (approve), "undecided," "disagree" (disapprove), and "strongly disagree" (disapprove). Weights are arbitrarily assigned to the various categories of responses according to the degree of favorableness that the category represents. For example, when five categories are used, the least favorable might be given a weight of 1 and the most favorable a weight of 5. Each respondent's score is equal to the sum of the weights given to the responses to the statements.

Up to this point the scale has not been established; respondents have simply been asked to categorize their reactions to each question in a list of almost 200 statements. The next step in establishing a scale to be used in the main research problem involves selecting from this large list of 200 statements those statements which discriminate best between respondents having favorable and unfavorable attitudes. In order to test the discrimination, the responses of those respondents having the highest and the lowest total scores are considered. Usually the highest 25 per cent and the lowest 25 per cent of the respondents, as judged by the total scores, are used for this purpose, but there is nothing sacred about these particular proportions. An analysis is then undertaken in order to select statements

¹ Rensis Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, New York, Vol. XXIII, No. 140, pp. 1-55, 1932.

for which those members of the highest group have given favorable reactions and for which members of the lowest group have given unfavorable reactions. Such statements are said to discriminate best between respondents with favorable attitudes and those with unfavorable attitudes. This process of analysis is called "item analysis."¹ About 25 statements—those which show the highest "resolving" or discriminating power—are actually used in the final survey. The final rating of individuals in the survey is secured by summing the weights attached to their responses to each of these 25 statements.

This technique is somewhat simpler than the method of equal-appearing intervals, and it does not involve the assignment of scale values to particular statements—a process which has been questioned by some psychometricians. However, it does involve the assignment of arbitrary scale values (e.g., 1, 2, 3, 4, and 5) to *reactions to statements*; hence the arbitrariness inherent in these scaling methods is not eliminated. In spite of such differences, however, results of using the two methods have been found to be about the same. That is, the ranking of the same individuals is about the same, and the correlations between the responses of the matched forms of each method have been about equally high.

A third technique in current use is that developed by Professor Louis Guttman and is known as "Scale Analysis." Whereas the Likert and the Thurstone methods involve the initial collection of a large number of statements from which a small number are selected to represent a scale, scale analysis starts with a small number of statements, usually 10 to 12, which are then tested to determine whether they are a scale. This list of statements is selected on the basis of intuitive judgment as to whether the statements really refer to the same attitude. The rest of the process consists in testing the soundness of this initial intuitive judgment.

The procedure will not be described in detail here. The main idea is logically similar to the Likert technique, except that instead of discarding statements which do not discriminate well between people who make high and low scores, this method modifies the method of classifying responses to the 10 or 12 statements which are used. In the end, the result is much the same as with the Likert technique. Objective criteria are set up for testing the soundness of the initial judgment. Final scores are calculated just as in the Likert method.

The Guttman technique has one feature not found in the Likert or Thurstone methods. All three permit a ranking of individuals according to attitude. The first two do not permit the classification of individuals

¹ For some of the specialized techniques used in carrying through this step the reader is referred to J. P. Guilford, *Psychometric Methods*, pp. 426-437, McGraw-Hill Book Company, Inc., New York, 1936.

into such categories as "favorable," "indifferent," "unfavorable," and the like. For example, it is possible to rank a group of individuals according to their attitudes toward a stimulus without knowing whether *all* might be favorable (or unfavorable). That is, qualitative labels cannot be attached. Guttman has provided for this in two ways; one by an analysis of his ordinary data and the other by use of a supplemental question on intensity of feeling.¹

There are other methods of scaling. Lazarsfeld's latent-structure theory, developed during World War II in the Research Branch of the Information and Educational Division of the War Department, is one example. The theory of these other methods is still evolving, and the volume of empirical work based on their use is small.²

The Thurstone, Likert, Guttman, and other methods of attitude measurement are all potentially useful for marketing research. If attitude scaling is to be used, it would seem that a large part of the decision as to which method to use would turn on such issues as convenience and cost.

3. *Ranking Stimuli According to Attitudes toward Them.* The methods of making scales discussed so far have been designed to rank individuals according to the degree of favorableness or unfavorableness of their attitudes toward a single stimulus such as a single advertisement, single product, etc. Such methods are useful when it is desired to determine the effect of a proposed marketing activity or decision upon consumer's attitudes and hence (potentially) upon their purchase behavior.

Another group of techniques, however, is better for the purpose of selecting the best of proposed alternatives. Whereas the Thurstone, Likert, Guttman, and other techniques result in placing individuals along a scale according to the attitudes of each with respect to a single stimulus, this second group of techniques results in placing competing stimuli along a scale according to the attitudes of individuals toward them. The ranking of stimuli in this way provides an efficient means for selecting that stimulus which can result in the largest volume of sales.

Probably the most widely used technique involves the selection by each respondent of the most preferred of a group of stimuli. It is not unusual for marketing researchers to request that each individual in a sample of respondents select the "best" or "the one which appeals to you most" of a group of proposed packages, flavors, advertising appeals, etc. Although

¹ The interested reader will find an excellent account of this technique, together with actual examples, in *Measurement of Consumer Interest*, edited by C. West Churchman, Russel L. Ackoff, and Murray Wax, pp. 64 ff., University of Pennsylvania Press, Philadelphia, 1947.

² For a rather complete discussion of the various theories of scaling see Chap. 1, Stouffer *et al.*, *Measurement and Prediction*, Vol. 4, *Studies in Social Psychology in World War II*, Princeton University Press, Princeton, N. J., 1950.

almost no rules have been laid down so far, it is probably justifiable to say that this technique should be used with reluctance when respondents must select from a group of more than two stimuli. That is, respondents should seldom be asked simply to select the best, or most preferred, from a group of three or more. If respondents do indicate only a first choice from a group of three or more, their responses are ambiguous.

Suppose, for example, that respondents were asked to select the most preferred of three packages, one of which consisted of red cellophane, one of colorless cellophane, and one of opaque waxed paper. Further, assume that 30 per cent expressed preference for the red cellophane, 30 per cent for the colorless cellophane, and 40 per cent for the opaque paper. A researcher would be running a serious risk of misinterpreting these results if he were to think that the best package to select was the one composed of opaque paper. It is possible that this is the correct interpretation, but it is also possible that 60 per cent of the people preferred cellophane to opaque paper and that they made a random choice between the red and the colorless cellophane. If this latter possibility were true, more than half of the respondents preferred either of the cellophane packages to the opaque package. In any case where respondents like equally well two or more alternatives, all of which are preferred to the remaining alternatives, this sort of misinterpretation is possible. The phenomenon which gives rise to this ambiguity is known as "vote splitting" and is the primary reason why this technique should be used with great caution.

One way of avoiding this ambiguity is to request respondents to rank the stimuli in order of their preference. Appropriately enough, this is known as the "rank order" method. After each respondent has ranked the stimuli in order of preference, it is necessary to combine the individual rankings into a combined ranking which represents the extent to which each stimulus is preferred by the group of respondents. There are two ways of doing this.

The first way of treating these data is the method of successive intervals. This method involves a consideration of both the median ranking of each stimulus by all respondents and the dispersion of ranking of each stimulus in constructing the scale for the stimuli.¹ The method of successive intervals results in arraying the stimuli along the scale which reflects the extent to which each stimulus is approved relative to each other stimulus. The ambiguity which was present when only first choices of respondents were recorded is avoided.

It is also possible to treat the data secured by the rank order method

¹ The technical procedure by which the method of successive intervals is used to determine scale values is explained in M. A. Saffir, "A Comparative Study of Scales Constructed by Three Psychophysical Methods," *Psychometrika*, Vol. II, No. 3, pp. 179-198, September, 1937.

according to the law of comparative judgment which was originally formulated by Professor Thurstone in 1927.¹ This law permits the determination of scale values for a group of stimuli whenever each stimulus is judged by each respondent relative to every other stimulus. The law of comparative judgment is therefore useful in constructing scales when the original data are collected either by the rank order method or by the method of paired comparisons, which will be discussed below. The technical procedures by which the law of comparative judgment is applied may be found in Thurstone's article cited above.

The method of paired comparisons is a second unambiguous method of collecting data reflecting the attitude of respondents toward a group of stimuli. It is a very simple method to understand but often a very time-consuming one to use. Each respondent is required to express a preference for each stimulus relative to every other stimulus. In other words, the respondents express a first choice for one stimulus out of each different possible pair of stimuli. It is essential that a preference be expressed for each pair. When a respondent claims that each stimulus of the pair is equally well liked, it is still important that one of the pair be chosen. Sometimes this choice is secured by presenting the pair again at a different time. At other times the respondent is requested to make a choice at random. If the selection is made at random, then the whole group of respondents will choose each of equally well-liked stimuli about 50 per cent of the time, on the average.

The major difficulty with the method of paired comparisons is that the number of individual judgments which each respondent must make is very large when several stimuli must be judged. The number of judgments which are required is given by the formula $\frac{n(n-1)}{2}$, where n equals the number of stimuli to be judged. Thus if 10 stimuli were to be judged, each respondent would be required to make 45 separate judgments. A task so tedious and time-consuming might result in a high rate of refusal from respondents or the expression of offhand judgments without proper reflection. One way around this obstacle is to give each respondent only a sample of all the possible pairs, yet plan the interviewing so that each possible pair of stimuli will be compared by an equal number of respondents. For example, if 10 stimuli are to be compared, one could divide the 45 possible pairs into three random samples of 15. But even this procedure involves costs and administrative difficulties, which usually make it unwise to use the method of paired comparisons when more than five or six stimuli must be compared.

¹ L. L. Thurstone, "A Law of Comparative Judgement," *Psychological Review*, Vol. XXXIV, No. 4, pp. 273-286, July, 1927.

The rank order method, discussed above, is a close substitute for the method of paired comparisons in that the method of rank order permits the investigator to construct a table of paired comparisons. A ranking of all stimuli permits the researcher to see how each respondent would (if he were consistent) judge each stimulus relative to every other stimulus, and this is all that the method of paired comparisons accomplishes.

Finally, it is also possible to use the method of successive intervals in a different way than before.¹ Each respondent assigns each stimulus to one of a small number of categories. A typical labeling of these categories would be "dislike very much," "dislike," "indifferent," "like," "like very much." More categories could be used, and other descriptive phrases could be used to designate them. It is essential, however, that the descriptive phrases represent judgments ranging from the least to the most favorable. Data secured in this way can be processed by the method of successive intervals² to provide scale values for all the stimuli judged. This method requires even less time for respondents than does the method of rank order, and obviously much less time than the method of paired comparison. Just as the use of rank order is to be preferred to paired comparisons when the number of stimuli exceeds five or six, so the method of successive intervals is to be preferred to the method of rank order when the number of stimuli to be judged exceeds perhaps 10 or 12. It is obviously much easier to assign 15 or 20 stimuli to one of five categories than it is to rank them in order of merit.

So far these three methods of evaluating stimuli—rank order, paired comparison, and successive intervals—have been judged solely with reference to the convenience of using them. And convenience is probably the only criterion besides those already mentioned for preferring one of these methods to the others. Although investigations of the properties of scales derived by these methods are still in their elementary stages, the evidence to date suggests that all methods produce equally accurate scales.³

The following can be said in summary of this section upon the mechanism of communication involved in measuring attitudes:

1. Even the most willing and capable respondents cannot provide accurate and unambiguous information on attitudes unless an adequate mechanism of communication is provided.
2. There are two basic types of information which attitude measurements provide. The first refers to the attitudes of a group toward a single stimulus. The second refers to the relative degree of approval which a group of respondents feels toward a group of stimuli.

¹ The method of successive intervals is used both to collect and to analyze data.

² See footnote reference, p. 283.

³ See Saffir, *op. cit.*

3. The most widely used technique of securing attitudes toward a single stimulus is the single-question technique. There is hope of securing accurate and unambiguous information if highly trained interviewers ask simple questions to which respondents reply in their own words.

4. There are various methods of establishing attitude scales whereby individuals can be arranged according to the degree of favorableness which they express toward the stimulus being investigated. These scaling techniques, although more complicated than the single-question technique, provide information of greater accuracy and less ambiguity.

5. There are several methods of arraying stimuli according to the degree of approval expressed toward them by a group of individuals. Three methods (rank order, paired comparisons, successive intervals) are currently used. They result in the establishment of scales indicating the degree of approval expressed for each stimulus. The choice among these three methods is determined by the convenience of each and the circumstances attending a particular application.¹

2. Willingness of respondents. Even when respondents are completely willing and able to give information on attitudes, serious difficulties beset attempts to get such information because of the problem of establishing accurate and unambiguous means of communication. The discussion so far in this chapter has dealt only with the mechanism of communication. We shall now turn attention to the willingness of respondents to provide information on attitudes.

In the preceding two chapters there is a rather lengthy discussion of the factors affecting the willingness of respondents to provide information on behavior and intentions. All that seems necessary here is to point out the differences in respondents' willingness to provide information on attitudes as compared with intentions and behavior. The generalizations which will be suggested are those which seem reasonable and which have not been contradicted by the varied experiences of personnel in marketing research. None of these generalizations has been subjected to rigorous testing.

In the first place, there seems to be some basis for believing that requests for opinions or attitudes create a more sympathetic and cooperative attitude on the part of respondents than requests for information about what has been or will be done. Requests for opinions are often considered

¹ No mention has been made in this chapter of the various projective techniques for determining attitudes, since these techniques are, as yet, of unproved usefulness for marketing research. They do seem promising. For a discussion of some of these techniques see Irving R. Weschler and Raymond E. Bernberg, "Indirect Methods of Attitude Measurement," *International Journal of Opinion and Attitude Research*, Vol. 4, No. 2, pp. 209-228, Summer, 1950.

flattering and a reflection on the importance of the respondent. Requests for information on past behavior or upon intended behavior, on the other hand, probably are less often considered flattering and more often an invasion of one's privacy. Finally, opinion questions are often interesting to the respondent.

There are some subjects which respondents have shown a reluctance to discuss in terms of either attitudes or behavior. Sex, religion, and politics are examples of subjects upon which many respondents have refused to be interviewed. The range of subjects which are typically investigated in marketing research do not usually involve stirring the emotions and probing fundamental beliefs as much as do these broader issues. It is not usual, therefore, to encounter refusals to express opinions on the rather limited range of subjects of concern in marketing. When refusals do take place, it is usually not because of an unwillingness to have one's attitude known but rather because of an unwillingness to take the time to express it fully. It is hard to believe that many respondents would be reluctant to have research personnel know which of several packages was preferred or which of several designs had the most appeal. It certainly is not inconceivable, however, that respondents would be unwilling to make the 190 statements of preference that would be necessary if the method of paired comparisons were used to select the most preferred of 20 alternative stimuli. Also it is conceivable that people would be unwilling to read the 20 statements of opinion which are often used as attitude scales in either the method of equal-appearing intervals or that of summated ratings. However, the range of devices available to researchers is great enough to make it unnecessary to use time-consuming expressions of opinion. The method of successive intervals, for example, can always be substituted for the method of paired comparison, and scale analysis can supplant the method of equal-appearing intervals.

3. The ability of respondents to provide accurate and unambiguous statements. The ability of respondents to provide accurate and unambiguous data on attitudes is limited by their capacity to understand the statements used in attitude scales or the questions used, by their capacity to consider adequately in forming their opinions the intangible attributes of the stimuli which they judge, and by their capacity to give coherent expression to opinions so formulated. Again the difficulties encountered are for the most part not new, a discussion of them having been given in the preceding two chapters. When the single-question technique is used, the difficulties of communication and of conception are quite similar to those involved in statements of intention. When attitude scales are used or when respondents are required to judge competing stimuli, it is obvious that the respondents are not required to formulate coherent expressions of

opinion. In the former case the respondents are merely required to endorse or reject or to indicate through the use of a check list the degree of approval or disapproval of statements. In the latter case respondents are merely required to express their approval of stimuli relative to each other.

Many factors which affect purchase decisions are not necessarily taken into account when the statements of opinion are made. The main factors are the place where the product can be purchased and the range of competing products that will be available. The problem is one of trying to acquaint the respondent with the relevant facts without biasing his response.

CHAPTER 16

SECURING INFORMATION ON KNOWLEDGE AND HABITS OF THOUGHT

16.1 Introduction

Under certain circumstances, information on knowledge and habits of thought may prove useful in forecasting consumer behavior. The following three circumstances seem to be necessary conditions for the usefulness of this type of information:

1. It is essential that there be some basis (either theoretical or empirical) for deciding what reactions a product or a way of selling the product should produce in potential purchasers. For example, there is good reason (both theoretical and empirical) to believe that the name of a product should be easily remembered. The name should, moreover, evoke associations which are considered helpful in selling the product. It is believed, justifiably or not, that associations which the consumer himself thinks are desirable would not be detrimental to the selling effort. One might therefore choose among alternative names for a product by testing (a) the ability of the respondent to remember the alternative names and (b) the associations which the alternative names evoke for the respondent.

Such a memory and association test (knowledge and habit-of-thought test) was used in England by an advertising agency in selecting a name for a detergent. The management wanted a name which would be easily remembered, and they believed they knew what associations were beneficially related to their product.

The practical problem of determining what reactions a product or selling method should evoke is much more difficult than the preceding comments suggest. Of course, it seems obvious that the name of the product should be easily remembered. A much more difficult problem is that of deciding, for example, what associations for the consumer the name should have. It is necessary, therefore, to consider a basis or bases for resolving a problem of this kind.

A perfectly plausible answer is that a product or selling method should evoke consumer reactions which lead to purchase of the product. Unfortunately, however, this answer simply rephrases the question. For it is still necessary to find out what reactions lead to purchases.

Frequently there are a priori beliefs as to the consumer reactions which

are favorable to the sale of a product. There are sometimes methods of testing such beliefs. The test must come through determining the relationship between sales and the supposedly favorable reactions. The most rigorous test involves some form of experimental sales test. Other kinds of evidence, too, are helpful. The use of split-run tests with magazine advertising has already been mentioned. Such tests could be used to determine the relative rates of coupon return with several advertisements, using names which create different associations.¹ Radio advertising too can be tested by a sort of split-run technique. That is, the advertising message can be differentiated between local audiences. If the message made some offer which required written responses from the audience, the comparison of the number heard from would provide the basis for an inference regarding the effectiveness of the advertisements. Probably the cheapest way to expose customers and potential customers to advertising, if only a small and localized group is involved, is through handbills and point-of-sale displays. One device involved handing out alternative forms of an advertisement to alternate persons entering a store. Each of the advertisements advertised the same product, and each was worth money when traded in for the advertised product. At the end of the day(s), the number of handbills of each type was counted. The difference between the number of handbills of each type was tested for statistical significance, and some conclusion was reached regarding the relative efficiency of the advertisements in effecting sales. Once the relative effectiveness of the advertisements has been determined in this way, it will be possible to see whether the advertisements differ significantly with respect to the knowledge they convey or the associations they evoke.

2. The second condition for the usefulness of information on knowledge and habits of thought is that practicable procedures should exist for obtaining the desired information on knowledge and habits of thought. In the example concerning the selection of a product name, practicable procedures did exist for determining (a) what names were easily remembered and (b) what associations were evoked by different names. To exemplify further, if it is considered desirable that an advertisement convey an understanding of how a product works, it must be possible to find some way of determining whether people actually have acquired this understanding.

¹ It should be pointed out, however, that the factors which influence coupon return may not coincide with the factors which influence actual sales. For example, if the coupon offers a prize or premium, a large number of children may send in coupons. Even if the coupon offers the product itself, perhaps for a slight reduction in price, this method of selling the product is not typical, except for a limited number of products, such as correspondence-school courses. In this sense, the test suggested here is of more limited usefulness than a sales test.

3. The third condition is that more direct measures must be (for some reason) unfeasible. These more direct measures have been considered in the preceding chapters, together with various reasons for their occasional impracticability. For example and in review: sales tests might be too costly; in the introduction of a new product there is no past behavior; intentions might be unknown to the respondent because of the nature of the product; or attitudes might be difficult to determine. A difficulty in attitude measurement, for example, is the tendency for a respondent to assume the role of an expert and to pronounce his conception of the general reaction rather than his own personal reaction. When, for any of these reasons (and the list is far from exhaustive), the relatively direct measures are impracticable, information on knowledge and habits of thought may prove to be useful. Sometimes, however, information on knowledge and habits of thought may be obtained at the same time as other types of information, in order to study the interrelationships among these different types of information.

In the following sections of this chapter we assume that the above essential conditions have been met and consider the techniques used to obtain information on (1) knowledge and (2) habits of thought.

16.2 Types of Knowledge

Three types of knowledge are of interest. These are (1) knowledge of an already adopted stimulus, such as an actual advertising theme (name or technique); (2) knowledge of stimuli introduced under test conditions—that is, stimuli which have been proposed but not yet adopted; and (3) *understanding* of sales messages or other communications about a product. While this third type of knowledge overlaps the first, a difference in emphasis warrants the separation. The first two types of knowledge are primarily the result of memory, while the third type primarily involves comprehension.

1. Knowledge of an already adopted stimulus. One aspect of tests for the effectiveness of an advertising program is the extent to which the advertising name has become recognized in the community. Such a measurement of the extent of knowledge of an advertised brand was used by George Brown in his study of the effectiveness of General Mills advertising in the *Chicago Tribune* for the calendar year ending January, 1950.¹

Brown's study employed, simultaneously, several different measures of advertising effectiveness. For example, two different measures of actual sales were used. One of these measures consisted of records of actual de-

¹ George H. Brown, "Measuring Consumer Attitudes towards Products," *Journal of Marketing*, Vol. XIV, No. 5, pp. 691-698, April, 1950.

liveries of the product into the test market. A second was obtained from records of purchases made by members of the *Chicago Tribune* Consumer Panel. In addition to these sales measures, attitude and preference measures were obtained; and, what is of relevance for our present discussion, measurement of consumer awareness of the advertised brand ("level of knowledge") was also obtained. The reason for using these latter measures was a desire to find out if such indirect measures of advertising effectiveness as preferences and awareness of the product would give an indication as to future changes in sales. Since it was felt that an important time lag might exist between changes in knowledge or attitudes and changes in purchasing behavior, a knowledge of the changes in awareness would be of obvious usefulness in forecasting. For example, if changes in preference or knowledge took place without a corresponding change in consumer purchases, it would suggest that the advertising should be continued for a reasonable length of time in order to see if such a time lag existed. For the present purpose, however, we shall consider only the measurements of awareness of advertising, since this type of information is discussed in this chapter.

Probability samples were used at different times during the year in order to determine whether awareness or knowledge of the advertised names had increased during the period, as we have just pointed out. Knowledge of advertised names (or copy) is certainly not by itself convincing evidence of advertising effectiveness and was not so used in Brown's study. It is quite possible that an advertisement would simultaneously increase people's knowledge of a product and also increase their reluctance to purchase it or would leave them unaffected. Or, conversely, advertising might increase the rate of sale of a brand without affecting its "awareness level" in the market. Basically, however, an advertisement must be perceived (heard or seen) in order to be effective, even though this may not ensure its effectiveness.

One test of the extent to which advertising has been perceived is the test for knowledge of the advertised name or theme. Brown tested the knowledge of *Tribune* readers of the advertised products and compared it with the knowledge of non-*Tribune* readers, who presumably did not have a chance to perceive the advertisements. The testing was done with both aided- and unaided-recall techniques. An example of the specific phrasing of the question based on unaided recall was the following: "When you think of flour, what brands do you think of?" The aided-recall technique was implemented by use of the familiar check list containing the names of all brands of flour sold in the Chicago market. In this particular bit of research, it was found that knowledge or awareness of the advertised

names did not increase through the year, and this was considered to be evidence, but not proof, of the ineffectiveness of the advertising.

There are, of course, other ways of testing factual knowledge. The theory of testing, itself, is complicated and could profitably occupy much space. There are good sources of such discussion, however, and we shall not duplicate that material here.

2. Knowledge of stimuli introduced under test conditions. This second type of knowledge is the type connected with the example concerning possible names for a detergent. To continue with this same example, the proposed names might be put on index cards and shown to the respondent in a random order. (Care must be taken to avoid bias due to the order of presentation, and different orders should be used with different respondents.) After a certain lapse of time, the respondent is asked to recall the names. The presentation of the stimuli and the memory test does not necessarily mean a break in contact with the respondent—that is, everything can often be done in a single interview.

One illustration of this general approach, sometimes known as the "Impact" technique,¹ requires a second interview. In this method respondents are exposed deliberately to a specially made magazine containing, among other things, advertisements to be tested, and then reinterviewed to see what is remembered of the contents of the magazine, including the advertisements being tested.

3. Understanding sales messages. Certain types of goods—such as consumer durable goods, style goods, pharmaceutical goods—have qualities and effects which are not readily apparent to the consumer but which are considered important by the seller. The selling effort, therefore, features these particular qualities and/or effects. The success of such selling effort can be determined in part by testing the respondents' understanding of these explained attributes. For example, the Studebaker Corporation may consider it important to its selling effort to indicate certain technical points about the transmission in its cars. Manufacturers of pharmaceutical goods occasionally consider it necessary to go into physiological explanations of the efficacy of their product. In both these cases, information on consumers' understanding of these sales messages is one sound basis for an evaluation of the effectiveness of the particular selling effort. Most of the standard testing devices are suitable for testing this type of knowledge: true-and-false questions, multiple-choice questions, simple questions to which respondents reply in their own words, etc.

Once again it is clear that making potential customers understand the nature of certain attributes of the product is not in itself enough to cause

¹ This method was developed by George Gallup and Claude Robinson.

people to buy the advertised product. It seems reasonable to assume, however, that such an understanding is probably frequently effective in increasing people's willingness to buy the product.

16.3 Habits of Thought

Information on habits of thought can be extremely useful under certain circumstances. We consider here a few of these habits of thought and their usefulness, primarily through the presentation of examples. The usefulness of the information should be apparent from the nature of the example.

1. Associations. The term "habits of thought" has to be stretched a little to include associations. Associations are the images and concepts that a stimulus evokes in the mind of the person exposed to it. It seems reasonable to assume that a brand name or advertising headline or theme ought to evoke associations that indicate the desirable attributes of the product rather than attributes not possessed by the product or the less desirable attributes of the product. Determining the associations for respondents of a name or theme is not very difficult. The names or themes or other stimuli which are being judged by the respondents are presented serially to them with the request, after the presentation of each stimulus, that the respondents mention the first things that occur to them.

We can explore this technique a little more fully by considering again the survey designed to select a name for a detergent. After informing the respondent that a new detergent capable of doing certain things was being put on the market, the researcher in this survey asked the respondent in essence the following two questions:

1. When you hear "Gleam," what do you think of?
2. Each name listed below refers to one of the kinds of products listed after each name. Pick the product to which you think each name belongs.

The answers to these two questions were then analyzed to see which name seemed to evoke the desired associations and which seemed to suggest best the type of product being sold.

2. Other reactions to stimuli. Free associations are not the only useful reactions to stimuli for marketing research. It is also sometimes useful to know whether respondents are interested or bored, approving or disapproving. Once again the information is useful only if there exists a valid assumption about the relationship between these reactions and purchases of the product being marketed. Once again, empirical evidence concerning these relationships is slight, and confidence must rest upon the apparent "reasonableness" of the assumption. And it seems plausible that

products which are approved will sell in larger volume than those which are not and that advertisements which interest consumers will be more effective than those which do not.

Some ingenious devices have been devised for measuring these reactions. Perhaps the best known is the Stanton-Lazarsfeld Program Analyser. This analyzer records and summarizes electrically the choices of an audience between approval and disapproval. Each member of the audience is provided with two buttons, one of which electrically records approval and the other disapproval. Indifference is registered by pressing neither button. As the stimuli proceed in fixed succession (as in movies, radio programs, or television programs), the device makes a continuous record of the reactions of the audience.

✓The Ad Detector Corporation and the Color Research Institute have another kind of device for determining reactions to stimuli. Their device measures and records the conductivity of the skin surface due to changes in the rate of perspiration, which, in turn, reflects emotional changes. The theory is similar to that underlying the use of the lie detector. The device permits a continuous recording as the respondent is exposed to the succession of stimuli which a radio program, television show, etc., present. (Later in this chapter we shall discuss the validity of the inferences based on such measurements.)

Such devices are obviously inappropriate for recording reactions to stimuli which do not hit the audience in some unalterable order and at some ascertainable rate of speed. This is true because the only dimension by which it is possible to connect uniquely a reaction to a stimulus is time. In other words, such devices are useless for determining reactions to the component stimuli of written advertisements.

✓The Ad Detector Corporation has created another instrument for measuring reactions to written advertisements. An arrangement of mirrors permits the observer to see an image of the respondent's eyes superimposed on the advertisement at which he is looking. The observer records electrically the order and time during which each part of the advertisement is looked at. Simultaneously, fluctuations in the skin perspiration of the respondent are measured. It is thus possible to determine the reaction to each part of the advertisement. It is possible, for example, that skin perspiration (interest) rises when the respondent looks at the brand name and declines while looking at the pretty girl in the bathing suit.

All these electrical devices pose serious sampling problems for the researcher. The Program Analyzer must rely on volunteer audiences which may differ importantly from other audiences. The Ad Detector must rely upon respondents who are willing to come to a downtown office for the

test. Clearly, probability samples are difficult or impossible to obtain, and the possibility of sample bias is present.

In addition to the devices already mentioned, there are available the techniques of formal interviews and depth interviews. No new problems are introduced in using interviews to obtain this type of information.

16.4 Ambiguity

So far, we have not judged the possibility of getting unambiguous and accurate information on knowledge and habits of thought. These criteria must be considered if the value of the information is to be judged.

For most types of tests of factual knowledge, the mechanism of communication can be so simple that ambiguity is negligible. This is particularly true when respondents are tested for memory of names, themes, or other stimuli which can be presented in the form of a simple check list when aided recall is used. Also, the recollection of names or themes without aided recall places so slight a tax on the power to formulate coherent responses that the possibility of ambiguity can probably be safely disregarded. With regard to information on associations, the problem of ambiguity is more serious. The difficulty arises because of the frequently great differences between the dictionary meaning of a word and the personal meaning or associations it may have for the respondent. These differences can cause an investigator to be seriously misled by the reported associations of respondents. For example, a respondent might associate "sea" with the stimulus, "waves." This might be interpreted as a pleasant association for the respondent, the investigator thinking of images of sandy beaches, golden sunshine, and beautiful women. In fact, the association may have involved images of chronic nausea and a painful sunburn. The example is a bit unreal, but it may serve to indicate the possible causes of ambiguity when associations from respondents are sought.

This ambiguity can be lessened by probing in order to determine the significance for respondents of their reported associations. This probing, if it is not to result in bias, must be skillful and relatively inobtrusive. One way to do this is to get a second "layer" of associations. That is, the fieldworker asks the respondent to indicate the associations to the associations for the original stimulus. Each association thus becomes the stimulus for further associations. If these associations were carried far enough, the result would be in some ways similar to the statements of a person being psychoanalyzed. It would be possible to understand the person's habits of thought well, better, perhaps, than they are understood by the respondent himself. Naturally, cost limitations and the limited skill of persons available for fieldwork make it difficult or impossible to carry the

chain of associations very far. It should, however, be possible and is often desirable to go beyond the first stage of association in the interest of diminishing ambiguity.

The devices for obtaining other kinds of reactions to stimuli are also subject to serious ambiguity under certain circumstances. For example, although the Program Analyzer records approval or disapproval, it does not always unambiguously indicate just what is being approved or disapproved. Although the approval or disapproval can be connected in time with the stimulus to which the reaction refers, it is sometimes difficult to assert with confidence which of the various stimuli present at any point in time caused the recorded reaction. For example, if approval (or disapproval) were registered at the time Jack Benny and Dennis Day were exchanging quips, it would be difficult to know whether the reaction referred to Benny, Day, or the substance of their remarks. Similarly, reactions to a musical program might refer to the composition being played or to the virtuosity of the musicians. Some people like to hear Bing Crosby whatever he sings; others like certain songs regardless of who sings (plays) them. The Program Analyzer alone does not permit distinctions between these different kinds of approval.

The devices of the Ad Detector Corporation are subject to ambiguity in that the heightened interest (called "arousal") which is recorded could be the result of either strong approval or strong disapproval. In attempts to overcome some of this ambiguity, respondents are usually interviewed later on the qualitative nature of these arousals.

16.5 Accuracy

Information on the factual knowledge of respondents can usually be considered accurate. Aside from general timidity and reluctance to be interviewed, there is seldom reason to believe that potential purchasers of a product would refrain from indicating their awareness of its existence or their familiarity with its attributes. Occasionally, of course, some products, the ownership of which may be thought to indicate that their owners are impoverished, diseased, or in some other way afflicted, may be more generally known than would be indicated by a general survey. Moreover, in order to impress the interviewer, people may claim factual knowledge they do not possess. This problem has already been seen in earlier chapters with reference to other kinds of information.

The major problem of accuracy arises when the fieldworker attempts to ascertain the associations of the respondent. Almost any train of associations will contain elements that respondents are unwilling to state. In other words, often respondents are thought to edit consciously their public

utterances. Probably all that can be done is to establish rapport and indicate, in so far as possible, that no judgment will be made of the morality or value of the respondent's comments. Undesirable censorship can sometimes be detected by pauses in the respondent's remarks.

There is no problem of accuracy for the electrical devices for determining reactions. The responses are recorded objectively and are essentially accurate if the machines are understood and correctly used.

CHAPTER 17

SECURING INFORMATION ON "REASONS WHY"

17.1 Introduction

Up to this point we have discussed ways of communicating with and observing people in order to find out what they have done, what they intend to do, how they feel, or what they know. This information has been sought in order to help in forecasting consumer behavior. Most of this information has been useful to some extent in making it possible to understand the factors that affect or determine this behavior. A consideration of statistical data on the sales of soft drinks may provide the basis for inferences regarding the influence of climate upon the consumption of these beverages. Similarly, statements of intention can cast light on some of the reasons why people do things. If, for example, significant differences were noted in the intentions of consumers to purchase a dress at various prices, some inference regarding the effect of price upon consumer behavior might be justified. Again, if the attitude measurement of a sample of respondents after an advertising campaign was significantly different from the measurement prior to the campaign, under some circumstances an inference regarding the influence of the campaign would be justified. The preceding chapters discuss types of information that indicate something about "reasons why," but they do it indirectly through some inferential process. So far in Part IV, however, we have not explicitly discussed methods by which "causal" relationships (reasons why) can be directly determined. This subject was discussed in Chapter 4, where we described experimentation and the use of observational data to test hypotheses. Now we shall discuss the securing of information on causal relationships by the direct questioning of respondents—the questioning almost always involving the use of the word "why."

Information regarding causal relationships which is secured by the direct questioning of respondents may be less ambiguous and more accurate than inferences regarding causal relationships which rest upon other kinds of evidence. In other words, it may be possible to achieve a sounder basis for understanding why respondents have done what they've done or why they feel as they do by asking them rather than by securing data relating to their behavior or attitudes and relating such data either to attributes of product or to consumers and their economic activity. It is possible, for

example, that one might understand better why people choose a particular brand of coffee by asking them than by inferring the determinants of choice from a consideration of data on rates of sale of various brands of coffee and of data on attitudes toward various attributes of these various brands. There are two primary reasons why direct questioning may sometimes be superior.

In the first place, statistical methods are often inadequate for identifying causal relationships among observational data. (Here the word "observational" is used as in Chapter 4 to refer simply to nonexperimental data.) It is well known that there is a high correlation between the birth rate and the stork population in certain Scandinavian cities. Only the most romantic, however, would accept this high correlation as evidence of some causal relationship between the two phenomena. In an example like this, it is easy to see that the correlation does not reflect causation. In marketing or economic research, it is often much harder to distinguish between mere statistical association and causal relationship.

Consider the example of the fairly common interpretation given to the high correlation between advertising expenditures and sales for individual firms. Frequently this correlation is interpreted as indicating the powerful effects of advertising. Another equally plausible explanation of this correlation is that advertising budgets are determined as a fixed percentage of sales. Both interpretations are consistent with the observed correlation. The statistical data alone cannot indicate which explanation—if either—is correct or more nearly correct. Direct questioning of respondents might supply the insight which correlation techniques are inadequate to provide.

Second, even when correlation analysis might otherwise be satisfactory, adequate data are often missing. The data may be absent altogether or may be so inaccurate or inappropriate as to make their use of little value. Under these conditions the best alternative available to an investigator desirous of understanding causal relationships may be the direct questioning of the people whose behavior is being studied.

In addition, direct questioning may be used to supplement or correct information that is available from other sources. Suppose, for example, that a high correlation was observed between the amount of money spent in advertising Fenster soap flakes and the total sales of the product. It is postulated that this correlation means that increases in advertising cause increases in sales. A conversation with the advertising manager of Fenster flakes might reveal that the company policy is to spend 3 per cent of total sales revenue on advertising. This gives good and direct evidence that at least part of the relationship underlying this correlation is the effect of sales on advertising rather than the effect of advertising on sales.

Consider another example. The design of the Fenster flakes package is changed at the same time that the product itself is changed by adding a new cleansing agent. Sales increase. It is postulated that the increase is the result of one of these changes. Conversations with housewives who have switched to Fenster flakes show that practically all of them attribute their change to Fenster's remarkable new cleansing qualities.

The foregoing discussion might even indicate that it is foolish ever to ask "what," "when," "where," or "how" when it is possible to ask "why." Most people would disagree with such a conclusion. Although the "why" question is widely used in marketing research, it is almost always used with fear and trembling. The reason for the almost universal reservation about dependence on "why" questions is that certain important difficulties are involved in any attempt to understand, simply by asking, why people do something or have certain attitudes or intentions. Thus, although there are certain advantages resulting from the successful use of "why" questions, there are various important difficulties in using them successfully. The difficulties referred to affect both the ambiguity and the accuracy of information on 'reasons why.' We turn, then, to a discussion of these difficulties and ways to overcome them.

17.2 Ambiguity

All the ordinary difficulties of communication which were discussed in Chapter 12 contribute to ambiguity here as well. There need be no further discussion of the wisdom of using simple familiar language and simple sentences, of avoiding abstractions, etc. There are some new difficulties, however. These difficulties have already been discussed very well by Paul Lazarsfeld.¹ Lazarsfeld found possibilities of difficulty at three stages in the communication between researcher and respondent: (1) getting the respondent to understand the exact meaning of the question concerning "reasons why" (motivation); (2) enabling the respondent to give a coherent and complete answer; (3) understanding the implicit as well as the explicit assumptions underlying the respondent's answer. We follow Mr. Lazarsfeld's organization in discussing the problem.

1. Getting the respondent to understand the exact meaning of the question concerning motivation. The major difficulty in getting a respondent to understand the precise meaning of a question on motivation lies in the fact that every activity (or attitude or intention or habit of thought) is determined by many things. Usually the investigator is not interested in understanding all the factors that determine a phenomenon but merely a very restricted range of determining factors. Some psychol-

¹ Paul F. Lazarsfeld, "The Art of Asking *Why*," *National Marketing Review*, Vol. 1, No. 1, Summer, 1935.

ogists have classified determining factors into three groups: (1) biological determinants, (2) biographical determinants, and (3) instantaneous determinants or "determinants of the first degree."

Usually most of the biological and biographical determinants of the act of a person would be understood without questioning by any investigator living in the same society as the person being investigated. One usually does not need to inquire as to the biological and biographical factors that cause children to be sent to school. On the other hand, it may be necessary to inquire as to the factors that cause a certain school to be selected, that cause schooling to begin or to terminate at certain times, etc.

These latter determinants would in most cases be determinants of the first degree, although it can be seen that the distinction between biographical determinants and determinants of the first degree often becomes very fuzzy. Again, in finding out why a person patronized a particular grocery store, one probably would not be interested in hearing that the person had to eat, that the person lived in the city and couldn't raise his own food, that the person lived in this particular city because his wife's brother-in-law ran a business in which there was a job, that the person lived in a particular district in the city because the schools were good and the rents low, etc., although all of these are determinants of the person's selection of a grocery store. Rather, the investigator is probably interested in the so-called "determinants of the first degree," which cause the selection of a particular store rather than a neighboring store or other stores which could conceivably be patronized by a person living under the circumstances in which the person being investigated finds himself. Some of these determinants of the first degree could also be considered as biographical determinants. For instance, the respondent, being an Italian, might choose an Italian grocery rather than the neighboring chain store.

We shall proceed on the assumption that marketing researchers are usually and primarily interested in determinants of the first degree. It must still be recognized that there are many such determinants and that it is a difficult task to get the respondent to answer with regard to the particular determinants in which the investigator is interested. If the investigator is interested in an exhaustive study, the respondent must be stimulated to provide comprehensive answers. If, on the other hand, the investigator is interested in only selected determinants of the first degree, (e.g., the effect of advertising on buying habits), the problem exists of making respondents understand which factors are of interest.

We shall consider a particular example, the purchase of a radio. The factors of the first degree determining the decision to purchase a particular radio can be included under the following categories: (1) factors relating to the attributes of the product, (2) factors relating to sources and types of

information about the product, (3) factors relating to the purpose for which the purchase was made. An example may indicate more precisely the meaning of this categorization and its usefulness.

Suppose an effort were being made to find out why people purchased the brand of radio which was in their home. Suppose further that the researcher thought that the way to find out why people chose a particular brand of radio was to ask them as directly and simply as possible. Such a question might be "Why did you buy a Blank brand radio?" It would not be unusual to receive responses such as the following: (1) "The reception on Blank brand is wonderful"; (2) "My brother George is an electrician, and he says that the Blank brand is the best"; (3) "Our kid is growing up and he wanted a radio of his own in his room, so we decided to give him one for his birthday"; (4) "I buy all my appliances at the Main Street Appliance Center, and they feature Blanks"; (5) "Our old radio was worn out. We couldn't hear anything more than 50 miles away"; (6) "It was cheaper than the other models that look about the same"; (7) "Their advertisements about the number of tubes and the duraflex filter seemed to make sense"; etc.

A little reflection upon the foregoing responses will indicate their ambiguity. It can be seen, for example, that the person who bought the radio for his son might have chosen a Blank radio because its reception was superior. Conversely, the person who stated that superior reception was the basis for selection might have purchased the radio in order to give it to a son. Thus the motivation of these two persons in buying the radio might have been about the same even though the incomplete responses which they gave failed to indicate the similarity. Again, the person who was impressed by the advertisement may have purchased the radio because of the superior reception which the advertisement claimed for the product. The person who stated that his selection was made on the basis of superior reception may have come to believe in its superiority because of advertising. The multiplicity of motives for each individual and the incompleteness of the statements concerning these motives leads to a sort of confusion and ambiguity that suggests the need for a categorization of responses.

