

A Historical Perspective on the Institutional Bases for Survey Research in the United States

STEPHEN E. FIENBERG and JUDITH M. TANUR¹

ABSTRACT

The basic theme of this paper is that the development of survey methods in the technical sense can only be well understood in the context of the development of the institutions through which survey-taking is done. Thus we consider here survey methods *in the large*, in order to better prepare the reader for consideration of more formal methodological developments in sampling theory in the mathematical statistics sense. After a brief introduction, we give a historical overview of the evolution of institutional and contextual factors in Europe and the United States, up through the early part of the twentieth century, concentrating on governmental activities. We then focus on the emergence of institutional bases for survey research in the United States, primarily in the 1930s and 1940s. In a separate section, we take special note of the role of the U.S. Bureau of the Census in the study of non-sampling errors that was initiated in the 1940s and 1950s. Then, we look at three areas of basic change in survey methodology since 1960.

KEY WORDS: Censuses; Cognitive aspects of survey design; Non-sampling errors; Probability sampling; Survey organizations.

1. INTRODUCTION

The development of survey methods in the technical sense can only be well understood in the context of the development of the institutions through which survey-taking is done. The purpose of this paper is to consider survey methods from this broader perspective in order to better prepare the reader for consideration of more formal methodological developments in sampling theory in the mathematical statistics sense that are described in numerous texts on sampling as well as in Rao and Bellhouse (1990). Although our viewpoint and organization is somewhat new, we have relied heavily on secondary sources which provide detailed expositions alternative to ours. Our paper focuses on the American experiences in the development of survey methodology, but it sketches some background of the much broader social science and institutional settings out of which survey methodology grew.

In the next section we present a very brief historical overview of the evolution of this institutional and contextual background, up through the early part of the twentieth century. We see two broad strands – social research and censuses. We begin with a short synopsis of the early history of European social research, turn to a brief overview of census-taking, especially in the context of the United States, and then take up the role of the International Statistical Congresses in the late nineteenth and early twentieth century in establishing the importance of sampling. Even following these congresses, the possible role of probability in sampling was not broadly understood. Further steps required an institutional base.

In section 3, we focus on the emergence of other U.S. institutional bases for survey research in the 1930s and 1940s. In particular, we note that a missing institutional ingredient was provided by the creation of the U.S. statistical agencies at the beginning of the twentieth century.

¹ Stephen E. Fienberg, College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890, U.S.A.; Judith M. Tanur, State University of New York, Stony Brook, U.S.A.

Then a number of factors, including the depression of the 1930s, the development of probability sampling methodology, and a U.S. federal statistical coordinating function came together to launch the modern era of survey methodology in the U.S. We also review market research and the universities as institutional bases. In section 4, we take special note of the role of the U.S. Bureau of the Census in the study of non-sampling errors that was initiated in the 1940s and 1950s. In section 5, we look at some of the basic changes in survey methodology since 1960, focusing on technological advances, the role of longitudinal surveys, and the recent movement to explore cognitive aspects of surveys.

2. A HISTORICAL OVERVIEW OF THE INSTITUTIONAL BACKGROUND FOR MODERN SURVEY METHODS

2.1 Institutional Bases in Early European Social Research

One set of roots of the U.S. tradition of survey methodology and data collection technology is in early European social research (*cf.* Lecuyer and Oberschall 1978, from whose work we have drawn).

In England that tradition can be traced to the seventeenth century. The research, dubbed political arithmetic, was based on administrative records (especially parish records) and personal observation. It was usually carried out by dedicated individuals, such as John Graunt who published his *Natural and Political Observations Made Upon the Bills of Mortality* in 1662. Until the beginning of the eighteenth century, the parish was the unit of local government and administration, so that it was sensible to use the clergy as informants for many inquiries. With the industrial revolution and the rise of cities, this convenient arrangement broke down, necessitating the institution of house-to-house surveys.

By the 1830s statistical societies were formed in England to investigate social problems. They organized committees, which in turn hired agents to go door-to-door to collect data. Although the statistical societies disbanded when the social problems seemed solved, similar procedures were revived towards the end of the nineteenth century when Booth (1889-1891) sent school attendance officers door-to-door to study London's poor.