How would the responses given in the example above fit into the three categories mentioned above? The first statement refers to an attribute of the product; the second to a source and type of information concerning attributes of the product; the third to the purpose for which the purchase was made; the fourth to an attribute of the product (its location and dealer); the fifth to the purpose for which the purchase was made; the sixth to an attribute of the product; the seventh to a source and type of information concerning attributes of the product.

If the investigator were interested in all three types of determinants,

an understanding of the categorization suggested here would provide a basis for completing the information secured from each respondent. Such a comprehensive coverage of the types of determinants could be secured either formally or informally. A formal questionnaire designed for this purpose could go a long way toward achieving this objective. For example, one way of doing this would be to ask three specific questions. To find out "factors relating to the purpose for which the purchase was made" one might ask, "Why did you decide you wanted a radio?" To find "factors relating to attributes of the product" one might ask, "What was there about the Blank brand that made you prefer it to other brands?" To find "factors relating to the sources and types of information about the product" one might ask, "Where did you find out about the Blank radio?"

It would be possible to subdivide these questions even further and make them more specific and less ambiguous. For example, to find out sources of information one might ask if the respondent talked to friends, neighbors, relatives, etc., if he read any advertisements, heard any radio programs, etc. The general plan of questioning is to proceed from general questions (which are relatively ambiguous) to specific questions (which are relatively unambiguous). The questions at the start serve the purpose of finding what the respondent feels to be most important and getting the respondent thinking about the general problem. The specific questions serve to fill in an unambiguous picture of attitudes or "reasons why."

This process is not automatic, however; skilled interviewing is required. Questions may not be completely understood, with a resulting confusion and lack of comprehensiveness in the answers. The job of making sure that the questions are understood without at the same time conveying a new meaning is challenging. The need for responsible, intelligent, and well-trained interviewers is particularly great when the difficult "why" questions are involved.

Rather than use the formal questionnaire, some researchers have suggested less formality and less direction in the interview. A questionnaire constructed along these lines might start with the simple question, "Why did you buy the Blank brand radio?" The answer to this question could be recorded under the proper category (of the three mentioned above), and subsequent questioning could be used to fill in information regarding determinants for the other categories.

Lazarsfeld points out a further difficulty in establishing unambiguous communication regarding the determinants of purchase behavior. Not only must the respondent understand which of the various types of determinants the questions concern, but there is also a possibility that the activity referred to in the question may not be clearly understood. Lazarsfeld uses as an example the question, "Why did you buy this book?" This

question could be understood to mean a request for information concerning the reason that the book was bought rather than borrowed or rented from the library; or it might be thought of as a request for the reasons why a particular book rather than other books was purchased; or it might be thought to be a request for reasons why money was expended for a book rather than for something else. Such ambiguity may at times be important enough to cause practical trouble. It can be eliminated only if the exact meaning of the question is understood by the interviewer and if the interviewer is able to detect misunderstanding by respondents. It can be reduced by proper inflection when the question is asked. Consider the following three inflections of the question: "Why did you *buy* this book?" "Why did you buy *this* book?" "Why did you buy this *book*?" These three inflections correspond with the three different interpretations just mentioned. However, in practice it is difficult enough to get interviewers to stick with the proper wording of the question, to say nothing of the proper inflection. To be doubly safe it is desirable to make the wording less ambiguous by lengthening the question slightly: "Why did you buy *this* book instead of some other?" But no matter how careful the wording and the instructions on inflection, the interviewer must still be on the alert to detect misunderstandings.

2. Enabling the respondent to give a coherent and complete answer. A second basic difficulty in establishing unambiguous communication concerning "reasons why" lies in the fact that for many types of activities people are unaccustomed to formulating explicitly their own motives. Thus, although the exact meaning of a question might be understood, the respondent could (and often does) have serious difficulties in making a coherent statement. The sort of difficulty referred to can probably be understood by anyone who takes the time to try to make a statement concerning the reasons why he uses a particular brand of soup, razor blades, or automobile tires or why he goes to the movies or why he listens to radio programs. People seldom do much thinking about the reasons for their actions, particularly such relatively minor ones as most purchase decisions. Moreover, most people would be inclined to lose patience if asked to do very much explaining "why." Again, the remedy is a well-trained interviewer who is able to stimulate, without biasing, responses. When a respondent says, for example, "I haven't any idea why I buy Fenster soap flakes; I just do," the interviewer must be able tactfully to stimulate the respondent to express her reasons (if any) without at the same time putting words in her mouth. Innocuous probes such as "Is there any special reason?" "How is that?" etc., are helpful, but much depends on the interviewer's manner and her ability to put the respondent at ease.

One solution to this problem is to abandon the free-response type of

question just discussed and use instead a check list or "fixed alternative" question. Then all respondents can be prompted to the same degree. It goes without saying that the check list of possible reasons, which is used in the fixed-alternative method, must be as exhaustive as possible. The only way this can be accomplished is by intensive pretesting with free-response questions. Rarely can the deskbound researcher think of all the important reasons for purchase. Even when a reasonably exhaustive list of reasons has been prepared, respondents should be given the opportunity to mention reasons not included on the list (there will almost always be some). Even when this is done, however, one must accept as an inevitable drawback of the check-list method that any reason not included on the list will be subject to relative downward bias.

Another drawback is that the lists may be long and unwieldy if they are really comprehensive, and it may therefore be hard to get respondents to consider the entire list. Alternation of the order of the reasons in the check list may be used to reduce the possible upward bias accruing to reasons near the top of the list (see discussion of aided recall, Chapter 13). Still another drawback is that inarticulateness may be forced underground; the respondent may check answers just to avoid embarrassment. It must not be forgotten that inarticulateness may be a sign of a lack of conscious reasons as well as an inability to express oneself. If people simply have no conscious reasons for purchasing a particular product, it is better to know this than to have a list of reasons (rationalizations) which were checked by respondents when they were given a check list. Later in this chapter we consider ability to give "reasons why."

Lazarsfeld believes that interviewers should be guided in their attempts to help the inarticulate respondent in discovering and stating his motives by formulating suggestions and questions in terms of the experience of the respondent. One way to do this is to help the respondent think through the actual act of purchase; indeed, information on this act would ordinarily be sought on any questionnaire which used questions concerning reasons for purchase. As the respondents give information on behavior, they may also focus their concentration on the purchase decision and be better able to perceive the factors which actually determined their decisions. It is doubtful, however, whether any conscious suggestion of possible reasons for purchase is warranted. Since most respondents have been through the unpleasant system of taking examinations in school, they are usually more than willing to accept the suggestions of the interviewer, regardless of whether or not the suggestions are really appropriate. This is less likely to be true when the questioning is concerned with issues of great importance to the respondent, but marketing research is seldom concerned with such issues.

3. Understanding the implicit as well as the explicit assumptions underlying the respondent's answer. The third difficulty in establishing unambiguous communication lies in the fact that even if a respondent understands exactly what is meant by an interviewer's question and is able to think of the information requested and to make some articulate statement of it, the interviewer (coder, analyst) may misunderstand what is said. We mentioned in Chapter 14 that the general problem of ambiguity in communication from respondent to interviewer is complicated when the statements concern intentions because the interviewer may not know the tacitly assumed contingencies underlying the intention. Communication concerning attitudes is complicated by the inevitable need for abstractions whose meanings vary among individuals. The complication for communications regarding "reasons why" also lies in the pervasive presence of abstractions.

Lazarsfeld states that this third type of difficulty rests upon the existence of tacit assumptions by the respondent which are not recognized by the interviewer. In every case that he presents and that we have been able to think of, the tacit assumptions have rested upon differing interpretations of abstract words that are used. In a Lazarsfeld example, some person responded to a question concerning the factors that determined the selection of shoes by stating that quality was decisive. Others stressed style; still others, the durability of the material of which the shoes were made. Some misunderstanding existed because the respondent's definition of quality differed in some cases from the definition of the researcher. Lazarsfeld explained this ambiguity by saying that the respondent who emphasized the importance of quality was tacitly assuming that everyone knew the quality was determined by style, or by material, or by workmanship, or by other factors or combinations of factors.

Such an assumption undoubtedly existed and undoubtedly accounted for the ambiguity. But such assumptions can probably be both more easily recognized and more easily eliminated if it is seen that the tacit assumption concerns the definition of an abstraction. Once the ambiguity is associated with an abstraction, its detection becomes almost mechanical and its elimination relatively easy. The trained interviewer would recognize that "quality" involved an abstraction (tacit assumption) and that some ambiguity had been introduced which needed attention. The sort of attention it should need is obvious. The interviewer should find out what the respondent means by "quality." In other words, the respondent should be stimulated to give an operational definition of an abstract term in order to make its meaning clear and precise. A simple probe, such as "What do you mean by quality?" might suffice.

There are two other causes of the failure to understand correctly the

clearly stated motivations of respondents. These additional sources of error are mentioned for completeness rather than because of their practical importance in marketing research. The first of these causes of misunderstanding derives from the differences between individual mentalities. Motives can be understood best when they arise from patterns of thought and scales of value similar to those of the investigator. When the differences are enormous, as they might be between occidentals and orientals, or between old people and young children, or between Eskimos and Central Americans, the difficulties of understanding each other's motivation may be insuperable even though language difficulties were unimportant. It is probably true that men are better able to understand other men than women and that women's intuition operates with the greatest accuracy when it concerns other women. Adults have difficulty in understanding children; businessmen are often mystified by the academic mind. Such misunderstandings and mystifications probably are greatest when communication concerns some complex or emotional subject. The mystifications and misunderstandings arising from this source are just barely worth mentioning when the communication concerns most of the subjects with which marketing research is concerned.

The other cause of misunderstanding also is of little practical importance. People sometimes have motives which are painful for them to recognize and which they therefore repress or exclude from consciousness. Such persons not only refuse to recognize such motives in themselves but they sometimes become unable to recognize these motives in others even when the existence of these motives is explicitly stated. The difficulty in understanding others is based not only on inevitable differences in the mentality of different individuals but also upon repressed forces within the individual interpreter. Such repressions may be an important cause of misunderstanding in studying some types of social behavior which are of great emotional significance, but it is highly doubtful that such repressions are significant when one is investigating the reasons why people buy one brand of soap rather than another, why they patronize one store rather than another, etc.

17.3 Accuracy

Up to this point, we have assumed that the reasons which the respondent assigned in his statement for an act, intention, or opinion were the real reasons. We now drop this assumption regarding the accuracy of statements on motivation and discuss the extent to which such statements can be expected to be inaccurate, the reasons for such inaccuracy, and some possible remedies. Although the problem of ambiguity is both important

and difficult, it is the problem of inaccuracy that has been the most serious deterrent to widespread reliance upon questionnaires concerning motivation.

1. Inaccuracy attributable to unwillingness. The first source of inaccuracy is of great importance. Deliberate deception, which was found to create inaccuracy in information regarding behavior, intentions, attitudes, or habits of thought, operates with even greater force when information on motivation is being sought. One frequently hears the clause, "But he meant well," appended to derogatory statements in order to excuse the faults referred to. This is true because people often are judged by their motives rather than by their deeds. Since this is true, it is not surprising that people are particularly careful that their reasons for action appear creditable to others. The need for deception in stating one's reasons is often, therefore, even greater than when reporting one's activities, for example.

An interesting example, taken from an actual investigation, illustrates this idea.¹ A few years ago a large commercial air line was interested in discovering the reason why people chose a particular means of transportation. The air line was interested in finding out why some people went by train, others drove, and others flew. At the conclusion of the investigation, the air line found that there were many reasons that people had given for not flying. Among them were the cost of flying, the inconvenience of reaching the airport, the fears of wives or husbands, etc. Only a very small number of persons said that they refused to fly because of their own fear. The research director of the air line did not believe the findings. His long experience had shown that air travel usually fell off sharply after spectacular air accidents. The director, therefore, refused to accept the research findings at face value and sought to find a way to test the validity of the apparent manifestation of fearlessness by the persons interviewed. The director made the assumption that he could learn something about people's motives by studying the motives that they ascribed to others. He therefore conducted more interviews and asked people not why *they* didn't fly but why their friends, neighbors, and relatives didn't fly. In a very much larger percentage of the cases than in the original interviews it was found that people were reported to abstain from flying because of a fear of accidents. The research director felt that this evidence confirmed his judgment based upon experience and that the air line would be well advised to give some emphasis in its advertising program to the safety of flying.¹

¹ It should be mentioned that the redoing of research because it fails to agree with one's preconceived ideas is dangerous. In this case, however, the original *method* seems rather clearly to be faulty; and there was also other evidence which conflicted with the first results.

It should not be overlooked in this example that, although people were willing to admit that they took a train, they were unwilling to admit what subsequent investigation revealed to be in many cases the true reason for the choice of train over plane. They were willing to give an accurate statement of their actions but were unwilling to give an accurate statement of the factors that determined that action. There are many other examples in marketing research where this same sort of deception has been found. People may claim to read a certain newspaper because of its editorials rather than its comics, or vice versa. A person might explain the purchase of whiskey by saying he had to do a great deal of entertaining when in fact he was fond of it himself. A housewife might attribute the purchase of a low-priced washing machine to its sterling qualities rather than its low price.

For many types of investigation this difficulty is practically insuperable. It is probably insuperable in the army, for example, when subordinates must report to superiors regarding the reasons for their (the subordinates') activities. The possibility of getting accurate information is particularly slim under these circumstances, because the person to whom the report is made has rather complete control over the welfare of the person providing the information. In other cases where a subordinate-superior relationship is involved, the chances of getting accurate information on motivation are probably small.

Another reason for deliberate deception (not entirely different from the one just given) is the existence of the power of the investigator to reward the respondent. Investigators frequently have this power even though no subordinate-superior relationship is involved. In order to see what is meant by this, one need only consider any of the so-called "give-away" programs on the radio. Here the directors of the show are not superior, in the ordinary sense, to the participants from the audience, yet one seems to see in many cases a yielding by these participants to the wishes of the sponsor in the expectation of reward. This conformance by the participants to the hopes or expectations of the "investigator" is most obvious when information is sought regarding the use of and the reasons for the use of the program sponsor's product.

There are two other situations in which it is probably naively optimistic to expect to get accurate information. In the first place, people have long shown a proneness to say what it is believed will be pleasing to the investigator, when such a statement seems to have little or no effect on the welfare of the respondent. In other words, people are willing to say what is pleasing when their own welfare is not affected. This tendency often weakens the accuracy of marketing data when interviewers let it be known that they are working on behalf of certain sponsors. An interviewer who in-

introduces himself by saying that he is conducting a survey on behalf of the Dipsy Wiget Company helps to ensure, by that introduction, that he will receive biased reports regarding the use of and the reasons for the use of Dipsy Wigets.

Finally, there seems to be relatively little chance of securing accurate information by the conventional survey methods of marketing research on subjects of great emotional importance to the persons being investigated. This point needs no special elaboration.

We have just listed several situations or circumstances which add to the normal difficulty of securing accurate information regarding reasons for behavior or attitudes, etc. It is worth seeing to what extent they are apt to be troublesome in marketing research.

Deliberate deception may be important when there is a subordinate-superior relationship between the investigator and the respondent. This relationship is apt to exist and be troublesome in commercial research primarily when employers are investigating their employees. Such investigations are normally not considered to be marketing research, although they are frequently conducted by marketing research personnel.¹

Deliberate deception may also be important if the investigator can reward the respondent for certain types of responses. The main ways of avoiding this difficulty are discussed under Unwillingness because of anticipation of reward or loss, in section 13.5.

Another circumstance making it difficult to obtain accurate information is the proneness of the respondent to try to please the investigator. This problem is discussed under Unwillingness for emotional reasons, in section 13.5, and presents no serious difficulty for marketing researchers.

Finally, subjects of great emotional importance will often not be frankly discussed. This problem is of relatively small importance for the typical marketing researcher, who is interested in product testing, price analysis, readership surveys, and the like.

These special difficulties, then, probably will seldom be serious in marketing research, but they seem worth mentioning for the few occasions on which they will be of importance.

Eliminating these special difficulties leaves just the ordinary difficulties of getting people to be accurate in reporting on anything. As has already been seen, respondents feel the need to place themselves in a favorable light even for such persons as a strange interviewer who is not known and will probably never be seen again. Just as respondents often feel compelled to overstate their readership of periodicals which reflect prestige,

¹ We do not discuss this rather specialized problem here. For a good discussion of this problem see Daniel Katz, "Employee Groups: What Motivates Them and How They Perform," *Advanced Management*, September, 1949.

so respondents feel compelled to overstate the importance of motives which reflect generosity, indifference to small monetary considerations, etc. It is sometimes possible to reveal the extent of these biases when respondents are reporting their behavior by making objective checks or by measuring accurately some other variable which is thought to have a certain relationship to the variable being investigated. Unfortunately, in studying motivation it is almost never possible to make such relatively objective checks on the accuracy of what is said. Therefore researchers have frequently found it useful to "sneak up" on people and learn indirectly of their motivations. The technique used most often is the so-called "depth interview," which finds its widest use in studies of motivation.

The depth interview is a device whereby the respondent is stimulated to talk informally and rather exhaustively about the subject being investigated and the associations which this subject has for him. The depth interview is known by other names. The subcommittee of the Marketing Research Techniques Committee reports the following commonly used terms: qualitative interview, intensive interview, detailed interview, non-directive interview, conversational interview, and informal interview. Each term emphasizes certain attributes of this process of interviewing. All the terms, however, refer to a process which is flexible, lengthy, informal, and conducted largely at the discretion of the interviewer. Through this sort of interview the trained researcher can possibly infer the nature of the forces which have been important in determining the respondent's actions, attitudes, or intentions.

The technique has been borrowed from the psychoanalyst, who frequently seeks to understand the structure and determinants of a patient's personality through a technique of free association. A person freely associates when he utters the thoughts that occur to him without censorship or editing. The process of censoring and editing one's thoughts before they are verbalized is so pervasive that the use of free association by a respondent or patient requires a special training and a special environment. The psychoanalyst provides the training by allowing the patient to re-educate himself through trial and error and long experience. The analyst provides the special environment by physically isolating the patient from distractions and by attempting to create in the patient an attitude of complete confidence in the analyst.

It is, of course, impossible for any fieldworker to provide either the same training or the same environment that is provided by an analyst. The most that can be hoped is that the fieldworker will succeed in establishing rapport with the respondent—that is, a feeling by the respondent of wishing to help and cooperate in an atmosphere of relaxation and ease during the interview. If rapport is established, frequently the interviewer by

skillful suggestion and probing can cause the respondent to reveal his motives

It is helpful to consider the techniques of suggesting and probing as a process of getting the respondent to answer a "why" question without actually asking a "why" question. Often the suggestions or probing take the form of noncommittal remarks that merely serve to keep the respondent's conversation going. On other occasions the interviewer, by reacting to a respondent's stated point of view with an expression of a different point of view, can get the respondent to try to explain the differences and their causes. Frequently the interviewer can best do this by attributing the different point of view to some earlier respondent and by wondering out loud about the reasons for this difference. This method may stimulate the respondent to explain his own point of view and, also important, to explain the opposing point of view. This explanation often reveals more about the person doing the explaining than about the person being explained.

Naturally, during any interview whose conduct is so much at the discretion of the interviewer, it is very important that the interviewer conduct himself so as to encourage confidence and expansiveness. It is extremely important that the interviewer not appear to judge the wisdom or foolishness of a respondent's actions or the rightness or wrongness of his motives. Any indication that the respondent is being judged would probably lead to a bias by the respondent toward that point of view which the interviewer appears to favor. The respondent can often be encouraged in the belief that he is not being censored if the interviewer can make himself seem to be a person much like the person being interviewed. The interviewer can do this sometimes by indicating that he and the respondent have much in common in their attitudes and have shared many of the same experiences. It is also important that the interviewer appear to be interested in what the respondent is saying. Probably nothing succeeds so well in damming up a tide of self-expression as the appearance on the part of a listener of boredom or indifference. Sometimes a respondent's conversation may seem to drift away from the subject of interest to the interviewer, and it then becomes a delicate task for the interviewer to redirect the thoughts of the respondent without appearing too blatantly to do so.

The task of the interviewer is not only to stimulate the respondent to relevant self-expression but also to record and interpret what is said. The task of ferreting out from a mass of what is often largely irrelevant material those few revealing sentences and statements that cast light on the problem must be left largely to the interviewer. It is usually difficult for the interviewer to record in detail what is said, as it has frequently been noted that a recording of a conversation inhibits the spontaneity and ex-

pansiveness of the respondent. Aside from this danger, however, it would be desirable to record answers as fully as possible. Mechanical recording devices have been experimented with, but their use is certainly not widespread.

In most cases, therefore, the crucially important task of ferreting out the most important remarks must be left to the interviewer. Sometimes, in addition, the interviewer is further asked to interpret her findings immediately after each interview. This procedure has the advantage of permitting an interpretation by the person who conducted the interview while it is still fresh. Even when this interpretive task is left to the interviewer, however, the interviewer should record either at the time of the interview or immediately afterward the statements by the respondent which led to the interpretation presented. These statements can then be subsequently analyzed by a more highly trained worker who will have the additional advantage of being able to compare the results of many interviewers in an effort to discern individual biases or common patterns. Such reexamination frequently leads to reinterpretation and in some cases permits generalization about the important motivating forces for some large group.

The task of coding or summarizing the results of a depth-interview study involves all the difficulties of the free-response questionnaire method and many more. If this is not done well, the results of the study are almost meaningless to anyone who does not take the time to read over the summaries of the original interviews. Qualitative or impressionistic conclusions obtained from reading these interviews are dangerous. Almost everyone is too inclined to attach too much importance to the questionnaires which tend to confirm what he believed already and to gloss over the ones which do not. Hence depth interviewing makes very great demands on interviewer, coder (or summarizer), and analyst, and the final results are difficult to interpret with any objectivity.¹

In summarizing the role of depth interviewing in marketing research, it should be said that probably the greatest usefulness of this technique is in exploratory investigations. The lack of objectivity of the data produced by this technique and the almost inevitably inadequate caliber of field-workers indicate that depth interviewing should usually be reserved for suggesting hypotheses rather than for testing their validity. Depth interviewing can also be used as a check on information provided by other techniques.

¹ "Content analysis," a technique widely used in communication research, probably has some applicability here. The technique involves counting the number of times that words or ideas occur or measuring the amount of time or space given to ideas. The classification of ideas is, of course, crucial to this process.

Some research techniques, common in other fields but relatively new to marketing research, are sometimes useful for the same purposes as depth interviews. These techniques are "projective tests," such as the Murray Thematic Apperception Test (T.A.T.), the Rorschach test, and modifications thereof. These tests, which have been applied on a limited scale¹ to marketing problems, are still open to the charge of subjectivity in interpretation, but at least the interview situation can be made relatively uniform. In these tests each respondent is shown a series of pictures which depict relatively ambiguous situations and is asked to invent a story which explains the picture. With the Rorschach test, for example, the pictures are ink blots or irregular shapes. None of the pictures has an unambiguous interpretation, and the respondent's interpretation will tell us nothing about the picture but a great deal about the respondent. In applying the method to marketing problems, the pictures might vaguely suggest the subject about which information was to be obtained. Or, instead of pictures, written statements could be used. In the example cited in the footnote, respondents were shown shopping lists and asked to describe the kind of woman the shopper seemed to be.

Whether or not projective techniques have real merit for marketing problems is not yet clearly established. Enough work has been done, however, to indicate (1) that projective tests can be administered to cross-sectional samples rather than only to volunteers, and (2) that the people who have explored marketing applications feel that the technique has real promise. The lack of a detailed treatment of projective techniques here is due to the authors' lack of detailed knowledge of the field rather than a desire to deemphasize the importance of the approach.

2. Inaccuracy due to inability. The second major source of inaccuracy is not the deliberate deception of an interviewer by a respondent but the unrecognized deception by a respondent of himself. More plainly, people often do not themselves know why they do things; unconscious as well as conscious motivations exist. Advertising, for example, is believed frequently to create subconscious motivations. It is entirely possible that people may know of a product because of an advertising message that has long been forgotten. Such people, having to choose between two or three brands, sometimes choose the name that is familiar because of the advertisement. The advertising message was an important determinant of behavior, yet the respondent would be unable to indicate this directly. Some psychologists would go even further and contend that some advertising messages are received "subliminally," that is, absorbed without conscious awareness.

¹ See, e.g., Mason Haire, "Projective Techniques in Marketing Research," *Journal of Marketing*, Vol. XIV, No. 5, pp. 649-656, April, 1950.

This source of inaccuracy in statements regarding motivation is, for some types of investigations, probably more serious than all the sources of inaccuracy previously taken up in this chapter, yet communicative techniques are often inadequate to cope with it. It is possible that the depth interview or projective test might provide some useful information in such investigations. For example, an interesting application of projective techniques has been reported by Mason Haire. His problem was to study the motivations behind the purchase or nonpurchase of soluble coffee. Haire devised two fictitious shopping lists, one of which included Nescafé coffee (soluble) and the other, Maxwell House coffee (conventional). The two shopping lists were shown to comparable samples of housewives, and respondents were asked to describe the personality and character of the woman who bought the groceries on the shopping list. The main conclusions were summarized as follows:

1. 48 per cent of the people described the woman who bought Nescafé as lazy; 4 per cent described the woman who bought Maxwell House as lazy.
2. 48 per cent of the people described the woman who bought Nescafé as failing to plan household purchases and schedules well; 12 per cent described the woman who bought Maxwell House in this way.
3. 4 per cent described the Nescafé woman as thrifty; 16 per cent described the Maxwell House woman as thrifty; 12 per cent described the Nescafé woman as spendthrift; 0 per cent described the Maxwell House woman in this way.
4. 16 per cent described the Nescafé woman as not a good wife; 0 per cent described the Maxwell House woman in this way; 4 per cent described the Nescafé woman as a good wife; 16 per cent described the Maxwell House woman as a good wife.¹

The techniques for stimulating attention and recall by the respondent, which are discussed in earlier chapters of Part IV, may also be applicable.

Another possible attack on this problem is through the sales tests. For example, it might be extremely important to figure out which of two methods of shelf display would yield the greater sales of coffee. One might psychoanalyze respondents for years and never find out if the shelf arrangement made them buy Hills Brothers rather than Chase and Sanborn. A properly designed sales test, by contrast, might do the job very well.

¹ *Ibid.*, p. 652.

PART V
ADMINISTRATION

CHAPTER 18

SOME GENERAL COMMENTS ON THE ADMINISTRATION OF MARKETING RESEARCH

18.1 Introduction

Part II of this book was devoted to a consideration of the nature of scientific method and of its applicability to marketing research. It was shown that scientific marketing research almost always involved getting evidence (directly or indirectly) regarding consumers' behavior and expected reactions to specified marketing practices. Very often this evidence consists of primary data concerning consumers—data which must be collected by marketing research activities. Getting these primary data involves two problems—the problem of specifying the sources of information (sampling) and the problem of getting information from the specified sources (communication and observation). The first of these problems was the subject of Part III of the book; the second problem was the subject of Part IV of the book. In Parts III and IV we considered primarily the theoretical problems of sampling and of establishing unambiguous and accurate communication and observation. Only incidentally did we discuss the practical difficulties which beset any attempt to apply the methods which were considered. Actually, of course, these practical difficulties are of great importance, since the failure to overcome them can render useless the best plans for research. In Part V, therefore, we focus attention on these practical difficulties and the means for overcoming them. We consider, in other words, the problem of administration.

This first chapter of this part includes a discussion of some very general problems of administration; the subsequent chapters consider the problems of administering particular jobs, such as planning, reporting, coding, tabulating, interviewing, etc., which are encountered in most marketing research projects. The precision and lack of qualification which characterized many of the statements concerning sampling—and to a lesser extent communication and observation—cannot be duplicated in this discussion of administration. Administration is, in part, the art of meeting unforeseen circumstances quickly and efficiently. Rules of thumb tend to replace more authoritative generalizations. Part V, thus, even more than the preceding ones, consists of opinions and unsupported assumptions. Some of these opinions and assumptions may be helpful in meeting practical

problems, but it should be kept in mind that the ideas suggested do not rest upon a systematic analysis of empirical evidence nor upon a well-developed theory of the sort which underlies sampling, for example.

18.2 Organization of a Marketing Research Group

Organizations of many different types engage in marketing research. These organizations range from the one-man research "department" to the large and complex organizations which are found in some industrial enterprises and in many marketing research and advertising agencies. The diversity of forms makes any general discussion difficult, but certain activities are common to almost all research organizations, and thus some general discussion is possible.

As a preface to this discussion, we divide the process of marketing research into three major functions: (1) coordination of research with the needs of executives, (2) the design, supervision, and interpretation of research projects, and (3) the execution of projects which have been designed. (This last function obviously includes many "sub-functions," and these are discussed in subsequent chapters.) Two basic questions must be answered for each of these functions: (1) Who is to perform them? and (2) How are they to be performed? We shall consider the questions in the order named.

18.3 What Functions Should Be Performed Internally?

Clearly, the one function that is most difficult to delegate to external personnel is the coordination of research with the needs of management. It is almost impossible to delegate to an outside agency the task of recognizing the need for the initiation of research. Even though management and marketing consultants are freely used, the recognition of the need to use them seldom comes at first from persons outside the firm. If any specialization for research is to take place, the first function—that of coordination between research and management—is the one that will be performed inside the organization at all times. Occasionally sales managers and other executives, whose major responsibilities are not primarily connected with research, recognize the need for research and see that this research is performed by outside agencies. Hence the most rudimentary form of specialization is of this sort and involves the use of only part of the time of a single person for research.

A greater degree of specialization for research involves the use of a man whose responsibilities are solely concerned with research. Again, the specialized person must understand the needs of the company for research and must assume responsibilities for seeing that the needed research is performed. It is possible that some of the other functions as well will be

performed from time to time, but these functions, unlike the first, can be delegated to outside personnel. A single competent research man could, under certain circumstances, perform the second function—the design, supervision, and interpretation of research projects. The research man could even execute some small projects without having to use external personnel. He might quite easily handle a research project which involved the collection of secondary data.

But when primary data are needed—when questionnaire-type studies, sales tests, etc. are to be undertaken—the lone research man clearly needs help. Should the research man “farm out” the functions of execution, or should he attempt to assemble temporary workers for these duties? It is clear, of course, that he could not maintain a field staff, an editing and coding group, tabulation section, skilled computers, or a clerical staff—all of which are needed to handle a large primary study. It is possible, however, that these people could be assembled as needed for specific projects—that is, that temporary workers can be used when the need arises. The other possibility is to “farm out” such work to outside marketing research companies which specialize in one or more of the jobs of execution.

For the small department with its typically small budget, it would be difficult to do otherwise than “farm out” most of the work of execution. The assembly and supervision of temporary workers to accomplish specific projects is always time-consuming. One task alone—that of training, supervising, and maintaining a regional or national field staff of part-time interviewers—would be a full-time job for at least two individuals, even if primary studies were undertaken relatively infrequently. The functions other than fieldwork could be handled more easily, but even these would be likely to require more time than one person has. Tabulating, coding, and editing would also be demanding on the time of a given individual. The amount of time needed for computational work is variable, though typically less than the functions already mentioned. Clerical work is of such a continuing nature that it would probably be handled, at least in large part, without recourse to outside assistance.

There are specialists who handle each of these tasks, and most marketing research companies are equipped to handle them all. Under these circumstances, the research man's duties in executing projects should probably be chiefly those of providing control of quality—checking the marketing research company to whom he has delegated the tasks of execution to see that the specified standards have been met. This job of quality control is not easy and cannot be done casually, hurriedly, or haphazardly. Even the best marketing research companies need checking. The research man can no more hand them specifications and sit back to wait for the final results than he could forego checking in his own depart-

ment if the specialists of the marketing research company were working for him for the duration of the study. The ultimate responsibility for the functions of execution is always with the research man himself, however much he may delegate actual performance. Nonetheless, by delegating actual performance he may free a large part of his own time for coordination and planning.

Some suggestions on the use of an outside marketing research company may be helpful:

1. The outside company should be carefully selected. The research man in a small department should study the strengths and weaknesses of the different research companies. It is not enough to ask for "bids" and to select the lowest bidder; the variation in quality is likely to be great. In order to judge the relative competence of different marketing research companies, the research man should himself be familiar with the theoretical and practical difficulties of research. For example, he should be familiar with the problems of interviewing and the supervision of interviewers, even if these duties are always farmed out to an outside company. Indeed, one of the most important skills a research man can offer his company is a knowledge of the strengths and weaknesses of different marketing research firms which specialize in the collection of primary data. By hiring one man who has this knowledge, a company obtains access to the best outside specialists in marketing research whenever they are needed.

2. The specification of what is to be done should be clear and complete. It is desirable that the research company be consulted in the early stages of planning so that the full advantages of their suggestions can be obtained and so that the research company can be given the opportunity to point out specifications that they consider impracticable or unnecessary.

3. Once the project is planned, the design should not be changed frequently. If specifications are changed frequently and for slight reasons, the marketing research company will find it difficult to adapt to them. Higher costs, extra time, and deterioration of quality are all likely to result.

4. The research man should arrange for checks on the quality of the work. If the research company is a good one, it will administer these checks willingly and competently, but it is also desirable that some checks be made by the research man himself and that it be understood by both the research man and the research company that these checks are to be made. For example, the research man might want to be present at the early periods of instruction for interviewers. Again, he might want to read over a sample of the questionnaires as they come in from the field in order to see that the specifications are being carried through. (Detailed suggestions for control of quality are given in Chapter 21).

5. The research man should tell the outside agency as much as possible about all phases of the study. Most marketing research companies are experienced in a broad variety of studies, and this experience can have transfer value if the research company is fully informed about the origin and purpose of the studies upon which it is working. Moreover, when specialists at the outside agency are consulted and told about the purposes of the project which engages them, they are likely to be more interested and to do better work.

We now consider what functions are appropriate for larger research departments. If the research department consists of more than one person, it will probably be possible for more of the research operation to be performed internally. Of course, the first function—that of coordination with the needs of management—must be performed internally, and any increase in the scope of internal research must come from incorporating the second or third functions—design and execution of projects. If only one additional function is to be performed, the second function—that of planning, supervising, and interpreting projects—should be included. Clearly, needless complications and expense would be incurred by having the fieldwork and clerical operations performed internally while the job of planning projects and supervising their execution was left to outside personnel.

Often it seems desirable to specialize the research personnel according to the first two functions. Thus one individual—the “research director”—would perform the first function. He would be a sort of foreign minister in charge of relationships between his department and executives of the company—finding out their problems, suggesting ways of solving them, initiating appropriate research, and reporting on its results. A second individual—a “research analyst”—could then focus his attention on planning research to meet the needs as specified by the “foreign minister” and on supervising the execution and recording of such research. This specialization does not imply that the research director should not be a thoroughly competent technician. He should be this and more. But his time should be taken up primarily in the task of coordination.

As the research department grows, the most desirable organizational setup becomes more debatable. There may be a strong case for bringing some of the specialized functions of execution into the research department. It may be desirable, for example, to have a librarian in charge of locating, assembling, and organizing secondary data. If any of these functions are brought into the department, full-time supervisory personnel are needed, even though most of the actual work is done by temporary employees who are assembled as the need arises. Even for tasks such as editing, coding,

and tabulating, which to the neophyte might appear simple, at least one skilled full-time supervisor is needed if moderately large-sized surveys are undertaken.

Another direction of expansion is to bring a second (third, etc.) research analyst into the department and continue to farm out the functions of execution. A compromise is to bring in research trainees to supervise fieldwork and internal processing, do secondary research, etc. This gives flexibility, since the research trainee can be assigned the work of the research analyst when the need arises and as he becomes qualified. Moreover, such a procedure provides excellent training for future research analysts. There are, however, drawbacks. A field-staff supervisor should be a specialist. It usually takes years of experience before a person becomes professionally competent. If research trainees are to be given such highly responsible work, they are likely not to do it well at the beginning; yet when they become well trained, they are likely to be made analysts, leaving the supervision of the field staff to other inexperienced persons. Such a result would probably have unfortunate effects on the quality of the fieldwork. If fieldwork is to be undertaken at all by the marketing research department, it probably should be assigned to an experienced and qualified person who is willing to specialize more or less permanently. This person should be a kind of "permanent undersecretary" who should remain at his (her) post through successive changes of administration "at the top." For other functions of execution, the compromise of using research trainees may be much more desirable. Editing and coding, especially, might be handled by a capable research trainee.

Another direction for expansion of the research department is the hiring of part-time consultants who are expert in one or more phases of research. This expedient is often a practical way of meeting specialized and technical problems, such as are involved in area sampling.

There are other alternatives to these methods of expanding the research department. The choice of any particular alternative depends on the factors already mentioned and upon other things such as the qualifications of available personnel, the caliber of accessible outside research organizations, etc. By way of general comment two things can be said: (1) The second function—design of projects—should not be undertaken until the first is adequately performed, and the third—execution of projects—(aside, perhaps, from nonspecialized clerical work)—should not be undertaken internally until the second is adequately performed internally. (2) The work should be performed where it can be done most cheaply, but the search for savings should not lead to forgetting that visible savings in dollar expenditures may be more than offset by invisible losses in the quality of the research.

As the research department continues to expand, new organizational problems are introduced. In general these problems are not importantly different from those associated with other large-scale organizations. One major difficulty arises in measuring the quality of much of the work that is produced—a measurement which makes it possible to evaluate and supervise employees. Ways in which this problem can be handled are suggested in subsequent chapters. Some of the problems which are present only in rudimentary form in the small research department, however, exist in greater complexity and with greater importance in the larger department. It is appropriate at this point, therefore, to consider two instruments of control which are useful for research departments, of whatever size, but which are especially important for the larger departments. These instruments are the research budget and the cost estimate.

18.4 Two Instruments of Control

The research budget is an accepted estimate of the magnitude of the routine expenditures for research during any specified period. It is the amount which management is willing to commit to research. Aside from the importance of having enough funds to support needed research, it is extremely important that there be some certainty about how much funds will be available for research if the research is to be efficient or effective. In other words, if research is to be efficient and if it is to be used to meet problems other than those which have become crises or emergencies, there must be some firm financial commitment which can be used as the basis for long-range planning.

Some research projects can be anticipated in sufficient detail to permit their underwriting long in advance, but many projects cannot be foreseen in such detail in spite of the utmost desire by management to make long-range plans. In these circumstances, the unforeseen research is typically provided for by specific appropriations which do not appear in the budget. The budget, however, should provide for adequate funds for exploratory work designed to indicate the need for such research. The consideration by management of the need for particular research projects as they arise probably results in a smaller volume of research than would result from a blanket underwriting of all research. The review by management of the merits of individual projects is often a stimulus to well-conceived and carefully planned research, although a blanket underwriting would to some extent facilitate long-range planning.

The cost estimate is the second instrument of control. It is the estimate of cost of particular research projects. These estimates provide the basis of the specific appropriations referred to in the preceding paragraph. Such estimates are essential to the intelligent use of research, since it is

clear that research should be undertaken only when the value of the information which it would provide is felt to be at least as much as the cost of obtaining it. Cost estimates are useful not only in indicating whether research should be undertaken but also in controlling and estimating the efficiency of the various operations which the research involves.

Before going into some rules of thumb which may be helpful in achieving accurate cost estimates, it is important to point out that estimates have very little usefulness unless they are related explicitly and specifically to time requirements and specifications of quality. The sooner information is available the more valuable it is. Also, research provides information of varying accuracy. A statement of the expected cost of estimating the magnitude of current sales to consumers would be of limited value unless there were also some statement of the expected accuracy of the estimate. The accuracy depends upon the sampling process, the care with which communication and observation take place and are supervised, the procedures for doing the tabulation and other clerical work, and many other things as well. A project which based its estimate of sales on information from a poorly designed sample, which employed poorly trained and unsupervised fieldworkers, and which made inadequate provision for supervising and checking the quality of the clerical work involved would probably be substantially less accurate than a project employing a well-designed sample, thoroughly trained fieldworkers, adequately supervised clerical workers, etc. The importance of time and quality thus make it essential that any cost estimate be accompanied by detailed specifications of the way in which the research is to be conducted and the expected time of its completion.

A by-product of this detailed specification of the research project and its attendant costs is that the researchers must give prior careful consideration to the entire research operation. Such careful consideration of each step of the proposed research can be expected to result in better research than would result from the initiation of research without such careful prior planning. Although there is much to be said for freedom from the restrictions of detailed specifications of what will be done in the exploratory stages of research, once a specific project has been decided, the existence of detailed plans and specifications is often a guide and spur to careful research.

It would be reasonable and prudent, perhaps, at the time of the cost estimate to have the research specified or planned in the following respects:

1. The objectives of the project should be precisely defined and stated.
2. The questionnaire, if one is to be used, should be in its final form. In other words, the questionnaire should already have been pretested and revised.
3. The sampling design should be specified in detail. For probability sampling

this means the specification of clusters (especially counties) that are to be sampled. For nonprobability sampling, this means the specification of the quotas or procedure of selection.

4. The coding specifications should be as nearly complete as possible before the receipt of the questionnaires from the field. If IBM machine tabulation is to be used, this would mean that the use of the columns in the cards should, in a general way, be established.

5. The main tabulations should be specified. This requires planning, prior to the execution of the research, the tables which will be used in the final report. Not all needed tabulations can be planned in advance, since the consideration of the actual data will often suggest ideas and hypotheses not thought of in the planning stages of the research.¹ Usually, however, in well-planned research most of the tables that will be needed in the final report can be anticipated.

Although such detailed planning is often difficult or impossible, the making of accurate cost estimates is difficult without it. Such planning can be slighted if there is some previous study of a similar nature which can provide the basis of the estimate. Unfortunately for the easy making of cost estimates, however, studies in marketing research are remarkably varied.

When such planning is feasible, it becomes necessary to estimate the costs involved at each step. The best basis for doing this is the previous experience of the research department. Although the previous experience of the department is usually inadequate to make possible an over-all estimate without a detailed analysis of the proposed research, this previous experience can usually provide the basis for an accurate estimate of the various steps of a carefully planned project.

For example, data may sometimes be available on the various costs of interviewing, such as (1) the verbal exchange between respondent and field-worker, (2) traveling between interviews, (3) the training of fieldworkers, (4) field editing of the questionnaires and correspondence with the central office, (5) "unallocated time," and (6) other things. The more detailed the breakdown, the greater the likelihood of an accurate estimate. This breakdown and use of previous experience cannot be mechanical; it requires knowledge of the process of research. It would be foolish, for example, to apply figures on the cost of interviewing a judgment sample to the task of estimating the costs of interviewing a probability sample of the same population. The probability sample would typically involve higher costs per interview because of the necessity of call-backs, etc.

Naturally, the use of previous experience is facilitated or made possible by the existence of records based upon intelligent cost accounting. This,

¹ Such hypotheses cannot be rigorously tested by the data which suggest them but they are nonetheless of great value.

in turn, rests upon the existence of some record of the time and money spent by the various types of employees in performing their various functions. This might mean that interviewers, for example, would have to estimate the average time required to complete an interview or even, when individual interviews are long, to record the time needed for each interview. By comparing this information with the total time required to finish the field assignment, it would be possible to obtain the division of time between interviewing proper and all the other components: travel, training, correspondence with home office, and "unallocated." Training is rather easy to assess directly, and correspondence with the home office can be estimated roughly. The residual is travel and "unallocated."

But obtaining this information is not enough; it is necessary to analyze the cost relationships for different types of studies and different types of interviewing conditions (time of year, rural or urban, etc.). From this one may hope to obtain valuable evidence for forecasting the cost of future studies. It is often possible to obtain supplemental evidence from the pretest. For example, the pretest can indicate well the average time required for conducting an interview.

By such analysis, one may not only hope to predict more accurately, but one may also detect poor performance and improve the performance in future studies. Suppose, for example, that it was found that one interviewer spent an average of two hours per questionnaire in actual interviewing, while the average was twenty minutes for other interviewers working on the same study. The supervisor could then investigate this particular interviewer to find out whether (1) two hours is required to complete the questionnaire correctly; or (2) the interviewer misunderstood instructions and included traveling time in her estimate of the length of each interview, was in the habit of following each interview with a long social conversation, didn't understand the questionnaire very well, or was simply dishonest. No matter which of these possibilities is really right, an investigation is justified for the sake of better performance on future studies.

Cost analysis of the type that has been sketched here in rough outline is expensive and time-consuming. The costs of making such analyses can be reduced to some extent by sampling procedures. One might, for example, study only a sample of the interviewers working on each investigation. Sampling methods have wide applicability to any kind of administrative problem, and marketing research is no exception. Nonetheless, even with the enlightened use of sampling, the analysis is likely to be expensive.

Is it really worth while to go into such detailed analysis? If standards of quality are to be maintained, time requirements to be met, and cost limitations not to be exceeded, such cost analysis is essential. If, for the

small or medium-sized marketing research department, cost analysis seems unfeasible, then it would probably be wise for the research department to delegate the functions of execution of research to outside marketing research companies which are large enough and which have sufficiently continuous work to make these "luxuries" possible. But once the functions of execution are undertaken, the responsibilities for efficient execution should not be shirked. The prediction, analysis, and control of costs are required.

No matter how detailed and scientific the analysis of costs which underlies the cost estimate, one should never forget to make adequate allowance for contingencies. Some experienced practitioners typically add 30 per cent to the cost estimates which are submitted in order to allow for the unexpected. Obviously, such rules of thumb are never completely satisfactory and must be tempered by knowledge of the particular problem. Yet there should be definite allowances for expenses that have not been anticipated in the detailed estimate. Accurate cost analysis of the experience of previous studies can reduce the magnitude of unanticipated contingencies, but almost inevitably some unexpected expenses will be incurred.

Occasionally an unfortunate by-product of the liberal provision for contingencies is the tendency to neglect close supervision and efforts at efficiency in the belief that any excess of expenses over the estimate can be charged to the contingency fund. Certainly the routine establishment of a contingency fund lends itself to such abuse. Nevertheless, it is probably better to attempt to control this abuse (perhaps by requiring that any unplanned expenditures receive the approval of a nonresearch official) than to underestimate the cost of research by failing to allow for the unexpected.

CHAPTER 19

COORDINATING RESEARCH WITH THE NEEDS OF MANAGEMENT

19.1 Introduction

This chapter is concerned chiefly with the function of coordinating marketing research with the needs of management. This function can be subdivided into two major steps: exploration and interpretation for management. The discussion is focused on the performance of this function in the research department of an enterprise whose primary business is manufacturing, although with obvious modifications the ideas presented will apply to independent marketing research agencies, advertising agencies, etc.

19.2 Exploration: Defining the Problem

In Part II, on scientific method, there was discussion of the procedures and criteria by which research could indicate the existence of problems. At that point the theory of problem recognition and the role of problem recognition in scientific inquiry was of primary interest; at this time the administrative processes by which that theory can be implemented are of primary interest.

Almost certainly, exploration receives less time and effort, both in practice and in theoretical discussions, than it deserves. The major reason for this inadequate attention in practice is the disparity between the pace of business and the optimum pace of exploratory work. In business research, research is speeded up rather than business slowed down. There is pressure to produce concrete results: to do surveys, to write reports. The relatively undramatic and often apparently sterile work of preliminary exploration is slighted. In abstract discussions of this sort, the work of preliminary explorations is often slighted because of the absence of any clearly defined theory or any empirically tested body of procedures. Although the following discussion, too, suffers from these deficiencies, it is hoped that the crucial importance of the subject will be appreciated.

Preliminary exploration takes place in order to clarify the nature of recognized problems and to ascertain what information is required for their solution. We shall now proceed to some suggestions for ways of doing these things.

1. Knowledge of the enterprise, the industry, and the general business environment. It is easy to overemphasize the importance of method and technique and to understate the importance of factual knowledge. The temptation is probably greatest for students and inexperienced practitioners whose knowledge of techniques usually far exceeds their knowledge of the facts of the company's operations, the structure of the industry, and the status of the general business environment. Professor Hayek makes this point strikingly:¹

Today it is almost heresy to suggest that scientific knowledge is not the sum of all knowledge. But a little reflection will show that there is beyond question a body of very important but unorganized knowledge which cannot possibly be called scientific in the sense of knowledge of general rules: the knowledge of the particular circumstances of time and place. It is with respect to this that practically every individual has some advantage over all others in that he possesses unique information of which beneficial use might be made, but of which use can be made only if the decisions depending on it are left to him or are made with his active co-operation. We need to remember only how much we have to learn in any occupation after we have completed our theoretical training, how big a part of our working life we spend learning particular jobs, and how valuable an asset in all walks of life is knowledge of people, of local conditions, and special circumstances.

The marketing researcher must acquire an intimate knowledge of "the particular circumstances of time and place" if he is to realize the full potentialities of the application of scientific method in attacking business problems. He must become as much an expert in the particular field to which he applies his techniques as he is in the techniques themselves.

There is no easy way of acquiring such knowledge. Academic courses in economic theory, monetary theory, institutional economics, economic geography, etc., can be helpful. Knowledge of the industry can be obtained in part from studies of the industry made by economists or governmental bodies and the information given by trade papers for the industry, and in part by experience. Knowledge of the particular company may be obtained in part by a study of its records and published history, but much valuable knowledge can probably best be secured by experience. A few words on the nature of "experience" may be helpful.

Experience in working in the research department on research studies is helpful, but it is not enough. Since marketing is primarily concerned with selling, experience in selling is likely to be extremely valuable for the marketing research man. Any kind of selling is helpful, but selling for the particular company for which the research is done is most helpful. Some research departments make arrangements for research trainees to

¹ F. A. Hayek, "The Use of Knowledge in Society," *The American Economic Review*, Vol. XXXV, No. 4, pp. 521-522, September, 1945.

spend several months or a year "in the field" with the operating divisions of the company. When this is not possible, the research man should make every attempt to take short tours of duty with the selling force. When specific selling problems arise, he should try to spend a little time in the field so that he has a keener appreciation of the difficulties involved. One experienced research director admitted that in two days of selling he had found out things about the product his company sold and the problems of selling it that he hadn't learned in months of research experience. It works the other way, too, of course: the research man finds out many things by his specialized techniques of obtaining information that "practical" executives and salesmen may not have learned in years of experience.

In addition to learning about the practical problems arising in selling, the research man should learn as much as he can about the organization of his company. He should find out who is responsible for what, who has the real authority, and how things get done. Reliance solely on formal organization charts (if there are any) should be avoided, since they often are only idealized fiction. There may be great discrepancies between the formal authority indicated by the organization chart and the informal authority that is operative.

2. Relationship between research and management. The most difficult part of research work is establishing and maintaining a close relationship with the executives of the company. This is sometimes difficult because some executives are suspicious of research and regard research as an infringement (or potential infringement) of their own prerogatives. There are only two ways to overcome this suspicion. The first is to have a high executive officer issue an order. This may be helpful as a means of drastic reform, but it is not likely to create genuine rapport between the executives and the research department. The alternative is diplomacy, which gives an opportunity for face saving whenever unpleasant information is unearthed and which is backed up by a record of good and helpful performance.

Sometimes the drastic solution is both necessary and feasible, but we shall focus attention on the alternative. Here are some suggestions for making that approach work:

1. The research director must know the executives personally and spend time with them as a matter of routine. The same injunction applies with less emphasis to the research analysts. (In the succeeding paragraphs "analyst" is interchangeable with "research director.") Informal contacts are usually much more useful than formal ones and should be encouraged in every way.

2. The research director must learn as much as possible about the problems faced by the executives. In this way he is in a much better

position to make realistic suggestions which will command the respect of the executives.

3. The research director should not give advice unless it is requested. The function of research is gathering and evaluating information, not telling executives how to discharge their responsibilities.

4. Whenever a specific problem is being investigated, the research director should find out everything the executives know about it before he does anything else. This requires interviewing which is as difficult as any encountered in a survey. Executives typically do not formulate verbally everything they know, and the research director must be patient and skillful in his interviewing. Needless to say, good personal relationships are essential.

5. Executives should be encouraged to take an active interest in research methods. Such an interest can create not only a more sympathetic view of the difficulties of research but can be the source of helpful suggestions regarding the techniques of research. Naturally, any suggestions by executives concerning research methods should be heard and weighed carefully, but not given special deference solely because of the source.

6. The executives should be informed frequently, either orally or in writing, of the progress of the preliminary investigation. If it is necessary to delay this investigation in favor of a more pressing problem, they should be told rather than left in suspense with the general impression that things are proceeding on schedule. Whenever they express any doubts as to the way in which the investigation is being conducted, they should be given a full explanation and asked for suggestions and comments. If the doubts arise from a general uneasiness, a simple explanation is usually enough. If the doubts arise because an executive has specific suggestions, these suggestions should be considered as if they had arisen within the research department.

7. Whenever it appears that research is likely to unearth information that will put an executive in an unfavorable light, this information should be shown to the executive concerned rather than broadcast. When such information is found, moreover, the executive should be given credit for initiating the research which led to these findings. This credit usually goes a long way toward reducing any stigma arising from failure to investigate sooner.

8. The executive should be given information in the form he best likes. If a full-blown study is recommended as a result of the preliminary investigation, he should be given as much detail as he wants. Questions should be encouraged, and answers should be given in language he will understand. No effort should be spared to get the executive's support for the project *before* it goes out in the field.

9. The research man should not make snap judgments and quick suggestions, even when asked, unless he is very sure of his ground. There are two good reasons for this suggestion. The first is that an executive who has been thinking about a problem may be suspicious of an outsider who has a lot of glib answers and suggestions. The second is that hasty judgments are often wrong and have to be revised later. Too frequent changes of direction by the research man may diminish the regard in which he is held.

These suggestions are helpful but not complete: many situations will arise which are not covered by them. Dealing with each individual executive is a special problem and must be recognized as such. The real danger is in spending most time with the executives who are congenial to research while neglecting the ones who are more difficult to deal with. The ones who are hard to deal with can ultimately be persuaded, if at all, by a demonstration of concrete results and, equally important, by seeing that the research department is not a threat to them. But in the meantime, they should not be ignored simply because they are uninterested in or even hostile toward research.

We hope that we have not unwittingly given the impression that the personnel of the research department should be like the popular conception of "striped-pants" diplomats, always displaying good manners and never becoming involved in unpleasantness or controversy. Many times the research department personnel will find it necessary to take positions that will be unpopular with some or all of the company's executives. If the research department always agreed with all the executives, it would have little or no justification for existence. It is important to realize, however, that trivial disagreements can diminish mutual confidence between line executives and the research department, diminishing the efficiency of both.

3. Sources of secondary data. One of the first steps of any preliminary exploration should be an investigation of relevant secondary data. In order to avoid a wasteful duplication of research expenditures, it is essential that the records of previous research be examined. Often these records of previous research require diligent investigation. For this reason, the following comments seem appropriate.

The need for secondary data often is fairly urgent, and the task of finding it can be time-consuming unless the research director has been far-sighted enough to make the bulk of possibly useful information readily available. The best way to do this is to have a librarian. As the research department is expanded, the hiring of a librarian should have a high priority. This person should be familiar with the requirements of research, be well acquainted with the marketing problems of the company, and, of course,

be familiar with the sources and nature of as much relevant secondary data as possible. The understanding and skill of the librarian should be adequate for the task of summarizing those sources of data which might be useful. Much of this data should be assembled and filed by appropriate categories. Naturally, the important basic sources such as the *Industrial Arts Index*, *The Reader's Guide to Periodical Literature*, the publications of the U.S. Bureau of the Census, etc. should be on hand. Other sources of bibliographical material and secondary data, such as the trade papers (e.g., *Advertising Age*, *Printers Ink*, *Sales Management*), academic journals (e.g., *The Journal of Marketing*, *Public Opinion Quarterly*, *The International Journal of Opinion and Attitude Research*), the reports of the publishing companies, etc., should be regularly scanned for information on methods of research, bibliographical material, and secondary data.

4. Exploration in the field. If secondary research can be left to a capable librarian, the research director or research analyst should get into the field as quickly as possible to find out things by direct conversation, observation, and experience. If the secondary research cannot be delegated to a specialist, the research director should probably explore the most fertile secondary sources immediately and then go to the field; the remaining secondary research can be sandwiched in between field explorations.

There is no set pattern for field investigations. The initial step is to talk with executives in the company who are interested in the problem being explored, in order to find out everything (or as much as possible) of what they know or think about the problem. From then on, the research man's experience and ingenuity are his main guides. We shall suggest certain approaches which may be fruitful.

1. Exploratory Interviews with (or Observation of) Consumers, Dealers, Salesmen, etc. After talking with the company executives to find out what they know about the problem, the research man should outline a course of informal exploration. In doing so he should first write down as many questions as possible. This first attempt will be disorganized and incomplete, but once made, it will serve as the basis for further refinements. As information comes in, it should be examined for further suggestions as to questions and tentative answers.

For most consumer products which are sold through retail outlets, the next step is usually to visit retail stores, to browse around, observe and talk to consumers, talk to retail clerks, store managers, etc. There are various ways in which this can be done. When the research man is still in the early groping stages of problem exploration, it is perhaps best for him to pose as a consumer himself. For some products, such as women's underwear, this approach may not be very helpful because the research

man will be a sufficiently unusual customer to receive atypical attention from salespeople. The success of this kind of informal shopping around depends on the research man's ability to appear as a *bona fide* customer and to keep from arousing suspicion. He may be as inquisitive a customer as he desires, but he should be a realistic one.

This kind of preliminary exploration serves chiefly to acquaint the research man with the background of the problem he is looking into, to suggest further questions that need answering, and to suggest tentative answers to the questions he has already formulated. The time that can be devoted to this exploration will vary; it is desirable that it continue until the research man feels that he knows enough about the subject to ask a sequence of intelligent questions. At this point it is well to write these questions down in rough questionnaire form and turn this part of the investigation over to trained interviewers who can conduct further dealer and consumer interviews with this questionnaire.

It may be necessary not only to visit stores but to call on consumers in their homes. Many things about the problem can be found out in this way which cannot be discovered by browsing around stores. Browsing can give an idea why people buy and why dealers think they buy, methods of retail selling employed, etc., but it may indicate little about consumers' experiences with the product. Knowledge of such experience is important in almost all problems, and the researcher must talk directly to people who own the product and have used it.

The next stage is talking to important people in the distribution process: chain-store managers, store owners, wholesalers, company salesmen, etc. It is important that this part of the exploration be postponed until the research man knows something about the subject from his browsing in stores, talking to consumers, and reading secondary information which has been accumulated. For example, if the research man can relate an anecdote or two about consumers he has talked to and in other ways demonstrate his familiarity with the field, he is much more likely to get information from an executive officer in a large wholesale house. For one thing, the executive will respect the research man for knowing something about the field, and for another, the executive may feel that the research man can give him some valuable information during the conversation.

The length to which the preliminary explorations should be prolonged depends on the point at which new questions cease to occur to the research man. The objective is not to arrive at answers unless the answers are exceedingly obvious; rather, it is to formulate the questions that need answering. The research man will inevitably feel that he has learned a lot of the answers, but he must be cautious about this. Personal impressions formed in such an unsystematic way are seldom the final answers.

If such unsystematic investigation were adequate, formal research studies would seldom be needed. Many problems require formal as well as informal study.

Even in such casual exploration, however, there is a problem of sampling. To find out about the sales of tape and wire recorders, the investigator would not confine his explorations to "downtown" department stores. On the other hand, of course, he would not be much interested in a probability sample. It would suffice simply to visit different kinds of retail outlets known or thought to carry tape recorders: appliance stores, department stores, radio stores, etc.; large stores as well as small ones; stores located in poor neighborhoods as well as in well-to-do neighborhoods; small-town stores as well as city stores. As always in sampling, the quality of the sample as well as its size must be considered in determining its usefulness. Similar requirements apply to consumer interviews in this exploratory stage: the sample should be "broadly distributed," although no attempt need be made to secure a probability sample.

5. Decision as to whether formal research is needed. The end of the preliminary exploration of a marketing research problem comes with the decision to conduct or not to conduct a formal study. The information gained from background knowledge of the company and the industry, contact with company executives, the survey of secondary sources, and preliminary work in the field should all be used in coming to this decision.

It is not at all necessary that every problem given to the research department should be subjected to formal study. The preliminary explorations of the kind that we have suggested on the last few pages may often suffice to indicate that what was conceived at first to be a problem is not really a problem or that it is a problem that does not require formal marketing research. As an example, suppose that in the early postwar period there had been widespread consumer dissatisfaction with washing machines and that this dissatisfaction had been shown by the preliminary investigation to be caused by certain substitute (plastic) materials which were being used in the manufacture of the agitator. This problem would have been solved as soon as government allocations of critical metals could be eased; there would be no need for marketing research to pursue the matter further.

Of course, there is a danger that the marketing research man will conclude from his preliminary investigation that the answers are obvious when in fact they are not. In spite of this danger, however, the marketing researcher should always ask himself whether a full-blown study is necessary, even in cases in which everyone assumes that such a study is needed. If the preliminary investigation results in evidence which seems to be conclusive—if personal contacts and secondary sources all point unambig-

uously toward the same answers and it is judged that these sources are reasonably satisfactory and extensive—there may be no need to go further. Moreover, frequently few deeds can more enhance the position of the research man within his own company than his saying that no further investigation is needed, assuming that he is right.

19.3 Design of Projects

If the preliminary exploration has shown that further research is really needed, the next step is to design the study. The preliminary exploration, if conducted properly, will be invaluable in facilitating study design. Most important of all, it will indicate in detail and with precision the important questions that need to be asked and answered. In addition, it will give hints as to the methods appropriate for getting the answers.

Before becoming intensively preoccupied with the details of study planning, it is important to outline in detail the information needed. It is probably wise to set up the outlines of the tables that will be the basis for the final report. If the preliminary exploration was adequate, this will be easy; the main questions that need answering will be clearly indicated. It may even be helpful in doing this to fill in imaginary answers in order to see how the final report would appear if these answers were right. Even though this step requires time and money at the start of the study, it will probably save time and expense in the long run by focusing the study more sharply on relevant problems and by eliminating waste motion. It is usually not possible to anticipate all the analyses of data procured in a study, but much can be done.

The next step is to consider all the promising methods of obtaining the information that is needed. Research men, like any other specialists, develop fondnesses for techniques that have been useful in the past. One researcher will prefer the consumer survey, a second the sales test, and a third an analysis of internal accounting records. There are similar preferences for alternative approaches to sampling: some researchers will insist on nothing but probability samples, while others will use only judgment samples. And so it goes. There is much to be said for selecting the techniques that have appeared to be most useful in the past, but it is also important to realize that each research investigation is unique in certain ways. No problem is just like an earlier problem, even though the resemblance may be close. Moreover, research techniques are improved as time passes. The probability sampling designs of today may be too expensive for consideration on many problems in marketing research. Next year or the year after, sampling experts may develop probability sampling designs that can be applied to such problems. The psychological techniques of today may offer little insight into subconscious motivation, but

developments of projective testing methods may permit progress in investigations of motivation.

There are no objective procedures for selecting the "best" research technique. The researcher's own scientific and business judgment are essential ingredients in a wise choice, but equally essential is familiarity with all the currently available methods of analysis. This latter requirement does not mean that the research man must be expert in all techniques of research, but it does require that he be familiar with them and that he know where expert advice can be obtained when his own knowledge is meager.

In continuing our comments on the design of studies, we shall focus on only one of the research techniques which might be used—the consumer survey. It should be evident that we do not believe that this approach is the only or even the most important tool of marketing research. It is, however, one of the most complex to administer and for this reason will serve well as an illustration of the administrative problems that arise in marketing research. We shall assume, moreover, that the research department itself is planning to execute the study. The project design must then be carried through in somewhat greater detail than otherwise.

1. Rough sketch of the sampling design. Before the detailed planning goes too far, it is important to set up the outline of the sampling design. This usually involved, first, the choice between probability and judgment sampling. Once this decision is made, the size of the sample should be considered and set tentatively. The guiding consideration is the accuracy of the information: how much accuracy is needed and how much accuracy the sampling design will yield. Even though there is a large element of personal judgment in setting standards, the resulting specifications are likely to be more satisfactory than they would have been if the question had not been considered. When sampling methods are suggested to replace complete enumerations, as in sampling inspection rather than complete inspection of industrial output, there is usually resistance because of the fear of a loss in accuracy. For most marketing decisions, however, there is room for a certain amount of inaccuracy in the data used in making the decisions. If the researcher recognizes this fact and understands the particular problem with which he is dealing, he can often make reasonable decisions as to the required degree of accuracy.

Often a survey is designed to obtain several types of information, and the question of determining the required accuracy is complicated. On some of the items of information, only a very low degree of accuracy is enough; on others, much higher standards must be met. Perhaps a useful procedure is to use as a standard the degree of accuracy required for the most important individual item of information being sought.

Sometimes (perhaps typically) the exact or approximate amount of the funds available for the research investigation is fixed in advance. In this case one can proceed immediately to the step of estimating how much accuracy the sample will *give*, rather than speculate on how much accuracy is really *needed*. Even then, however, it is desirable to speculate on the accuracy really required; for if the funds allotted seem to be either more than ample or less than enough, the research department should recommend that the allocated amount of money be changed.

Once the researcher decides, then, the necessary degree of accuracy, he must estimate the degree of accuracy that the tentative sampling design will yield. This also is difficult. If the only errors arising in sampling surveys were sampling errors (*i.e.*, if response errors were absent), and if a probability sampling design were used, the researcher could estimate in advance the degree of accuracy that would be obtained, provided that he had some rough information about the population from which the sample is to be drawn. In practice, however, response errors are likely to be important, and they can be estimated only crudely. Information gained from talking to consumers in the preliminary exploration of the problem may be of some assistance, and later, when the pretest is carried out, it may be possible to obtain a better estimate of response error.

When judgment rather than probability sampling designs are used, an estimate of sampling error is really impossible, at least by objective methods. As pointed out in Part III, the degree of accuracy of a judgment sample cannot be appraised solely from a knowledge of the method by which the sample was chosen. If, however, a judgment sample appears to be the only satisfactory alternative, it is necessary to make a guess at the degree of accuracy obtained. Errors of response, of course, will be present in judgment sampling as well as in probability sampling.

It is plain from what has been said here and from the discussions in Parts III and IV that any discussion of the accuracy (required or obtainable) of a marketing research survey is unfortunately not rigorous. Even under the best of circumstances, there is need for guesswork and judgment, particularly with regard to the determination of how much accuracy is needed. Nonetheless, it is desirable and useful to make a rough decision at the start before the planning of the survey goes too far. For once the sample size is set, it will be possible to make a very rough estimate of cost, and this estimate can be shown informally to management in order to see whether the figure is considered reasonable. If there are objections by management, it is cheaper for revisions in plans to take place at this stage in the study rather than later.

When the rough design of the sample has been approved, the planning of the survey should move ahead at a quickened pace. The next steps

are the development of a questionnaire, the revision of the questionnaire, the final selection of the sample, planning of the fieldwork, etc.

2. Developing the questionnaire. Questionnaire writing might seem to be one of the simplest tasks in marketing research, yet it is in fact extremely difficult. Indirectly, we devote a large part of this book to questionnaires (Part IV). It may be helpful to draw upon the general principles discussed in Part IV and to make some practical suggestions for questionnaire writing.

1. An essential first step is to list explicitly the information which the preliminary exploration has indicated as being necessary for the solution of the problem being investigated. Next, it is essential to devise questions which will provide this information without repetition or gaps. An alternative approach is to consider the tables which the preliminary investigation has indicated as being necessary as a basis for the final report and to consider what questions will provide the data for those tables. If the preliminary investigation has been competent, the major problem is not that of deciding what questions should be asked but of deciding on the wording of the questions.

2. The sequence of questions usually should be from the general to the specific in order to make as clear as possible to the respondent the rationale of the questionnaire and to focus his attention on the subject being investigated. For this reason, general questions are sometimes included even though the answers to them are of little importance. The desirability of making the organization of the questionnaire clear to the respondent makes it wise also to include questions which indicate the transition from one subject to another.

3. The questionnaire should be no longer than necessary. Although there has been no experimentation to determine the relationship between the length of questionnaires and the extent to which respondents cooperate, there are many experienced researchers who have felt that their experience indicated the deterioration of cooperative effort as interviews became longer.¹ Thus there is probably good reason to curb the temptation to ask questions in the vague hope that they may somehow provide useful information. Effort should be concentrated on the questions which are essential to obtain the information which is clearly required by the study design. Brevity of questionnaires is usually more important for marketing research than for surveys in other fields, since the interest of most respondents in the subjects which are investigated by marketing

¹ The Bureau of Labor Statistics, The Survey Research Center, and other agencies with much research experience have had what they feel to be satisfactory results with questionnaires requiring more than an hour to complete. This is longer than most marketing research agencies have thought it wise to take.

research is usually much less than in politics, sexual behavior, or family relationships, for example.

4. The language of the questionnaire should be familiar and simple.

5. The questionnaire should be free of ambiguities and should be tested to this end. This testing is usually in the form of a pretest. (The pretest serves other ends, too.) Prior to the pretest, however, it is possible to detect some ambiguities and other flaws by submitting the questionnaire to members of the office staff either of the research department or of other departments.

3. Pretesting. Pretesting is the use of the questionnaire on a small sample of respondents in order to find weaknesses in the questionnaire rather than to secure information on the subject of the questions. The pretest also provides information on the cost of interviewing which can be extremely useful in making cost estimates. Moreover, it may be used to try out the sampling method, training of interviewers, interviewing approaches, etc. Since written instructions for interviewers are often needed ultimately, it is very helpful to prepare them in advance of the pretest so that they can be tried out.

Pretesting is of great importance, and it should not be omitted unless the urgency of the project is such that pretesting is impossible. Almost never will a pretest fail to indicate the possibility of substantial improvements in the questionnaire or in other aspects of the survey. Sometimes more than one pretest is required when the first pretest indicates the need for great revisions in the original questionnaire. The revised versions should be pretested until they are satisfactory.

Usually from one to four interviewers are used in the pretest. Often the research analyst participates in the interviewing, but it is important that persons other than those who formulated the questionnaire conduct interviews. The authors might be insensitive to the weaknesses of their creation. It is also useful to use an interviewer who is very good and one who is only mediocre. The good interviewer will be acute in recognizing flaws and in making constructive suggestions. The mediocre interviewer will, by the difficulties which her completed interviews indicate, suggest changes that should be made if other interviewers are to avoid similar difficulties.

It is desirable that the research analyst discuss with the interviewers the experience of the pretest, questionnaire by questionnaire. The good interviewer, of course, will have independent suggestions of her own and may be capable of presenting a summary of her experiences. Nonetheless, the questionnaires should be reviewed one by one, both to check on the adequacy of the interviewers' summary of their experiences (interviewers may be unduly impressed by one difficult interview or one particularly

critical respondent) and to give the research analyst a more concrete impression of the problems encountered. For the mediocre interviewer, this detailed review of the work is especially important. She should be queried on mistakes so that the reasons underlying these errors (either ambiguities in the instructions or ambiguities in the questions themselves) can be uncovered and corrections made.

The sample of respondents interviewed on the pretest need not be large (often, or even typically, not more than 50), and there is no need for elaborate methods of selection. Usually (though not invariably) the pretest is confined to a small geographic area—usually the city where the research department is located and its surrounding small towns and rural areas. It is usually important, however, that an effort be made to concentrate the pretesting among the less literate respondents, particularly those in foreign-born groups and the lower economic classes, since the major interviewer problems on the survey will probably arise among these groups. It would thus be unwise to pretest questionnaires solely among such limited groups as office employees of the company in which the research department is located. Many of the weaknesses of the questionnaire will not be apparent when such a group is used. If time is pressing and no other arrangement is possible, such restricted pretesting is preferable to none, but genuine field pretesting is much more desirable.

When pretesting has been conducted until all but the most minor aspects of the questionnaire have been tried out in the field, it is time to cast the questionnaire in final form and to polish up the written instructions.

4. Final selection of the sample. Concurrently with the development and pretesting of the questionnaire, the actual sample to be used in the final study should be selected. The task is not difficult for judgment samples. For probability samples it may or may not be difficult.

If the research department has previously used probability sampling, it will usually be able to incorporate many features of earlier designs into the present sample. For example, the same counties or primary sampling units may be used for several studies. The past experience in selecting samples, together with the fact that much of the work (such as assembling county and city maps) has already been done, makes sample selection relatively simple.

If probability sampling has not been used before, and if the research department is inexperienced in probability sampling, the main worry is that sufficient time may not be available to select the sample before the study is ready to go into the field. If time is not available for this, there are just two alternatives: (1) buy an already selected design; (2) use a judgment instead of a probability sampling design. The use of probability

sampling methods usually requires long planning in advance of specific studies.

5. Scheduling fieldwork. As the sample is selected, the scheduling of fieldwork should begin. In the next chapters we consider in detail some of the problems which must be solved; here we simply enumerate them. The final preparation and reproduction of all necessary documents (instructions, questionnaires, contracts, letters to interviewers) should be arranged. Arrangements should be made for mailing out these materials in time for the study, and for notifying interviewers well in advance as to when the study will take place. The traveling schedule for field supervisors should be carefully planned to permit time for adequate personal instruction and supervision. Provision should be made for timing the fieldwork itself in order to give the interviewers enough time to complete their assignments. Communication with interviewers (telephone, telegraph, mail, face to face) should be arranged. Plans should be made in order to meet the inevitable breakdowns (cases in which interviewers fail to complete assignments on time), and some time should be allowed for this. As on other phases of the research investigation, the fieldwork must be planned in advance if it is to proceed smoothly and efficiently and be completed in time.

6. The preliminary code. As soon as the questionnaire is in final form, it is possible to set up an outline of the final "code." The code is a document that tells how answers to questions are to be classified into larger groupings for ease of tabulation and interpretation. The code cannot be completed until final questionnaires come in from the field in substantial numbers, but it can be set up in outline with the aid of the questionnaires from the pretest.

There are two reasons why it is desirable to set up a preliminary code. In the first place it will save time when the questionnaires are arriving from the field. More important, however, is the fact that detailed plans for tabulation can be made once a tentative code is available.

7. Detailed specifications for tabulation. With the aid of a tentative code and the general outline tables that were prepared earlier, it will be possible to make plans for tabulating the results of the survey. These plans are much more useful if made before the study actually gets under way, because the cost estimate for the study can then be made up more accurately. Moreover, time is saved if the plans for tabulation are made early. As we have emphasized before, some revisions of these plans will usually be necessary when the data come in from the field.

8. Some organizational problems in project design. Efficient preliminary exploration and project designing raise an interesting problem concerning the organization of the research department that was not

considered in Chapter 18. By this is meant the problem of solving the way in which research analysts divide up the work of planning. One approach is to have each analyst (where there are several) specialize in one phase of the work. One may be an expert in sampling and statistical problems; another in questionnaire construction and coding; a third (perhaps the research director) in the preliminary planning of the study. The essence of this approach, then, is that a series of specialists work on different phases of the study. Of course, over-all coordination must be achieved.

Another approach is for a single analyst to carry the project through from beginning to end and be directly responsible for all phases of the work. He may have to consult frequently with other members of the department for technical advice on matters with which he is not so familiar as they, but he is the initiator of all the work that gets done.

The first approach, the use of specialists (psychologists, economists, statisticians, etc.), is usually feasible only in large research departments. Ordinarily in medium or small departments each analyst must be versatile. Even when it is possible to maintain specialists, one analyst (or the research director) should always have the over-all responsibility for the study. This is the only way to localize responsibility to the end of seeing that commitments are met.

19.4 Reporting to Management

We shall skip over the problems of execution (*e.g.*, interviewing, coding, tabulating, etc.) for the time being and turn now to the problem of making a report of findings to management.

1. Interim reports to management. Management should be aware of the nature of the work of the research department and its progress on particular projects before the final report of findings. Knowledge of the general nature of projects is obtained in the discussions which, in part, constitute the preliminary exploration. Further information is obtained through the cost estimates. Management should additionally be informed of the progress of the study, particularly when unexpected events cause some significant deviation from original plans or the original time schedule. It is often wise to submit periodic reports of progress so that management can have an accurate idea of when desired information will become available.

Often there is the temptation to release to management preliminary findings. Only great urgency usually justifies releasing preliminary data when there is much chance that subsequent data will cause a significant revision in the conclusions suggested by the preliminary data. If preliminary findings and conclusions are released and subsequently modified,

there is a danger that management may be confused or lose confidence in the seemingly uncertain progress of research. Preliminary findings, if understood, can be of great value, but sometimes executives show a tendency to seize upon preliminary findings as definitive in spite of precautions to the contrary.

2. The written report. All studies should be summarized in written form. Simple memoranda often serve as well as leather-bound reports with gloss paper. Frequently the main presentation of the findings will be oral, but there should always be a written record as well. The written record is an aid to the oral presentation, and, more important, it can be referred to in the future when the basic data of the study have been stored away. Moreover, written reports are often valuable in that they increase the circulation of research findings.

There are some suggestions that may be helpful in preparing written reports:

1. The researcher should always use the simplest language possible. He should avoid technical words, long or complicated sentences, and jargon of all kinds. It has often been observed that children succeed better in written expression than adults because they use short simple sentences without artifice or pretentiousness.

2. In writing the report, the research man should completely reorient himself and attempt to put himself in the position of the executive. Research men usually take a strong interest in the technical aspects of an investigation, an interest that is intensified by preoccupation with technical problems during the time the study is being carried through in the field. By contrast, while executives should be encouraged to take some interest in the technical problems of a study, their primary interest is in the practical importance of the findings. Most executives are willing to permit the research men to worry about, and bear the responsibility for, the technical problems.

3. While no report should be fitted into an inflexible form, the following main divisions are usually desirable:

- a. A short statement of the objectives of the study. Seldom would this require more than a page.

- b. A short statement of the main findings, perhaps no longer than two pages and preferably shorter.

- c. A more detailed statement of the main findings and important qualifications, including some of the more important tables. In this section, quick reading can be facilitated by capitalizing or underlining the main conclusions so that the executive can read the supporting data only where he is interested.

d A brief statement of the method of the study—again about one or two pages in length.

e There may be one or several appendices. One such appendix might describe the technical methods in as much detail as would be needed for another research man to judge the competence and skill with which the study was conducted and even to repeat the study himself if this were required. (When studies are repeated at infrequent intervals, such a technical appendix is especially useful.) A second appendix may contain all the basic information that was tabulated for the study. This appendix is useful not only in connection with the study to which it is attached but as a source of information in answering future questions that are not covered in the main body of the report. It may be neither necessary nor desirable to issue the appendices to the study in the same volume as the main report. Many people to whom the main report is of vital importance would have no interest in the appendix. Moreover, the preparation of the appendix usually requires a rather long time, and it is needless to hold up the main report while awaiting the final typing of the appendix.

f. Sometimes there may be a statement of recommendations. Whether or not this part of the report will be prepared depends on whether or not the research department is empowered officially to make recommendations.

4 The report should be prepared on time. Promptness and speed by the research analyst will help, as will a typist who can convert a rough draft into a smooth, correct final copy. Such a typist should at least not introduce errors into the final draft that were not in the rough draft, at best she should be able to ferret out errors and correct them.

5. The report should be proofread with care and patience. Proofreading is dull and undramatic. Errors in final reports, however, may be of major importance. Important decisions may be based on the erroneous conclusion or finding. Equally embarrassing to the research man, though less serious to the company, is the error that is found by the executive before a decision is made.

3. Different methods of reporting to management. There should always be a written report, even though it is not the main channel of communication to management. Variations of oral reports, however, are possible and should be used whenever the executive to whom the information is given prefers such a presentation. Usually in a business office it is much easier to listen than to read, and oral presentations may often be more effective than written ones.

There is little need to go into the details of the different kinds of oral presentations. Presentations vary from informal telephone or personal conversations to formal meetings equipped with large "flip charts," or

easels, in which the speaker develops his presentation around a series of pictures or graphs, each of which contains a key idea or conclusion in writing.

The method of reporting should never be unvarying, either from one executive to the next or from one problem to the next. The whims of executives need not be compelling in the planning or execution of the study, but almost every whim should be indulged in presenting reports in the way each executive prefers.

4. Following up. It is most difficult to see that the study really gets the attention it deserves in future months or years. Nothing is easier than to forget the study when it is finished and plunge into the next one. The trouble is that not only research men but executives are likely to forget the past studies. The research department should be a staff department with no authority to make or suggest policy except on direct request. Nonetheless, research men cannot retreat from the problems of persuasion and seeing that their findings are given adequate attention in policy formation. The distinction between fact finding and policy formation is indeed much less clear than one might think at first. Every summary of conclusions of a research report partakes of some of the characteristics of a statement of policy recommendations. Often factual information may clearly indicate one particular policy. Perhaps more often, however, the factual information is only one of the elements to be considered in the final policy decision; business judgment would be needed even with a million-dollar research department grinding out the best factual information possible. The research man should never lose sight of this.

Ultimately the problem of seeing that research findings are adequately considered and the deeper problem of deciding where research ends and policy formation begins must be solved by close cooperation between the research department and management. Personal confidence of one in the other is likely to smooth over most of the potential trouble. Ultimately, then, the success of the final stage of a research study depends as strongly on close rapport between research and management as does the success of the first stage—preliminary exploration of the problem. Many of the suggestions made at the beginning of this chapter on ways to encourage and maintain good relations with management apply equally at this stage.

CHAPTER 20

INTERVIEWING

20.1 Introduction

In this book most of the space has been devoted to the question, "What should be done?" rather than, "How should things be done?" Even in Part V, where we have explicitly focused attention on "how" rather than "what," the emphasis thus far has been on the planning of primary research rather than on its execution. We have distinguished three major functions—coordination with management, planning, execution—and have so far considered only the first two. Now it is time to consider the hitherto neglected but vitally important problems of execution.

The execution of marketing research studies is just as important as the other functions and is often much more difficult. For example, few jobs in marketing research are more difficult than interviewing, unless it be the supervision of interviewing. Interviewing requires a rare combination of intelligence and skill in social contacts. Almost any intelligent person can select an acceptable sample or write a good questionnaire, if he is given proper training, but many intelligent people cannot become good interviewers with the best of training. And the problem of supervising interviewers is also impressively difficult. It involves the supervision of employees who are beyond the reach of observation and who produce a nonstandardized product of a quality that is very difficult to determine.

In the next two chapters we attack the problem of interviewing and the supervision of interviewing; in the final chapter, the problems of execution arising internally—editing, coding, tabulating, etc.

20.2 Interviewing

The job of the interviewer varies from study to study. The work is so diversified that it defies concise description: there are different kinds of interviews, different sampling methods, different types of information being sought, and different degrees of supervision and training. Each job has its peculiarities, and few interviewers get to the point where novelty is completely missing. Despite the differences between individual studies, however, it is possible to some extent to generalize about the nature of the work.

Fieldwork in marketing research almost always brings the interviewer

into contact with many different kinds of people. Except in purely observational studies, the interviewer must talk to these people, almost all of whom will be strangers, and ask a series of questions. While most of the people to whom the interviewer talks will be pleasant and cooperative, a few will be rude and some even insulting.

Interviewing is usually hard physical work. The interviewer is on her feet, either walking or standing, for long periods. Interviewing must often be done in bad weather and in inaccessible parts of the city and country. Even though the interviewer talks to many different people, there is a certain loneliness about the work, for the investigator typically works on her own with little or no close supervision and seldom sees the other interviewers who are working on the same study.

Offsetting the rather drab aspects of interviewing just mentioned, there are compensating features. Interviewing is interesting and challenging to many interviewers. Sometimes the interest is primarily in the questionnaire and findings; more often in the respondents. There are few better ways for learning about people and the way they live and think. Interviewing is challenging in that a new study will usually present new problems.

1. Qualifications for interviewers. Job specifications, whether for interviewers or anyone else, usually list all known human virtues without defining them very precisely. Hundreds of nouns denoting these virtues have been listed at one time or other, but the essential qualifications for interviewing can probably be compressed under the following headings:

1. The interviewer should be intelligent enough to follow complicated instructions.

2. The interviewer should be willing to follow directions precisely, even when she is not closely supervised. More candidly, she must be honest.

3. She should be able to make all kinds of people feel at ease with her quickly.

4. She should realize that objective reporting can be obtained only by constantly fighting against her own ideas as to what the answers should be, and she should be willing to keep fighting her prejudices.

5. She should have persistence and imagination in the face of difficult and unpleasant interviews and unexpected situations.

6. She should like to interview.

7. She should have the required physical stamina.

8. She should have time for the work, especially for peak loads.

9. She should be "mobile," and preferably she should have access to a car.

A much more concise expression of the same requirements is that the interviewer (1) must be able to get the information that is wanted and (2) must be able to record that information for use by someone else. But regardless of how one organizes the necessary qualifications, it is much

easier to list these qualities than to find people who possess them. Not every capable, honest, industrious person will make a good interviewer. Not all experienced interviewers possess these qualities. It is hard to tell even whether the interviewers already working possess them; it is even harder to tell about some new person being considered for an interviewing job. We return to problems like these in the next chapter.