In France, where the government was more highly centralized, early social research was carried out by the government. District administrators were used as informants to fill out questionnaires on the demographic and economic conditions of their districts. By the mid-eighteenth century what we might consider an early study of the effects of mass communication was carried out in France. Administrators were instructed to spread rumors of increases in taxes and of military conscription and to report on the reactions of the populace.

In the Napoleonic period following the revolution, the French government established a national office responsible for gathering survey-like data on population, social situation, agriculture, and industry and commerce (Bourguet 1988). While this effort was not fully successful, and while it fell short of census methods as we now understand them, it did set in place an institutional structure. During the nineteenth century, France continued the tradition of government responsibility for statistical functions through reporting of data by prefects and in its Bureau de Statistique. The Napoleonic effort also launched a social science data enterprise in France that explicitly rejected the ideas from the theory of probability as it was then known. The French interest in social statistics affected many scientists, such as Laplace and Quetelet (a Belgian who studied in France under Laplace), who in turn contributed in major ways to the art and science of census-taking, attempting to reintroduce ideas from probability, through the use of what we now know as ratio estimation (see Stigler 1986, Chapter 5). After

the revolution of 1830, the Académie des Sciences Morales et Politiques sponsored prize competitions that encouraged statisticians to undertake their own research.

In Germany, the origin of “statistics” (collection of data on the state) was in the universities as early as the end of the seventeenth century. By the early nineteenth century this work was split into three parts, with descriptive political science and historical/quantitative political economy remaining university-based but statisticians collecting data in census bureaus and other government agencies.

In 1872, the Verein for Socialpolitik was founded – part pressure group, part professional organization, part research organization. It drew up questionnaires to be answered by supposedly knowledgeable informants such as landowners, ministers, and notaries. Problems of informants’ possibly inaccurate information, haphazardly grouped and imprecise questions, and low response rates dogged these efforts. By the early twentieth century, Levenstein (1912) published what was probably the first large scale attitude and opinion survey, for which he used a snowball sampling technique. At about the same time Max Weber attempted a survey of industrial workers, planning to get some information directly from respondents but finding that the majority did not care to cooperate.

2.2 Censuses: A Prelude to Survey-Taking

Another set of roots of survey methodology is intertwined with the history of methods for census-taking and thus we present a brief overview on censuses and census-taking infrastructures. Many others have observed that the origins of the modern census are found in biblical censuses described in the Old Testament (Madansky 1986) as well as in censuses carried out by the the ancient Egyptians, Greeks, Japanese, Persians, and Romans (Taeuber 1978). The emphasis in the biblical accounts of censuses seemed to be on the results of the enumeration, rather than on how the counting was done, although in several instances we are told about the rapidity of the process. For most practical purposes we can skip from biblical times to the end of the eighteenth century and the initiation of census activities the United States of America, although there is some debate as whether Canada, Sweden, or the United States should be credited with originating the modern census (Willcox 1930).

In the United States, the first census was taken in 1790 (in 1990 the U.S. government will take its bicentennial census) by State officials who were then reimbursed by the Federal government. Then, in the next census of 1800, the enumerators were deputies or assistants to Federal marshals (Duncan and Shelton 1978). It was only with the 1880 census that the central Census Office gained control over field operations and secured the authority to appoint enumerators.

Prior to 1850 the U.S. decennial census considered the family as the unit of interest and reported few data on persons. The change to an individual-focus in census-taking was strongly influenced by the work Lemuel Shattuck, one of the of ASA’s founders who had earlier conducted the Boston census of 1845 (Anderson 1988, pp. 36-37), as well as that of Quetelet, who helped to organize the 1846 Belgian census (Willcox 1930).

Progress on the methodology of census-taking continued, as every 10 years, a special operation was mounted to fulfill the constitutional obligation of an enumeration of the U.S. population; however, there was a clear lack of continuity from one census to the next (American Economic Association 1899). It was only after the first 12 censuses had been taken that the Bureau of the Census was created in 1902 as a permanent agency. Over this period there was a steady expansion of the number of censuses of other sorts and the broadening of topics covered in addition to simple enumeration.