2. How to interview. The quotation given below is a reproduction in entirety of the *Manual for Interviewers* used by the Division of Special Surveys, Bureau of Agricultural Economics, in the U.S. Department of Agriculture. The manual as presented here formerly provided the general written instructions or outline of basic training for the fieldworkers of that organization and is the best statement of its kind that has come to our attention.¹ As this book goes to press, the Division of Special Surveys is completing a substantial modification and amplification of the manual. We are sorry that we could not delay publication in order to incorporate this more up-to-date version in our book. The revision will include considerably more material and have a different format. It will represent the procedures and policies which are now actually in effect with respect to the field interviewing of the Division of Special Surveys.

We are anxious to acknowledge the great helpfulness of the Division and particularly its Director, Dr. Forrest E. Clements, in making the quoted material available to us.

As the introduction to the manual explains, the studies for which this manual was written are in the domain of public policy rather than marketing research. Nonetheless, the principles apply to marketing research with only minor qualifications and amplifications. We shall make these qualifications and amplifications when appropriate. (Our comments, in brackets, are interspersed in the text of the manual.) The manual follows:

INTRODUCTION

The Division's Function :

About 10 years ago, several men in the Department of Agriculture became interested in ascertaining at first hand what farmers were thinking about agricultural policies and programs. It was felt that in this way the programs could be formulated more effectively to meet the needs of the farmers. From this basic idea sprang the present Division of Special Surveys.

With the advent of the war, the Division's functions were expanded to include studies of public attitudes among urban people as well as rural people on many problems in which Government agencies were interested. The broad function of

¹ *Manual for Interviewers*, U.S. Department of Agriculture, Bureau of Agricultural Economics, Division of Special Surveys, Washington, May, 1950. (This is not an official publication of the Bureau of Agricultural Economics.)

the Division at present is briefly as follows: A Government agency is faced with a problem, the solution of which requires information on how people feel about the problem. The Division is sometimes requested to gather the information by conducting interview surveys among a cross section of the population.

The approach to public attitudes as used by the Division varies from that of the usual survey in that the questions used are usually but not always quite broad and give the respondent (the person being interviewed) an opportunity to discuss his feelings at some length. Another important aspect of this approach is that the respondent is also brought to discuss *why* he feels as he does and to give the context out of which his attitudes arise.

An analysis of these interviews by the Division gives the requesting Government agency a picture of how people feel about certain issues and why they feel that way.

The Interviewer's Responsibility:

To serve these purposes, the interviewing process must result in a true picture of what the respondent thinks and sometimes what he knows about certain issues, with no suggestion from the interviewer to influence his thinking. For the interviewer's function is neither that of an educator nor missionary, but that of an impartial observer of public opinion as he finds it. No matter how much he may privately disagree with the attitudes he encounters, it is not his business to enlighten or convert but to listen and understand.

The discussion of the interviewing procedure in the following pages is based on the methods which experience has shown are most successful in carrying out the Division's aims. In order that our work be as uniform as possible, it is important that the interviewing methods discussed here be conscientiously followed.

[Few things are so hard to learn as the injunction that the interviewer should be an impartial observer of things as she finds them. This goes wholly counter to one's usual habits of conversation and action, and genuine neutrality in interviewing requires training, practice, and continual awareness of the danger that one's ideas and opinions will creep into the interview]

SETTING THE STAGE FOR THE INTERVIEW

To be able to get accurate information from the respondent, the situation throughout the interview must be such that he feels willing to talk freely and frankly in providing the information desired. In every interview there must exist a feeling of mutual trust and confidence between the interviewer and his respondent. The respondent must be certain of the interviewer's identity and he must feel friendly enough toward him to be willing to talk with him frankly. He must also understand clearly what the survey is about and he must be convinced that it is important enough to warrant his time and serious attention.

[The interviewing problems confronted by the Division of Special Surveys are in most respects more difficult than those typically encountered in marketing research. Marketing research does, however, present one obstacle that is not encountered to the same extent in opinion research on questions of public policy: the intrinsic interest to respondents of the subject matter of many marketing re-

search studies is discouragingly low. While it is possible to keep respondents' interest on opinion studies for more than an hour, in many marketing research studies it may be hard to do so for even a quarter of an hour.

The questionnaires prepared by the Division of Special Surveys are chiefly of the "free response" type. That is, the respondent is encouraged to express himself freely and fully, and his answers are recorded verbatim or nearly verbatim. However, the advice given for interviewing on this type of question applies, with certain obvious exceptions, to interviewing on "fixed alternative" questions.]

Your first task then in every interview is to create a situation in which the respondent will talk freely and frankly in giving answers to the schedule questions.¹

Gaining Acceptance

Fortunately, most people are friendly and with a little effort on your part, you will be readily accepted. It may be easier for you in approaching the person whom you want to interview if you keep in mind that the chances of being refused will be very slight. Actually, few people refuse to cooperate on a Division survey; the number on most surveys is less than two percent of the people contacted.

Show a genuine interest in your respondent. This is one of the best ways to make friends with him and gain his confidence. When you approach a farmer, for instance, pick out something about the farm in which he is likely to be interested and show interest in that yourself. There are many things around any home or farm which indicate the person's interest; look for those things and call attention to them. In the case of the farmer, an indication of his interest might be a nice herd of cattle or a well terraced farm. Show interest in those things which you think are of interest to the person whom you want to interview.

Another way to make friends with the person is to identify yourself in some way with him. You can do this by calling attention to things you both have in common. This might be the fact that you are both from the same part of the country, or you know someone who attended the same school, or you both may be interested in the same sport or you may smoke the same kind of tobacco.

See and understand your respondent's point of view. Understanding a point of view does not necessarily mean agreeing with it; you can understand how a person feels about an issue and why he feels as he does without necessarily agreeing with him. It is not your job to disagree, but it is your job as an interviewer to see and understand how the other person feels about things. The habit of seeing and understanding a point of view is very important in conducting opinion surveys, and is particularly important in encouraging the respondent to talk freely once he has started expressing his opinions on the question. To let him see that you understand how he feels and that you sympathize with his right to feel that way is one of the very best means of getting him to talk freely. Everyone likes a good listener.

Putting the Respondent at Ease

The more natural the interview situation is, the more likely the respondent is to talk freely and frankly. A part of the process of creating a natural situation is

¹ A "schedule" is roughly the same as a "questionnaire." (Our footnote.)

putting the respondent at ease. He will be most talkative when he is relaxed and at ease.

The best way to make him feel at ease is to be at ease yourself. Act naturally, just as you would in talking to a friend. Talk in the same tone of voice as you ordinarily do and use the gestures that you usually use in talking with a friend.

Allow the respondent time and give him a chance to develop poise and get ready for the interview. Open the discussion by bringing up pleasant topics and then encourage him to talk. Talk about the weather or whatever the person happened to be doing when you approached him.

Aim to make your respondent feel that you are an ordinary person with whom he can talk as an equal. This is a state of mind which every respondent must have if you are to get a good interview. One way of encouraging him to feel your equal is to encourage him to talk about things that he knows about. Ask him questions about his method of farming, if he is a farmer, or about his job and his work if he is a worker, and then encourage him to talk about these subjects. This works well with anyone but is of greatest importance with inarticulate people, people who have difficulty in expressing their feelings normally. It not only helps make them feel at ease but it also helps them develop the habit of talking freely.

Making the Introduction

The second step in setting the stage for the interview is making the introduction. This involves identifying yourself as a representative of the United States Department of Agriculture, explaining the purpose of the survey and making it clear to the respondent why he was selected for an interview.

Introducing Yourself: Each respondent should be sure of your identity as a government representative and understand that you are visiting him on official business. This can usually be accomplished by saying something like: "My name is John Doe. I am working on an opinion survey here in Plainsville for the United States Department of Agriculture."

You should show your identification card at some time during each interview, but whether it is shown here or later depends on whether you feel that showing it at this point is necessary to establish your identity.

Explaining the Purpose of the Survey: The instructions for each survey will contain a statement of the purpose of the survey. You need to know the purpose of the survey and have clearly in mind the ideas about it which you want to put over to the respondent. A recent Division survey conducted among a cross section of the adult population of the United States was designed to find out how people felt about the international situation. Topics covered were reactions to the proposed loans to England and Russia, reactions to the United Nations organization as an agency to keep peace and so on. As in most Division questionnaires, the first few questions were of a general nature, the first being, "How do you feel about the way the different countries are getting along together since the end of the war?" A satisfactory explanation of the purpose of this survey would have been:

"The Government is interested in finding out how people feel about the way the

different countries of the world have been getting along together since the end of the war, and we are conducting a survey in different cities and counties all over the country, talking with all kinds of people to get their opinions on this. We have a list of topics that we bring up with each person we talk with just to get their opinions on them."

Another survey was conducted among a cross section of American housewives and the purpose of that survey was to find out their fiber preferences in selecting different garments of clothing, and certain household articles such as curtains. The first series of questions was to get fiber preferences in ready-made house dresses. The explanation of this survey might have been given as follows:

"The Government is interested in getting the opinions of women consumers on what they like and dislike about the materials used in certain types of ready-made clothing, such as house dresses, and in things they buy for use in keeping house. So we are conducting an opinion survey among women in different cities and counties all over the country to get their opinions about these things. We have a list of topics about different garments of clothing and household articles which we discuss with each person we talk to in order to find out what housewives look for in selecting these articles."

In the fall of 1944 the Division conducted a survey among cotton farmers to find out what role they would like to see the Government follow after the war in helping to stabilize cotton prices. The questionnaire was designed to find out first what role the respondent wanted the Government to follow and then to get his reactions to specific proposals by which the Government might help stabilize cotton prices. The first questions were designed to get the respondent's attitude on the postwar outlook for cotton prices. Here is one way the purpose of this survey might have been given:

"We are talking with cotton farmers in different parts of the cotton belt to find out how they feel about the outlook for the cotton business after the war and to find out what kind of role they would like to see the Government play in the cotton business. The Government is interested in finding out how cotton farmers, like yourself, size up the situation, and they also want the opinions of the cotton farmers about what kind of Government program they would favor after the war."

It will be noted that there are a few things which all of these statements have in common. The first is that the purpose of the survey is given in general terms, without naming any of the specific topics to be taken up except that in the first question in the questionnaire. This is important, because if the interviewer had mentioned any of the specific topics that were to be taken up in the interview, he would have conditioned the response to that or other questions. As an example, if a respondent on the cotton survey had been told in the introduction about the different ways in which the Government might help support cotton prices after the war, that would have influenced his answer to the question, "What role would you like to see the Government play in the cotton business after the war?"

Secondly, although the explanation is in general terms, it gives the respondent a good idea of the purpose of the survey.

Thirdly, the use of such phrases as "talking to all kinds of people" and "talking to cotton farmers all over the cotton belt" helps establish a feeling that the survey

is important and worth spending some time on. Mention that "the Government is interested in" also helps add importance to the interview. Such a statement may often be repeated with profit at different points during the interview.

Fourth, mention of "opinions," "how people feel," and "how cotton farmers size up the situation" help to make it clear that there are no right or wrong answers to the questions since all that is wanted are people's opinions.

And finally, in each statement the content of the first question in the questionnaire is indicated. This helps to lead easily from the introduction into a discussion of the first question.

Explaining How the Respondent Happened To Be Selected: Some people will wonder why they were singled out for interview by a government representative, and the purpose in explaining the method of selecting respondents is to reassure such people by making clear that the selection is purely random. This can be accomplished without a detailed explanation of our sampling methods. A satisfactory explanation is a general statement of how the sample dwelling units are selected in area sampling. One way of stating this to an urban respondent would be:

"Perhaps you wonder how I happened to stop here. You see, in trying to find out how people all over the country feel about something, we can't talk to everyone, but we try to talk to people in all walks of life—men and women, young and old, people with different incomes and so on. The way we find people to talk with is to select certain cities and counties all over the country and then in each city we select certain blocks in different parts of the city and then stop at certain houses in each block and talk with someone."

For a rural resident the explanation might be:

"In doing a survey like this to find out how people feel about different issues, we can't talk to everyone, but we try to talk to some men and some women, in different age groups, with different incomes and so on. The way we locate people to talk to in rural areas is to select areas of land scattered all over the country and then go to those areas of land and talk to someone at each house. That way we get to talk with all kinds of people in all parts of the county."

How the Interview Will Be Conducted: Though it may seem minor, a necessary ingredient in the recipe for a good introduction is to tell the respondent what to expect in the interview. It is important to put across the idea that the interview is a place to **discuss** things rather than merely answer a set of questions, and that there is no right or wrong answer to the questions. All that is wanted is a frank expression of the person's opinions. The reasons for putting over these ideas are (1) to overcome resistance based on an unwillingness to "answer a bunch of questions," and (2) to help prepare the way for a free and full discussion of opinions on the topics covered by the survey. These ideas can be worked in anywhere in the introduction; they follow very well the statement of the purpose of the survey, but may be worked in after the explanation of the sampling method. The point that there is no right or wrong way to answer the questions can often profitably be made occasionally during the interview, particularly when the respondent seems reluctant to express himself freely.

Some Special Situations

As was pointed out in the foregoing discussion, most people are friendly and cooperative and when given an adequate explanation of the purpose of the survey are willing to be interviewed. They generally like to talk with a sympathetic person about things which concern them, even though these concerns may be highly personal and the sympathetic person a total stranger. They are usually pleased to have their wishes consulted by a government representative.

However, every interviewer, no matter how competent, will at times run into resistance. The purpose of this section is to deal with some of the more common types of resistance and to offer suggestions to the interviewer on how he can best overcome them. It is difficult to tell an interviewer exactly what to do when he runs into resistance because the cause for it varies so much. We can only point out some of the more common types of resistance and enumerate some of the possible causes. However, it is up to the interviewer to diagnose the cause for resistance when he runs into it and to decide what should be done to overcome it.

The Suspicious Respondent: The suspicious person may show his suspicion in various ways. He may show it by asking many questions of the interviewer as to who he is, what agency he works for, why he wants the information, and so on. Or he may show that he is suspicious by remaining quiet and merely staring at the interviewer. The causes for suspicion are varied. The person may think that the interviewer is a salesman or he may think that he is some kind of investigator. Whatever the cause, the fact that a person is suspicious indicates that he is not convinced about the interviewer's identity or he does not have a clear understanding of the purpose of the survey or both. It seems then that the first thing to be done in any case of suspicion is to elaborate on the purpose of the survey to give the person a clear understanding of the purpose of the call and to convince him of your identity by showing your identification card.

Threatened Refusals: There are several "symptoms" of threatened refusals. The person may say that he just doesn't have time to talk because of some matter of pressing importance at the moment. Some people may make the statement: "Oh, I don't know anything about those things. You don't want to talk with me."

The causes for these forms of resistance are also varied. The person may infer that the interviewer is a salesman or an investigator; in other words he may be suspicious. The person may not think the survey is of any importance and therefore does not want to take time to talk with the interviewer. The person may feel incompetent because of lack of education, information, etc., and doesn't want to show up in an unfavorable light and for that reason tries to get out of the interview. Or the person may legitimately have other more pressing matters to attend to.

Whatever the cause of "threatened refusals" you should always spend some time in carefully explaining the purpose of the survey, why the respondent happened to be selected and make clear your own identity. If the cause of the resistance is suspicion, you will probably overcome this and get an interview by the additional explanation. If the person is trying to get out of the interview because he is not

interested, this additional explanation should be aimed to make him feel interested. If you think that the person is reluctant to be interviewed because he feels that he is not well educated enough to give "good" answers, you should stress the point that there are no right or wrong answers to the questions but that it is just a matter of the person expressing how he feels about certain issues. You should also spend some more time in helping this person feel at ease and develop the habit of expressing himself so that he will feel able to express his opinions on the questions.

As for the person who says he is too busy for the interview, you will have to use your own judgment in determining whether he is really too busy or whether he is merely putting you off because he doesn't want to be bothered. If you feel that the person is really too busy at the moment to give serious consideration to the questions, it is always best to make an appointment to see him at a later time. In making an appointment, you should agree on a definite date before leaving the person.

The Angry or Belligerent Respondent: This type person may react in various ways to the interviewer, depending in some cases on the cause of his anger. A common reaction is to reproach the interviewer by asking him if he "can't find anything better to do" or to belittle his work in some way. If the cause for the anger is dissatisfaction with the government, he may react by raving against the people in Washington.

The correct way to deal with this type person is to give him a chance to overcome the anger and then the respondent is willing to participate in the interview. In dealing with a person like this, the art of listening is of greatest importance. It is important that the interviewer not lose his head, but remain calm. A **must** in dealing with such people is to never argue with them; that only serves to further antagonize them.

["Setting the stage for the interview" requires the highest level of skill on the part of the interviewer. The statement is made in this section of the Manual that "on most surveys the number who refuse to cooperate is less than two per cent of the people contacted." This low proportion of refusals is probably attributable at least in part to indoctrination of interviewers in these excellent suggestions for obtaining cooperation.¹ The importance of keeping the rate of refusal low has been more and more recognized in marketing research, particularly as probability sampling designs have come to be more frequently used. The danger of "passive resistance" or token cooperation has received less attention, but this danger should be materially reduced by the procedures suggested here.]

[The detailed nature of these instructions for gaining cooperation may be surprising to many marketing research people. One might get the impression that it takes fifteen minutes merely to introduce the interview; this is not so, but the time devoted to introduction is undoubtedly longer than is customary in marketing research. For a short interview, it is often possible to go right into the interview

¹ See also Herbert G. Heneman, Jr. and Donald G. Paterson, "Refusal Rates and Interviewer Quality," *International Journal of Opinion and Attitude Research*, Vol. 3, No. 3, pp. 392-398, Fall, 1949.

after a one sentence introduction and to have the interview over before the respondent has even had a chance to refuse. This procedure is, in fact, often recommended and often used in marketing research, particularly when respondents are interviewed on the street or other public places. It is possible to use this "blitzkrieg" approach only for short questionnaires containing the simplest nonpersonal, factual questions. If the questionnaire is long, if the questions are at all personal, or if they require careful reflection and consideration, it is better to take the time to introduce the survey thoroughly before jumping into specific questions. Otherwise, even if the respondent doesn't refuse to be interviewed, he is likely to give only token cooperation and to be thoroughly suspicious of the interviewer. A common experience, when the introduction has been of the one-sentence variety, is to have the respondent interrupt the interview about half-way through to ask, "Say, what's this all about, anyway?" A respondent in this frame of mind is not likely to answer questions well.]

[Another important point is the fact that many respondents who are initially suspicious or opposed to the interview may be thoroughly converted by a good introduction. In marketing research so much of the work has been done with quota samples, where typically a refusal is not considered serious, that relatively little emphasis has been placed on persistence in getting an interview with a reluctant respondent. Yet recognition of the need to obtain information from reluctant as well as willing respondents has grown. Skill in introducing oneself reduces sharply the number of refusals from reluctant respondents.]

ASKING THE QUESTIONS

Once you feel you have established yourself with the respondent and have gotten him interested in the survey, you are ready to start with the main part of the interview, asking the questions. In conducting the interview, you should keep the questionnaire in sight, and read over each question quickly before asking it to be sure that you recall the exact wording.

There are a few general rules on asking the questions. These are necessary to ensure that the data we collect will be valid and reliable.

Ask Each Question Exactly As It Is Worded

Each question is worded so that it will have the same meaning to each respondent and so that it will not bias or influence his answer. Once the question form has been determined, it is extremely important that it be asked of each respondent exactly as it is worded. There are two reasons for this rule. First, if you change the wording of the question, you may change the meaning that it is intended to convey to the respondent. Second, you can seriously influence the response of the person interviewed by the slightest change in the wording.

During the war, the Bureau conducted a survey among cotton farmers, one objective of which was to find out what they thought would happen to cotton prices in the first two years after the war. The question which was designed to meet this objective was phrased as follows: "What do you think will happen to cotton prices in the first two years after the war?" Notice that we were asking for the re-

spondent's opinion about **cotton prices in the first two years after the war.** Now let us demonstrate how the meaning of this question can be completely changed by a slight change in the wording. "What do you think will happen to cotton prices after the war?" "What do you think will happen to cotton prices in the next two years?" In these phrasings of the questions, the time period has been changed which changes the meaning of the question. The answers to either of these questions would, therefore, not be comparable to answers to the original question. "What do you think will happen to cotton prices in the first two years after the war—think they'll go down like they did after the last war?" The addition to the original question here biases the answer by suggesting the depression in farm prices following the last war.

State All Explanations That Appear in the Questionnaire to Each Respondent: Frequently the questionnaire contains an explanation or a statement preceding a question or series of questions about a topic, as in the two examples below.

"Reports show that some people are putting their money into bank accounts, cash savings, and securities instead of bonds. Why do you think they're doing that instead of putting it into bonds?" "Some people would like to see our Government keep to itself and not have anything to do with the rest of the world. How would you feel about that?"

When such statements as these appear, they are to be stated to each person interviewed **exactly as worded** as they are part of the question.

Ask Each Question That Is Applicable

The reliability of the questionnaire depends on having each respondent answer exactly the same questions. So, you are expected to ask each question of every person you interview except when the question is clearly not applicable. There are two conditions under which the question might not apply and might be omitted:

First, when the question has been adequately answered previously. As an example, in the following extract from an interview, the respondent's reply to Question 4 included information which fully answers Question 6.

Question 4. "Do you think the prices of things you have to buy will stay where they are?"

Answer: "No, I don't see how they can. What with the shortages and with labor wanting higher wages every day, it looks to me as if prices are going to get sky-high the way they were in the last war. And that's certainly going to be hard on people like us whose salaries stay the same. I had a small increase two months ago, but there's no chance of getting another for a couple of years."

Question 6. "Do you think your income will change any in the next six months?"

Since Question 6 was covered by the respondent's answer to Question 4, it may be omitted.

If there is any doubt in your mind as to whether the question has been answered, by all means ask it. It is not difficult to ask a respondent a question which he has already talked about and which he may think he has answered, if you preface

the question with some such statement as, "You've talked some about the next question, but I want to ask it to see if you have anything more to say." Such a statement indicates to the respondent that you remember his having touched on the question, and by doing this, in most cases, you will avoid any resentment. It is surprising how often you will get an elaboration from the respondent on the topic by stating the question in this way.

Secondly, there are questions to which we refer as contingency questions, which are asked or omitted depending on the answer to a preceding question. Here is an example of contingency questions taken from the questionnaire of one of our surveys:

Question 26. "Have you bought any bonds since the first of January?"

IF NO

Question 26b. "How does it happen that you are not buying bonds now?"

Question 26c. "How did you happen to stop when you did?"

The latter two questions are contingencies and would be asked only if the person being interviewed answered Question 26 by saying he had not bought any bonds since the first of January. When contingency questions are used in a questionnaire, they are always "set off" in some way from the questions that are to be asked of everyone. This is done as an aid to you in helping you to locate the questions easily.

Ask the Questions in Order

We know that some questions will influence the answer to other questions, so in formulating the questionnaire this is kept in mind and the questions are arranged in an order which will minimize this influence. To illustrate, we can take this example of a series of questions from a recent questionnaire.

14. "There has been some talk about this country making a loan to England; what would you think of that?" "Why?"

15. "Do you think this country would have anything to gain from making such a loan?"

To have asked a person first whether he thought there was anything to be gained by this country in making a loan to England might have influenced his attitude toward the loan itself. He may have thought of the loan entirely as a political issue, and to have asked whether he thought we had anything to gain by making the loan before getting his reaction to the loan could seriously influence his answer. He may not have thought of the loan at all in terms of this country's gaining something in return, and the purpose in giving the above order to the questions was to find out exactly in what terms or frame of reference he did think of it.

We could take similar examples from any Division questionnaire to demonstrate the importance of the order of the questions, but that is not necessary. You can discover this for yourself by looking at any questionnaire and studying the order in which the questions are arranged so that one does not influence the answer to succeeding ones.

[To a person just learning about marketing research these suggestions—ask each question exactly as it is worded, ask each question that is applicable, and ask the questions in order—will sound so obvious as to be hardly worth mentioning. Yet

few things seem harder than to follow these suggestions or to get another person to do so.]

[In part, the difficulty in doing so may be because of a poor questionnaire. If an interviewer is supposed to ask the respondent an obviously foolish or obscure question, the temptation is strong to change the wording, especially in repeating the question after the respondent has reacted to the original wording with a blank stare or a bewildered expression.]

[The second possible difficulty is that the interview may be hurried. If the interviewer feels that she must rush through the interview, she is likely to leave out questions, vary the wording, and alter the order.]

[But more fundamental than either of these difficulties is the fact that an interview is a conversation, and the conversation would be terribly awkward if the interviewer were to say nothing but the questions written on the questionnaire. This fact is recognized explicitly in the sections, "Manner of Asking the Question," and "Probing" and suggestions are given for ways to make the interview more conversational without injecting the interviewer's ideas into it.]

[The suggestion to omit questions when the respondent has clearly answered it already is realistic and important. Nothing seems more irritating to respondents than to have to repeat the same answer several times, and even with the best of questionnaires, respondents will give answers that will anticipate later questions.]

Manner of Asking the Question

As we pointed out earlier, the person being interviewed is most likely to talk freely and frankly in giving answers to the questions when the interview is a natural, conversational situation. The manner in which you put the question to the respondent can help in making the interview a conversational situation and help to draw full discussion, or it can have the opposite effect. You should try to ask each question in a manner which will draw out the fullest discussion from the respondent. There are several ways of doing this.

First, state the question clearly and slowly enough to make sure that the respondent catches the meaning the question is intended to convey. At this point, you should concentrate on the question and on getting the meaning over to the respondent and not on what the answer is likely to be or whether the respondent is going to object to answering.

Second, ask the question in a natural, conversational tone of voice.

Third, do not imply the answer to the question. It is easy to influence an answer by the tone or manner in which the question is asked. If you indicate by your tone of voice that one alternative in a question is better than another, you can easily suggest a choice to the respondent, especially when he does not have strong feelings either way.

Making a Conversational Delivery. Being able to ask the question conversationally comes about through experience and effort. It is a good idea to read a new schedule aloud several times before taking an interview or to try the questions out on several friends in order to work out a conversational delivery. In asking the questions, it is a good rule to change your tone of voice now and then as you

would in ordinary conversation. Experienced interviewers frequently introduce the schedule questions by some neutral statement such as, "In your opinion," or "Another thing the Government is interested in is," followed by an exact statement of the schedule question. Here's how they do it.

"In your opinion, why do you think the Government is interested in getting people to buy war bonds?"

"Another thing the Government is interested in is, what do you think the foreign demand for that will be like during the first two years after the war?"

You will notice that these prefatory statements were neutral; they added nothing to the question nor changed it in any way, but they served the purpose of making the questions sound more conversational, and doubtless made the interview more informal to the respondent.

Conversational Aids in the Questionnaire: The questionnaire is designed deliberately to help you in conducting a conversational interview. Here are some of the aids which you will find in the questionnaire:

1. Conversational questions: Each question is worded so that you can ask it easily in a conversational manner.

2. Conversational sequence: The questions on the schedule are arranged in what the pretest showed to be a natural and usual conversational sequence. The question sequence represents the order of associations most commonly made by the people who were interviewed when the questionnaire was being tried by experienced interviewers.

3. Special introduction for difficult questions: There are certain questions, particularly those designed to find out how much a person knows about a topic or issue, which can cause embarrassment or resentment if not phrased properly. For instance, suppose we wanted to find out how well informed people were about the purpose of the United Nations Organization. If we asked bluntly, "What is the purpose of the United Nations Organization?" we would run a chance of embarrassing the person who was not too well informed and might not find out what he did know, and the well-informed person might resent being asked for information so bluntly. We want to avoid either type of reaction because it tends to create an unfriendly feeling which is not conducive to full discussion of the remaining questions. So to help you in avoiding these kinds of reactions, we phrase the question in some such way as this: "As you understand it, what is the United Nations Organization set up to do?" Stated thus, it is easier for the poorly informed person to tell what he knows about the UN or admit that he knows nothing about its purposes without embarrassment, and the better informed person responds more readily without feeling resentful about being asked a question which may sound simple to him.

Transitions

The questionnaire for any survey will deal with several topics, and between the main topics it is usually advisable to use a transition statement which serves to close one topic and introduce a new one. By use of transition, the interviewer can help the respondent mentally bridge the gap between the completed topic and the

new one and this is often necessary, particularly in the case of an inarticulate respondent. A transition statement brings the respondent up to date so that he can begin to think in terms of the new topic, and helps avoid the problem of having questions on the new topic answered in the same frame of reference as the previous one. The transition also makes the interview more of a natural, friendly discussion and less of a question-answer situation, and is therefore another technique used in making the interview natural and informal.

The procedure for making transitions is to briefly summarize the general aspects of the topic covered and then give a general introduction to the new topic. Two points need to be kept in mind in introducing the new topic. The first and most important is to be careful not to say anything about the new topic which would influence the person's answers to the questions that are to be asked. The new topic should be presented in general terms, just as the purpose of the survey is explained. Another point is to phrase the transition statement so that it leads easily into the first question of the new topic. An example will serve to illustrate how this is done.

On one survey the questionnaire began with a series of questions designed to find out the respondent's preferences in selecting ready-made house dresses, followed by a similar series of questions to get the same kind of information about summer street dresses. The transition between these topics might have been stated this way: "So far we've been talking about the things you look for in selecting a ready-made house dress. Now I'd like to talk with you some about the better type of dresses that are worn in the summertime." The interviewer would follow up immediately on this statement by asking the first schedule question on the new topic.

GETTING ADEQUATE ANSWERS

If the respondent always made perfect answers to the schedule questions, the interviewer's job would be relatively simple. However, more often than not the response to a schedule question will only partially answer it or will evade the objective of the question altogether. When this occurs, it is necessary that the interviewer make use of special devices to get the respondent to give additional information in order to get an adequate answer to the question.

Meaning of an Adequate Answer

By an adequate answer, we mean one that is complete, clear and specific. For example, an adequate answer to an opinion question is a complete and clear expression of opinion on the specific topic. An adequate answer to an information question is a complete and clear statement of what the person knows about the specific topic or issue.

Examples of Adequate Answers:

Question: "Some people say they would like to see this country keep to itself and not have anything to do with the rest of the world. How would you feel about that?"

Answer: "That is the isolationist line. (Pause) I don't hold to that view, and

I never did. The time of isolation has passed, that was brought about by the aeroplane; no one country can be isolated now. All countries must try to live together and to do that they must cooperate by trading and by working together for peace. That is the only hope I see for peace. A policy of isolation would get us in trouble, just like it almost caused us to lose this war by getting in too late."

This is a complete and clear expression of opinion on the specific topic.

Question: "As you understand it, why do you think the Government is anxious for people to buy War Bonds during the war?"

Respondent: "Well, they need money to finance the war and the bonds are about the only way the Government has of raising the money. It takes lots of money to finance a war as big as this one, and it will take more to take care of the injured long after this war is over."

Interviewer: "Any other reason?"

Respondent: "Yes, there is another that you hear a lot about on the radio and in the papers, and that is to put your money in bonds and not buy things you don't need as that will help to keep prices from going too high."

This answer is also quite adequate. It shows clearly what the person knew about the reasons the Government was pushing bond sales during the war. However, had the interviewer not probed for "Any other reason?" the answer would not have been complete since it would not have revealed whether the respondent was conscious of the argument for buying bonds to control inflation.

It should be pointed out that the adequacy of an answer is not judged by its length. For instance, one word such as "Absolutely" can quite adequately answer a question if that is the respondent's real feeling about an issue. "Don't know" is also a satisfactory answer, if the respondent really does not have an opinion or if he does not have the information to answer a question.

Examples of Inadequate Answers: In this section we shall take up examples of the most common types of inadequate answers. These examples should be studied carefully so that you will be aware of the kinds of answers on which you will need always to probe.

1. Ambiguous answers: One of the most common responses which does not adequately answer a question is the ambiguous or vague answer. We have in mind here the response related to the specific question but which is so vague that one is not certain of its meaning. Here are examples:

Question: "What are the most important things you look for in buying a house dress?"

Answers: "Quality and style."

"I like one in good style."

"I like good quality."

These responses are related to the particular question and the respondent, no doubt, felt that she answered the question. However, these answers are not meaningful because "Quality" and "Style" mean different things to different people and it is not clear from the answer just what the respondent meant by the terms. Let us take another example:

Question: "If you heard of a Home Demonstration Club meeting around here, would you attend?"

Answer: "Yes, if I could get there."

In this example, the respondent has qualified her answer but the qualification is not clear. For instance, she might mean one of several things. She might mean if she felt well enough she would attend. She might mean that if she had some means of transportation she would attend. Or she might possibly mean that she would attend if she had someone to keep her children.

2. Partial answers: Frequently, the first response to a question is merely a tentative expression of the person's real feelings on an issue or is only a partial statement of what he knows about a topic. Unless the interviewer gets an elaboration on such responses, the objective of the question is only partially met. The following example will clarify what we mean by partial answers.

Question: "What do you think of the idea of rationing?"

Answers: "Well, I think probably we gotta have it."

"I guess it is all right."

"Well, I don't hardly know; it looks to me like a lot of this stuff that they are puttin' on us during war time we could do without; I may be wrong but that is the way I feel about it."

None of these answers represents a complete or full expression of the person's real feeling about rationing. They imply that each had some reservation about it, but the answer does not indicate what their real feeling was.

3. Evasive answers: A third kind of inadequate answer is the one that evades the objective of the particular question. Evasions take several forms but there are three rather general kinds. One kind is where the respondent makes a very general response that is not directly related to the objective of the particular question. Here's an example:

Question: "What would you think of the Government gradually reducing the support price for wheat and eventually letting supply and demand set the price?"

Answer: "That supply-and-demand idea; it worked all right in the horse days but now with all this machinery I don't know about that. Right around here there's too many vacant farms. What's really hurt is plowing up all this good grass land here in the West and putting it to wheat. It's set us off our balance. There are too many big farmers now. In Western Kansas, they farm 6 to 10 sections of land. When a man farms 10 sections, he can make a good profit on a low price, and he buys machinery and more machinery and rents more and more land and pretty soon he owns the whole damn country. That's just what I think, because I've seen it happen. Now a small farmer, if he's gonna stand up to that, he's got to be diversified."

The respondent has not made a specific answer to the question. How he feels about the proposition of removing Government price supports for wheat and leaving the price to supply and demand is not clear in this answer. The interviewer made a mistake by not getting a specific answer to the question by either probing or repeating the question.

Sometimes a respondent evades the question unintentionally by giving it a

slightly different interpretation than the interviewer intended. The following three examples are common types of evasions of this kind.

Question: "What do you think will happen to wages during the next year?"

Answer: "I hope they stay up."

Question: "How does your income now compare with what it was a year ago?"

Answer: "I have a good income now, I'm working full time."

Question: "How do you think rayon curtains compare with cotton ones for looks?"

Answer: "I've seen some awfully pretty cotton curtains."

Each of these answers is unsatisfactory because none are specific answers to the questions. In the first place, for instance, the fact that the person hopes wages will remain up tells us nothing about what he *thinks* will happen to wages during the next year.

In the second example, we know that the respondent is working full time now and has a good income, but that information did not answer the question.

The answer in the final example is also evasive because the respondent has not made the comparison called for by the question.

Probing

You should never leave a question until the respondent has adequately answered or until you are convinced that he cannot answer. To get adequate answers, it is necessary to make frequent use of probes. A probe is anything the interviewer says to get the respondent to add to or explain what he has said or to direct his attention back to the subject when he has strayed from the question. There are good probes and poor probes. A good probe is one that gently encourages the respondent to amplify or explain a remark without in any way suggesting an answer or line of thinking to him. Examples are, "How?", "Why not?", "Why so?", "In what way?", "What do you have in mind?" A poor probe is one that suggests answers to the respondent.

Encouraging the Respondent To Talk: One effective way of probing is the use of devices to encourage the respondent to talk once he has made a brief response to the question. There are two devices which are especially useful for this purpose.

1. **An expectant pause** on the part of the interviewer. The interview should be a casual, easy situation in which the respondent has time to think things over without feeling that he must snap back with an answer. On some of the questions, he may not have well-formulated opinions and may want to "weigh the evidence" before answering. An expectant pause is often sufficient to convey the impression that the respondent has merely begun answering the question, and to invite him to amplify on the topic.

2. **Brief assenting comments** such as "Uh-huh," "Yes," "I see," or "That's interesting," which are very natural to social conversation and fit into the interview very conveniently, may be used frequently. By using these comments, the interviewer encourages the respondent to "tell more." He shows that he understands how the person thinks and feels, and indicates by his remarks and by his general

attitude that he can sympathize with or accept the way the person feels. He never argues, debates, or shows disapproval of **anything** which the respondent says. By doing these things, the interviewer provides an opportunity for the respondent to talk without biasing or influencing his answers in any way.

Repeating the question is never wrong as a probe and is particularly useful in the following situations:

First, when the respondent partially answers the question and then wanders off to some unrelated subject.

Second, when the first response to the question is an evasive answer of the type discussed above.

Third, when the response is meaningless. Sometimes the respondent does not hear all of the question, or misunderstands it and gives a meaningless answer. In such cases, the question should always be repeated.

Fourth, when the answer to the question is "Don't know," if the interviewer feels that the respondent did not understand the question.

Neutral Questions: The most frequently used method of probing, and the best for the inexperienced interviewer, is the use of neutral questions. "What do you have in mind?", "How is that?", "How do you mean?" are examples. Such questions are especially useful for getting the respondent to elaborate on a statement or an idea and to get him to explain something he has said.

Here is an extract from an interview which demonstrates how neutral questions are correctly used as probes.

Schedule question: "Do you think the United States has made any mistakes in dealing with other countries since the end of the war? (If Yes) What?"

Respondent: "Well, yes, I think this country has made some mistakes since the war along that line."

Interviewer: "What do you have in mind?"

Respondent: "Taking for instance our dealings with Russia, I think we have made a mistake there."

Interviewer: "In what way?"

Respondent: "I think we have taken the attitude that Russia should take the initiative and we seem to follow them instead of this country taking the initiative and them following us. Russia seems to want to have everything their own way and not give in to the opinions of this country and when there is a controversy they seem to want to be right and have us be wrong."

Interviewer: "Do you think this country has made any other mistakes since the end of the war?"

Respondent: "Yes, I think of another mistake which I think they have made."

Interviewer: "What is that?"

Respondent: "I think these secret conferences they've had are a mistake. They've made secret treaties that have been kept from the people and I think that is a mistake."

Interviewer: "How so?"

Respondent: "Because secret treaties are bad; when the people do hear about

them. they do not like them, no matter how good they are, because they still wonder if maybe something about the treaty is being kept from them. They don't like having treaties made without their knowledge, no matter how good the treaty is."

Notice how the respondent was encouraged to add to and explain his original answer by the use of neutral questions stated in a conversational manner. Notice also that the probes, "What do you have in mind?", "In what way?", "What is that?", and "How so?", in no way influence the respondent's answer. They do not "direct" his thinking in any particular way nor do they "suggest" answers to him.

Reassuring Statements: Sometimes all that is necessary to get an adequate answer is to reassure the respondent of his ability to answer. There are some people who hesitate to express themselves because of a feeling of inadequacy. For example, the poorly educated person may be conscious of his lack of information on the topics and for that reason hesitates to express an answer when he has the information. And there are people who have difficulty expressing themselves without considerable encouragement from the interviewer. The following examples of probing show how such people can sometimes be encouraged to make adequate answers.

Schedule question: "What do you think of the idea of rationing?"

Respondent: "Well, I guess it is necessary and all right, but of course I don't pretend to know so much about it."

Interviewer: "From what you do know about it, what do you think of the idea of rationing?"

Respondent: "Well, I think that the food rationing is necessary and it is all right to have it because if it wasn't rationed, I'm satisfied that the poor people wouldn't be able to buy what they need because the big man would get all of it. But the rationing on this farm machinery, I don't think that that is necessary, and it don't work."

Avoid Probes That Suggest Answers: One of the most common errors in probing is suggesting an answer to the respondent. Here are examples of such errors.

Schedule question: "What do you think will happen to farm prices in the first two years after the war?"

Respondent: "They're going down, no doubt about that because there will not be the demand for all this stuff after the war that there is now."

Interviewer: "Don't you think the Government will support farm prices as they agreed to do?"

Respondent: "Well, I had not thought of that, but that will make a difference; maybe they will not drop too badly."

Schedule question: "What are the most important things you look for in buying a house dress?"

Respondent: "I look for one made of good quality material and one that is made in good style."

Interviewer: **"What about the price?"**

Respondent: "Oh, sure, I consider the price too, that's important."

Schedule Question: "What do you think is the best thing that the United States could do to help keep peace in the world?"

Respondent: "I don't know; that is beyond me."

Interviewer: **"Do you think it would be a good thing for us to help feed the hungry people over there?"**

Respondent: "Well, yes, that would make them feel friendlier toward us I feel sure and that would help make peace. I did not think of that."

Schedule question: "In buying a house dress, would you rather have one made of cotton or rayon? Why?"

Respondent: "Well, for a house dress, I'd rather have cotton."

Interviewer: **"Because cotton wears better?"**

Respondent: "Yes, it wears better and is more serviceable for a house dress."

These four probes have one thing in common—they all suggest an answer to the respondent. The final answers cannot be compared to the answers other interviewers got by using non-suggestive probes because these respondents have been "enlightened" on the issues by the interviewer.

[This section on "Getting Adequate Answers" is of great importance and applies to marketing research without modification. Learning to probe correctly is difficult. If the interviewer puts words in the respondent's mouth, he will talk, but his views are likely to be influenced by what the interviewer has said. Indeed, the best way to get a full response (without regard to quality) would be to start an argument with the respondent on every opinion question or question on "reasons why." But probing should not be done in this way; it is necessary to use the "neutral" probes of the kind suggested here.]

RECORDING THE INTERVIEW

Contents:

A good interview is one in which the respondent's answers are reported in full and in the words he himself used to express them. The correctly recorded answer not only reveals the meaning of the respondent's comments but also indicates something of the kind of person he is and how strongly he feels about the particular issue. Here is an example of a correctly recorded answer and an example of the same answer incorrectly recorded.

Question: "Suppose that surplus wheat is produced in the next two years, do you think the Government should try to do something about it or not?"

Answer Correctly Recorded

"Well, there are some who say that when the Government stepped in be-

Answer Incorrectly Recorded

"Can't leave it to the farmer. He would over-produce. Got to have

fore it did not help much. They say the farmer is an individualist and that he likes to do his own running. Them folks tell you to leave it to the farmer and if he produces too much wheat he will cut down. Well, I say the farmer is an individualist and if he was turned loose on this wheat proposition he would produce too blame much. You got to have Government control because by God the farmers would just glut the market with wheat if we don't have that."

Government control or farmers would glut the market with wheat."

Summarizing or Paraphrasing a Respondent's Answers Should Be Avoided.

Use of this technique creates an artificial and unnecessary step between the source of the data and the final interpretation, and it is likely to result in oversimplification. It is easy for the interviewer using this method to distort, however unintentionally, the emphasis of an answer or to obscure the respondent's frame of reference or level of thinking. This is clearly demonstrated by the incorrectly recorded answer immediately below.

Question: "What are the most important things you look for in buying a summer street dress?"

Correct Recording

"Well, for a summer street dress, I want a cool material, that is the first thing I look for. And I look next at the price of the dress. I don't want to pay an outrageous price for a street dress, but I'll pay more for one of cool material."

Incorrect Recording

"Cool material and price."

Preliminary Remarks Made by the Respondent before Making a Definite Answer to the Question Should Be Recorded if Related to the Topic. Frequently such remarks help to explain the answer, as in the correctly recorded answer below.

Question: "Suppose surplus wheat is produced in the first two years after the war, do you think the Government should try to do something about it or not?"

Correct Recording

"I got fined once for raisin' too much wheat and I didn't think it was fair then and I don't think it's fair now. They didn't tell me either if I'd keep

Incorrect Recording

"No, I think those things are jus' a whole lot better left to the farmers. When we got a surplus, we got a low price, and then we cut down without

it off the market it was all right. I could have fed it to my hogs if I'd known that. But they didn't tell me that. So I paid my \$60 fine and then I could sell the rest. I think that's jus' some kids' trick playin' around. So I say no; those things are jus' a whole lot better left to the farmers. When we got a surplus we got a low price and then we cut down without the Government and price went back up."

the Government and the price went back up."

[The importance of correct recording cannot be stressed too strongly. Often when the interviewer condenses, she uses her words rather than the respondent's, and often the interviewer's ideas as well as her words creep into the recorded answer. Another advantage of insisting on a complete recording in the respondent's own words is that it is almost impossible for an interviewer to falsify interviews when each answer must display all the individuality of the individual respondent. The interviewer would have to be a literary artist, with a great talent for projecting herself into other people's personalities, to be able to falsify interviews under these circumstances.]

Mechanics:

To facilitate analysis, it is necessary that the interview write-ups follow a uniform procedure. The following rules for the form of an interview should, therefore, be observed.

1. The writing should be legible.

2. Use standard abbreviations for recording probes. We know that it is time consuming to record probes, but we want to know the content of each question you use as a probe and we are indicating below standard abbreviations for the more common probes.

<i>Probe</i>	<i>Abbreviation</i>
Why is that?	(Why?)
Why do you feel that way?	"
Why do you think so?	"
How?	(How?)
How so?	"
How is that?	"
Anything else?	(Else?)
What else?	"
How do you mean?	(Mean?)
What do you mean?	"
In what way?	(Way?)

When you ask your respondent to explain a word or phrase, you should indicate the probe by enclosing the word or phrase in parentheses. For example, "What do you mean by a good store?" would be (Good store?), etc.

3. To differentiate clearly between the answer of the respondent and the comments of the interviewer, all **probes, comments or explanations made by the interviewer should appear in parentheses in the write-up.**

4. **Cross-reference should be used** to call attention to material in one question which contains an answer, in whole or in part, to another question.

Specimen Schedule and Interview

For the guidance of the new interviewer, an example of part of a schedule and the answers to the questions are given here. The interview will demonstrate what we mean by full recording of answers and also the correct way of recording probes and the use of cross reference.

Notice how the interviewer has employed techniques to draw out the respondent and to secure sufficiently complete discussion of the topics, and how these devices are indicated in parentheses at the points at which they were introduced. Note also the use of cross references in the answers to questions 3 and 6.

Specimen Schedule:

1. Are you getting along better or worse than before the war began?
2. How about six months from now—how do you think you will be getting along then?
3. Do you feel that the war is making any difference in your chances of getting ahead? How?
4. Do you think prices of the things you have to buy will stay where they are? Why?
5. If prices were to go much higher, would you be able to get along the way you are now?
6. Do you think your income will change any in the next six months?
7. In your opinion why have prices been going up?

[The specimen questionnaire appears at first sight to contain some ambiguous questions, such as, "Are you getting along better or worse than before the war began?" Actually, the study appears to be a study of businessmen, and the introduction undoubtedly made clear the economic context in which this question was to be interpreted.]

Specimen Interview:

John Smith, New York City, Int. No. 6, Study 80, 1/22/43

1. I'm getting along worse. (How is that?) It's due to the lack of a lot of merchandise. You just can't get a lot of things. Some of them aren't being made anymore; some are held back for some reason; and some are frozen by the government. Some things are rationed for the wholesalers but not for the consumers. That makes it hard on us because people buy up more than their fair share.

2. It's hard telling. It depends on how many of the things will be available

then. (Do you think it will be better, or worse, or about the same as now?) The way it looks to me things will be worse. They may force us to drop out of the picture entirely. The handwriting is on the wall for a business like mine. It's only a question of time.

3. (See 1 and 2.)

4. That depends on how the government is going to stop the rising of prices. Judging from the way they have been going, I think they will go up. (Why do you think this is?) The farmer will compare himself with the skilled worker in the defense plant. Unless he can get more for what he raises he will go out and work for wages. (Are there any other reasons?) There's a slight other reason. It's that there is a black market in almost every field. It's just working on a small scale with us, but unless the government steps in, the picture will get worse.

5. Things would get worse. The earning power of any person is only just fixed so that would mean that he could buy less. When people buy less our profits are less. Profits are bigger if we can sell more, but today we must persuade people to buy less. It makes quite a difference in our work.

6. Yes, it will decrease. (See also 2.)

7. Prices have gone up due to profiteering. (What sort of profiteering?) It starts from the top and goes on down. In my field the top is the wholesaler or manufacturer. That's passed on to us and we must pass it on to the consumer on the ground that prices have gone up for us. The cost of production sends prices up too. If a man gets higher wages at the bottom, then the article he is making goes up. They say that wages have got to go up to meet the cost of living and then one thing just chases another.

GATHERING THE FACE SHEET DATA

For purposes of analysis, it is necessary to have knowledge of certain facts about each person interviewed. To this end the interviewer is supplied with a check-list or "face sheet" on which the desired data are to be recorded. They include such items as age of the respondent, his marital status, income, education and the like.

This material is best obtained after the respondent has answered the last question on the interview schedule. The transition from the interview to the face sheet can be readily accomplished by the interviewer's approaching it in a manner somewhat as follows: "Well, I guess that about covers the things I came to discuss. Thanks very much. Now before I go, there are some facts I'd like to ask for. As I was saying earlier, we are trying to talk to people of different ages and various occupations. To know that we're accomplishing this, we need to know a few things about the people we talk to, such as their occupation, income and so on." This gives the respondent a logical reason for our desiring this information and shows him why his cooperation will be of help. Mention of specific items such as age, occupation and income helps to prepare him for giving his sort of information. If there seems to be need for further reassurance, the interviewer may add: "We aren't interested in your name, of course. Each interview has just a number, and in our reports we confine all interviews to a summary."

The general idea is to set up a new interviewing situation, but without "protest-

ing too much." People are used to giving information about themselves to various Government agencies, so that gathering such data represents much less difficulty than new interviewers often imagine. If the interviewer is matter of fact in his approach, he will not encounter many problems.

[Face-sheet data, or classification information, is an essential part of almost every marketing research questionnaire. Often, however, marketing research questionnaires do not attempt to obtain the two most resented (yet highly useful) items: age and income. One method that may be used when this information is needed and the researcher does not want to ask the questions directly is to present the respondent with a check list in which age or income is shown only by broad classes, such as "30-39" or "\$4000-\$4999." Check lists are often used for other questions as well when the fixed-alternative rather than the free-response questioning method is used. For example, if an interviewer wanted respondents to rank five radio programs in order of preference, she might hand each respondent a card with the programs written on it. The interviewer would name the programs as she asked the question, and the respondent would be aided in his response by the card.]

CLOSING THE INTERVIEW

It is important to leave the respondent with a friendly feeling toward the interviewer and toward the agency he represents. The respondent should have the feeling that his time has been well spent and that he has contributed something to the Government by his answers. A few words of thanks will make the respondent feel that his efforts have been appreciated. It is also important that any questions or doubts about the interview which the respondent might have should be cleared up before the interviewer leaves. This can usually be accomplished by elaborating on the explanation of the purpose of the survey which was given at the beginning of the interview.

[One point not mentioned in the section on Closing the Interview is that the departure should not be delayed too long. Sometimes one encounters talkative respondents who want to prolong the conversation after the interview itself is finished. It is important for economic reasons to terminate such interviews, with tact and yet with firmness, as quickly as possible.]

3. Sampling problems in fieldwork. So far nothing has been said about the problem of *finding respondents*; only the problems of *interviewing them* have been considered.

The problem of finding respondents is different in probability sampling and in nonprobability sampling. With probability sampling, each respondent to be interviewed is located by an address, a map, or an aerial photo, by some rule for traveling, or by a combination of these methods. Usually these requirements will be somewhat complicated and will require intelligence to follow, but the only real technical skill that is needed is the ability to read a map or an aerial photo. This skill is quite rare among the adult population, even for people of high general intelligence. We shall not discuss map reading here beyond pointing out that real skill requires not only

theoretical familiarity with the subject but some actual field experience with the kinds of maps to be used on the survey.

Another problem in area sampling is that of "prelisting." If prelisting is specified, the interviewer would be required to list all the families living in a particular block, for example, as a preliminary to the selection of a subsample of these families for actual interviewing. Sometimes, when the interviewer is required to interview every n -th family in the block, the prelisting is in effect accomplished simultaneously with the interviewing. In any event, all the families in the block must be located in order to find the particular ones who are to be interviewed. The interviewer must then be a kind of detective in order not to miss families living above garages, in houses in the center of the block, and other unusual dwellings.

Still another problem in area sampling is that of deciding which member of a given family is to be interviewed. As has already been pointed out, it is fortunate that most marketing research studies require talking to the

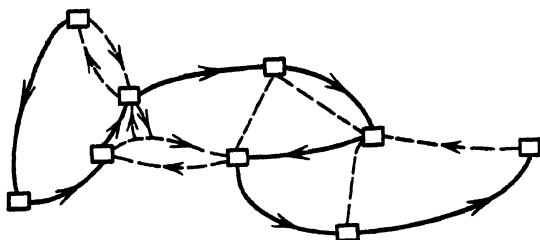


FIG. 3.

"housewife" or "female head of family," to obtain the needed information. Occasionally, however, special members of the family, sometimes randomly selected members of the family, are to be reached. The interviewer must thoroughly understand the procedure that is established for doing this.¹

In area sampling, the responsibility of the interviewer in selecting respondents is only that of following instructions carefully and intelligently. In arranging for call-backs on people not at home on the first call, however, the interviewer needs initiative and ingenuity if she is to accomplish her assignment efficiently. She must route herself so that she accomplishes the necessary call-backs at favorable times of the day or evening without too much costly traveling. In addition to ingenuity, this requires both training and an understanding of maps.

One procedure that tends to facilitate the making of call-backs is a routing that keeps "looping back" on itself. For example, suppose that nine blocks in a city were selected in an area sample, the position of the blocks

¹ See for example, Leslie Kish, "A Procedure for Objective Respondent Selection within the Household," *Journal of the American Statistical Association*, Vol. 44, No. 247, pp. 380-387, September, 1949.

being represented schematically by the squares in Fig. 3. By following this "snaking" pattern, the interviewer would be able frequently to go only slightly out of her way to make call-backs on people visited earlier. (Some possible call-backs are shown by the dotted lines.)

Route planning by itself is not enough. The time of day at which call-backs are to be made is also important, and the interviewer should be encouraged to arrange call-backs in the late afternoon or early evening or on Saturday and Sunday.

With nonprobability sampling, most of the problems of respondent selection that we have been discussing are typically of smaller or negligible importance, unless the nonprobability sample contains features of probability sampling (see Chapter 11). The interviewer is usually given rather general instructions as to whom she is to talk to and then permitted to select respondents who fulfill these general requirements. The most typical requirement is that the interviewer obtain certain numbers, or quotas, of men and women, rich and poor, young and old, white and Negro, etc. Particularly with respect to quotas for economic status, the interviewer must know precisely what is meant by "upper," "upper middle," "lower middle" and "lower," or whatever terms happen to be used, and she must recognize people who actually fit into these quotas. To be able to do this well, the interviewer must have an objective knowledge of the different classes of people to be found in the area in which she is to interview; in particular, she must recognize that there are many people who aren't at all like herself in economic status (either much wealthier or much poorer, since most interviewers probably are "middle class"), and she must be willing and able to find these people when the study calls for it. For a fuller description of the problems of selecting respondents when quota samples are used, see *Interviewing for N.O.R.C.*¹

¹ National Opinion Research Center, *Interviewing for N.O.R.C.*, Chicago, 1947.

CHAPTER 21

SELECTING, TRAINING, AND SUPERVISING INTERVIEWERS

21.1 Selection of Interviewers

In all personnel work the selection of the "right" people to fill jobs is a baffling problem. In spite of the new scientific and semiscientific techniques that have been evolved in recent years, there is still a high degree of uncertainty as to whether a new employee will succeed in the job for which he has been hired. This uncertainty certainly exists in hiring interviewers. The selection of interviewers often has to be done under conditions that do not permit adequate recruitment, interviewing, and testing. Sometimes the pressure of time may be so severe as to allow only a hurried opportunity for finding the interviewers and no chance at all for interviewing or testing them. Moreover, training and supervision of fieldworkers also is often conducted under far from ideal conditions, with the result that it is hard to know the qualities and characteristics that distinguish good interviewers from bad ones. This lack of knowledge makes the task of selection even more difficult. One must consider not only the problem of selection in the abstract, that is, under conditions that permit the use of all the ordinary tools of personnel administration, but also the ways in which it is possible to make these tools available. For this reason, we shall begin with a discussion of sampling that at first sight may seem distant from the problem of selecting interviewers.

1. Sampling design and selection of interviewers. To make this discussion concrete, consider an extreme but not unrealistic example. A national survey, involving interviewing in 75 cities and towns scattered across the country, is to be made. In many of these cities—say, 45—there are interviewers who have worked on previous studies, and these interviewers are available to work on the present study. In the other 30 cities and towns, however, interviewers must be selected and trained if the study is to be done. The time allotted for the selection and training of interviewers is only six weeks. The supervisor of the field staff and her assistants (if any) must travel to all these towns and find, select, and train interviewers in six weeks. The supervisor will have to travel from one strange town to another during this time. In each town she will have about two days to locate, select, and train the interviewers needed for the study. Even with careful advance planning, she will usually need at least one of these days

simply to locate and select the interviewers needed. If anything goes wrong, part of the second day will also be needed for selection of interviewers. At best, little time is left for training. Finally, as soon as one town is "set up," the supervisor will have to rush off on a train, plane, or bus in order to arrive at the next town in time to repeat the act.

It is easy to see that under these conditions the supervisor must use catch-as-catch-can personnel methods. She must make the snappiest of snap judgments in the selection of new interviewers. After a few weeks of strenuous routine like this, she is lucky to be able to do more than go through the motions.

Sampling design affects this process. In the above example, it was assumed that there were many new interviewing areas for which interviewers were not available and that it was necessary for this reason to recruit interviewers under conditions of extreme pressure. Under such pressure there would be hardly any possibility for applying scientific techniques in the selection of interviewers.

Suppose, however, that things had been a little different. Suppose that the national sampling design had been selected five years before and had been used repeatedly on national studies during the intervening time. Except for a few towns where the interviewers were ill, or busy, or had moved since the last study, the interviewers who had worked on previous studies would be available to work on this one. The problem of selection of interviewers could then be viewed in an entirely different light.

In the first place, the number of new interviewers to be recruited would be small, and therefore more time could be spent in recruiting each one. It would then be reasonable to think about applying more objective standards of selection.

It would also be possible to reexamine the existing group of interviewers. In areas where the present interviewers failed to meet a given standard of performance, new interviewers might be selected. In this way the quality of the existing field staff could be improved by setting up standards of reselection as well as standards of initial selection.

But what about the feasibility of using the same sampling design on repeated studies? This procedure is used by many organizations engaged in interviewing studies. For governmental work, the U.S. Bureau of the Census is able to use the same cities and towns, and therefore the same interviewers, on many different types of studies. For general social and opinion research, the Survey Research Center at the University of Michigan also uses the same sampling design (or subsamples from it) on repeated surveys. In marketing research, several companies have developed sampling designs that are used repeatedly.

In all these cases, the sampling design is a probability sampling design.

It is possible, of course, to select a single probability sampling design that will yield unbiased estimates for any characteristic whatsoever. It is not clear, however, whether it is possible to design a probability sampling design that will be reasonably *efficient* for the diverse studies that have to be made in marketing research.

One difficulty is the differences in the geographical variation of the various things commonly studied in marketing research. Geographical variations in the competitive conditions faced by flour mills are probably widely different from the variations in competitive conditions faced by shoe manufacturers, for example. This would mean that the sampling design which was most efficient for studying the sales of flour might be quite inefficient for studying the sales of shoes.

A second difficulty exists because many marketing studies concern regional or small-area characteristics. The particular regions or small areas of interest might not even be included in the national design. As a matter of fact, since almost any practicable national sample would contain a very small percentage of the number of areas of possible interest, it would be surprising if the national sample contained the areas being investigated.

Even with these difficulties, however, it is possible that the loss in statistical efficiency that might come with the use of an all-purpose sampling design would be offset by the improvement in the quality of interviewing that comes with the opportunity for more leisurely and critical selection and reselection of interviewers. Moreover, there is a way of selecting the all-purpose sampling design that may lessen the importance of the disadvantages that have been mentioned.

The idea is to select first a very large number of primary sampling units (counties)—say, 400—in such a way that either all the counties or only subsamples of them can be used. For many national studies no more than 30 to 50 counties would be needed, and these could be staffed with resident interviewers immediately.

When a national study arose that called for a larger number of primary sampling units, the 30 to 50 sampling units could easily be expanded into a larger sample containing, for example, 60 to 100 counties. Interviewers could then be recruited for new sampling units, and of course, the interviewers employed on the basic sample of 30 to 50 counties could be used on the new study.

When a regional or local study arose, still more primary sampling units would probably be of interest. For those primary sampling units that were already a part of the national sample mentioned before, interviewers would already be available. It is possible that some of the interviewers in those counties in the basic national sample could travel to other nearby counties

in the region. If it were unlikely that another intensive study of the same area would be made, it would be particularly desirable to do this. At any rate the number of new interviewers to be recruited would be kept as small as possible.

This last paragraph suggests an entirely different approach to the problem of selection. Up to now we have been talking solely of the system of employing resident, part-time interviewers (either directly or through local organizations or individuals) in different parts of the country to handle interviewing assignments in their own localities. This is the traditional and by far the most common method, but there is an alternative.

The alternative is to have interviewers travel from a central location (or several central locations) to the different parts of the country in which interviewing must be done. Of course, the travel expenses of interviewers are paid by the company. This alternative permits centralized recruitment and training with their attendant advantages. The net effect on the cost and time required for the study is not clear.

By requiring interviewers to travel to a certain central point rather than having supervisors travel to train each individual interviewer, the cost of traveling by interviewers is increased but the cost of traveling by supervisors is reduced; the net is probably an increase in the total cost of doing the study, although the percentage increase may be relatively small.

The effect on the total time needed for the study is even more difficult to guess. On the one hand, the total time of training and selection is almost certainly reduced. On the other hand, if one interviewer must work in several counties rather than one, the time needed for actual interviewing will be somewhat increased. The net effect could be either an increase or decrease.

It is quite possible that a compromise between resident and traveling interviewers will yield many of the advantages of central selection and recruitment without materially increasing either the cost or the time of the study. That is, interviewers can be recruited and trained at "substations" throughout the country rather than in one central point. The interviewers can travel for relatively short distances from their home counties to neighboring counties included in the sample.

Under some circumstances, then, it seems desirable to use a fixed sampling design for many studies, and under some conditions it is better to have interviewers travel from area to area, at least within a limited geographic region, than to have a resident interviewer in each county of the sample. Such a procedure would increase the chances of adequate selection and training. Better selection could be made initially because fewer interviewers would have to be hired. Reselection is equally facilitated

because it is easier to evaluate the work of a relatively small number of interviewers than a large number. Training and supervision are similarly made easier when the number of interviewers is smaller.

The all-purpose sampling designs in use today are, so far as we know, probability sampling designs. For the purpose of economizing the number of interviewers to be selected, it would make no difference whether the design was a probability or nonprobability sample. If, however, an all-purpose sample is to be set up for the first time, there would seem to be little reason for not selecting the *counties* or primary sampling units by probability sampling methods. The added expense of probability sampling comes chiefly in the selection of the sample *within counties*. The cost of drawing a probability sample of counties is nominal, and this sampling method assures an unbiased design for any study whatever, up to the point of subsampling within counties. If the counties are selected by judgment sampling, they may be quite satisfactory for one study but badly biased for other studies. If the counties are selected by probability sampling, there is no reason why nonprobability sampling cannot be used *within counties* for specific studies for which probability sampling would seem to be too costly.

The discussion of the interrelations between the problems of sampling design and the problem of recruitment of interviewers is incomplete at one point. It makes a great deal of difference whether the group conducting the study has a preexisting field staff or not. If a marketing research company, for example, is beginning from scratch, there is little doubt that the best method is to select a probability sample of counties and then to recruit interviewers to "activate" this sample. If the company already has a field staff, it still seems clearly desirable to use a fixed sampling design and cut down on the total number of interviewers that must be dealt with. For such a company, there is some question as to whether the fixed sampling design should not be a judgment sample of counties that employs only counties in which there are already interviewers. One compromise in this situation is to make a gradual transition from the old to the new until finally the fixed sample of counties is a probability sample.

2. Finding interviewers. Marketing research has had most of its growth during the 1940's, a period in which there was no serious unemployment in the United States. Because of the "overemployment" of the labor force during most of this time, it has been exceedingly difficult even to *find* people who are willing to interview, to say nothing of being selective in considering alternative applicants. It is, of course, hard to speculate on the future, but it seems reasonable to believe that the finding of prospective interviewers will continue to be an important problem in marketing research.

An interviewing supervisor will often find herself in a strange town in which she knows no one and where she must find one or more competent interviewers within a few hours. There are certain advance preparations that she may make before arriving at this town; and, as we have seen, if an effort is made to use the same all-purpose sample from one study to the next, it is possible that more time may be allowed for locating interviewers.

If it is possible to obtain, in advance of the trip, the names of people with interviewing experience or who are interested in obtaining interviewing experience, this should be done. The main sources for finding interviewers are likely to be leaders of local organizations. These people can often be located in advance of the trip to the city or town in question. If a letter is sent to them in advance, they can become acquainted with the nature of the jobs to be filled and can be told when the interviewing supervisor will arrive. Ideally, these people could locate candidates and be able to furnish names to the supervisor immediately on her arrival in the town. More often, however, the letter will serve only as a means of facilitating the establishment of working relations between the supervisor and the local community leaders. Some kinds of local sources that may be helpful are listed below:

1. The local newspaper(s) or radio station(s). Newspapers and radio stations often conduct marketing research studies of their own and may therefore know of interviewers or people interested in interviewing. The advertising department is usually the best source, since much of the marketing research done by communication media is conducted under the supervision of the advertising department. In addition to advertising people, however, editors, society editors, etc., usually know a great many people in the local community and are therefore able to give suggestions.

2. Leaders, such as officers of women's clubs, church groups, community chest, farm organizations, chamber of commerce, etc.

3. School officials and librarians, including local colleges, if there are any.

4. Marketing instructors of colleges and universities.

5. Anyone with previous or current interviewing experience. One interviewer in a community will very likely know others, so as soon as one is found she may serve as an entering wedge for finding many others.

6. "Help wanted" advertisements in local newspapers, and employment agencies.

When the ordinary sources fail, the supervisor may try casual sources, such as the hotel manager or the waiter in the hotel restaurant. Persistence is usually rewarded, even though a great deal of time may be needed. When all else fails, however, the best thing to do is to go to a nearby town (preferably a larger one) and start all over. The interviewer finally selected can then travel to the town in which the interviewing must be done.

When the sample requires interviews in very small towns, it is frequently

wise to look for interviewers in larger nearby towns, even though the interviewer may have to travel 20 or 30 miles to get to the areas in which interviewing is to be done. With area sampling the entire county is usually the sampling unit. Interviews will typically be scattered in blocks and open-country segments throughout the entire county, so that interviewers will have to do a great deal of traveling. It is then desirable to go to the largest town in the county in the attempt to find interviewers, although smaller towns, such as college towns, may have special advantages. If the county doesn't possess a large town, towns in immediately adjoining counties should be used as sources of interviewers.

3. Selection of interviewers. We have now arrived at the point at which we can discuss the problem of selecting, rather than the problem of locating, interviewers. The care with which selection can be made depends on the number of prospective candidates in relation to the number of interviewers needed, and upon the time available for selection. We assume both that there is a reasonable number of candidates and that the time available for selection is adequate and suggest what can be done under these conditions. When these conditions are not met, obviously less elaborate procedures are in order.

In order to select wisely among prospective interviewers, it is necessary to know the characteristics of good interviewers and to distinguish people who possess these characteristics from those who do not. A general list of desirable characteristics is given under The Employment Interview, later in this section.

There are numerous and conflicting assertions by experienced marketing researchers as to the kind of people who have these characteristics. Some assert that women are the best interviewers; others prefer men. Some claim that students are best, while others claim that students are very often too immature and sometimes dishonest. Some feel that previous experience is essential; others that it is positively undesirable.

We shall not enter into these controversies beyond certain specific suggestions that will seem warranted later. By and large, however, it seems unwise to single out any particular group of people as being especially qualified or unqualified. The group from which an individual comes is at best only partial evidence of the fitness of the candidate, and it should be considered as such only when much experience has indicated that members of the group do make particularly good or particularly bad interviewers. It has been shown, for example, that Negro interviewers get more complete information from other Negroes in attitude research on social issues than do white interviewers; it has not been shown, formally or otherwise, that this carries over to the kinds of problems investigated in marketing research.

One clue to the solution of the problem of selection comes from an earlier

comment that *reselection* is as important as initial selection. Since interviewers are typically rehired, the researcher is free to reconsider each interviewer on the basis of her actual performance (in so far as the quality of her performance can be realistically judged). It is much easier to rectify errors in initial selection of interviewers than errors in the selection of individuals for full-time jobs, where "firing" is usually painful and time-consuming. It therefore seems desirable to focus attention on the problem of reselection rather than initial selection. In this way the quality of the field staff can be continually improved (regardless of how good or bad the methods of initial selection), if one can distinguish differences in the quality of completed interviews.

For example, suppose that a study required the use of 100 interviewers. One could pick out the 10 (or any other number) interviewers whose work seemed poorest (assuming, of course, that the work of all 100 was not satisfactory on an absolute scale). One then would have the choice either of retraining these 10 interviewers or not using them again and hiring new interviewers in their place.

Suppose, for illustration, that 10 new interviewers were to be hired in their place. Assuming that the methods of selection were no better than when the 100 interviewers were picked in the first place, it is reasonable to expect that on the average one or two would be as bad as the interviewers they replaced, but that the remainder would be better. Hence, even if the methods of initial selection were not particularly good, it would be possible to improve continually the quality of the existing field staff.

Fortunately there are methods for judging the quality of work that has been completed in the field, and we suggest these in detail in section 21.3. For the present, however, we shall concentrate on the other approach to the problem, that of initial selection of interviewers.

Many techniques can be used in the initial selection of interviewers: application blanks, employment interviews, references, and personnel tests. All these techniques are useful to the extent that they indicate the presence or absence in the prospects of those characteristics deemed essential for the jobs being filled. Let us consider each of the possible methods.

1. *Application Blanks.* Many people who hire interviewers require the completion of an application blank as a preliminary to actual hiring. The application blank actually serves a double purpose: it must provide information useful in initial hiring and it must obtain facts that will be useful later on. There is, of course, overlapping. In the latter category falls such information as name, address, telephone number, social security number, etc. We shall consider only information useful in initial hiring:

ABILITY TO FOLLOW INSTRUCTIONS. Usually the ability to follow instructions is greater for people who have progressed farther in school.

Therefore the amount of formal education is relevant. Previous jobs also provide an important clue, although the vagueness of job titles and descriptions makes it difficult to evaluate previous employment records. Previous interviewing experience, in particular, is very difficult to appraise. Sometimes experience is a disadvantage because the interviewer learned poor techniques. In order to appraise previous experience, it is essential to know the exact kind of interviewing experience and for whom the interviewing was done. Finally, the ability to follow directions can be partially tested by the manner in which the employee fills in the application blank itself.

WILLINGNESS TO FOLLOW INSTRUCTIONS. The kind of information obtainable on the application blank seems almost useless in evaluating willingness to follow instructions unless the applicant does a very poor job of filling out the application form because of unwillingness rather than inability.

ABILITY TO ESTABLISH RAPPORT. Many characteristics—age, sex, education, etc.—are related to this important ability. Evidence is largely based on hearsay and personal opinion. In most types of marketing investigations women appear to have some advantage over men in putting respondents at ease and encouraging them to give full answers to questions. Similarly, it is rather commonly believed that very young people are at a disadvantage in establishing rapport with respondents. Neither of these statements is supported clearly by enough evidence to make it conclusive, but both may be useful as guides.

PHYSICAL STAMINA. This quality is seldom investigated on application blanks, yet it is important and probably very easy to ascertain. A question like the following, "Do you think you would be able to walk five miles and stand around most of the rest of the time in a day's work?" would probably screen out many obviously unfit candidates by discouraging them from going further with their application.

TIME AVAILABLE FOR THE WORK. The applicant should be asked about other part- or full-time jobs (including housework, studying, and civic activities). She should be asked, also, both the *average* number of hours per week and the *maximum* number of hours she can in any ordinary week devote to interviewing. Because fieldwork is usually intermittent, the second question is often more useful than the first.

MOBILITY. The applicant should be asked if she can drive and if she has access to a car (the first question being more important, since it is usually possible to rent cars). She should also be asked the extent to which she would be willing to travel, and, in particular, whether or not she is willing to stay overnight in a strange town.

It is possible to consider tests and names of references as a part of the

application blank. We discuss these devices separately, however, later in this section.

2. The Employment Interview. The employment interview provides the employer with information about the prospective employee and provides the applicant with information about the job that she is being interviewed for. The applicant should be told as precisely as possible what the work of interviewing is like. If the advantages are overstressed and the applicant accepts the job with a definite misunderstanding as to what the work is like, the risk is high that the new interviewer will quit on the first day's work and that the job of selection will have to be redone. On the other hand, the advantages should also be indicated. In short, the applicant must be given as objective a picture as possible of the job she is applying for in order that (a) she won't accept the job and quit almost immediately and (b) she won't refuse to consider the job because of having been told only about its unpleasant features. While there is little question of the desirability of giving the applicant an objective picture of the job, interviewers are sometimes hired under conditions that make it hard to do so. If it is necessary to hire two interviewers in a strange town within a very short period, there is a temptation to "oversell" the job in order to be sure that two interviewers get hired within the allotted time.

It is difficult to describe how one gets information about applicants, since judgments formed in the course of conversations necessarily are highly subjective and depend greatly on one's own past experiences. The only general rule that is likely to be of much help is this: the applicant should be encouraged in every way to talk freely about herself, her experiences, her interests. She should be put thoroughly at ease. The supervisor, in this part of the interview, should ask only questions that will keep the applicant talking about relevant points. The length of the interview is likely to be dictated by the amount of time available for hiring, but even if the time available for the interview is very short, every attempt should be made to prevent the applicant from feeling that she is being given only a hasty screening.

The information needed about the applicant concerns chiefly those characteristics that are hard to obtain information about on the application blank. Here are some important ones:

1. The ability to follow instructions, which is chiefly a function of general intelligence, can usually be judged rather easily by personal conversation. It is easy to observe how quickly the applicant understands new ideas, how quickly she grasps the nature of the job she is applying for, how well she expresses herself, etc.

2. The willingness to follow instructions can probably be fairly well evaluated by the employment interview, although this evaluation is highly

subjective. The important thing to be found out is whether the applicant is interested solely in picking up a little extra money by part-time work, or whether she has some interest in interviewing. In addition, it may be possible to steer the conversation so as to find out about the applicant's persistence and conscientiousness in carrying through other jobs and duties that she has undertaken.

3. The ability to make people feel at ease can also be judged by conversation, but again, it is hard to say in objective terms how this can be done. The hardest task, perhaps, is to allow for either bravado or timidity, each of which tends to be brought out in different people by the circumstances under which employment interviews take place.

The remaining characteristics—physical stamina, time available for the work, and mobility—are all reasonably objective. The interview should serve only to verify the information given on the application blank. One should make sure, for example, that the applicant really can devote as much time as she says, since people frequently have been found to overestimate the time they will be able to devote to part-time activities.

3. *References.* It is possible on the application blank to obtain the names of people who know the applicant and can provide information that will be useful in the task of selection. Ordinarily, however, there is not enough time actually to contact the references before a decision must be made on hiring the candidate. Even if time is not available to check the references, it may be worth while to obtain references simply to find out whether or not the applicant will list former employers as references. If she is willing to list former employers, the chances seem high that her performance on earlier jobs was reasonably good.

When there is time to get in touch with references, the main problem is to obtain candid opinions. People are often evasive or noncommittal about the poor qualities of someone they are asked to evaluate. Therefore references should be asked pointedly and specifically about the applicant's weaknesses. Indeed, the main reason for getting information from references is to find out about possible weaknesses. A job applicant will usually list as references people who have a high opinion of her, but it is sometimes possible to find out serious shortcomings, if there are any, by asking directly about them.

4. *Personnel Tests.* From the description of the conditions under which fieldworkers must be hired, it should be clear that extensive personnel testing is usually not feasible. It is possible, however, that simple intelligence tests can be administered, possibly as a part of the application blank itself. The ability to follow instructions can then be more accurately ascertained than by conversation and other less direct objective evidence.

There is little doubt that a more elaborate battery of personnel tests

could be fitted to the requirements of interviewing and could materially aid further in the weeding out of people who have little or no chance of becoming competent interviewers. The listing of desirable qualities under Qualifications for Interviewers, in section 20.2, suggests the areas of personality that such tests might explore. If time permitted, such tests might be tried. To our knowledge, however, there has been little or no systematic work in devising personnel tests for hiring interviewers, and for this reason we shall not consider them further.

A different kind of test that is both useful and feasible is to have the applicant conduct some interviews. If the prospective interviewer is going to be trained for the task of interviewing, her work in this training period is an excellent indicator of what she can do. The ideas in section 21.3 on Evaluation of the Quality of Fieldwork are applicable in judging the quality of work done in the training or trial period. There are certain cautions, however, in using this method.

In the first place, the interviewer should be given the benefit of training before she is asked to do a "work test." No one can be expected to interview well without instruction. Even interviewers with previous experience may have learned many things incorrectly.

In the second place, the applicant can display only what she is capable of doing. If later she does not try as hard as she did on the work test, she will naturally not do as well. Moreover, she may not perform as well as she might with additional training and experience. The chief finding that can come out of a work test, then, is that a person *can* interview well. If a person does not do well on the test, this does not mean that she *cannot* interview well.

At first sight it might not seem practicable to attempt to combine selection and training. If several applicants are accepted for a training test, it is possible that all might be satisfactory and that therefore it would be embarrassing to select one and tell the others that they had not qualified. Moreover, many applicants, particularly those with previous experience, might conceivably resent such a method of selection. Neither of these objections is of great practical importance.

In the first place, it is rare that one would end up with a larger number of competent interviewers than were actually needed. More often the opposite result occurs. Often interviewers who seem promising during the preliminary screening and selection will drop out after the work of the training period. More often, applicants who seemed good on paper will do poorly in actual fieldwork.

But even if the unusual should happen and more interviewers survived the training test than were needed, the job that has to be done could usually be divided up among the qualified applicants. If, for example, the

supervisor wished to hire one interviewer for three weeks' work in a particular county, she could always hire two interviewers and give each of them one and one-half weeks' work. Then in future studies there would be added assurance of getting work finished in that county because of the known presence of two trained interviewers.

The other possible objection—that many applicants will resent having to go through a testing period along with training—is equally spurious. It is true that some prospects may resent such an apprenticeship. However, if the applicants are paid for the training period, the only ones likely to resent the training test are those among the veteran interviewers who are convinced that they already know almost everything about interviewing. These applicants should probably be eliminated unless there is unequivocal evidence that they really are the superb interviewers that they claim to be. For the others, the training test is a great advantage. In the first place, they are paid for the work. In the second place, they are given a chance to find out what the work is really like and to quit gracefully if they don't like it, rather than to have to give up on an important assignment.

Thus the problem of selection merges with the problem of training, and we shall now turn specifically to the problem of training.

21.2 Training of Interviewers

In practice, interviewers are trained in many different ways and in widely differing degrees. Interviewers have too often been hired on sight, given a pile of questionnaires, and told to fill them in. In other instances the training period may consist of a half-hour's oral instruction and two or three practice interviews. In still other cases, training may occupy several days and include not only careful instruction on all phases of the work but practice interviews followed by critiques that are themselves followed by still more practice interviews, etc.

From the interviewing manual quoted in Chapter 20, it should be obvious that interviewing is not a skill that can be expected to exist without training. Organizations engaged in general social research, such as the Survey Research Center at the University of Michigan and the National Opinion Research Center at the University of Chicago, sometimes devote several days to the training of new interviewers. In marketing research the training is usually much shorter, but there is much to be said for lengthening the time spent in this essential stage.

There are two problems in training: initial training and special training for specific jobs. If the initial training is well done and if subsequent work is closely supervised, the training for specific jobs is relatively simple and can often be carried on satisfactorily by mail. There is, however, no substitute for thorough initial training.

We first consider initial training and then turn to the problem of training for specific studies. There is actually some overlapping between these two stages, since interviewers are almost invariably hired initially with specific studies in mind. Nonetheless it is helpful to consider the two problems separately.

1. Initial training. The first step in training a new interviewer is to acquaint her with the general nature of the work: what marketing research is about and what the particular study is designed to do. She should be then told that there are two main things she will have to do: find respondents and interview them. If area sampling is being used on the study, the problem of finding respondents will require some special instruction in map reading. Training in the interviewing process itself can be secured by having the trainees interview the instructor, who points out errors and commends correct performance. Up to this point the training has been in the "classroom." It is necessary to have this kind of training at the start, but there is little hope that it will, by itself, make a very deep impression on the trainees. Doing is necessary.

The next step is to send the trainees into the field to conduct interviews. Ideally, the trainees should initially be accompanied by the instructor who can not only catch and correct bad habits in the incipient stage but can conduct one or two interviews herself to show the trainee how interviewing should be done.

Finally, if time permits, the trainee should be sent out on her own. On her return her work should be gone over thoroughly and each interview should be discussed. It is usually at this stage that most of the questions will be asked by the trainee.

Although some confusion may exist regarding the location of respondents because of difficulties in map reading or in placing respondents in the proper quota, the major problems will almost always be in the interview itself. These problems should be discussed thoroughly, both for the purpose of instruction on the issues raised and to impress the trainee with the importance of careful work.

2. Training for individual studies. Written instructions for interviewers are an essential preliminary to the training of interviewers for specific assignments. It is always desirable to write down the interviewing specifications on paper even when the major reliance in training is on oral instruction. Written instructions serve also to aid in the oral instructions, much as textbooks supplement lectures in schools. Most important of all, however, written instructions serve as a reference for interviewers to use after the supervisor has departed. Many new questions arise after the interviewing begins, and often these can be answered by a good set of written instructions.

When personal instruction is impossible, complete reliance must be placed on written instructions. If the interviewers have been well trained initially, this practice is probably satisfactory. Every effort must, of course, be made to produce general instructions that are clear, unambiguous, comprehensive, and interesting; and if this has been done, there seems to be no reason that a well-trained interviewer cannot learn from written instructions concerning a particular study what is required of her. Some provision, however, should be made for occasional personal contact with supervisors (say, at least once a year).

If interviewers are not well trained initially, however, one cannot hope to train them for specific studies by means of written instructions alone. An interviewer who has never received adequate personal instruction can hardly be expected to read instructions carefully enough to understand them as well as she should. More important, she will not understand the basic requirements of interviewing well enough to realize when she is not following the instructions properly.

The instructions should themselves include in a prominent place the instruction that they are to be read carefully. Most research organizations pay the interviewers for the time taken in reading instructions, and if this is done, this information should be included in a prominent place at the beginning of the instructions themselves. The main points to be included in instructions are listed below.

1. Purpose of the study. The interviewer must have a clear understanding of the purpose of the study, if she is to be expected to take a serious interest in it. Naturally it isn't necessary to go into great detail, but the description of the purpose of the study should be accurate and reasonably complete. No attempt should be made to talk down to the interviewer or to distort the true purposes of the study. It may be necessary to give her some information that might be regarded as "confidential," such as the name of the company that is conducting the study. Obviously, results might be biased if respondents knew that a particular automobile company was conducting the study on preferences in automobile design,¹ but it is not always true that the best way to keep this information from respondents is to keep it from the interviewer. If the interviewer is at all curious, she can probably get the desired information on her own. If she finds out by herself, she is more likely to tell the respondents the truth if they ask her who is conducting the study. If, on the other hand, the interviewer is told

¹ Moreover, it may be important to keep from competitors the fact that one particular company is conducting a survey (such as a test of a new product) that may presage an important policy change. We suggest that confiding in interviewers is a good way of preventing "security leaks," but there are many who would disagree with this position.

in advance but told also that it is essential to keep the information from respondents, she is much more likely to keep the information to herself. The interviewer must, however, be given a stock answer to the respondent's question, "Who are you working for?" Usually the interviewer should state only that she is working for a research organization or an advertising agency.

2. *General method of the study.* To encourage intelligent and interested participation by the interviewer, she must be told not only the purpose of the study but its general method as well. Technical detail is not necessary, but it is usually possible to explain the sampling design and interviewing methods in a general way so that the interviewer will have some idea of how the sample was picked and how the interviewing will be carried on. This is important not only to help ensure the interviewer's interest but to enable her to answer intelligently and accurately the questions that will be asked by curious respondents.

3. *Instructions on finding respondents—sampling.* The necessary instructions on finding respondents are usually quite detailed. With quota sampling designs, the interviewer must be given as much specific detail as possible on the general areas in which interviewing is to be done and the qualifications for respondents in each of the economic, ethnic, educational, etc., groups that make up the quotas. While inevitably a rather wide range of discretion must be left to interviewers when nonprobability sampling is used, no more discretion should be permitted than is necessary. If, for example, interviews are not to be conducted by telephone, the interviewer should be told. Otherwise, she is perfectly within the letter and spirit of the instructions if she completes her interviews in this way. Similarly, it may be necessary explicitly to write down prohibitions of interviewing in office buildings and downtown areas or directions for interviewing only at certain times of day or on certain days of the week. The general rule to follow in writing instructions of this kind is to put oneself in the place of the interviewer and ask if the instructions are really clear and complete on what is expected.