2.3 International Statistical Congresses

The move from censuses to sample surveys was slow and laborious. Kruskal and Mosteller (1980) trace some of this movement, especially as it was reflected in the discussions regarding surveys that took place at the meetings of the International Statistical Institute (ISI), and our exposition here owes much to their work. The groundwork for these meetings was laid in the 1850s by Quetelet who helped to organize the first of a series of International Statistical Congresses in 1853. After nine such Congresses from 1853 to 1876, the ISI was founded in 1885. It is interesting to note that there is only one index entry for sample surveys in Stigler's (1986) history of statistics before 1900 – to 1830s work of Quetelet linked to a census method suggested by Laplace – and only two index entries in Porter's (1986) history – one to a 1900 paper by Karl Pearson and the other to the work of Kiaer and the ISI.

As early as the 1895 ISI meeting, Kiaer (1895-1896) argued for a "representative method" or "partial investigation", in which the investigator would first choose districts, cities, *etc.*, and then units (individuals) within those primary choices. The choosing at each level was to be done purposively, with an eye to the inclusion of all types of units. That coverage tenet, together with the large sample sizes recommended at all levels of sampling, was what was judged to make the selection representative.

The idea of less than a complete enumeration was widely opposed, but Kiaer presented arguments for it (with some members agreeing and others disagreeing) at ISI meetings in 1897, 1901, and 1903. Towards the end of this period, the idea of probability sampling entered the discussion, but the topic of the representative method seems absent from the records of the ISI meetings until 1925. By then the record suggests that the representative method was taken for granted, and the discussions centered around how to accomplish representativeness and how to measure the precision of sample-based estimates (Bowley 1926; Jensen 1926). Notions of clustering and stratification were put forward, but purposive sampling was still the method of choice.

It was not until Gini and Galvani made a purposive choice of which returns of an Italian census to preserve and found that districts chosen to represent the country's average on seven variables were, in that sense, unrepresentative on other variables, that purposive sampling was definitively discredited (Gini 1928; Gini and Galvani 1929). Soon thereafter Neyman published his groundbreaking 1934 paper that demonstrated, among other thing, the virtues of probability sampling.

3. THE DEVELOPMENT OF INSTITUTIONAL BASES FOR SURVEY RESEARCH IN THE UNITED STATES

Survey research in the United States grew from a blending of the same three institutional bases that had been influential in Europe – private individuals acting as entrepreneurs in the private sector, universities, and the government. Early social research in this country (before World War I) seems to have followed the earlier British model, being carried out by social workers, public health workers, and reformers. An early university involvement was the hiring by the University of Pennsylvania in 1899 of W.E.B. DuBois to carry out his study of the Philadelphia Negro, conducted as a house-to-house survey. Starting in the 1930's, and especially in the period after World War II, the U.S. experienced a flowering of survey methodology in the three broad institutional bases: market research and polling, universities, and government. But before we describe that flowering we shall take a step backwards and note the establishment of the U.S. government statistical agencies.

Recently, Jean Converse (1987) has written an extremely scholarly and graceful study of the roots and emergence of survey research in the United States, with special focus on market research and polling and on universities. Our exposition on these bases closely follow hers. We have separated out the institutional bases both to reflect a social reality and to structure our exposition. But there is another social reality that we ask the reader to bear in mind; the membranes separating the institutions are permeable. They not only permit the flow of cross-fertilizing ideas and methods in all directions; they also permit a somewhat lesser flow of people, as individuals move from one sector to another over the course of their careers.

3.1 The Establishment of U.S. Statistical Agencies

The establishment of American statistical agencies effectively began in 1863, when the newly created Department of Agriculture released the first crop and livestock report to provide information on Union food supplies during the Civil War. This report was based on data from a purposive sample of 2,000 farmers in 22 states. This agricultural statistical reporting activity has existed in the Department of Agriculture on a continuing basis to the present day and is now known as the National Agricultural Statistical Service. By the late 1920s, correlational and regression methodology was well established in the work of agricultural statisticians (Duncan and Shelton 1978).

In 1884, Congress voted to establish a Bureau of Labor (later renamed the Bureau of Labor Statistics, BLS) to “collect information” on the earnings and the working conditions of “laboring men and women.” Under the leadership of Carroll Wright, the first Commissioner, BLS expanded its statistical activities to cover such issues as depressions, strikes and lockouts, women’s wages, marriage and divorce, and the domestic liquor trade (Norwood and Early 1984).