In area sampling, the instructions are much more explicit and usually more technical. The interviewer is given no leeway as to the people or families to be interviewed, and the instructions for finding these people inevitably require considerable detail. If the interviewer has never worked before on an area sample, she should be given oral as well as written instructions, no matter how skilled she may be. Some points that should be covered in both the written and oral instructions are these:

a. The importance of calling only on the families designated and making no substitutions, unless substitutions are a part of the design for one reason or another.

b. The fact that while her county may apparently include a nonrepresentative sample, the national sample will still be representative. Most interviewers who have worked only on quota sampling assignments don't get an accurate impression of the types of people who live in their areas, and they may need reassurance that the area sample will really work even though their particular interviews seem to be with an atypical group. There is usually no need to convey technical ideas to interviewers, but they should be reassured that their work will be merged into the work of other interviewers in order to make a good national cross section. Otherwise morale may suffer if the interviewer feels that the total sample is no more accurate as a picture of the entire country than her particular interviews are as a description of the county in which she lives.

c. The interviewer should be told precisely how to find the blocks and open-country segments in which she is to conduct her interviewing. Almost invariably these areas will be marked off on a map in order to help her. Here, of course, is another reason for giving personal instruction to an interviewer who is undertaking her first area sampling assignment. She will have to locate areas from the map, something that few people can do without special personal instruction and supervised practice.

d. The interviewers must be given precise instructions as to whom they are to interview in each block or open-country segment in their assignment. These instructions will vary greatly from study to study, since different variations of subsampling within these areas may be used. If, for example, all the families in the block are to be prelisted before the subsample is to be chosen, the interviewer must be told exactly how the prelisting is to be done. She must realize that *all* families living in the area must be listed, even though they may live in back alleys or apartments above garages or stores. She should be given instructions as to what to do in most of these commonly occurring "emergencies."

e. The term "family" must be defined clearly—a sometimes difficult task—and the interviewer must have unambiguous instructions as to whom she is to interview in each family. This is particularly important for studies in which certain members of the family may be the only ones qualified to give the required information accurately. If it is necessary to talk only to the husband, the housewife, or a teen-age daughter, as the case may be, the interviewer should be told.

f. Instruction on call-backs must be given. The typical requirement is to have a specified number of call-backs on families not at home on the first try, but at other times the interviewer may be given more discretion as to how many call-backs are to be made. Almost always, however, substitutions are not permissible and the interviewer must know this.

4. *Instructions on interviewing.* General instructions on interviewing

were discussed in Chapter 20. Each interviewer should be familiar with the contents of an interviewing manual of the kind reproduced there, and she should have had previous training under personal supervision. Usually it is unnecessary, therefore, to repeat general instructions on how to interview, although some of the most important rules may be repeated briefly for review and emphasis.

The interviewing instructions that should be included concern the particular questionnaire that is to be used. The interviewer must be told everything she needs to know in order to complete her interview in the desired manner. In particular, each question should be explained in as much detail as is necessary, and any instructions for skipping certain questions that are not to be asked of all respondents should be plain. The explanation of individual questions should tell what information is to be obtained in response to each. Where the question is completely self-explanatory, this is, of course, unnecessary. While good questions should be unambiguous, there is often some possible ambiguity in spite of the most painstaking pretesting. For this reason, the question's meaning should be spelled out whenever it seems possible that alternative interpretations could arise in the field. It is also helpful to give the interviewer the benefit of all the things that were learned on the pretest. Often, words used in the questions will not be understood in the same way by all respondents, and the interviewer should be told of any common difficulties that arose on the pretest because of this fact.

Sometimes questions are included that appear foolish or unnecessary to the interviewer. In these cases the instruction should explain the need for the particular question, so that the interviewer will feel that this question really is important even though on first glance it appears to be superfluous or a simple duplication of an earlier question.

Certain questions will require especially persistent attempts to probe, while for others no probing at all may be necessary; the interviewer should be told what is expected on each question.

When the recording of the answer is at all complicated—as, for example in precoded questionnaires—the interviewer should be given a sample questionnaire already filled in and accompanied by comments explaining each entry.

Instructions on communication with the home office, due dates for completion on the material, etc., must also be given. Interviewers should be encouraged to ask questions when they encounter situations not covered explicitly by instructions. It is important that the interviewers feel that the home office attaches great importance to following instructions precisely.

It is not easy to generalize about the length of instructions. Very im-

posing documents probably discourage some interviewers from a thorough reading and study. On the other hand, abbreviated instructions often leave out many things that the interviewer must know. The important consideration is probably not so much the length as the extent to which the information contained is helpful to the interviewer. If she is told the most obvious things with much repetition, and if the things that are important but not obvious are omitted, the interviewer will naturally find it difficult to put much effort into studying the instructions. If, on the other hand, she is given the benefit of all the experience gained in developing and pre-testing the questionnaire, the instructions will be helpful and length should not be a serious drawback. Interviewers who haven't the patience to read reasonably detailed instructions should be weeded out; mere brevity in instructions is not likely to encourage these interviewers to give the instructions the necessary attention.

21.3 Supervision of Interviewers

The problem of supervising interviewers has undoubtedly attracted more discussion than any other phase of survey work, aside from sampling. Much of this discussion has revolved about one particular issue which in its most candid form can be phrased, "How can interviewers be kept from cheating?"¹ Actually the problem of cheating is a relatively easy one to overcome by comparison with other problems that arise in supervision. Moreover, supervision itself is certainly no more difficult or crucial than the two steps that must precede it, selection and training.

The main job in supervising interviewers is to see that the fieldwork on a study is carried out as competently as possible. It is usually not feasible to do this by directly watching the fieldworker in the performance of her work. The only way this could be done would be to have one supervisor accompany each interviewer during all her work, and this practice is obviously impractical. At best, the supervisor will see the interviewer only about once a day; and usually supervision is not even this close, being limited to telephone, telegraph, or mail correspondence. Supervision is largely indirect.

At first it would seem an impossible task to exercise indirectly any degree of control over a far-flung field staff. The task is in fact made possible only by reasonably good procedures of training and selection. If the fieldworkers are incompetent or badly trained, supervision will be quite inadequate to achieve competent fieldwork. If selection and training are reasonably good, supervision can be very helpful. The most important part

¹ "Cheating" is the process of filling in questionnaires as though in accord with instructions when actually instructions are knowingly violated. In its most extreme form, cheating consists of filling in questionnaires without doing any interviewing.

of the job is the ability to evaluate the quality of work done in the field and to do this fairly quickly. Evaluation of the quality of the fieldwork is by far the most important part of good supervision. For if one is able to tell good work from bad work, one is then in a position to solve not only the perplexing problem of cheating but the less spectacular yet more widespread problem of substandard interviewing.

1. Evaluation of the quality of fieldwork. The interviewer has two main jobs, finding respondents and interviewing them. The supervisor must be able to tell how well each of these jobs is done without being on the spot to judge by direct observation.

How can the supervisor tell if interviewing has been done correctly? As with many problems of survey technique, the answer depends on whether probability or nonprobability sampling is used. In nonprobability sampling, if the interviewer is not required to take the name and address of the respondent, there is no way by which the supervisor can tell if the sampling instructions have been followed. Often it is not desirable to record the respondent's name for the sole purpose of checking up on the interviewer, because many respondents feel that the loss of anonymity constitutes too great an invasion of privacy. Sometimes, however, it may be necessary to know the respondent's name in order to get information from the same person at a future date, as in panel studies. (Sometimes an address may suffice.) In any event, if a respondent has been identified in such a way that the same respondent can be reached again, it is possible to reinterview the respondent to find out if the interviewer has really followed sampling instructions. It is possible to reinterview a sample of each interviewer's work and check not only the way in which sampling instructions were followed but also the other aspects of the interview: whether the person was really interviewed, how accurate the reported information about the respondent was, the way in which the initial interview was conducted, etc.

Reinterviews are expensive, however, and for this reason are seldom made. The more common procedure is some variation of the "post-card check." This requires sending a post card to a sample of the respondents to find out about the interview. This method, however, is useful more as a policing measure than as a check of quality. Many of the respondents will not send back post cards at all. The post cards that do come back are of relatively little use. About all they can indicate is that the interview was made; almost nothing can be learned about the adequacy of the interview or whether the respondent really fitted the quota to which the interviewer had assigned him. Moreover, some of the post cards are likely to reach persons other than the persons actually interviewed, and the supervisor may then get false evidence to the effect that the interview was not conducted.

With area sampling it is much easier to see if the sampling instructions were carried out accurately. The instructions themselves are much more precise and much less subject to different interpretations by different interviewers. Interviewers may be required to send in brief sketches of the blocks or open-country segments in which they interviewed. Moreover, there may be independent information about the blocks or segments (*e.g.*, approximate number of dwelling units) that can be compared with the interviewer's findings. In other words, in area sampling there is much less likelihood that the interviewers will depart from instructions, and it is easier to detect the departures that do take place. If a supervisor wishes to check up directly, she does not have to talk to the respondent as in quota sampling to find out which quota he should have been assigned to; she need merely visit the block in which the interviewing was done and see if the addresses listed on the interviews were correctly chosen. (In area sampling one would usually record the address of the respondent, although the name usually would not have to be taken.)

The central task of evaluating interviewers, however, is to evaluate the interview itself. One approach has already been discussed: the interview itself can be repeated (usually on a sampling basis) by another person. This approach is good, but it is also expensive and time-consuming, even on a sampling basis. An alternative goal is the evaluation of the interviewer's work by an evaluation of the finished product—the questionnaire. At first this task might seem impossible. In almost all respects a good interview looks almost the same as a bad one when it is recorded on paper. A clever interviewer can invent interviews and make most people believe that the resulting questionnaires refer to real interviews. It has been alleged that on simple interviews that require only checking of "yes's" and "no's," cheating is widespread and difficult to detect. Instances of low-quality but honest interviewing are likewise hard to detect.

Fortunately, however, there are methods by which both dishonest interviewing and poor interviewing can be detected for certain kinds of questions. These methods are not only quite promising theoretically but they have had limited trials that indicate their practicability. These methods may be applied either to all, or to only a sample, of the interviews sent in by each interviewer. In the following discussion, consider the questionnaires under scrutiny to be those of a single interviewer. Three types of procedures are possible, and usually all should be used.

1. The first step is to read through the questionnaires to detect failures to follow instructions. The interviewer may fail to record an answer for a certain question that should have been asked; she may not follow instructions properly in recording answers; or she may have made other more or less obvious errors in following instructions. To speed this check, one

may take a sample of the questions: one might pick out, for example, five different points on the questionnaire at which it was particularly important to follow instructions and at which it is particularly obvious when the instructions have not been followed. A check of this kind, while not complete, is useful, for it usually shows up quite clearly interviewers who will not or cannot follow instructions properly. (Sometimes, of course, it will also indicate where the instructions themselves were not satisfactory.)

One cannot, however, often detect deliberate violations of instructions by this method. In particular, interviewers who cheat are often likely to fill in the questionnaires according to the instructions, since doing otherwise might attract attention to their work. Checks of this kind indicate something about the completeness with which instructions were followed but relatively little about the skill with which the interview was conducted.

2. A second type of test is aimed directly at the cheating problem. Two methods are the reinterview and the post-card check; the first is expensive and the second is inconclusive. Both are time-consuming. An alternative approach is quick and inexpensive. Its use has been too limited to warrant confident statements about its efficacy, but the tentative results have been encouraging.

The assumption behind this approach is simple: an interviewer who cheats tends to fall automatically into certain nonrandom patterns of filling in her questionnaires. If there are statistical tests that are sensitive to nonrandom patterns and if valid responses have been shown to follow a random pattern, one can often identify the questionnaires of such interviewers.

The problem is more complicated, however, because many of the phenomena encountered in actual interviewing are likely to be met in nonrandom sequences. In area sampling, for example, an interviewer might interview four Negroes, five Italians, and three native-born whites in that order without cheating in any way. For the first block in which she was required to obtain interviews might have been in a colored neighborhood, the second in an Italian neighborhood, etc. It is necessary, therefore, to find phenomena that will be encountered in a random sequence in ordinary interviewing. Often this is not difficult to do: in effect one has the task of finding a characteristic that is distributed "effectively" at random. (See Chapter 10 for a discussion of effective randomization.) Usually on each study it is possible to find some characteristics of this kind.

An example will be taken from a study of gasoline purchases. One of the questions referred to the brand of gasoline purchased on each of the five previous purchases. For the purpose of this check, the answers were classified in two ways: the respondent either bought the same brand on the five previous purchases or he did not. Let the symbol X stand for a

motorist who was reported as having purchased the same brand on each of the last five purchases, and let *O* stand for a motorist who was reported as having purchased at least two different brands. One might obtain sequences like the following from different interviewers:

XXXXXXXXXX000000XXXXXXXX00000000
 XOXOXOXOXOXOXOXOXOXOXOXOXOXOXOXOX
 OX000X0XX000X0XX00X0XXOX0000X

The first two patterns show strong evidence of nonrandomness in the sequence of *X*'s and *O*'s. The third, by both casual observation and statistical test, shows little evidence of nonrandomness. The actual statistical tests themselves are exceedingly simple to apply in practice.¹

One caution is essential. If a nonrandom pattern is discovered, there may be three explanations: (1) the phenomenon thought to be effectively random really was not; (2) the observed pattern is accidental; (3) the interviewer really was cheating on that particular question or the whole questionnaire. The risk of being misled if the second possibility is correct may be held within certain desired limits by statistical techniques. The first assumption, however, is based on one's own intuition and may or may not be correct. If the assumption is wrong, one might erroneously conclude that many honest interviewers had been cheating. Therefore evidence of dishonesty according to this test should be regarded only as grounds for further investigation. Perhaps 5 interviewers in 75 might be deemed "suspicious" by this test. Further checking of the same kind or of the kind to be suggested below might exonerate three of these interviewers. Then a personal field check, including reinterviews, could be undertaken for the other two.

3. The third possible method possesses the advantages already mentioned for the first two and, when it can be applied, is probably more useful than both combined. Its use is limited, however, to free-response questions in which the interviewer asks a general question and records the respondent's comments as nearly verbatim as possible. If such questions are asked, one can not only form an impression of the extent to which the interviewer followed instructions but obtain, in addition, a direct indication of the quality of the interview. The procedure is somewhat more subjective than the first two, but again, limited experience suggests that it is useful.

Again one uses a sample of each interviewer's questionnaires. (If the above two tests are being applied also, the same questionnaires might be used for all three tests.) The answers to certain selected free-response questions are read for all these questionnaires. The average length of the

¹ The interested reader is referred to Paul G. Hoel, *Introduction to Mathematical Statistics*, pp. 177-184, John Wiley & Sons, Inc., New York.

responses gives a good idea as to the interviewer's ability to put the respondent at ease and obtain full answers. The extent to which the answers are really in the respondent's own words (and this is usually very easy to tell) gives a direct clue as to whether or not the instructions about recording answers were really followed. For when interviewers condense responses or paraphrase them, they usually introduce their own phrases and expressions. These are usually easy to recognize, particularly when one has been reading the work of many interviewers. Only the most skilled "cheaters," those who have literary talent, can hope to reproduce the variety of responses and modes of expressions that one obtains from a cross section of respondents on a free-response question.

If interviewers are instructed to write down their probes, one may obtain also some indication of possible bias introduced by interviewers. If, for example, the interviewer probes incorrectly, this may be detected immediately and the deficiency may be pointed out to the interviewer so that she may overcome it.

The free-response question, then, is not only an extremely useful way of obtaining information, but it is also an extremely valuable tool of administrative control in the supervision of interviewers. There is much to be said for including one or two such questions, even when they are not otherwise called for, simply as a tool of supervision.

All three of these tests permit an evaluation of an interviewer's work almost as quickly as it comes in from the field. The same tests can be used, with certain limitations, in training new interviewers. The supervisor should be able to make many helpful suggestions to interviewers on the basis of these checks alone.

It is quite easy, also, to work out grading systems to rate the performance of each interviewer on each study. The idea is simply to work out point systems in which a certain number of points is assigned for completeness and a certain number for the quality of free-response answers. In such grading systems one may also try to work out ratings for other aspects of the interviewer's work. The average number of interviews per day (after possible adjustment for the difficulty of interviewing conditions, etc.) would be one such project. It is undesirable to put much of a premium on extreme speed, because quality is likely to suffer, but it may be worth while to grade interviewers down slightly for extremely slow work. Another possible criterion is the percentage of completed interviews (non-refusals) in area sampling designs. Here again, it is clear that some adjustment must be made for differing interviewing conditions, particularly the difference between large cities and rural areas. But it is clear that a high percentage of completed interviews does tell something about the interviewer's skill.

With the grading system, either informal or formal, it is possible to identify the poorest interviewers and either retrain them or replace them. In this way, good supervision can lead to a constant improvement of the interviewing staff. That is, once it is possible to rate performance in some meaningful way, it becomes possible to improve the quality of the field staff even in the absence of the knowledge necessary to do a good job of initial selection.

There are other approaches to the evaluation of interviewing performance. One of the most interesting is the method of giving interviewers comparable interviewing tasks (by use of random selection) and comparing results among different interviewers. Experiments using this approach have been described by Mahalanobis, and it has been used in more recent investigations in public-opinion methodology. Another method, of course, is direct observation, which is often feasible at least in the stage of initial training. Recording devices have been worked out which permit the interview to be recorded, sometimes without even the knowledge of the interviewer!

Still another helpful method is to encourage criticisms and suggestions from interviewers. Sometimes a formal questionnaire may be used for this purpose. It is possible not only to obtain in this way invaluable suggestions for improvement of questionnaire design and instructions but also to learn a lot about the interviewers and, particularly, their interest in the work.

2. Relations with fieldworkers. A quick evaluation of the quality of the work done by fieldworkers must precede progress in bettering performance in the field. It is equally important, however, to be able to deal with interviewers in a way that will encourage their best performance. Many problems, of which the method of compensation is only one of the more important, are involved here.

The first issue is whether interviewers should be employees or independent contractors. In most states, the usual conditions under which interviewers work are regarded as conditions of independent contracting rather than of employment. In some states, however, different regulations governing conditions of employment make it doubtful as to whether or not interviewers working under normal conditions would be considered employees or independent contractors. If interviewers are considered independent contractors by the state, the company may still wish to give them the legal status of employees rather than independent contractors. If they are given such status, the financial obligations of the company—such as employment insurance—are greater, and it may be more difficult to get rid of inferior workers. On the other hand, there are possible advantages for morale and discipline if interviewers are actual employees. By

and large, however, the prevailing practice is to deal with interviewers as independent contractors.

A second issue, closely related to the first, is whether or not an attempt should be made to build up a small staff of permanent workers rather than a relatively large staff of part-time workers.

The second alternative is overwhelmingly used, but the first method has been tried out by several companies, such as the General Motors Corporation and Kenyon & Eckhardt, Inc. The cost of the permanent staff is likely to be high when work is intermittent, unless the interviewers can be transferred to other tasks during slack periods. Similarly, travel costs are likely to be high, since the small staff of permanent interviewers may have to travel to all parts of the country to do a study. Nonetheless, superior interviewers can probably be secured by this method. Again, a compromise might be warranted: a small staff of full-time interviewers might be supplemented by a larger staff of part-time interviewers when the need arose. The full-time interviewers could then be used not only for interviewing but for field supervision as well.

The third issue is the method of payment. The two main methods, of course, are piece and hourly rates, but various crosses between the two are frequently used in practice. Straight payment for each interview is open to serious criticism. If the interviewer is rewarded financially solely on the basis of total output, there is every incentive to sacrifice quality. Since quality is perhaps harder to define in interviewing than in many other kinds of work, special efforts must be made to preserve high standards. As has been suggested, it is probably important to have some system for discouraging extremely slow work, but little distinction should be made between an average rate of production and an extremely high rate. Attention should be focused on quality instead.

The piecework method has arisen chiefly when some assurance was needed that costs would definitely stay within narrow limits. Sometimes this assurance is imperative. For example, census enumerators are paid on a piecework basis because of the seriousness of exceeding the total appropriation for the census. In private business, and in smaller scale government projects as well, it is perhaps better to take some small risk that the field costs will exceed the estimate and to concentrate primarily on maintaining standards of quality.

Even when fieldworkers are paid on an hourly basis, however, the work may be administered in such a way that the system of payment naturally becomes a piecework system. If, for example, one ostensibly pays the interviewers by the hour but lets them know that the work should take five days, then in effect the interviewers are being paid by the interview. For the interviewers who take less than five days will probably report that the

work took five days. If the five-day requirement is ample, the interviewing will cost too much; if it is not adequate, quality and morale are likely to suffer. Needless to say, it is extremely difficult to guess in advance a time requirement that will be just right, even if a separate guess is made for each interviewer. A better system is not to tell the interviewers the length of time required, except possibly as a very wide range, and to put the major stress on quality. It should be mentioned, however, that most interviewers must be given a rough indication of the length of the job before they are willing to accept the work.

The fourth question is the amount of payment. In general, the rate of payment for marketing research interviewers is low. Probably a typical (modal) figure is slightly more than \$1 per hour. Frequently research people in survey work, marketing research or elsewhere, deplore this low rate of pay, but little experimentation has been done with higher rates of pay. Obviously it would do little immediate good simply to increase the rate of pay for all interviewers.

If, however, good interviewers were rewarded by pay increases, it might be possible not only to motivate the best workers but to attract promising people into the work if they knew that it would be possible to attain more substantial rewards than are now possible. As in any other work, the basic criterion is that the interviewers like their jobs; at the same time, however, one cannot lose sight of the fact that people are also motivated by money and that they would enjoy pleasant work more (and be willing to do better work) if they felt that monetary as well as psychic rewards were forthcoming.

We have suggested devices by which merit and quality of work can be identified and roughly measured; devices of this kind are absolutely essential. The difference between good interviewers and average interviewers is very great. If a staff consisting almost wholly of good interviewers can be obtained by use of salary increases for good workers, the interviewing cost for work of a given quantity and quality might be considerably less than for a staff of more variable quality paid a lower average wage. There is little reason why the hourly rate for the best interviewers should not be at least comparable to that for junior analysts or skilled statistical clerks.

Before leaving the question of payment, it should be pointed out that prompt payment is sometimes neglected. Everything about the survey is hurried except mailing out the checks to the interviewers. Needless to say, few practices are more irritating to the interviewers.

The final issue in supervision of interviewers might be called "human relations." Much attention has been given in modern personnel practice to treating people like people rather than like soldiers in an army, or slaves, or second-class citizens. Obviously the same rules apply to the handling

of interviewers. If the interviewer can be made to feel that the company for which she works is interested in her and considerate of her welfare, she is likely to do much better work.

Pay is an important part of this, but other things count, too. Interviewers should be warned in advance that a job is coming up, so that they can make the personal plans that are necessary. They should be given a reasonable time to complete the work. Their questions should be encouraged, and attention should be given to their comments. The company should commend them for good work, prompt completion of assignments, work done under difficult weather conditions, etc. They should occasionally see some representative of the central office, even though instructions are usually by mail. Possibly an occasional all-expense tour to the central office for "training" may be worth even more in good will than in technical training. Little touches like house organs may help. Christmas cards from the company may be helpful. Sending the interviewer a copy of the final report, where possible, may also help. The main point, perhaps, is that the specific form of recognition is unimportant so long as the interviewer feels that the company is interested and considerate. Since many of the interviewers may work for several companies on a part-time basis, interviewer good will may often be the decisive element in being able to complete work of good quality on short notice. The interviewer is likely to give priority to the company that she regards most highly.

Many experienced practitioners have said—and with probable justification—that improvement in the quality of interviewing has the greatest potential for improving the quality of marketing research.

CHAPTER 22

PROCESSING MARKETING RESEARCH DATA

22.1 Introduction

This chapter is concerned primarily with the "office work" of marketing research. Again we shall consider the problems that arise in executing surveys rather than in other types of research. The limitation in scope will not seriously affect the completeness of the discussion, since almost all the problems that arise in processing data arise in processing *survey* data.

Administering the processing operation is much simpler than administering fieldwork. The major difficulties in administering fieldwork arise because the employees who are "administered" are beyond the reach of continuing supervision and produce a product (questionnaires) whose quality can be appraised only roughly and with difficulty. Further problems are created because the employees are hired only part-time, at irregular intervals, and on short notice. These problems are absent, or present to a much slighter degree, when the task is the administration of internal employees. The fact that the task of internal administration is easier has caused it to be overlooked too often, with the consequence that well-executed fieldwork is followed by delays and confusions in the office. Careful planning is needed if the processing is to proceed smoothly and on schedule.

The processing operation is designed to convert raw data into a form more suitable for analysis and presentation. The process consists chiefly of classifications. Answers to individual questions must be edited, then classified (coded); and then answers falling into each of the classes must be counted (tabulated). Next the relationships between these primary classifications must be explored (cross-classified). In addition to the logical process of classification, there are the mechanical tasks of preparing final reports. This involves tedious but essential typing, chart preparation, etc.

22.2 Coding

Any marketing research study contains more information than is actually used. For example, in the interest of clarifying the over-all picture, the analyst is usually willing to sacrifice some of the detail of the individ-

ual responses; he will group individual answers for each question into a relatively small number of categories. Within each of these categories there will be small differences in wording and meaning, but for the purpose of the study the answers are considered to be alike. Thus the several hundred or several thousand individual answers that were obtained in the field might be reduced to, say, six categories. A great deal of the original detail is lost by the classification process; but in exchange, if the six summary categories are intelligently chosen, the basic information which the question was designed to obtain will be indicated. Without this classification, the detail would be overwhelming.

In coding, the coder is guided both by logical principles of classification and by practical requirements. Before going into the practical requirements, it will be helpful to examine some of the guiding logical principles.

1. Each individual response should fit into *at least one* category. In other words, the scheme of classification should be exhaustive. This principle is obvious enough, yet it is not always observed. For example, often in surveys no answer is obtained for some questions. The interviewer may have forgotten to ask the question or she may have forgotten to write down the answer; sometimes, too, the respondent may have refused to answer.

The correct way to handle this type of response is to set up a category called "no answer" or, better, "not ascertained." Sometimes, however, this is not done, and it may then be impossible to tell from the final report of a survey how many questions were unanswered. The "not ascertained" category usually contains a high proportion of difficult-to-interview and uncommunicative people who may differ significantly with respect to the characteristic being studied from the respondents who did answer the question.

2. Whenever possible, each answer should fit into *only one* category. That is, the scheme of classification should be mutually exclusive. Sometimes this rule cannot be followed, as in the case of the so-called "multiple answer," in which the respondent may correctly give more than one response to a single question. For example, in questions on automobile ownership there will usually be a few multiple answers because a small per cent of motorists own two or more cars.

Many apparent instances of multiple answers, however, especially in questions that require "reasons why" or expressions of attitudes, arise because of poor questions. In these cases, treating the responses as multiple answers obscures results. For if the question itself is ambiguous (*e.g.*, "Do you think you are better off than before the war?"), respondents may interpret the question in many ways, and some respondents may interpret it in more than one way. In Chapter 14 we called this "multidimension-

ality" and suggested ways of avoiding it. To the extent, however, that there is ambiguity in the phrasing of the question, the remaining ambiguity should be diminished by classifying responses as if, in fact, two (or more) distinct questions had been asked. When this is done, the latent ambiguity of the question is brought out into the open, and the problem of multiple answers is eliminated.

3. The scheme of classification is determined by the question being investigated. If the researcher wished to determine attitudes toward a particular advertising theme, the categories of responses would be relatively clear: he would not classify answers by their length (although this might be useful in judging the work of interviewers) but according to the types of comments made about the advertising theme and possibly also according to intensity of feeling.

4. The detail to which the classification is carried is a compromise between the obvious need for summarization and the need for knowing the differences in individual answers. If there are too many categories with a small number of responses in each, it is hard not to lose sight of the salient features of the findings. If there is a very small number of categories with a large number of responses in each, almost all the detailed information and possibly the broader picture as well is lost. The first situation is usually better than the second because small categories can often (though not always) be consolidated satisfactorily.

Now we turn to some of the "practical requirements."

1. Precoding. Before the questionnaire is sent to the field, it is often desirable to anticipate many of the problems of coding that will arise. Where responses are likely to be standardized or stereotyped, the response categories can be set up in advance in such a way that the interviewer has only to check or circle the appropriate alternative. This might be done, for example, on questions concerning income, age, education, or family size. The possible responses are included in the questionnaire itself. By doing this it is possible to save the interviewer some work in the field by reducing the amount of writing she has to do (although this saving is minor for questions like the ones just mentioned). Much time is saved in coding the questionnaires, since the coding on questions like these is in effect done by the interviewer.

For many types of information, however, it is difficult or impossible to establish in advance of the fieldwork appropriate and exhaustive categories for coding. The difficulty arises primarily because of the complexity and number of possible responses and the impossibility of foreseeing with sufficient accuracy the nature and variety of the responses. An attempt to overcome this problem by establishing categories of a very general nature results in increasing both the reliability of the coding process and the

ambiguity of the summarized information. For types of information for which precoding is impossible or subject to these difficulties, it is better to forego precoding and accept the more time-consuming and expensive alternative. The alternative is to have fieldworkers record as nearly as possible verbatim responses to nondirective questions, leaving to specialized coders in the office the task of classification. This alternative has the great advantages of placing coding under supervision, which can eliminate to a large extent variations in coding decisions between different workers; leaving the coding to specialists whose proficiency can be expected to be higher than that of most interviewers; and delaying the establishment of categories until the data have been collected.

2. Editing. Editing is closely linked with coding. When questionnaires are first received from the field, they must be "checked in" by the supervisor of the field staff in order to be sure that the field assignments have been properly fulfilled. At this time any checks of the quality of fieldwork are performed (these usually must be done immediately if it is going to be possible to take any action upon the findings). The next step is usually editing.

Editing might be defined as the correction of obvious errors and omissions in the questionnaires received from the field. Many errors, of course, can never be detected once the interview has been finished, and for this reason interviewers should be trained to check questionnaires before leaving the respondent. Even with the best of interviewers, however, there are usually more or less obvious errors such as misspellings, indistinct handwriting, failures to fill in certain questions whose answers are obvious from the answers to other questions, etc.

For example, in a readership study the interviewer might fail to check the fact that the respondent had seen the advertisement. Subsequent questions on detailed parts of the advertisement might make it obvious that the respondent had seen it. The editor could then fill in the omitted questions on seeing the advertisement (using a pencil of different color from that used by the interviewer, in order to fix the responsibility for the changes). Again, a respondent might anticipate the answer to question 10 when she was asked question 3. The interviewer, properly, might not ask question 10 when she came to it. But she might, improperly, forget to mention the fact that question 10 had already been answered in the response to question 3. The editor, who would read the entire questionnaire as part of the task of editing, should notice this omission and fill it in.

Inconsistencies in responses are much harder to handle. If these inconsistencies simply reflect inconsistency on the part of the respondent and it is desired to find out the extent to which respondents are inconsistent, nothing should be done, either by the interviewer or by the editor. At

other times, however, inconsistencies should not be permitted. The primary responsibility, then, is with the interviewer, who should diplomatically call inconsistencies to the attention of the respondent and ask the respondent for a reconciliation. If the interviewer fails to do this, the responsibility for making the best of the questionnaire falls to the editor. She must do her best to find some reasonable way of resolving the inconsistency: that is, she must find some interpretation that seems plausible and make the necessary correction. Obviously, this is often impossible to do. If the respondent says at one point in the questionnaire that she buys Pillsbury flour regularly and at another point that she has used Gold Medal flour exclusively for the past two years, there is usually nothing the editor can do to resolve the confusion. Occasionally, however, there might be some other internal evidence in the questionnaire to suggest that the reference to Pillsbury was due to a slip of the pencil by the interviewer and that the respondent really had used nothing but Gold Medal. The editor must really be a kind of detective, and she should excel, as did Sherlock Holmes, in observation, memory for details, and deductive reasoning.

If the questions are chiefly of the fixed-alternative, precoded variety, editing is much simpler though still necessary. Sometimes editing is combined with coding—that is, the answers are classified at the same time that the questionnaires are edited. There is some difference between the skills involved in editing and coding, and usually there seems to be a gain in performing the tasks separately.

3. Final coding. The next operation is the assignment of answers to categories. Two steps are involved: the clear definition of categories and the decision as to which of the established categories is appropriate for each particular response.

The defining of the categories is called "setting up the code." The beginning of this process should take place very early in the planning of a study, when a tentative outline of the code may be sketched. Further work is usually done by reading through the early questionnaires coming in from the field in order to see the kinds of patterns the answers are falling into. Then, on the basis of this perusal, a code is established, which can subsequently be modified to the extent that later questionnaires indicate as being necessary. The process can be speeded up considerably if categories that are expected are indicated in advance. The information gained on the pretest and general information about the problem as gained from the exploratory stages usually make it possible to write a preliminary code that will have to be modified only to a relatively small degree.

There is still another advantage to having categories tentatively established in advance. When a code is established solely on the basis of the early questionnaires, it is possible to get so immersed in the details of the

individual responses that the logical principles of classification (see above) that are so essential to the proper handling of raw data are overlooked.

Setting up the code in advance permits putting in categories that may later turn out to include no responses at all. Such information can be very important. Without an advance code, however, it could be overlooked.

Finally, if the main categories are tentatively established in advance, the modifications and additions that are necessary as a result of the actual returns serve to remind the researcher forcibly of the ways in which his advance guesses were wrong. This is highly desirable. If his attention were not drawn explicitly to the ways in which the empirical findings deviated from his expectations, he would not be so certain to notice these deviations. Even for research people, the task of changing opinions and preconceived notions in the light of evidence is not easy. The task is facilitated if the differences between the factual picture as revealed by the study itself and one's first picture formed on the basis of preliminary exploration are explicitly indicated.

No matter how adequate the advance preparations, the researcher is confronted with the difficult task of getting the code into final form on the basis of the information contained in the first questionnaires coming in from the field. Usually the formulation of the final code can begin as soon as, say, 50 or 100 questionnaires have come in from the field. The usual procedure is to read through the questionnaires and to see how well the responses for each question fit into the previously established categories. Almost invariably some unanticipated responses will be found, and it may therefore be necessary to add categories. Moreover, it is possible that some of the categories anticipated in advance were not defined with adequate precision.

It is usually desirable to read as many questionnaires as time permits before "freezing" the code, that is, before putting the code into the final form that will be used for all the remaining questionnaires. Actually, until the study is finished the code is never (or hardly ever) put into a form that cannot be changed slightly as the need arises. But frequent changing is undesirable, once the final coding gets under way. Sometimes, for example, earlier questionnaires would have to be recoded if the changes in the code are major.

Establishing the code should be under the supervision of the research analyst in charge of the study unless all the questions involve fixed-alternative responses. No matter how ingenious and intelligent the coder, he cannot be expected to know as much of the background of the study and the uses to which the data are to be put as the research analyst. Ideally, the research analyst should spend two or three days actually working with

the coders in order to assure that the code is in the form he wants it. (Often, under pressure of other activities, the research analyst may be tempted to regard the code as an unimportant clerical routine that is not worth much of his attention.)

One important practical problem that comes up at this stage is whether the categories, as defined, really mean the same thing to different coders. It is desirable to test this as early as possible by having several coders code each of the first questionnaires to come in from the field. If coders are almost unanimous in their classification of responses, the categories can be considered to be unambiguous. If, on the other hand, there is considerable disagreement among different coders, the source of disagreement should be determined and removed. It is helpful to have "round-table" sessions between the analyst and the coders after each day's work in this early critical period. All kinds of questions of interpretation come up in the early stages of coding, and it is important to agree on a single reasonable interpretation rather than have each coder make his decision in isolation.

The preparation of the code would be considerably easier if the first questionnaires from the field represented a random sample of the entire group of questionnaires eventually to be coded. Then the major categories of response could be more definitively set up, since random sampling would make it reasonably sure that all major categories of answers would be illustrated in the first 100 or 200 questionnaires. Unfortunately, the first questionnaires are usually localized and represent the work of a small number of interviewers. This seems inevitable, and as a result, important changes in the code sometimes have to be made even after a large fraction of the questionnaires have been coded.

Now we shall consider in a little more detail some of the more technical phases of code construction.

1. Amount of detail. If responses are classified into a very large number of categories, the details can cause the investigator to lose sight of the major findings. Some detail must be removed by grouping together categories of responses that are only slightly different. This process of grouping must end before so much of the detail is lost that the broad categories remaining are of little use. To state the mere need for a compromise between these two extremes is not enough; we must now direct attention toward some practical aids in arriving at an intelligent compromise.

In general, too little detail is much worse than too much. If there are too few categories, the only remedy is to go back to the original questionnaires and recode them completely. Not only is recoding costly and time-consuming; in addition there are many occasions in which it is impossible. Often the need for greater detail may become apparent after the original

study has been put in final form. It may even be that a new problem arises that requires information from the earlier study and greater detail of response is needed. The questionnaires by this time may be difficult to find or may have been discarded. If the response categories are too numerous, on the other hand, the remedy is usually simpler. All that is necessary is to consolidate similar categories by adding together the number of responses falling in the categories to be combined. This can be done with an adding machine in a few minutes and requires no reference to the original questionnaires.

In effecting the compromise between detail and generality, there is the temptation to lean toward excessive detail, since an error in this direction can be so easily remedied. It is important, therefore, to define categories in such a way that it is not hard to consolidate them. Consider an example.

The research analyst is interested in obtaining housewives' opinions of a particular brand of scouring powder. Responses are classified into 25 categories. Later it is decided to obtain a rough scaling of over-all attitudes into five categories: (1) like strongly; (2) like mildly; (3) indifferent; (4) dislike mildly; (5) dislike strongly. The original free-response answers might have permitted such a coding, but the 25 established categories might not fit well into such a scheme of classification.

There are two ways of getting around problems like these. (Again the need for advance planning will be apparent.) One method is to code the same question in two ways: one which gives the scaling of over-all attitudes and the second which gives the 25 detailed response categories. Both types of information might be important, and yet it might be impossible to work them into the same code. The other approach would be to set up "hierarchies" of classifications (*e.g.*, five main categories and a large number of finer subdivisions within each one).

In trying to attain a reasonable amount of detail in response categories, it is important to remember not only that excessive detail is exceedingly difficult to interpret but also that the cost of coding (and tabulation as well) increases and the inaccuracy of coding usually becomes greater as the number of categories increases. These practical factors operate as a strong brake on a desire for detailed categories. The larger the number of categories, the finer the nuances of difference between them and the more difficult the problem of classification. If the coders have a large number of response categories, they will have many more borderline cases to decide and more time will be spent in indecision. Moreover, the decisions will tend to become more and more subjective as the objective criteria of difference becomes less and less pronounced. The reliability of the coding—that is, the consistency with which any given response would be identically

classified by different coders—becomes less and less. Therefore, even if the proliferation of categories were not unnecessary and confusing in the stage of interpretation, there would be limitations imposed by expense and by the fact that coding tends to lose a great deal of reliability as the difference between response categories becomes less pronounced.

2. *Multidimensionality.* Even with the greatest care in planning and designing questionnaires, some questions may be interpreted differently by different respondents. In other words, some questions may prove to be ambiguous. It is imperative that the coding take account of this when it occurs. The answers to interpretation A, for example, should not be merged with the responses to interpretation B.

3. *Format of the code.* The code finally emerges as a written document for the guidance of coders in the actual work of coding. For each question, the code lists the categories into which responses are to be classified. The way in which the code is arranged has a great deal to do with the ease of coding. If it is difficult to find the appropriate category because of a poor format, coding is slowed up and accuracy may be diminished. Response categories should be grouped in such a way that it is easy to locate each individual category.

There are several ways in which this can be done. Certain recurring categories, such as "don't know," "no opinion," "no answer," and "all other" (the last being synonymous with miscellaneous), should always be placed in the same position for each question, usually at the bottom of the list of response categories.

Secondly, the categories can be grouped in a more or less logical sequence so that the coder can quickly learn where to look for the category that is likely to fit the specific response she is coding. A logical sequence also helps the coders to memorize the code, wholly or partly, and thereby to speed up the coding.

Finally, little tricks of arrangement of the code are useful. For example, a great deal of time is ordinarily required to turn pages of the code. If large-sized pages are used, this time can be reduced substantially. The ordinary considerations of preparation of printed forms—legibility, compactness, pleasing appearance—quite obviously apply to the code.

4. *The "all other" classification.* One of the most perplexing problems in coding is the handling of miscellaneous responses, or, more specifically, deciding when to throw an answer into "all other" and when to create a separate category for it. It is hard to generalize as to the danger point, but usually if more than 10 per cent of the responses are being classified as "all other," it is worth while to check to see if some important answer is being buried. If the coders are warned about this point, they will usually learn to watch closely the kinds of responses that are going into "all other"

and to call attention to answers that are sufficiently similar to warrant separate classification.

Ideally, the "all other" category should consist of a wide variety of answers that defy useful grouping into larger categories. Even when this has been achieved, however, it is well not to forget these miscellaneous answers completely. There will be no individual type of response of great numerical importance, but there may be some answers of great general importance. For example, if respondents were asked for suggestions about the improvement of a product, one respondent might make a useful suggestion that would be buried in the "all other" category. Marketing research is, of course, primarily concerned with patterns of similarity in behavior and ideas. Nonetheless, the unique patterns of behavior and ideas may be exceedingly useful. For this reason, the "all other" responses should be canvassed carefully for answers that might otherwise be lost. Coders should be alerted to look for such answers and report them.

5. Question-by-question codes vs. over-all codes. Ordinarily questionnaires are coded question by question. That is, there is a separate list of response classifications for each of the questions appearing on the questionnaire, and each question is considered separately. If all questionnaires could be constructed so that there was no overlap from question to question, and if all respondents could be steered into exactly the patterns of thought dictated by the sequence of questions, then question-by-question coding would obviously be the only method to use.

Neither of these ideals is attainable completely. Respondents will repeat earlier responses and anticipate future questions. The interviewer is, of course, responsible for indicating that a given question has already been answered earlier, but this responsibility applies only when the interviewer did not deem it necessary to ask the already answered question. Otherwise, when the respondent says the same thing in several places on the questionnaire, or says something relevant to question 7 in her answer to question 10, the interviewer's job is only to record it. A first-class job of editing can catch most of the irrelevant answers and move them to their correct place, but duplications are harder to compensate for by editing. An alternative or supplementary approach is to use an over-all code.