With the creation of the Bureau of the Census in 1902, there were three major U.S. agencies in place, each with a mandate to collect national data on a regular basis. During the first three decades of the twentieth century, the role of government statistical agencies expanded considerably and, at the time of the stock market crash of October 29, 1929, data on various facets of economic and social life were available. As late as 1932, however, there were few examples of probability sampling anywhere in the Federal Government (Duncan and Shelton 1978).

Difficult though it is to conceive in a period when we are used to receiving reliable readings on the unemployment rate monthly, there was no comparable survey data resource available in the 1920s and early 1930s. Except for selected monthly non-survey data gathered by BLS from most manufacturing industries and some nonmanufacturing industries, there were no regular national unemployment figures. In the 1920 census the question on unemployment was dropped because of statistical concerns regarding the accuracy of the resulting data. This question was restored to the 1930 census because of the wide-spread concerns regarding the employment situation. The extensive controversy that surrounded the 1930 unemployment data (Van Kleeck 1930) and those from the special January 1931 Unemployment Census was especially acrimonious (Anderson 1988), and played a role in the 1932 presidential election campaign.

3.2 The ASA-SSRC Committee and the Institutionalization of Probability Sampling: An Early Bridge

Thus, at the beginning of the Great Depression of the 1930s in the United States, the federal statistical agencies had difficulty responding to the demand for statistics to monitor the effects of the programs of President Franklin Roosevelt’s New Deal. In 1933, Secretary of Labor Frances Perkins asked Stuart A. Rice, the ASA president, to set up an advisory committee on

the programs of BLS. This committee grew into the Committee on Government Statistics and Information Services (COGSIS), sponsored jointly by ASA and the Social Science Research Council (SSRC). Duncan and Shelton (1978) give a detailed account of the activities of COGSIS, and for our discussion here two outcomes are worthy of note.

First, in 1933, COGSIS recommended the creation of a Central Statistics Board (CSB) to help coordinate government statistical activities. With the groundwork laid for a coordinated federal statistical system, COGSIS and CSB proceeded, in early 1934, to arrange for an interagency agreement through which Census would collect basic data on production and labor for BLS.

Second, COGSIS helped to stimulate the use of probability sampling methods in various parts of the Federal government, and it encouraged research on sampling theory, to be done by employees of statistical agencies. For example, to establish a technical basis for unemployment estimates, COGSIS and CSB organized an experimental Trial Census of Unemployment as a Civil Works Administration project in three cities using probability sampling, carried out in late 1933 and early 1934. The positive results from this study and the interagency arrangement mentioned above led in 1940 to the first large-scale, ongoing sample survey on employment and unemployment using probability sampling methods. This survey later became the Current Population Survey.

Another somewhat indirect outcome of the COGSIS emphasis on probability sampling took place at the Department of Agriculture Graduate School where W. Edwards Deming organized a series of lectures in 1937 on sampling and other statistical methods by Jerzy Neyman (1938). These lectures had a profound impact on the further development of sampling theory across the government as well as in universities.

What we see happening in this period is the confluence of a number of factors that served to launch the use and development of sampling methods in the U.S. government statistical agencies. A key prerequisite was the existence of the agencies themselves. A second was the methodological advances in sampling theory as encapsulated in Neyman's landmark 1934 paper. What was required to bring these together was the Great Depression, a new administration hungry for quality data to assess the impact of its social programs, and the joint ASA-SSRC Committee on Government Statistics and Information Services.

3.3 Market Research and Polling

The institutional base of survey methodology in U.S. market research and polling traces its own pre-history to election straw votes collected by newspapers, dating back at least to the beginning of the nineteenth century. Often publicity and circulation boosting were more important than accuracy of prediction. Converse (1987) points out, however, a more serious journalistic base; election polls were taken and published by such reputable magazines as the *Literary Digest* (which had gained a reputation for accuracy before the 1936 fiasco). Then, as now, election forecasting was taken as the acid test of survey validity. A reputation for accuracy in "calling" elections was thought to spill over to a presumption of accuracy in other, less verifiable areas.

There was a parallel tradition in market research, dating back to just before the turn of the century, attempting to measure consumers' product preferences and the effectiveness of advertising. It was seen as only a short step from measuring the opinions of potential consumers about products to measuring the opinions of the general public about other objects, either material or conceptual. By the mid 1930s there were several well established market research firms. Many of them conducted election polls in 1936 and achieved much greater accuracy than did the *Literary Digest*. It was the principals of these firms (e.g., Archibald Crossley, George Gallup, and Elmo Roper) who put polling – election, public opinion, and consumer – on the map in the immediate pre-World War II period.

Data collection technology developed broadly in the market research and polling organizations in this era. Sampling was either by purposively selected groups or by quota. Samples were large, with the size enlarged sequentially until the law of large numbers caused the mean or percentage being estimated to stabilize. Some questionnaires were very informal with the interviewer instructed to bring certain topics into a conversation – what we might now call an unstructured interview. Others were more standardized, but shorter, actual forms. The progression seems to have been that as interviewers became more distanced from the primary investigators – in space, in education, in training, in identification with the research project, and perhaps in their very numerosity – the interview became more standardized.

The same kinds of validity issues that interest survey researchers today surfaced in the period. What should be the balance between open and closed questions? (Practice seems to have favored a combination; the device of the “opinion thermometer” to calibrate answers was first developed by the *Literary Digest* in 1925.) The pollsters tackled the problem of how to ask sensitive questions – about age, income, occupation, and home owning – by providing check lists, functioning much like contemporary visual aids. Experiments in question wording were carried out in the polling houses.

Market research in this early period, as now, of necessity put a premium on the timeliness of results. Then, as now, this tended to create some tension between academics and market researchers, with academics believing commercial workers to be corrupted by money and thus too far from basic science and commercial workers believing academics were overly concerned with the abstract. It is noteworthy, however, that one of the earliest homes of public opinion and market surveys was the Psychological Corporation, an organization of academic psychologists committed to plowing part of their profits back into the research process. The Psychological Corporation carried out its surveys from its Market Surveys Division, organized and run by Henry C. Link.

3.4 The Universities

But the universities were hardly totally above the polling movement. As early as 1911 the Harvard Graduate School of Business established a Bureau of Business Research to carry out consumer research. Such household names of social science as Paul Lazarsfeld, Hadley Cantril, and Rensis Likert moved to university affiliations and attached research institutes. Lazarsfeld came to the United States in 1933 determined to bring the techniques developed in market research to the basic scientific endeavor. He went on to form the Office of Radio Research, later to be called the Bureau of Applied Social Research, at Columbia University. His myriad contributions included the use of panels and a system of causal analysis.

Hadley Cantril was an academic who early on collaborated with Lazarsfeld on research on radio listening. When the two had a falling out, Cantril established the Office of Public Opinion Research at Princeton University. Here studies were carried out to improve data collection techniques. For example, in investigating the effects of question wording, Rugg and Cantril (1944) found that in 1940 – 41 over a six-week period, the percentage of Americans who favored “giving aid [to Great Britain] even at the risk of war” varied between 56% and 78%. At the same time, the percent in favor of “entering the war immediately” ranged from 8% to 22%.

Rensis Likert started out teaching at New York University and with a connection to the surveys of the Psychological Corporation. Moving to business, he carried out a survey of life insurance agents’ attitudes, comparing qualitative and quantitative (mostly questionnaire) methods. He then became Director of the Division of Program Surveys at the Department of Agriculture. There he worked to standardize questionnaires. When Likert left the Department of Agriculture after World War II, he brought his group to the University of Michigan to form the Survey Research Center.