With an over-all code, the coder has a list of response categories before her as she reads the entire questionnaire. The over-all code may of course be used in conjunction with a question-by-question code and often is.

A cross between the two methods would be a code that applied to groups of questions rather than to an individual question, at the one extreme, or to the entire questionnaire, at the other. This combination has sometimes been given the name "block code."

Over-all codes are most helpful, and indeed almost essential, if depth

interviews are used. In the depth interview, it will be remembered, the questioning does not follow a set procedure, the job of the interviewer being that of guiding the respondent into the relevant areas of discourse. Without a good over-all code to organize the answers obtained, and without a skillful coder to do the coding, it is nearly impossible to summarize properly the information that has been obtained. All one can do in the absence of an over-all code is to report qualitative impressions, and these are notoriously subject to conscious or unconscious bias and great inaccuracies in memory.¹

6. *Relation to machine tabulation.* Finally, the code should be planned with the requirements of tabulation in mind, particularly when machine tabulation is to be used. These requirements are discussed in section 22.3.

7. *Other remarks.* So far the discussion has centered on the problems in setting up the code rather than the problems arising in the actual process of coding. A good coder should have what is popularly thought of as a "legal mind." She must be able to see the logic by which the response categories have been set up, and she must be able to weigh borderline cases much as a judge does when considering the essential similarities and differences between an event and other events with reference to dicta or existing statutes. She must have an excellent memory in order to retain all the precedents that inevitably must be established during the process of coding in order to create a uniform procedure in borderline cases.² These precedents should, if possible, be reduced to written form in order that the findings can be properly interpreted, but during the actual process of coding they are seldom written. The more familiar the coder becomes with the code, the quicker the actual process of coding, since references to the written document are made more easily and less frequently. When over-all codes are used, a good memory is indispensable. The coder must also be able to resist the tendency toward carelessness that comes as the coding becomes routine and monotonous. Finally, she must be an expert at deciphering obscure handwriting.

It is often helpful to set up sampling "spot checks" of coding as work progresses. One coder will then recode, preferably independently, a sample of the work of another. This serves not only to catch errors and to indicate

¹ It has been suggested that content analysis can provide a quantitative, reliable, objective means for summarizing and describing the responses to nondirective questions and depth interviews (Bernard Berelson and Paul Lazarsfeld, *The Analysis of Communication Content*, preliminary draft of unpublished manuscript, 1948). Unfortunately, content analysis has not been used sufficiently in coding operations to permit an appraisal of its feasibility.

² It is possible that all borderline cases could be referred to a single coder to achieve such uniformity.

the need for more careful work when errors creep in, but it serves to indicate cases in which different coders have placed a different interpretation on a particular borderline response. An important by-product of spot checking is a salutary effect on conscientiousness and diligence. If a coder knows that his work is to be checked, he is likely to be much more careful than if he knows errors will never be discovered.

In coding, as in all clerical routines, one may apply the newly developed statistical techniques of quality control and sampling inspection in order to improve the quality of work. These methods can be thought of as scientific extensions of simple spot checks. Many good texts on these subjects are readily available.

A major organizational question that arises in coding is whether, assuming that a question-by-question code is used, coders should specialize in the coding of certain individual questions on the questionnaire or each coder should code the entire questionnaire. The chief pros and cons are rather simple. If coders specialize on individual questions they can probably develop greater speed, since they can memorize the code more rapidly. Further, they will probably treat borderline cases more uniformly, since relatively few coders (perhaps one) handle each question. Moreover, the most skilled coders can be used for the most difficult questions. On the other hand, if the coder handles the entire questionnaire she can get a good picture of each individual respondent and pick up the cases in which question 7 was answered in question 3. This is, of course, a part of the editing function, but a good coder can serve as a check on the editor.

Perhaps the most persuasive of all the arguments for coding the entire questionnaire is that such a process makes it relatively easy to tell which questionnaires have been coded and which have not. This may sound unimportant; but as a practical matter, coding inevitably involves numerous piles of questionnaires, and the mere task of funneling these questionnaires to different specialist coders is formidable and complicated. Closely related to this point is the fact that the total amount of paper handling is much reduced if each coder handles entire questionnaires.

22.3 Tabulation

The experienced marketing researcher will notice one curious omission in the discussion of coding: we have said nothing about the interrelationships between coding and tabulating. We have preferred to postpone this issue until now, since we wanted to focus attention on the fact that coding is essentially a problem in classification of responses and that it exists as such regardless of the form of tabulation that is finally used. Actually, there are two chief methods of tabulation now in current use, and the coding

procedure looks, on the surface, quite different depending on which of these methods of tabulation is used. The two methods of tabulation are hand tabulation and machine tabulation.

Hand tabulation usually involves sorting questionnaires into piles corresponding to the main categories of response for each question and then counting the questionnaires in each pile. In effect, the function of sorting into piles corresponds properly to coding rather than tabulation, even though no formal code is prepared. In practice, the two functions are usually merged. It is not necessary that this be so. There is no reason why questionnaires could not be coded in the ordinary way—marking code numbers corresponding with response categories on the questionnaires themselves—and then sorted into piles and counted.

Machine tabulation clearly brings out coding as a separate step. In effect the information in each questionnaire, in summarized and coded form, must be translated into a card which can then be sorted (and counted) mechanically according to the categories of response represented in the answers on the original questionnaire. One of the most popular methods of doing this is the Hollerith or IBM (International Business Machines) method. The card used looks like Fig. 4.

Suppose, for example, that the response category “don’t know” on question 11 of a questionnaire was given the code number 63-0. A coder encountering a “don’t know” on this question would mark 63-0 either on the questionnaire itself or on a separate form used for this purpose. A rectangular hole would be punched in the card in the 63d column at the place where the number 0 appears. (See illustration of card opposite.) Then an IBM machine called a “sorter-counter” would sort out very rapidly all cards punched with the number 63-0 (or any other number) and count the number so punched. Since about 400 cards per minute can be sorted in this fashion (an average of considerably fewer cards per minute, counting the time that the machine is idle), the advantage in speed is quite obvious.

Hand tabulation and machine tabulation each offer certain combinations of advantages and disadvantages. If the number of questionnaires is relatively small (for rough guidance the number 800 is sometimes used), or if the cross-classifications of answers to different questions are few and simple (or if both conditions hold), hand tabulation is likely to be superior. With a larger number of questions, or numerous and involved cross-classifications, or both, machine tabulation usually becomes relatively more efficient.

Machine tabulation has some advantages other than speed. For one thing, the necessity of tabulating by machine forces people explicitly to go through the process of coding. This means that much more careful attention is likely to be given to the vital function of classifying individual responses. Surprisingly enough, very frequently marketing research stud-

ies that have been tabulated by machine are much easier to analyze because of the necessity of planning the coding carefully in advance.

Another advantage of the machine method—one that is really inherent in the method itself—is the fact that the information can be much more economically stored. Small piles of IBM cards are equivalent to mountains of questionnaires. If special tabulations are desired in the future, they may be easily obtained without the need for storing bulky questionnaires, provided that the original coding was adequate for the future need.

With this brief comment on the comparative merits of hand and machine tabulation, we now examine the two methods in a little more detail.

1. Hand tabulation. The reason for the superiority of hand tabulation under the circumstances indicated above is relatively simple. Machine

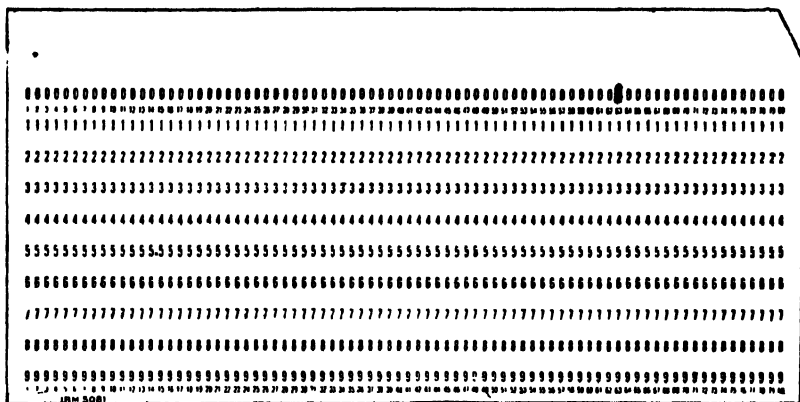


FIG. 4.

methods of tabulation sort and count the different responses to each question very rapidly, but before machine tabulation can be used, the information contained in the original questionnaires must be transferred or punched into IBM cards. Hand tabulation, of course, dispenses completely with this step, and in spite of the fact that the actual sorting and counting of questionnaires is slower by hand than by machine, the over-all time may be less for hand tabulation if the total amount of sorting and tabulation is relatively small.

There are two different ways in which hand tabulation can proceed. The first is to do what we have suggested above—to sort the questionnaires into piles corresponding to the different categories of response on each individual question and then to count the number of questionnaires in each pile. The second approach is to examine each questionnaire, check the appropriate response on a tally sheet, continue until all the questionnaires have been examined, and then count the checks for each category. With

this approach it is possible for a single person to examine questionnaires and check appropriate responses; or, alternatively, one person may examine the questionnaire and call off the answers while another person records.

The first method is probably less subject to error—or rather, it is easier to verify the results. For example, one pile of questionnaires will correspond to the “don’t know” answers on the question being tabulated. The counting of these questionnaires goes very rapidly and may easily be done twice as a check on the accuracy of the count. When discrepancies are found between the first and second count, they are easily localized. For example, of all the individual response categories for a given question, the first and second count may be identical except for one category. Then the questionnaires in the pile corresponding to that category may be recounted as a check; nothing need be done with the remaining questionnaires.

When the second method is used, the only way to check is to examine the entire set of questionnaires and make a new check list. If there is any disagreement between the first result and the second result, the only recourse is to repeat the entire process a third time.

The “examine and check” or “hen scratching” method is more subject to error than the “sort and count” method in another respect. The manual operation—checking—involved in the “examine and check” method seems to introduce a slightly greater chance for error than the corresponding mechanical operation of sorting. For example, interruptions may cause the checker to make errors in the number of checks written down.

As ordinarily practiced, there seems little doubt that the “sort and count” method of hand tabulation is superior. It is less subject to error and, for that reason, considerably faster if the same standards of accuracy are maintained for both methods. The “examine and check” method, however, has one possibility that the “sort and count” method does not have. The “sort and count” method is limited to a single question at a time, and if multiple answers are possible on an individual question, it does not work well even on single questions. The “examine and check” method potentially at least makes it possible to tabulate answers to several questions simultaneously. Mechanical methods are also available to speed both the examination and the checking. One interesting development is a conveyor belt that saves the time-consuming task of picking up and placing down each questionnaire to be checked. Another mechanical possibility is an automatic counting device that shortens the operation of checking off each response. People who have specialized in developments of this kind report that surprising speed is attainable.

2. Machine tabulation. Machine tabulation is at the same time the pride and the terror of modern survey work. On the one hand, the tabulation of large studies is much faster with machine methods, and the most

elaborate cross-tabulations (or cross-classifications) become possible. On the other hand, the very ease of making cross-tabulations often encourages needless proliferation of these analyses with resulting confusion for all who must use the results.¹

Used with restraint and foresight, however, machine tabulation is an extremely efficient way to tabulate survey data. If advance planning is inadequate or wholly lacking, machine tabulation will not only turn out a lot of indigestible information, but the simple mechanics of tabulation may become a major stumbling block to the completion of the survey. Once again, the value of advance planning becomes clear. The additional tabulations that become necessary after an examination of the preliminary tabulations should not be numerous if the study was planned properly in the first place.

Mechanically there are two main steps to machine tabulation, transferring information from questionnaires to IBM cards ("punching") and tabulation proper ("running"). In order to plan machine tabulation properly and to work smoothly with the specialists who do this work, it is necessary to know something of the problems involved. We shall consider only the chief problems and shall omit the minor technical issues that invariably arise.

1. *Punching.* Punching is the process of transferring coded information from questionnaires to IBM cards. Considered by itself, punching is a mechanical procedure somewhat analogous to typewriting. In order to facilitate the mechanics of punching and to expedite the work of tabulation later on, it is necessary to master some of the mysteries of the IBM card.

At the end of the coding process, the individual responses on each questionnaire have been assigned to summary categories. Suppose that a particular question has the following summary categories: (1) prefer product A to B; (2) prefer B to A; (3) like A and B equally well; (4) no opinion; (5) not ascertained. Each respondent's answer to this question can be classified into one and only one of these categories. This information can be transferred to an IBM card (for picture of an IBM card, see Fig. 4). There are 80 "columns" in a card (see the numbers across the top of the card), and within each of these 80 columns there are 12 "digits."² Thus there are 960 (*i.e.*, 80×12) possible places where an IBM card can receive a punch. Each of these possible punches is designated first by the column

¹ McNemar comments on this point: "If he [the researcher] is fortunate enough to have punched card equipment, it becomes the misfortune of his professional contemporaries to find the literature filled with the results of cross tabulations which are so lacking in rationale as to be nonsensical." Quinn McNemar, "Opinion-Attitude Methodology," *Psychological Bulletin*, Vol. 43, No. 4, p. 293, July, 1946.

² In addition to the ten numbered spaces, 0, . . . , 9, there is room at the top of each column for two additional punches, usually designated "x" and "y."

number (from 1 to 80) and second by the digit within the column (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, *x*, and *y*). For example, 57-4 corresponds with the 4th digit in the 57th column.¹

In the above example, the following digits could be assigned to the answer categories listed above.

Prefer A to B	57-1	No opinion	.	.	57-9
Prefer B to A	57-2	Not ascertained	...		57-0
Like A and B equally well	57-3				

The first thing to notice is that all the response categories to the particular question are assigned to the same column, column 57. This is extremely important; for in the subsequent tabulation using a sorter-counter, it will be possible to count and sort simultaneously on only one column at a time. That is, it will be possible to run all the cards through a sorter-counter once and sort all the cards into one of the five categories above and at the same time count the number of cards falling in each category.

Another point worth mentioning about the above procedure is that the numbers are not consecutive. In particular, we see that "no opinion" is classified as 57-9 and "not ascertained" is classified 57-0 because it makes coding more convenient. Recurring response categories such as "no opinion," "don't know," "all other" and "not ascertained" should be given the same digit throughout the card. That is, "no opinion" on other questions might be 34-9, 71-9, 17-9, etc. This helps the coder a great deal in memorizing the code numbers corresponding with each response category, and it simplifies all subsequent processing. The digits employed for these standard answers are usually 9, 0, *x*, and *y*.

Another important point is that column 57 should not be cluttered up with answers to another question, even though some space would be saved in this way. Digits 4, 5, 6, 7, 8, *x*, and *y* are not used at all, but by not doing so, a great deal of mechanical and administrative trouble is avoided.

Finally, it is necessary to consider what would have happened if there had been more than 12 response categories to the question under study. The only thing to do, of course, would be to overlap into column 58 and use up some of the digits in this column. There are two ways in which this could be done. Each of the twelve digits in each of the two columns could correspond to one response category, yielding altogether 24 response categories. Or it is possible to consider the digits in combination with one another. In this way it would be possible to obtain a total of 144 response categories from the two columns. For example if digit 1 were punched in

¹ Variations of this symbolism are encountered in practice, but we shall not mention them here.

column 57, there would be 12 distinct categories corresponding with each of the 12 digits in column 58 that might be punched. A similar statement applies to the remaining 11 digits in column 57, so that in total there would be 144 possible combinations of punches in the two columns. If there are not more than 24 response categories, it is simpler to punch the two columns individually, since the subsequent tabulation will then be easier. If there are more than 24 response categories, however, it is often better to use combination punches rather than to spill over into a third or fourth column. Two columns can accommodate 144 response categories, while if individual punches were used, 12 columns would be necessary.

There is one remaining problem, the so-called "multiple punch." So far, the answer to each question has corresponded either to a single code number in a column (57-3) or to two code numbers in two columns (like 57-1, 58-2). In either case, there would be only one punch per column. Suppose, however, that the question had been one that often resulted in more than one answer from each respondent. We have mentioned multiple answers and pointed out that multiple answers often indicate an improper phrasing of the question. Multiple answers of this kind, which reflect different interpretations of a question, should be handled in coding as answers to separate questions or, better, should be forestalled in advance by asking questions that are not subject to different interpretations (the requirement of unambiguity, which is spelled out in great detail in Part IV). But there will always be some questions that do permit legitimate multiple answers, and these present difficult problems in tabulation.

One simple solution is to put as many punches in the column as the respondent gives answers. Suppose, for example, that a particular question had been assigned 8 response categories: 34-1, -2, -3, -4, -5, -6, -9, -0. A respondent might give answers corresponding to numbers 34-2 and 34-6. His card would then receive two separate punches in column 34, one for each of these digits. This solution is often used in marketing research, but it has severe disadvantages. The people who ordinarily do the mechanical work of punching and tabulating are confused and slowed down by multiple punching. The reason is simple. Most work done on IBM machines is accounting work, and multiple punching is hardly ever used for this work. As a result, most IBM operators have not had experience in working with multiple punches. There are only a few tabulation firms that specialize in marketing research tabulations, and these are probably the only ones with experience in multiple punching. Even for IBM operators experienced in dealing with multiple punching, there is little doubt that punching and tabulation are slowed down by multiple punching. Thus multiple punching should be avoided whenever possible. If there is no way to avoid

multiple punching, it is necessary either to be sure that the IBM operators who will process the questionnaires are already trained for this kind of work or to be resigned to a long period of errors and slow work.

Multiple punching becomes almost unavoidable when there is a large number of questions with multiple answers. If multiple punching is not used, it is necessary to assign several columns of the IBM card to each of these questions. An example would be a question regarding the brands of flour with which the respondent is familiar. There might be, say, five major brands of flour, and the question might be coded as follows:

Brand A	32-1	Brand E	32-5
Brand B	32-2	All other	32-9
Brand C	32-3	None	32-0
Brand D	32-4	Not ascertained	32-y

Suppose a respondent was familiar with brands B and E. Digits 32-2 and 32-5 would be punched if multiple punching was being used; that is, there would be two punches in column 32. If, on the other hand, multiple punching was to be avoided, it would be necessary to allow at least 2 columns and probably more in order to accommodate responses to this question. For example, columns 32 to 35 might be devoted to the answers to this particular question. In each column the digit designations would be the same. Brand A, for example, would be 32-1, 33-1, 34-1, etc. The respondent's first answer would be punched in column 32, the second in 33, the third in 34, and the fourth in 35. If a respondent gave more than four answers (say she had used all five brands listed), her answers could not be accommodated without devoting still another column to this single question.

Hence the amount of space of the IBM card used in answering this single question would be several times as great with single punching as with multiple punching. Often all the information in each questionnaire could be punched into a single IBM card if multiple punching was used but not if single punching was used.¹ The use of more than one IBM card per questionnaire itself raises problems in the actual "running" (sorting, counting) of cards when cross-tabulations are to be made. (The extra cost of the extra IBM cards would ordinarily be negligible.)

It is worth reemphasizing that multiple answers can be reduced in frequency by careful planning of questions to remove ambiguities. Even in cases like the one illustrated above, one may be able to reduce the magnitude of the problem by listing only the two brands with which the respondent has had the most experience. If only a few respondents have had

¹ While it is desirable to compress information into a single card, this objective is only one of many, and some of the others may conflict with it.

experience with more than two brands, for most purposes very little information would be lost by not listing more than two brands. Moreover, by using the method of single punching it would be possible to distinguish between the brand with which the respondent was most familiar and the one with which she was next most familiar. For example, column 32 might be devoted to the brand of greatest familiarity and column 33 to the second most familiar brand.

Whether or not multiple punching is used, questionnaires sometimes require more than one IBM card per questionnaire. On very elaborate studies, as many as 5 or even 10 IBM cards may be necessary to accommodate the information contained in each questionnaire. The only real problem occasioned by this procedure occurs when it is desirable to obtain a cross-tabulation between two or more questions and the answers to each of these questions are punched into different cards. If it is possible to anticipate in advance all the cross-tabulations that will be required, it is often possible to arrange the punching in such a way that this will not happen.

Often many cross-tabulations will involve the same question. Often, for example, researchers wish to tabulate answers to many questions by the income of the respondent, or age, or city size, or region of the country. For this reason it is usually desirable to repeat the information on income, age, city size, etc., on each of the cards. Thus the columns from 5 to 20 on each card might be reserved for this so-called "demographic" or "classification" data. Usually the first four columns are reserved for the questionnaire number. Questionnaires are usually numbered as they come in from the field, and these numbers are punched into the card. For example, questionnaire 1390 would be punched as follows in the first four columns:

1-1; 2-3; 3-9; 4-0

This procedure is indispensable if anyone is ever to check back from an IBM card to the questionnaire to which it corresponds, and it is often necessary to do this to check up on apparent errors in punching and for other reasons as well.

Whenever the same punches are to be placed on more than one card, there is a special IBM machine, usually called a "gang punch" or "reproducer," that will accomplish this task very quickly and economically.

There are many fine points about the actual mechanics of punching, many of which can be learned only by experience, but there is one important issue that can be discussed here. Punching of information into IBM cards is a clerical operation, and, like any other, it is subject to errors. With inexperienced or unskilled operators these errors are worth worrying about. As a result it is customary to "verify" all punching, unless the operator is known to be (consistently) practically errorless. The verification consists of putting the original card through a machine that looks very

much like the puncher that was used to punch the card. The person doing the verification goes through a series of mechanical operations that are identical to those he would use if he were punching the card for the first time (he must, of course, have the original questionnaire in order to do this). Instead of actually punching, however, the machine does nothing but give a signal whenever the verifier's punch disagrees with the original punching of the card. The verifier can then look to see if the original card is wrong or if he has made a mistake himself. In either case the procedure serves as a check on the accuracy of the initial punching.

The check is not perfect, because it is conceivable that the verifier and the original puncher would make the same error. The probability of this is quite small if errors in punching are independent. However, there is one kind of error that is not independent: if the original operator punches the digits correctly but puts them in the wrong columns, the verifier might very well make the same error. For example, suppose that column 6 was to be left blank if there was no answer to the question being recorded in column 6. The operator might inadvertently forget to skip this column and record the digit that belonged in column 7 in column 6 instead. Then, except for those which were right by accident, all the remaining columns would be punched incorrectly. If the person verifying the original punching made the same error in column 6, she would not detect anything wrong with the entire card. If multiple punching is used, the possibility of such errors is enhanced. The best way to guard against this error is to design the code in such a way that there will always be one punch, but never more than one, in a column. This requires, of course, that single rather than multiple punching be used, and this is one of the arguments in favor of single punching. It requires that "no answer" or "not ascertained" be assigned an actual digit rather than being recorded by omitting a punch in the appropriate column. If this procedure is followed, verification becomes an almost airtight procedure.

On very large studies it may sometimes be worth while to verify only a sample of each operator's punching. The U.S. Bureau of the Census has applied this procedure to the punching of census information and has used the methods of industrial quality control in order to maintain the desired level of accuracy in the punching.

2. Running. Once the IBM cards have been punched and verified, tabulation begins. The machine most frequently used for tabulation is the sorter-counter. It automatically counts and sorts the cards according to the way they are punched. The important thing that has to be understood about this process, short of the mechanical details, is that this machine can count and sort *one column at a time and no more*. That is, if cards are put through the machine once, the machine will (1) count the number of times

each digit in a particular column, say, column 17, has been punched and (2) sort the cards into as many as 12 piles according to the digit that is punched. Both these processes are carried on simultaneously. If, however, a particular column has more than one punch (*i.e.*, if double punching or multiple punching has been used), the process of sorting becomes somewhat more complicated and we shall not here consider these complications.

A disadvantage of single punching with respect to tabulation is now obvious. If there are several punches in a single column (*i.e.*, multiple punching), these punches are counted simultaneously as the cards are run through the machine a single time. If, on the other hand, single punching is used and several columns are needed in order to record the information on the multiple-answer question, the whole set (or "deck") of cards must be run through the machine once for each column that is used. Hence the time required to tabulate is greater with the single-punching method than with the multiple-punching method *if no errors are made by the machine operators*. Usually, however, the errors made by operators unfamiliar with the method of multiple punching are much more serious with multiple punching than with single punching.

Often it is necessary simply to count, and then the process of sorting is superfluous. The need to sort exists because of the need for cross-tabulation. Suppose, for example, that the investigator wished to tabulate the purchase of oriental rugs by income groups. He would first sort the cards by income groups and then run each of these groups through the machine to obtain a separate count on the number of purchases of oriental rugs per income group.

Cross-tabulation becomes more complicated when more than one card per questionnaire is needed. The process is simplified if the questions between which cross-tabulations are to be made are punched on the same card. If this is not possible the cross-tabulation can be accomplished only by rather complicated procedures. One involves the use of an IBM machine called the "collator." The second involves the use of the gang punch. This machine makes it possible to punch in another card (not necessarily in the corresponding columns) the information contained in a given card or any part of the given card. Thus it is easy to punch from the original cards an entirely new deck of cards such that the two questions to be cross-tabulated appear on the same card. The simple method of tabulation can then be used.

When more complicated cross-tabulations are needed, the basic principles are no more difficult but the actual machine procedure gets more intricate and there are more opportunities for short cuts. It is worth saying that skilled IBM machine operators usually learn a great many "tricks" and short cuts that the average marketing researcher can seldom hope to

know. When the tabulations become complicated, there are great advantages in having a skilled and intelligent machine operator who understands the marketing research study as well as the IBM machines. Such an individual should usually be consulted before punching begins in order to get his advice on the "layout" of the IBM cards. For example, before deciding to use multiple punching, the IBM operator should certainly be consulted, since he is the one who will eventually have to suffer with the actual work of tabulation.

The sorter-counter is the most common IBM machine for use in tabulation. The collator and the gang punch also have specialized uses. Still another machine, the tabulator, can be adapted to the problems of tabulation. This machine can accomplish the same thing as the counter-sorter in counting the number of punches in cards. However, instead of being limited to a single column at a time, as is the counter-sorter, the tabulator can count several columns at once. As an offset, however, the tabulator handles cards more slowly than the counter-sorter and is somewhat more expensive.¹

In actually recording the results of tabulating, the tabulator prints the results automatically, while the sorter-counter requires that the answers be recorded manually by the operator. If the sorter-counter is used, it is highly desirable to use a standardized form on which to record the answers (IBM operators usually have these forms). These forms then serve as the rough tables for use in planning and writing the final report of the study. Only two things need be done to them before they can be used for this purpose. First, they must be "decoded"—that is, the numerical designation such as 32-2 must be supplemented by its equivalent, "has used Brand B flour." Second, percentage figures must be supplied so that per cents may accompany the absolute numbers counted by the machine.

3. Other methods of tabulation. Thus far we have discussed the most commonly used methods of tabulation in marketing research. Other approaches have been and can be used. The Thurstone Method of Edge-marking has possible applications.² The procedures of mark sensing, often used in mechanically scoring examinations, have been used experimentally in census work and may have eventual application to marketing research. The McBee Card enables one to use a procedure similar to the IBM process just described, except that the sorting and counting is manual rather than mechanical. Moreover, many new methods of tabulation are now in the

¹ All IBM machines must be rented from the International Business Machines Company. The rent on a tabulator is several times that on a counter-sorter.

² L. L. Thurstone, "The Edge-Marking Method of Analyzing Data," *Journal of the American Statistical Association*, Vol. 43, No. 243, pp. 451-462, September, 1948.

exploratory stage. A full account of these possibilities would require too much detail.

22.4 Calculations

The ordinary calculations required in marketing research consist only of converting the absolute numbers produced by tabulation into relative or percentage form. In bulk, at least, this operation is by far the most important in the typical survey. It is usually accomplished by the use of any of the ordinary desk calculating machines made by the Friden, Monroe, or Marchant companies. Skill in the use of these machines for "percentaging" is relatively easily acquired, and ordinarily, a clerical worker who is capable of coding or hand tabulation can quickly learn to calculate percentages. There is only one important trick that is worth mentioning: it is much faster to multiply than to divide. The process of percentaging should therefore be reduced to a series of multiplications by the use of reciprocals, which can be obtained from a table of reciprocals such as "Barlow's Tables" or which can be computed directly from a single division.

Sometimes more elaborate computations are needed in order to carry through some of the statistical operations mentioned at various points in Parts II, III, and IV of this book. For this work desk calculating machines are usually adequate, but a much higher level of skill and training is needed. The training manuals put out by the companies manufacturing calculating machines are almost worthless for research calculations, but fortunately a statistician has taken the trouble to write a manual that should meet this need: *Machine Computation of Elementary Statistics*, by Katherine Pease, Department of Psychology, Barnard College (Chartwell House, Inc., New York, 1950). The skill required, in addition to training, seems to be chiefly an understanding of elementary arithmetic, which, unfortunately, many otherwise competent adults lack.

Even for good computers, it is desirable to check all computations unless automatic checks are available. Ideally such a check should be done by a different person, but if the same individual carries through the operations a second time without reference to the first computation, the end result is probably about as good. (If the person is not a competent calculator, he can probably repeat the calculation ten times without coming to the same answer twice.)

The researcher should always check calculations by scanning to see if they are approximately right. (A slide rule is invaluable for this.) Gross errors, particularly those due to misplaced decimal points or incorrectly taken square roots, are usually quite apparent. The ability to locate such errors often stands between the researcher and the embarrassment of seri-

ous errors that go undetected. The same ability to scan is also useful in checking the reasonableness of tabulations.

When very elaborate or time-consuming statistical operations are essential (and they should be resorted to only after deliberation and with a clear understanding of what is to be accomplished by them), the researcher may wish to explore beyond the desk calculator for methods of getting the work done. In regression analysis, for example, the IBM tabulator, referred to in the discussion on tabulation, can be used to speed one exceedingly time-consuming stage. Moreover, the modern developments in methods of calculation may become more and more accessible to the man with relatively modest computing problems that nonetheless are somewhat tedious with the use of the ordinary desk calculator.

22.5 Mechanical Preparation of Reports

The typing of reports is a much more highly skilled job than the typing of ordinary business letters. The research typist should be more capable than most, and she should be intelligent enough to learn something about research. Given only general instructions, she should be able to translate a scrawled manuscript containing tabular material and possible charts as well into an attractive final report that contains no errors that did not appear in the rough draft with which she worked. Ideally, she should be able to detect errors in the draft itself and call them to the attention of the research analyst. Few typists are able to meet the first requirement, and fewer still the second. Nonetheless, every attempt should be made to obtain a typist who can meet at least the first, since the time of the research analyst is perhaps better used if he does not have to spend his time in correcting typing errors.

The analyst himself can make important contributions to the quality of the typing of the final report. First, he can prepare the rough manuscript in as legible a form as possible (or dictate it as coherently as possible). Second, he can give adequate instructions on format. Usually it is desirable to set up a more or less standard format in order to achieve both good results and uniformity. Third, he can proofread the final manuscript thoroughly and critically. Few tasks are so distasteful yet so rewarding, because there are often serious errors (*i.e.*, other than misspellings or incorrect punctuations), and these can cause a lot of trouble if they are permitted to pass.

Sometimes, especially when an easel or "flip chart" presentation is desired, charts or drawings are needed. It is often worth while, especially in large research departments, to have an individual who can do respectable art work as the need arises. This individual should understand research as well as have some modest proficiency in art work. Since there is

seldom a steady flow of art work, even in relatively large departments, the "artist" should be able to do other jobs in the department in between his artistic assignments.

The final problem that requires discussion is that of reproduction. There is no need for a technical discussion, but it should be pointed out that there is a considerable difference in the attractiveness of results, given comparable skill on the part of typists. For especially attractive work, for example, processes like multilith or Ozalid are worth while. One should not count too closely the cost of reproduction, since it is almost always a minute fraction of the total cost of the study and since the quality of reproduction has a lot to do with the interest with which the research report is received and read by the executives of the company.

A caution about preparing reports by the method of carbon copies is worth while. This is the only method of reproduction that permits no flexibility about the number of reports that can be prepared. No more than six or seven copies can be prepared (except with an electric typewriter), and if more than this number are eventually needed, the whole study has to be typed over. If, on the other hand, the report had been prepared on hectograph or mimeograph, a request for more copies than had originally been prepared could be met at slight extra cost. In general, the demand for copies of a research report is difficult to predict and it is embarrassing to be "caught short."

SOME SUGGESTIONS FOR FURTHER READING

The readings listed in the following pages are not intended to exhaust all the excellent sources of further information on the topics discussed in this book. Each one, however, is introduced with a special purpose in mind, and in many instances we have stated precisely what this purpose is. Usually footnotes to the text of the book are *not* duplicated in this section. After references have been listed for the chapters of this book, a few references will be given to works which include vastly more comprehensive listings than we have included. Also included at the end are the names of professional journals which contain most of the relevant developments in the basic social sciences.

CHAPTER 1

In this chapter we mention that a knowledge of elementary statistics is assumed throughout this book. One of the best books on this subject is *Statistics*, L. H. C. Tippett (Oxford University Press, New York, 1947). This treatment is non-technical and can be read very rapidly. A second and more careful reading, however, will usually be rewarding. A slightly more technical treatment is given by *Elementary Statistical Methods*, William A. Neiswanger (The Macmillan Company, New York, 1943). Chapters 1-3, 5, 6, 8-10 are particularly relevant.

CHAPTER 2

Marketing Research Practice, edited by Donald M. Hobart (The Ronald Press Company, New York, 1950), contains many case histories of marketing research studies that can be read profitably in conjunction with many chapters of the present book.

How to Conduct Consumer and Opinion Research, edited by Albert B. Blankenship (Harper, New York, 1946) contains descriptions of different kinds of continuing, widely used marketing research services, each described by a representative of the company selling the service. Much can be learned of the operation of consumer and store panels, various devices for measuring radio audiences, etc.

How You Can Get Facts about Your Market (Chamber of Commerce of the United States, Washington, 1948) contains a list of sources of basic marketing data, primarily from governmental agencies and publishing companies.

CHAPTER 3

An Introduction to Logic and Scientific Method, Morris R. Cohen and Ernest Nagel (Harcourt, Brace and Company, Inc., New York, 1934), is a standard text on scientific method and contains a substantial bibliography.

CHAPTER 4

Many statistical techniques are mentioned in this chapter, some with nontechnical explanations and some without. The marketing researcher can enhance his ability to attack practical problems by learning more about these techniques. There are excellent references at different levels of mathematical difficulty. For those without college mathematics, the best reference we know is *Statistical Methods*, George W. Snedecor (Iowa State College Press, Ames, Iowa, 1946, 4th ed.). As a text for those with a background in elementary calculus, an excellent choice is *Introduction to the Theory of Statistics*, Alexander Mood (McGraw-Hill Book Company, Inc., New York, 1950). *Mathematical Methods of Statistics*, Harald Cramér (Princeton University Press, Princeton, N. J., 1946), is excellent at the advanced level.

For those interested in a nontechnical discussion of statistical principles of experimentation, the first three chapters of *The Design of Experiments*, R. A. Fisher (Oliver & Boyd, Ltd., Edinburgh and London, 1950, 5th ed.), is recommended. Two sources of further technical information on experimentation are the remaining chapters of Fisher's book, just mentioned, and *Experimental Designs*, William G. Cochran and Gertrude M. Cox (John Wiley & Sons, Inc., New York, 1950).

CHAPTER 6

Two interesting discussions of sampling history are given in the following articles: "History of the Uses of Modern Sampling Procedures," F. F. Stephan, *Journal of the American Statistical Association*, Vol. 43, pp. 12-39, 1948; and "Use of Sampling in Great Britain," C. A. Moser, *Journal of the American Statistical Association*, Vol. 44, pp. 231-259, 1949. A very worth-while introduction to sampling problems in marketing is given by "On Sampling in Market Surveys," Philip M. Hauser and Morris H. Hansen, *Journal of Marketing*, Vol. 9, pp. 26-31, 1944.

CHAPTER 7

The statistical references suggested under Chapter 4 will be found useful for a further treatment of the statistical concepts introduced in this chapter. A very good nontechnical discussion of the concept of sampling error is given in "Some Criteria for Judging the Quality of Surveys," W. Edwards Deming, *Journal of Marketing*, Vol. 12, pp. 145-157, 1944.

CHAPTERS 8 AND 9

Two textbooks on sampling from human populations have recently appeared. In general these texts give considerably more detail than the presentation in the present book. On a relatively nontechnical level is *Sampling Methods for Censuses and Surveys*, Frank Yates (Hafner Publishing Company, New York, 1949). In addition to its excellent treatment of sampling, this book also treats many of the other practical problems connected with surveys. The second book, more technical on the whole, is *Some Theory of Sampling*, W. Edwards Deming (John Wiley & Sons, Inc., New York, 1950). Part I of this book is quite nontechnical and provides stimulating reading for any serious student of marketing research. Interesting examples of sampling are found throughout the book.

An excellent treatment of sampling theory is the following mimeographed manuscript: *Sample Survey Techniques*, W. G. Cochran (Institute of Statistics, North Carolina State College, and Bureau of Agricultural Economics, U.S. Department of Agriculture cooperating, Raleigh, N. C., 1948).

Numerous articles on the theory and practice of sampling have appeared in the various journals and trade publications in recent years. We list those which will probably be particularly useful in complementing this book.

1. "On the Theory of Sampling from Finite Populations," Morris H. Hansen and William N. Hurwitz, *Annals of Mathematical Statistics*, Vol. 14, pp. 333-362, 1943. Although this article makes considerable explicit use of mathematics, the verbal explanations will be clear to readers on almost all levels of mathematical competence. It summarizes the early development in area sampling made by the technicians of the U.S. Bureau of the Census.

2. "On Some Census Aids to Sampling," Morris H. Hansen and W. Edwards Deming, *Journal of the American Statistical Association*, Vol. 38, pp. 45-49, 1943. This article gives some idea of the kinds of "sampling materials" needed and available in the practical implementation of area samples.

3. "The Problem of Non-Response in Sample Surveys," Morris H. Hansen and William N. Hurwitz, *Journal of the American Statistical Association*, Vol. 41, pp. 517-529, 1946.

4. "A Procedure for Objective Respondent Selection within the Household," Leslie Kish, *Journal of the American Statistical Association*, Vol. 44, pp. 380-387, 1949. A discussion of the problem of getting a sample of individuals, as contrasted with a sample of families.

5. *Recent Development in Sampling Theory in the United States*, W. G. Cochran, International Statistical Conference, 1947 (mimeographed).

6. "Systematic Sampling and its Relation to Other Sample Designs," Lillian H. Madow, *Journal of the American Statistical Association*, Vol. 41, pp. 204-217, 1946.

7. "Commercial Use of Probability Sampling," Warren N. Cordell, *Journal of Marketing*, Vol. 14, pp. 447-449, 1949.

Some applications of area sampling are given in the following articles:

1. "On the Design of a Sample for Dealer Inventories," W. E. Deming and Willard Simmons, *Journal of the American Statistical Association*, Vol. 41, pp. 16-33, 1946.

2. "Sampling for the 1947 Survey of Consumer Finances," Roe Goodman, *Journal of the American Statistical Association*, Vol. 41, pp. 173-189, 1946.

3. "Problems and Methods of the Sample Survey of Business," Morris H. Hansen, W. N. Hurwitz, and Margaret Gervey, *Journal of the American Statistical Association*, Vol. 41, 173-189, 1946.

4. "On a Population Sample for Greece," R. J. Jesson *et al.*, *Journal of the American Statistical Association*, Vol. 42, pp. 357-384, 1947.

CHAPTERS 10 AND 11

Discussions of nonprobability sampling have been widely scattered throughout trade journals and popular articles, but there has been little tendency to discuss

the subject theoretically. For a thorough knowledge of the techniques of non-probability sampling, the reader is urged to investigate the printed sources just suggested and, even more important, to talk to experienced practitioners.

Many comparisons of probability and nonprobability sampling have been made. One of the best for the practicing marketing researcher is "Selecting Sampling Methods in Commercial Research," Peter G. Peterson and William F. O'Dell, *Journal of Marketing*, Vol. 14, pp. 182-189, 1950.

CHAPTER 12

One of the best treatments of the ground covered in this chapter is "On Errors in Surveys," W. E. Deming, *American Sociological Review*, Vol. 9, pp. 359-369, 1944.

An interesting and pioneering effort to measure response errors is found in "Response Errors in Surveys," Morris H. Hansen, William N. Hurwitz, Eli S. Marks, and W. Parker Mauldin, *Journal of the American Statistical Association*, Vol. 46, pp. 147-190, June, 1951.

CHAPTER 13

Gertrude Bancroft, "Consistency of Information from Records and Interviews," *Journal of the American Statistical Association*, Vol. XXXV, pp. 377-381, 1940. This brief note contains the results of interviews with respondents who had previously provided the required information in official records.

Archibald S. Bennett, "Some Aspects of Preparing Questionnaires," *Journal of Marketing*, Vol. X, pp. 175-179, 1945.

CHAPTER 14

Probably the most useful and interesting discussions and data on statements of intentions are in the annual Survey of Consumer Finances, conducted by the Board of Governors of the Federal Reserve System. See the articles on expenditures for durable goods in the following issues of the *Federal Reserve Bulletin*: June, 1948; June, 1949; and July, 1950.

CHAPTER 15

Of particular interest is the discussion and bibliography of Quinn McNemar, "General Review and Summary of Opinion-Attitude Methodology," *Psychological Bulletin*, Vol. 43, 1946.

The following articles are of general interest to the student and practitioner of opinion and attitude research:

1. Stanley Payne, "Some Opinion Research Principles Developed Through Studies of Social Medicine," *Public Opinion Quarterly*, Vol. 10, 1946.

2. Don Cahalan, Valerie Tamulonis, and Helen Verner, "Interviewer Bias Involved in Certain Types of Opinion Survey Questions," *International Journal of Opinion and Attitude Research*, Vol. 1, No. 1, March, 1947.

3. H. C. Link and A. D. Freiberg, "The Problem of Validity vs. Reliability in Public Opinion Polls," *Public Opinion Quarterly*, Vol. 6, 1942.

The following references are especially useful in understanding the theory of attitude measurement:

1. L. L. Thurstone, "Theory of Attitude Measurement," *Psychological Review*, Vol. 35, 1929.
2. J. P. Guilford, *Psychometric Methods*, McGraw-Hill Book Company, Inc., New York, 1936.

CHAPTER 17

Paul Lazarsfeld, "The Why Question in Marketing Research," *The National Marketing Review*, Vol. 1, 1935.

William F. Brown, "The Determination of Factors Influencing Brand Choice," *The Journal of Marketing*, Vol. 14, 1950.