4. FROM SAMPLING THEORY TO THE STUDY OF NON-SAMPLING ERROR

As we have seen above, the introduction of probability sampling into government surveys in the mid-1930s came at the time of rapid development in many areas of statistics, and the development of a foundation for experimentation and inference more broadly under the leadership of such statisticians as R.A. Fisher, Walter Shewart, Jerzy Neyman, and Egon Pearson. Among those who worked on the probability-sampling-based trial Census of Unemployment at the Bureau of the Census were Calvert Dedrick, Morris Hansen, Samuel Stouffer, and Frederick Stephan (Anderson 1988; Duncan and Shelton 1978). Hansen was then assigned with a few others to explore the field of sampling for other possible uses at the Bureau, and went on to work on the 1937 sample Unemployment Census. After working on the sample component of the 1940 decennial census (under the direction of Deming), Hansen worked with others (e.g., Jerome Cornfield, Lester Frankel, William Hurwitz and J. Steven Stock) to redesign the unemployment survey based on new ideas on multi-stage probability samples and cluster sampling (Hansen and Hurwitz 1942, 1943). They expanded and applied their approach in various Bureau surveys, often in collaboration and interaction with others, and this effort culminated in 1953 with the publication of a two-volume compendium of theory and methodology (Hansen, Hurwitz and Madow 1953). The recent interview with Hansen (Olkin 1987) and the Duncan and Shelton (1978) volume provide interesting and detailed descriptions of the developments during this period.

Virtually independent and often complementary contributions to sampling theory came via the statistical sampling work in agriculture by P.C. Mahalanobis and students in India and by Frank Yates and William Cochran in England. Cochran's 1939 paper is especially notable because of its use of the analysis of variance in sampling settings and the introduction of superpopulation and modeling approaches to the analysis of survey data (see Fienberg and Tanur 1987, 1988 for related discussion on the design and analysis linkages between sampling and experimentation). In the 1940s, as results from these two separate schools appeared in various statistical journals, we see some convergence of ideas and results.

The 1940s saw a rapid spread of probability sampling methods to other government agencies. It was only after the fiasco of the 1948 presidential pre-election poll predictions (Mosteller *et al.* 1949) that market research firms and others shifted towards probability sampling. Even today many organizations use a version of probability sampling with quotas (Sudman 1987).

Amidst the flurry of activity on the theory and practice of probability sampling during the 1940s, attention was also being focused on issues of nonresponse and other forms of non-sampling error. In a review of work on errors in surveys, Deming (1944) listed 13 factors affecting the ultimate usefulness of surveys (note that most of these are nonsampling errors):

1. variability in response;
2. differences between different kinds and degrees of canvass;
3. bias and variation arising from the interviewer;
4. bias of the auspices;
5. imperfections in the design of the questionnaire and tabulation plans;
6. changes that take place in the universe before tabulations are available;
7. bias arising from nonresponse (including omissions);
8. bias arising from late reports;
9. bias arising from an unrepresentative selection of date for the survey, or of the period covered;

10. bias arising from an unrepresentative selection of respondents;
11. sampling errors and biases;
12. processing errors (coding, editing, calculating, tabulating, tallying, *etc.*);
13. errors in interpretation.

Most of the errors described in this list either had been or would become the focus of research by statisticians at the Bureau of the Census.

A milestone in this effort to understand and model non-response errors was the development of an integrated model for sampling and non-sampling error in censuses and surveys, in connection with planning for and evaluation of the 1950 census (Hansen, Hurwitz, Marks and Mauldin 1951). This analysis-of-variance-like model, or variants of it, has served as the basis of much of the work on non-sampling error over the past 35 years, both inside and outside the Bureau of the Census. An excellent qualitative analysis of the error structure of the Current Population Survey is given in Brooks and Bailer (1978), and reviews of the non-sampling error literature are given by Mosteller (1978) and Fienberg and Tanur (1983). Finally, we note that Groves' (1989) recent book gives an updated approach to a variant of this census model, making a careful distinction between random and fixed components that arise from the various sources of error.

The paper by Bailer (1990) in this issue contains a detailed discussion on non-sampling error from the perspective of the Bureau of the Census.

5. CHANGING DIMENSIONS OF SURVEY METHODOLOGY AFTER 1960

The decades of the 1960s and 1970s saw polls and surveys becoming an all-pervasive fact of American life, beginning with the hard-fought presidential election of 1960 in which both candidates (Kennedy and Nixon) commissioned and relied on private polls of the electorate. Here we focus on three major areas of innovation during recent decades. We refer the reader to other presentations for such important topics as imputation for incomplete data and the ever-present controversies surrounding inferences from survey data (*e.g.*, see Fienberg and Tanur 1983, 1986).

5.1 Mode of Interviewing: The Role of Telephones and Computers in Surveys

The development and diffusion of technology, especially telephones and computers, strongly influenced survey practice in these decades. U.S. telephone coverage, which was estimated to have been only 35% in 1936 and hence contributed to the *Literary Digest's* problem (Massey 1988), reached 75% by 1960 and 88% in 1970 on its way to around 93% in 1986 (Thornberry and Massey 1988). Thus telephone surveys, often based on random digit dialing (RDD) techniques, became increasingly prevalent and accurate. The movement began among commercial survey researchers, with governmental and academics lagging behind because of their concerns over differential coverage by such variables as income and race (Trewin and Lee 1988) and accompanying fears of lack of "representativeness". Indeed, most government uses of telephone interviewing remain as follow-ups of initial in-person contacts (as in the Current Population Survey which has been using telephone interviewing for households in later months of the survey since 1954). Only recently has there been a marked shift towards the use of RDD for government surveys. Groves and Kahn (1979) provide a review of work on telephone interviewing and, by and large, they document the comparability of survey results through comparisons of data gathered by personal interviews and by telephone.

The advent and proliferation of the computer meant that the tasks of analyses of survey responses could be carried out much more rapidly and broadly than ever before. This led to an increase in the number of surveys carried out under all institutional auspices. In retrospect it seems only natural that computer technology should be combined with telephone technology to produce systems of computer assisted telephone interviewing (CATI). These systems provide automated questionnaires that carry out skip patterns and display the appropriate question on a monitor screen, schedule (and often actually place) calls and callbacks, carry out randomizations, and automate data entry, in addition to other functions. CATI systems were developed by U.S. market research organizations in the early 1970s in part to keep track of respondent characteristics and thus ensure that quotas are precisely and efficiently met (Nicholls 1988). Chilton Research was one of the commercial CATI pioneers, using a CATI system for surveys intended to determine the level of customer satisfaction with services provided by the telephone companies (Nicholls and Groves 1986). Largely independently, university survey organizations began to develop their own CATI systems in the mid-1970s, and introduced them to the larger statistical community with an emphasis on their usefulness for documentation, standardization, and interviewer flexibility. While government agencies exhibited early interest in CATI, they have only recently begun to actually employ systems, sometimes on an experimental basis and often in tandem with other data collection methodologies, as in panel designs where the first interview is carried out in person. At this writing we see the beginnings of a movement to the use of computer-assisted personal interviewing (CAPI), a development made possible by the technological advances that produced truly portable laptop computers.

5.2 Longitudinal Surveys

While panel surveys were conducted in connection with the 1924 and 1940 U.S. presidential election campaigns (Rice 1928; Lazarsfeld *et al.* 1944), interest in over-time survey data did not really become fashionable in social research until the 1960s. This is all the more surprising when we realize that the Current Population Survey has traditionally had a rotating-panel structure and, since 1953, many respondents are interviewed as many as 8 times over a 16 month period. This rotating-panel structure was originally intended to produce estimates of change in aggregate quantities that had smaller variances than those from repeated cross-sections but, in principle, the CPS could have been analyzed in panel form on a regular basis. The fact that the CPS is a survey of sample addresses and not individuals or households is a major obstacle to the use of it as a panel survey (see related comments on the National Crime Survey in Fienberg 1978), but this has not prevented the elaborate use of the CPS to study gross flows in individual employment status (*e.g.*, see Abowd and Zellner 1985 and Stasny 1988).

Not all survey attempts to measure change need be based on longitudinal data; often repeated cross-sections can do at least as well if not better in measuring aggregate change. By the 1970s the Gallup Poll and others had developed a tradition of asking the same questions repeatedly and reporting the results in newspapers. These established time series became incorporated into the burgeoning Social Indicators movement. In 1972 the National Opinion Research Center first fielded the General Social Survey (GSS), funded by the National Science Foundation. GSS was designed by a broadly based group of academics to provide periodic readings on social indicators and to provide an original data set for use by students and academics doing modestly funded research. For purposes of continuity, the designers incorporated into GSS many questions first developed by Gallup and other commercial pollsters, yielding a fruitful cross-institutional collaboration (*e.g.*, see Smith 1975).