CHAPTER 18

Of the many books that have been written on the problems of administration in business, one of the best is *The Functions of the Executive*, Chester I. Barnard (Harvard University Press, Cambridge, Mass., 1938.) Another excellent book is *Administrative Behavior*, Herbert A. Simon (The Macmillan Company, New York, 1948.)

CHAPTER 22

An extremely stimulating book on problems in marketing research is *Say It with Figures*, Hans Zeisel (Harper & Brothers, New York, 1947). It is mentioned at this point because Part I on Problems of Classification is a very good elementary discussion of coding problems.

An excellent discussion of the techniques of quality control and sampling inspection, alluded to in this chapter, is *Technological Applications of Statistics*, L. H. C. Tippett (John Wiley & Sons, Inc., New York, 1950).

The works listed below are of interest in themselves, and they contain much more extensive bibliographies than we have presented:

1. Mildred B. Parten, *Surveys, Polls, and Samples*, Harper & Brothers, New York, 1950. This book is well written and worth reading. One of its most impressive features is a bibliography of 1145 items on all phases of survey research.
2. Donald M. Hobart, ed., *Marketing Research Practice*, The Ronald Press Company, New York, 1950.

The following journals are of particular interest for marketing research:

1. *Journal of Marketing*
2. *Journal of the American Statistical Association*
3. *International Journal of Opinion and Attitude Research*
4. *The Public Opinion Quarterly*

INDEX

A

- Ability to provide information, 205-207
 - on attitudes, 287-288
 - on behavior, 217, 224-243
 - on intentions, 254-256
 - on "reasons why," 315
- Accuracy, in expressions of attitude, 266-288
 - ability of respondents to provide, 287-288
 - scales used in ranking, of individuals, 273-282
 - of stimuli, 282-286
 - single-question technique, 266-273
 - willingness of respondents to provide, 286-287
- of information, on behavior, in communication, 224-240, 243-252
 - in observation, 219
 - in self-observation, 222
 - on knowledge and habits of thought, 297-298
- meanings of, 205-207, 269
- of "reasons why," 308-316
- requirements for, effects of, on choice of sampling design, 188-189
- of statements of intentions, 262-264
- Ackoff, Russell, 165*n.*, 178*n.*, 282*n.*
- Ad Detector Corporation, 295, 297
- Administration, organization of discussion of, 8, 319-320
- Advertising, measuring effectiveness of, 50-66, 209-211, 216, 299-301
- Advertising Age*, 335
- Advertising agencies, marketing research by, 12, 14
- Aerial photographs, use of, in area sampling, 131
- "Affirmative bias," 231
- Agencies performing marketing research, 12-15
- Agricultural marketing research, *in.*, 11-12
- Aided recall, 232-234
- Allen, Fred, 85
- Allocation, optimal, 149
- Ambiguity, of communication, 240-243
 - in securing information about intentions, 256-257
 - of information on knowledge and habits of thought, 296
 - meaning of, 205-207
 - of statements of intention, 262-264
 - in "why" questions, 301-308
- Analysis, content, 314*n.*, 416*n.*
 - of covariance, 65
 - item, 281
 - Scale, 281-282
 - of variance, 65
- Area sampling, acceptance of, in marketing research, 165-167
 - basic ideas of, 112-114
 - controversies connected with, 167
 - different methods of design, 115-116
 - as form of "cluster sampling," 112
 - independence of lists of population, 112-113
 - practical problems in application of, 158-166, 376-377
 - identification of people with areas, 159-160
 - nonresponse (not-at-homes and refusals), 160-163
 - prelisting, 113, 376
 - saving of travel time in, 114
 - (*See also* Sampling; specific sampling methods)
- "Arousal," 297
- Art work in research reports, 430-431
- Association, 229, 294
 - free, 236-237
- Attitude measurements, 77, 265-288
 - assumption underlying, 273
 - nature of, 268-269, 273
- Attitudes, 212-213
 - accurate and unambiguous expressions of, ability to provide, 287-288

- Attitudes, accurate and unambiguous expressions of, willingness to provide, 286-287
 and behavior, reaction to requests for information on, 286-289
 relative usefulness of information on, 212-213
 difficulties encountered in research on, 269
 of individuals, toward competing stimuli, 212, 265
 methods of scaling, 213, 282-286
 toward single stimulus, 265
 methods of scaling, 266-282, 285-286
 and intentions, relative usefulness of information on, 286-287
 measurement of, 77, 265-288
 assumption underlying, 273
 nature of, 268-269, 273
 and opinions, definition of, 213, 265
 quantification of, 273-274
 scaling of, 273-282, 287
 (See also Scaling; Scaling techniques)
 securing information on, 265-288
 by communication, summary of discussion on, 285-286
 usefulness of information on, 265-266
 "Audience," definition of, 216*n*.
 Audimeter, Nielsen, 219
 Audit Bureau of Circulations, 179, 216, 245, 246
 Averages, law of, 87
- B**
- Banks, Seymour, 265-266
 Barlow's tables of reciprocals, 429
 Behavior, and attitudes, reaction to requests for information on, 286-287
 (See also Attitudes, and behavior)
 information on, 211-212
 ability to provide, 217, 224-243
 cost of securing, 252
 techniques for securing, 207-208, 217-252
 usefulness of, in business, 215-217
 willingness to provide, 217, 243-252
 and intentions, relative usefulness of information on, 253-255
 and "reasons why," relative usefulness of information on, 299-301
- Berelson, Bernard, 416*n*.
 Bergson, 225
 Bernberg, Raymond, 286*n*.
 Bias, affirmative, 231, 270
 due to interviewers, 272
 lessening of, 193-197
 measurement of, 247
 negative, 235
 in observation, 219-220
 order of presentation, 230-231
 positional, 230-231
 sampling, 97
 from unwillingness of respondents, 245-247
 (See also Leading questions)
 Biographical determinants of actions, 302-304
 Biological determinants of actions, 302-304
 Brown, George, 196, 291-292
 Brown, Lyndon, 270
 Budget for research, 325
 "Buying Power Index," *Sales Management*, 58*n*.
- C**
- Cahalan, Don, 272*n*.
 Calculating machines, 429
 Calculations, 429-430
 Call-backs, 160-163, 188, 190, 376-377
 Cantril, Hadley, 217, 272
 Cardinal measurement, 269
 Categorization (see Classification; Coding)
 Causal relationships, 213-214, 299-300
 "other" factors in, 60
 problems in extraction of, from observational data, 66
 Causation, "direction" of, 61, 64, 65
 interpretation of, 60-61, 213*n*.
 "other" factors in, 60-61, 64, 65
 Cause, definition of, 213*n*.
 Census, 84
 sample, 145
 U.S. Bureau of the, 85*n*., 86, 119-120, 131, 132, 133, 379, 426
 Central-limit theorem, 103
 Cheating in interviewing, 396, 398-401
 Check list in questionnaires securing "reasons why," 306
 Chi-square test, 65

- Chicago Tribune* Consumer Panel, 15, 292
- Choice, involving risk, 274
prediction of, 279
- Churchill, Winston, 19, 20
- Churchman, C. West, 165*n.*, 178*n.*, 282*n.*
- Classification, cross-, 61-65
principles of, 407-408
problem of, 267-269, 411
(*See also* Coding)
- Clements, Forrest, 351
- Cluster sampling, advantages of, 114
applications of, 112-115
equal-sized clusters in, possibility of getting, 116
multistage sampling in, 129-134
numbers of clusters at different stages, 132-134
primary sampling units, 130-132
secondary sampling units, 130-132
nature of, 112
practical problems in (*see* Area sampling, practical problems in application of)
sampling unit in, 112
and stratified sampling, comparison of, 137-138
use of, together, 138-141
with subsampling, equal-sized clusters, 124-128
sampling distribution of the mean in, 125-126
unequal-sized clusters, 128-129
sampling distribution of the mean in, 128-129
vs. without, 126-128
without subsampling, equal-sized clusters, 116-123
method of formation of clusters in, 121-123
optimum size of clusters in, 118-121
sampling distribution of the mean in, 117-118
unequal-sized clusters, 123-124
finite population multiplier, 124
sampling distribution of the mean in, 124
theory of, 115-134
types of clusters, areas (*see* Area sampling)
- Cluster sampling, types of clusters, elements contiguous in time, 115
variable probabilities of selection in, 142-147
"Cluster size," definition of, 116*n.*
(*See also* Cluster sampling)
- Clusters, determination of number of, at different stages, 132-134
ways of forming, 114-116
- Code, block, 415
format of, 414
"freezing" of, 411
preliminary, 344
question-by-question vs. over-all, 415-416
"setting up" of, 410
- Coding, 267-269, 271-272, 302-304, 314, 319, 344, 406-417
"all other" category in, 414-415
amount of detail in, 408, 412-414
control of, 416-417
of depth interviews, 314
and editing, 409-410
final, 410-412
general principles of, 407-408
internal vs. outside agency for, 321, 323
multidimensionality of, 414
multiple answers, 407-408
"no answer," handling of, 407
"not ascertained," handling of, 407
postcoding (*see* final, above)
precoding, 408-409
qualifications for, 416
question-by-question vs. over-all, 415-416
relation of, to tabulation requirements, 416
reliability of, 413-414
specifications for outside, 327
(*See also* Code)
- Color Research Institute, 295
- Commerce, chambers of, 74, 383
U.S. Department of, 74
- Common sense vs. scientific analysis, 26
- Communication, 207-208, 217, 224-252
ambiguity in, 205-207, 240-243, 263-264, 296-297
on "reasons why," 301
control of response error in, 205-207
inaccuracy in, 205-207
due to memory errors, 225-240

- Communication, inaccuracy in, due to unwillingness of respondents, 243-252
 on "reasons why," 308-316
 when obtaining associations, 297-298
 kinds of errors in, 203-207
 measurement of error in, 201-202
 and observation as techniques, 207-208, 217
 applications of, 207-208
 criteria for judging, 208
 organization of discussion of, 8, 202, 210
 relative accuracy of, 219
 relative advantages of, 224-225
 relative costs of, 252
 supplementing each other, 220
 problem of, organization of discussion of, 8, 202, 214
- Comparative judgment, law of, 284
- Compensation, of interviewers, 403-405
 of respondents, 221, 223
- Conant, James, 26
- Conclusions in logic, 33
- Confidence coefficient (*see* Confidence interval, level of confidence of)
- Confidence interval, 103-108
 appropriateness in probability sampling designs, 108
 caution in use of formulas for, 109
 for cluster sampling, 118, 126
 (*See also* Cluster sampling, sampling distribution of the mean in)
 computation for sample proportion, 108
 inapplicability of concept in nonprobability sampling, 109
 interpretation of, 106-107
 level of confidence of, 106, 108
 in stratified sampling, 136-137
- Confounding, 60, 61
- Consumer panel, 207, 221-224
 Chicago Tribune, 15, 292
 "mortality" of, 193-194, 222
 remuneration of members of, 223
- Content analysis, 314*n.*, 416*n.*
- Contingency questions, handling of, in interviews, 361
- Control, instruments of, over research department, 325-329
 statistical, 47 (*See also* Randomization)
- "Conversational interview," meaning of, 312
- Coombs, Clyde, 274
- Correlation analysis, 65, 300
- Cost, of communicative techniques, 252
 of observational techniques, 220
- Cost criterion in judging techniques for obtaining information, 208
- Cost estimate of projects, 325-329
- Costs, of interviewing, 327-328
 in sampling, 189-191
- Count units, 132
- County Data Book*, 131, 139*n.*
- Covariance, analysis of, 65
- Cowan, Donald, 58*n.*, 216*n.*
- Criterion situation, 259
- Cross-classification of observational data, 61-65
- Crossley, Archibald, 174
- Crum, William, 174
- D
- Data, primary, 319
 processing of marketing research, 406-437
 secondary, 207
 (*See also* Secondary sources)
 types of, 15-18, 44-45
- Deduction, relation of, to prediction, 33
 as stage of scientific method, 27, 32-34
- "Definition of the problem," 31, 330-338
- Deming, W. Edwards, 178*n.*
- Department of Commerce, 131
- Depth interview, 236-237, 312-314, 416*n.*
- Descriptive statistics, 6, 30
- Design, of experiments, 49-50
 of projects, 338-345
- "Detailed interview," meaning of, 312
- Determinants, biographical, 302-304
 biological, 302-304
 first-degree, 302-304
- Deviation, standard, 90, 99
- Dishonesty in interviewing, 396, 398-401
- Disproportionate stratified sampling, 147-151, 185
 limitations of, 150-151
 reasons for use of, 148-150
- Distribution, sampling (*see* Sampling distribution)
 frequency, definition of, 96*n.*

- Division of Special Surveys, Bureau of
Agricultural Economics, 351
- Dollard, John, 259
- Double sampling, 153-155, 231-232, 248
- E
- Easel presentation, 430
- Ebbinghaus, Hermann, 226-227
- Econometrics, 66
- Edge-marking, 428
- Editing, 251, 271-272, 409-410
- Effective randomization, 171, 177
- Effectiveness of advertising, measure-
ment of, 50-66, 209-211, 299-301
- Efficiency, 101-102
 economic, 102
 statistical, 102
- Equal-appearing intervals, 275
 scaling by method of, 273-278
- "Equal-take" approach, 197
- Error, nonsampling, 84, 97, 202-207
 response, 202-207
 risks of, in statistical inference, 51-54
 (See also sampling, below)
 sampling, 84, 97, 201
 controlling and diminishing risks of,
 53-55
 (See also Sampling error)
 standard, of the mean, 99
- Estimate, of cost, 325-329
 (See also Cost)
 ratio, 157*n*.
 regression, 157-158
 sample, appraising accuracy of, 96-97
 unbiased, 97-99
- Estimation, methods of, 102*n*., 157-158
 of population totals, 145, 192*n*.
- Evaluation, simultaneous, of effects of
 several variables in experimenta-
 tion, 55-56
- Expenditures, establishing standards for
 frequently recurring, 69-73
- Experiment, definition of, 17
- Experimental data, 17-18, 44-45
 difference of, from observational data,
 57, 58, 258
 interpretation of, 50, 58
- Experimentation, 45-57
 administrative problems in, 57
 advantages of, 258
- Experimentation, basic prerequisite for,
 48
 costs of, 57
 definition of, 45
 effect of "other" factors on, 48-50, 54
 "experimental" and "control" groups
 in, 50
 limitations of, 56-57, 259-260
 reasons for restricted use of, in social
 sciences, 45-47
 relation of statistical inference to, 50-
 53
 risks of error in, 51-52
 control of, 53-55
 role of randomization in, 47
 simultaneous evaluation of effects of
 several variables in, 55-56
 use of stratification in, 55
- Exploration, preliminary, 330-338
- Eye estimates, 147
- F
- Factor analysis, 66
- Federal Reserve Bulletin*, 257
- Fieldwork, evaluation of quality of, 397-
 402
 problems in administration of, 406
 scheduling of, 344
 supervision of, 396-405
 (See also Interviewers; Interviewing)
- Finite population multiplier, 101*n*., 117*n*.
- First-degree determinants, 302-304
- Fisher, R. A., 97*n*.
- Fixed-alternative questions, 306
- Fixed costs in sampling, 190
- "Flip chart" presentation, 430
- Forecasting consumer behavior, useful-
 ness in, of information on attitudes,
 265-267
 of information on knowledge and
 habits of thought, 289-291
 of statements of intention, 256-262
- Forecasts, 74-75
 general business, 74-75
 sales, 40-41, 254-255, 265-266
 types of, 40
- Forgetting, 226-228
 (See also Memory)
- Free association, 236-237
- Free-response questions, 271-272, 305-
 306, 353

Frequency, relative, 91-92
 Frequency distribution, definition of, 96n.
 Freud, Sigmund, 28, 225
 Friedman, Milton, 274

G

Gallup, George, 174, 293n.
 Gang punch, 425, 427
 General Motors Corporation, 403
 Goodman, Roe, 192n.
 Guilford, J. P., 281

H

Habit recollection, 225-226
 Habits of thought, 213, 294
 and associations, 294
 and other reactions to stimuli, devices for recording, 294-296
 usefulness of information on, 294
 (See also Knowledge, and habits of thought)
 Haire, Mason, 315n., 316
 Hand tabulation, 418-420
 Haphazardness, 94
 Hardy, C. O., 261
 Hauser, Philip, 187
 Hayek, F. A., 331
 Hearst "trading areas," 59n., 130n.
 Hectograph, 431
 Heneman, Herbert, 358n.
 Historical data, 15-16, 58
 Hollerith card, 418-419
 Hooperating, 17, 85
 Human populations, sampling of, 110-197
 Hypotheses, supplementary, 33-34
 Hypothesis, definition of, 32
 formulation of, 26, 77
 as stage of scientific method, 30-32
 testing of, 43
 as stage of scientific method, 27, 28, 34-39

I

Impact technique, 293
 Inability to provide information, 205
 (See also Ability)

"Inaccuracy," in response, 205-207
 (See also Accuracy)
 Inadequate response, 365-367
 Independent recollection, 225-226
Industrial Arts Index, 335
 Inference, statistical, 36
 "Informal interview," meaning of, 312
 Information, types of, 209-214
 Instantaneous determinants, 302-304
 Instructions for interviewers, 351-375, 391-396
 Instruments of control, 325-329
 cost estimate, 325-329
 research budget, 325
 "Intensive interview," meaning of, 312
 Intention, statements of, accuracy and ambiguity of, 262-264
 usefulness of, in forecasting, 254-262, 299
 Intentions, and behavior, relative usefulness of information on, 254-256
 information about, 212
 securing information on, 253-264
 Internal processing of data, 406-431
 International Business Machines (IBM) card, 418-419
 International Business Machines Company, 428n.
 IBM machines, collator, 427
 gang punch, 425, 427
 reproducer, 425
 sorter-counter, 418, 426-427
 tabulator, 428, 430
 verifier, 425-426
 IBM tabulation, 418-429
 Interview in depth, 236-237, 312-314
 Interviewer bias, 272
 Interviewer discretion, in probability sampling, 163
 in quota sampling, 173
 Interviewers, compensation of, 403-404
 finding of, 382-384
 full-time vs. part-time, 403
 independent contracting vs. employment for, 402-403
 instructions for, 351-375, 391-396
 qualifications for, 350-352, 386
 relations with, 404-405
 reselection of, 385
 selection of (see Selection of interviewers)

Interviewers, supervision of, 396-405
 training of, 389-396
 Interviewing, 349-377
 angry or belligerent respondents in,
 358-359
 answers in, adequate, obtaining of,
 364-375
 evasive, handling of, 366-367
 partial, handling of, 366
 asking of questions in, 359-364
 biographical data in, 374-375
 cheating in, 396, 398-401
 classification data in, 374-375
 closing of interview, 375
 contingency questions in, 361
 costs of, 327-328
 depth, 236-237, 312-314
 discretionary selection of respondents
 in, 173, 195-196
 dishonesty in (*see* cheating in, above)
 evaluation of quality of, 397-402
 example of good interview, 373-374
 face-sheet data, 374-375
 gaining acceptance in, 353
 instructions for, 351-375, 391-396
 introduction to respondent in, 354-356
 nature of work of, 349-350
 pretesting in, 342
 probing in, 271, 313, 367-370
 problems in, in probability sampling,
 163-164
 special, 357-359
 putting respondent at ease in, 353-354
 qualifications for, 350-352, 386
 recording in, 272, 314, 370-374
 sampling problems in, 375-377
 (*See also* Call-backs; Prelisting)
 supervision of, 396-405
 suspicious respondents, 357
 threatened refusals in, 357-358
 training for, 389-396
 transitions between topics in, 363-364
 (*See also* Questionnaires)
 Inventory, pantry, 218
 Item analysis, 281

J

James, William, 229, 235
 Jesson, Raymond, 158*n.*, 197

Judgment sampling (*see* Nonprobability
 sampling)

K

k, definition of, 104
 Kelvin, Lord, 28
 Kenyon & Eckhardt, Inc., 403
 Kish, Leslie, 240*n.*, 367*n.*
 Knowledge, of already adopted stimulus,
 291-293
 and habits of thought, 213, 289-298
 information on, accuracy of, 297-298
 ambiguity of, 296-297
 conditions for usefulness of, 289-291
 level of, 292
 organization of discussion of, 291
 of particular circumstances of time and
 place, 331-332
 of stimuli introduced under test con-
 ditions, 293
 types of, 291
 and understanding of sales messages,
 293-294
 Koopmans, Tjalling, 66

L

Latent-structure analysis, 282
 Law of comparative judgment, 284
 Lazarsfeld, Paul, 301*ff.*, 301*n.*, 416*n.*
 Leading questions, 270-271
 Level of confidence, 108
 Likert, Rensis, 280
Literary Digest poll, 173-174, 192, 194
 Logic, 33
 premises in, 34
 propositions in, 32
 relation of, to scientific method, 34
 syllogisms in, 33
 Lucas, D. B., 232*n.*

M

McBee card, 428
 McCall Corporation, 178-179
 Machine tabulation, 418-429
 McNemar, Quinn, 187*n.*, 421*n.*
 Mahalanobis, P. C., 84-85
 Mail survey, 161, 163, 173

- Management, needs of, coordinating research with, 330-348
 and research department, relations between, 332-334
- Maps, reading of, 375-376
 use of, in area sampling, 131-132
- Mark sensing, 428
- Market Facts, Inc., 209
- Market Research Company of America, The, 221
- Marketing research, administration of, 319-431
 organization of discussion on, 8, 319-320
 agencies performing, 12-15
 basic methods of, organization of discussion of, 7-8
 basic nature of, 3-4
 business environment of, 18
 definition of, 3-4, 12
 "definition of the problem" in, 31, 330-338
 kinds of problems in, 67
 planning of, 326-327
 purpose of, 10-11
 question of functions to be performed internally in, 320-322
 relation of, to economic concepts of demand and supply, 11-12
 study of, justification for, 5-6
 training for, 5-6
 types of, 9-10
- Marketing research companies, 12-14
 selection and use of outside, 322-323
- Master Sample of Agriculture, 132, 147
- Matching, 194
- Mayo, Elton, 46
- Mean, in simple random sampling, sampling distribution of, 95-103
 standard error of, 99-102
- Media, selection of, 216
- Memory, 225-237
 aid to, by formulation of questions, 229
 effect on, of recency, 226-228
 of types of events, 237-239
 habit, 225
 independent recollection, 225
 stimuli to, 228-237
 "aided recall," 232-234
 recognition method, 230-232
- Memory, stimuli to, "unaided recall," 234-237
 variables affecting, 226
- Memory errors in communication (*see* Memory)
- Method, scientific (*see* Scientific method)
- Mimeograph, 431
- Mirror technique, "one-way," 219
- Motivation, information on, 299-316
 (*See also* Reasons why)
- Mu (μ), symbol for mean of population, 90
- Multidimensionality, 206, 414
- Multilith process, 431
- Multiple-question scaling techniques, 273-282
- Multiple sampling, 156*n*.
- Multistage sampling, 129-134, 156
- Murray Thematic Apperception Tests (T.A.T.), 315
- N
- National Opinion Research Center, 250, 272, 377*n.*, 390
- New-product research, 10
- Nielsen Audimeter, 219
- Nielsen Food Index, 218
- Nondirective counseling, 236-237
- "Nondirective interview," meaning of, 312
 summarizing and describing of, 416*n*.
- Nonprobability sampling, 88, 90, 165-180
 adaptations of, for marketing research, 192-197
 bias in, of the mean, 98
 measurement of, 186
 possibility of, 168
 comparison of, with probability sampling, 185-192
 controversies connected with, 167
 distinctive characteristics of, 168-171
 effectively randomized population in, 171-172
 formulas for, 101, 109, 168, 182
 freedom of choice of respondents in, 169
 origin of, 166-167
 probability vs., 93*n.*, 191-192
 rationale for, 171-180
 sampling error in, lack of estimate of, 168

Nonprobability sampling, use in, of
 judgment, 169-171
 of stratification, 193
 Nonresponse, problem of, 160-163
 Nonsampling error, 84, 202-207
 Normal deviate, 104*n.*
 Normal distribution, 102, 105
 "Not-at-homes," 160-163

O

Oblivescence, 227*n.*
 Observation, 207-208, 217-221
 ability and willingness of respondents
 to permit, 220-221
 accuracy of, 219
 and communication as techniques (*see*
 . Communication, and observation
 as techniques)
 cost of, 220
 self-, 221-224
 some devices used for, 218-219
 as stage of scientific method, 26, 29-30
 unambiguity of, 219-220
 Observational data, 44, 57-66
 difference between experimental data
 and, 57
 problems in analysis of, 59-61
 suggested handling of, 61-66
 sources of, 58
 Odell, William, 191*n.*
 "One-way mirror" technique, 219
 Open-ended questions, 271-272
 Operations analysis, 5*n.*
 Opinions, 212-213
 obtaining information on, 265-298
 (*See also* Attitudes)
 Optimal allocation, 149
 Ordinal measurement, 268-269
 Organization of marketing research
 groups, 320-325, 344-345
 Over-all vs. question-by-question cod-
 ing, 415-416
 Ozalid process, 431

P

Paired-comparison method, 284-285
 Panel technique, 221-224
 (*See also* Consumer panel)
 Pantry inventory, 218
 Parameter, definition of, 90

Pascal, Gerald, 229*n.*
 Paterson, Donald, 358*n.*
 Pease, Katherine, 429
 Per cent (*see* Proportion)
 Personnel, kinds of, in marketing re-
 search departments, 319, 321
 use of outside marketing research,
 321-325
 Personnel tests, use of, in selection of
 interviewers, 388-390
 Peterson, Peter, 191*n.*
 Pilot testing (*see* Pretesting)
 Policy, alternatives to present, 78-79
 Politz, Alfred, 161*n.*, 162, 163, 177*n.*, 190,
 195, 197
 Politz-Simmons technique for eliminat-
 ing call-backs, 161-163
 Population, 89-91
 definition of, 84, 89
 randomized, 171-173
 Population multiplier, finite, 101*n.*
 Populations, human, sampling of, 110-
 197
 Post-card checks on fieldwork, 397, 399
 Pound, Ezra, 34*n.*
 Precoding, 408-409
 Prediction, 253-262
 assumption underlying, 209, 211
 of choice, 279
 relation of, to deduction, 33
 as stage of scientific method, 27, 32-34
 usefulness of statements of intention
 in, 256-262
 Preference measurements, 265-288
 Preliminary planning of studies, 330-338
 Prelisting, 376
 Premises, logical, 33
 Presentation of findings, 345-348
 Pretesting, 228, 241-242, 342-343
 Primary sampling unit, 130
Printers' Ink, 335
 Probability, of error (*see* Error, risks of)
 meaning of, 91-93
 theory of, 36
 variable, of selection, 142-147
 Probability proportional to size, 144-147
 Probability sampling, 88, 90
 adaptations of, for marketing research,
 192-197
 comparison of, with nonprobability
 sampling, 185-192

- Probability sampling, controversies connected with, 167
costs of, 189-190
practical difficulties in, 165
types of (*see* Sampling; specific sampling methods)
- Probability sampling designs, selection of, 186
- Probing, 271, 305, 313, 365, 367-370
- Problem, definition of, in marketing research, 31, 330-338
- Problems in marketing, devising alternatives to established policy, 78-79
kinds of, 67-69
recognition of, 67-68
 concerning major policies, 69, 73-79
 involving frequently recurring expenditures, 68-73
use of research in solution of, 79-80
- Processing of data, 406-431
- Profitability, factors influencing, 73-74
- Program Analyser, 295, 297
- Projective techniques (or tests), 286*n.*, 315
- Proportion, calculation of confidence interval for, 108
- Proposition, definition of, 32
standard deviation of its sampling distribution, 100*n.*
- Psychoanalysis, 236-237
- Punch cards, 151
- Punching, multiple, 423-425
- Q
- Qualitative concepts in scientific method, 28
"Qualitative interview," definition of, 312
- Quantification, 27-28
- Question-by-question vs. over-all coding, 415-416
- Questionnaires, establishing rapport in, 250-252
length of, 341, 359
pretesting of, 342-343
self-administered, 221
writing of, 220, 241-243, 270-271, 341-342, 359-364 (Comments relevant to this subject are also scattered throughout Part IV and Chap. 20 of Part V.)
- Questions, contingency, 361
 having emotional impact, handling of, 245-248
 leading, 270-271
 wording of, 240-243, 304-305
- Quota sampling, 173-180
 basic data needed for establishing quotas for, 180
 and cluster sampling, 174*n.*
 effective randomization in, 177
 justification for, 180
 knowledge of size of quota in, 180
 relation of, to stratified random sampling, 174-175, 179-180
 selection of respondents by interviewers in, 173, 176
 typical example of, 176-177
- Quotas, sales, establishing, 69-73
- R
- Radio diary, 222, 224
- Random digits, tables of, 51, 93
 as randomized populations, 171
 use of, 94
- Random error, 53
 in response, 203-204
 (*See also* Sampling error)
- Random numbers, tables of, 51, 93
 as randomized populations, 171
 use of, 94
- Random observations, sequence of, 92-93
- Random sampling (*see* Probability sampling)
- Random sampling process, basic requirement of, 171
 confidence in, in making inferences, 107
- Random selection of markets, example of, 51-52
- Random sequence, 92
- Random walk, 195
- Randomization, 47, 50
 application of, to design of experiments, 50
- Randomness, concept of, 92-95
 effective, 171-172
 general definition of, 171
 tests of, 93

- Range, definition of, 90n.
- Rank order method, 283-284, 285
- Ratio estimate, 157n.
- Readers Guide to Periodical Literature, The*, 335
- Readership, measurement of, 216, 219, 245-246
- Reasons why, 213-214
 - securing information on, 299-316
 - deliberate deception by respondents, 309-311
 - by depth interviewing, 312-314
 - by direct questioning, 299-316
 - vs. indirect methods in, 299-301
 - establishing unambiguous communication, 205-207, 301-308
 - getting accurate answers, 205-207, 308-316
 - getting understanding of questions, 301-305
 - helping respondents to give coherent and complete answers, 305-306
 - inability to give accurate answers, 315-316
 - by indirect methods, 299
 - limitations of statistical analysis of observational data in, 300
 - projective techniques in, 315
 - understanding respondent's answer, 307-308
 - unwillingness to give accurate answers, 243-244, 309-315
- Recall, 225-237
 - aided, 232-234
 - facsimile in, 233
 - method of, 226, 230, 232-237
 - unaided, 234-237
 - aids to memory in, 235-237
 (See also Memory)
- Recognition method, 230
 - dangers in use of, 230-232, 234
 - vs. observation, 231
- Recruiting of interviewers, 378-390
- Redintegration, 229n.
- Refusals in interviews, 163, 357-359
 - partial, 249-252
 - threatened, 357-358
- Regression analysis, 65
- Regression estimate, 157-158
- Reiersøl, Olav, 66
- Relative frequency, 91-92
- Reliability, 84, 203-205
- Reporting to management, 345-348
- Reports, reproduction of, 431
 - typing of, 430
 - writing of, 346-347
- Representative sample, 94-95
- Reproducer, 425
- Research, marketing (see Marketing research)
 - meaning of term, 4
- Research budget, 325
- Research departments, of advertising agencies, 12, 14
 - of distributors, 12-13
 - organization of, 320-325, 344-345
 - of publishing companies, 12, 14-15
- Research personnel (see Personnel)
- Respondents, angry or belligerent, 358-359
 - careless, 249-252
 - finding of, problem of, 375-377
 - handling of (see Interviewing)
 - remuneration of, 221, 223
 - suspicious, 357
- Response, free, 271-272, 305-306, 353
 - recording of (see Interviewing, recording)
- Response errors, 202-207
 - ambiguity, 205-207
 - inaccuracy, 205-207
 - kinds of, 203-205
 - sources of, 205-207
 (See also Nonsampling errors)
- Risk of error, 87-88
- Robinson, Claude, 293n.
- Roper, Elmo, 170, 174
- Rorschach test, 315
- Russell, Bertrand, 225n., 227n.

S

- s, symbol for standard deviation of sample, 91
- Saffir, M. A., 263n.
- Sales analysis, 16, 58, 59, 216
- Sales Management*, 335
- Sales quotas, 11
 - establishing, 69-73
- Sales tests, examples of, 17, 211, 289-290
 - as experiments, 17, 45, 47

- Sales tests, limitations of, 43, 258-259
 observation in, 208
 "out-of-stock" problem in, 54*n.*
 uses of, 77, 289-290, 316
- Sample, definition of, 83-84, 90
 final selection of, 343-344
 of the Labor Force, 133
 nonprobability, 88, 90
 (See also Nonprobability sampling)
 probability, 88, 90
 (See also Probability sampling)
 quota, 173-180
 (See also Quota sampling)
 representative, 94-95
 size of, 181-185
 (See also Sampling)
- Sample census, 145
- Sample estimate, appraising accuracy of, 96-97
- Sample selection, 343-344
- Sample size, determination of, 181-185
 effect of, on risks of error, 53
 when estimates for subgroups are needed, 185
 when several items of information are sought, 184
- Sampling, area (see Area sampling)
 call-backs in, 376-377
 cluster (see Cluster sampling)
 double, 153-155, 231-232, 248
 instructions to interviewers on, 393
 judgment in (see Nonprobability sampling)
 justifications for use of, 87
 multiple, 156
 nonprobability, 88, 90
 (See also Nonprobability sampling)
 place of, in marketing research, 83-85
 probability, 88, 90
 (See also Probability sampling)
 problem of, 83, 84
 problems in, nonresponse, 160-163
 refusals, 163
 quota, 173-180
 (See also Quota sampling)
 sequential, 155-157
 (See also Sequential sampling)
 simple random, 91-95
 (See also Simple random sampling)
 stratified, 134-141
 (See also Stratified sampling)
- Sampling, systematic, 151-153
 (See also Systematic sampling)
 two-phase, 153-155
 two-stage (see Cluster sampling, with subsampling)
 without replacement, 94
 effect of, on standard error of the mean, 100-101
- Sampling design, adaptation of, for marketing research, 192-197
 "all-purpose," 379-382
 change of, effect on risks of error, 53-55
 choice of, accuracy needed in, 188-189
 cost, 189-191
 economic efficiency of, 101-102
 "fixed," 379-382
 outline of, 339-341
 relation of, to selection of interviewers, 378-382
 selection of, 185-192
 statistical efficiency of, 101-102
- Sampling designs, types of probability (see specific sampling methods)
- Sampling distribution, 95-103
 definition of, 96
 illustration of, 95
 of the mean, 96
 in cluster sampling, with subsampling, equal-sized clusters, 125-126
 unequal-sized clusters, 128-129
 without subsampling, equal-sized clusters, 117-118
 unequal-sized clusters, 124
 finite population multiplier, 101*n.*
 form of, 102-103
 in simple random sampling, 95-103
 the mean of, 97-99
 standard deviation of, 99-101
 in stratified sampling, 135-136
 of the proportion, 95*n.*, 100*n.*
- Sampling error, 53, 84, 97, 201
 comparison of, between simple random sampling and cluster sampling, 121
 between simple random sampling and stratified sampling, 137
 sources of, 201
- Sampling interval, 146*n.*
- Sampling ratio, 146
- Sampling theory, 89-109
 basic concepts in, parameter, 90
 population, 89-90

- Sampling theory, basic concepts in, probability, 91-94
 randomness, 91-94
 sample, 90
 statistic, 91
 confidence intervals in, 103-108
 (See also Confidence intervals)
 for estimation of the mean in simple random sampling, 97-108
 nature of, 85, 87
- Sampling unit, primary, 130
 secondary, 130
- Sanborn Insurance Maps, 147
- Savage, L. J., 274
- Scale Analysis, 281-282
- Scaling, 269
 construction of attitude scale, essential features in, 273-274
- Scaling techniques, for ranking attitudes toward one stimulus, Guttman method, 281-282
 Lazarsfeld's latent-structure theory, 282
 Likert method, 280-281
 Method of Summated Ratings, 280-281
 Scale Analysis, 281-282
 Thurstone method of equal-appearing intervals, 273-282
 for ranking stimuli according to attitudes, criterion for choosing between methods, 285
 paired comparisons, 284
 rank order method, 283, 285
 successive intervals method, 283, 285
 vs. single-question technique, 278-279
 usefulness of, in marketing, 279
- Schedule, definition of, 353*n.*
- Scientific method, 25-39
 relation of, to logic, 34
 stages of, 26-27
- Secondary sampling unit, 130
- Secondary sources, use of, 334-335
- Segments, open-country, 132
- Selection, of interviewers, 384-390
 effect of sampling design on, 378-382
 relation of sampling design to, 378-382
 use in, of application blanks, 385-387
- Selection, of interviewers, use in, of employment interview, 387-388
 of personnel tests, 388-390
 of references, 388
 of respondent within household, 239-240
 of simple random sample, 91-95
- Self-observation, 221-224
 accuracy and unambiguity of, 222-224
- Semirefusal, 249-252
- Sentences, definition of, 32
- Sequential sampling, 155-157
 applications of, 156-157
 limitations in use of, 155-156
- Sheatsley, Paul, 250*n.*
- Sigma (σ), symbol for standard deviation of population, 90
- Simmons, Willard, 161*n.*, 162, 163, 190
- Simple random sampling, 91-95
 amount of traveling in, 112
 vs. cluster sampling, relative efficiency of, 120
 cost of reaching people in, 112
 definition of, 50, 93-94
 dependence upon lists of the population in, 111
 determination of sample size in, 181-182
 limitations on use of, 110-112
 sampling distribution of the mean in, 95-103
 vs. stratified sampling, 137
 vs. systematic sampling, 152
- Single-question technique, 266-273, 286, 287
 bias in, due to interviewers, 272
 classification in, problem of, 267-269
 leading questions in, 270-271
 limitations of, 273
- Skewness, 103*n.*
- Snedecor, George, 56*n.*, 93*n.*
- Sorter-counter, 426-427
 IBM, 418
- Sources of County Data*, 131
- Sources, secondary, use of, 334-335
- Split-run technique, 47
- Split-run tests, 290
- Standard deviation, 90*n.*, 99
 computation of, in statistical inference, 107*n.*
- Standard error, of the mean, 99-102, 181-182

- Standard error, of the mean, in cluster sampling, 117, 124, 125-126, 129
 formula for, in simple random sampling, 100
 in stratified sampling, 136
 of the proportion, formula for, in simple random sampling, 100*n*.
 of sample estimate, 157
- Standards of sales performance, 69-73
- Stanton-Lazarsfeld Program Analyser, 295
- Star, Shirley, 242*n*.
- Statistical control, 47, 50
- Statistical inference, risks of error in, 51-52
 control of, 53-55
 role of, in testing of hypotheses, 36-37
- Statistical Research Group, 155*n*.
- Statistic, definition of, 90-91
- Statistics, descriptive, 6, 30
- Store audit as method of measuring retail sales, 115
- Store inventory, 218, 220
- Stouffer, S., 282*n*.
- Strata, 134
 criterion for forming, 137
- Stratification, 134-141
 comparison of, with cluster sampling, 137-138
 criterion for, 137-138
 definition of, 134
 desirability of, 135
 disproportionality in, 147-151
 effect of size of sampling unit on efficiency of, 179-180
 in experimental design, 55
 in nonprobability sampling, 193
 proportionality in, 135
 size of strata in, importance of knowledge of, 139-140
 sources of variation in, 137
 weighting in, 135*n*.
- Stratified sampling, 134-141, 147-151
 comparison of, with cluster sampling, 137-138
 disproportionate (nonproportionate), 147-151
 proportionate, 134-141
 relation of, to quota sampling, 174-175
 sampling distribution of the mean in, 135-136
- Stratified sampling, vs. simple random sampling, 137
 use of, in conjunction with cluster sampling, 138-141
- Stratum, 134
- Study design, 338-345
- Subliminal messages, 315
- Successive intervals, method of, 283, 286
- Summated Ratings, Method of, 280-281
- Supervision of interviewing, 396-405
- Supplements to the First Series Housing Bulletin*, 131
- Survey, of Consumer Finances, 257
 continuing, 17
 "one-time," 16-17
 panel, 17
- Survey data, 16-17
- Survey Research Center, 379, 390
- Syllogism, 33
- Systematic sampling, 141, 151-153
 dangers of, 152-153
 vs. simple random sampling, 152
- ## T
- Tables of random digits, 51, 93
 use of, 94
- Tabulation, 344, 417-429
 hand, 418-420
 machine, 418-419, 420-429
 punching, 421-426
 running, 426-429
 verification, 425-426
 mechanical, 151
- Tabulator, 428
 IBM, 430
- Tamulonis, Valerie, 272*n*.
- Taste testing, 172
- Test situation, 259
- Thematic Apperception Test (T.A.T.), 315
- Thurstone, L. L., 274, 279, 284, 428*n*.
 construction of attitude scales by, 274-278
 method of prediction of choice, 279
- Tribune, Chicago*, Consumer Panel, 15, 292
- Two-phase sampling, 153-155
- Two-stage sampling (*see* Cluster sampling, with subsampling)
- Typing of reports, 430

U

- Unaided recall, 234-267
 - aids to memory in, 235-237
- Unambiguity (*see* Ambiguity)
- Unbiased estimate, 97-99
- Unemployment, estimates of, 86
- Unequal probability of selection, 142-147
- Unit normal deviate, 104*n*.
- U.S. Bureau of the Census (*see* Census, U.S. Bureau of the)
- Universe, definition of, 84, 89
 - randomized, 171-173
- University of Chicago, National Opinion Research Center at, 250, 272, 377*n*, 390
- University of Michigan, Survey Research Center of the, 379, 390
- Unwillingness, basic causes of, 243-244
 - effect of reward or punishment on, 244-245
 - emotional factors affecting, 245-248
 - reluctance to spend time or effort, 249-252
 (*See also* Willingness)

V

- Validation, 186-187
 - definition of, 178
- Validity, 203-205
- Variable probability of selection, 142-147
 - procedure for selection of sample, 144*n*.
 - reasons for, 143

- Variance, analysis of, 65
- Verbatim response, recording of, 371, 409
- Verifier, 425-426
- Verner, Helen, 272*n*.
- Vote splitting, 283

W

- Watson, Alfred, 165
- Wax, Murray, 165*n*., 178*n*., 282*n*.
- Weighting, 194
- Welles, Orson, 238
- Weschler, Irving, 286*n*.
- "Why" questions, 299-316
 - ambiguity in, 301-308
 - (*See also* Reasons why)
- Willingness, definition, 243*n*.
 - to permit observation, 220-221
 - to provide information, 205, 206
 - on attitudes, 286-287
 - on behavior, 217, 243-252
 - on "reasons why," 309-315
 - (*See also* Unwillingness)

X

- $\bar{x}$, symbol for mean of sample, 91

Z

- Zero block, definition of, 163-164