The basic idea behind the conduct of longitudinal surveys of panels, however, is to measure changes over time, not by comparing the changes in aggregate quantities, but by focusing on individual change. Such surveys typically focus on changes in status, the duration of activities, and events occurring over time. The rise of interest in longitudinal panel surveys occurred primarily outside the government, and early examples are the Panel Study of Income Dynamics, conducted by the Institute for Social Research at the University of Michigan annually since 1968; the National Longitudinal Surveys of Labor Market Experience, sponsored by the Center for Human Resources Research at Ohio State University beginning in 1966, and currently funded by BLS; and the Longitudinal Retirement History Survey, sponsored by the Social Security Administration from 1969 to 1979. The 1970s saw expanded use of longitudinal panel surveys, especially under government auspices (e.g., see Boruch and Pearson 1988), but the basic survey methodology used often resembled that for traditional cross-sectional surveys. Only in the late 1970s did researchers begin to question the conventional wisdom about longitudinal survey design and analysis and to explore such fundamental issues as the definition of a longitudinal family (for a discussion, see Fienberg and Tanur 1986).

In the 1980s, interest in longitudinal panel surveys expanded and considerable attention was focused on aspects of non-sampling error such as attrition and on issues of data management and analysis. Kalton *et al.* (1989) includes a number of papers on these topics.

5.3 Cognitive Aspects of Surveys

As a result of systematic efforts to improve survey methodology over the past forty years, survey researchers have evolved a highly developed art of questionnaire design and interview procedures to reduce nonsampling errors, such as those described in Deming's list above (e.g., see Payne 1951), and they have carried out many scientific studies to test aspects of that art (e.g., see Sudman and Bradburn 1974, Bradburn and Sudman 1979, and Schuman and Presser 1981). Until recently, however, research on understanding the survey interview situation has been relatively unsystematic. The recent change came, in part, through the recognition that other fields, in particular cognitive psychology, had insights that would assist survey researchers in examining the interview process.

Among non-sampling errors are those occasioned by the cognitive processes that respondents and interviewers are required to exercise in the survey interview situation. Respondents must often recall events and make judgments or estimates, and must always face issues of comprehension of the questions asked – their meaning to respondents as well as their meaning to interviewers. Survey researchers are now beginning to draw on the concepts of cognitive psychology and the expertise of cognitive psychologists to investigate more systematically these issues of non-sampling error. We note especially that the exploration of meaning is not new to the enterprise of survey research. Indeed, Cantril (1944) devotes two chapters to reporting the results of experiments on the meaning and wording of questions. These experiments used many of the same probing and paraphrasing techniques used in today's cognitive laboratory.

This explicit movement to study cognitive aspects of surveys originated in a 1981 conference sponsored by the Bureau of Social Science Research and the Bureau of Justice Statistics that brought together cognitive psychologists and survey researchers to concentrate on the National Crime Survey. A more intensive 1983 conference, sponsored by the Committee on National Statistics (CNSTAT) of the National Research Council, concentrated on the National Health Interview Survey (Jabine *et al.* 1984). From the beginning the movement was, by design, a partnership between people from academia, from research institutes and other academic institutions, and from the government.

A direct outgrowth of the CNSTAT conference was the establishment of a Questionnaire Design Research Laboratory at the U.S. National Center for Health Statistics under the leadership of Monroe Sirken to do pretesting (in parallel with full scale field testing) of major government surveys. It employs government personnel, brings in visiting scholars, and contracts with academics and people in research institutes to carry out its mission. This has been followed by the establishment of similar laboratories at the Bureau of Labor Statistics and the Bureau of the Census. Another outgrowth is the establishment of the Social Science Research Council's Committee on Cognition and Survey Research, which is, itself, both cross disciplinary and cross institutional. The Committee has fostered research in such directions as the interactive process of the survey interview, the uses and pitfalls of retrospective memory, and issues in measuring pain in a survey context. Examples of other outgrowths of this movement are (a) an investigation by the OECD's Working Party on Labor Statistics of cognitive aspects of labor surveys, addressing such issues as the meaning of "looking for work" – a knotty conceptual problem within a culture, and even more problematic across cultures (Schwarz 1987), (b) work at combining the cognitive perspective with statistical work on the embedding of experiments within surveys (Fienberg and Tanur 1989), (c) international conferences on work at the interface of cognition and survey methods (*e.g.*, see Hippler, Schwarz and Sudman 1987).

At the same time that methodological techniques of the cognitive laboratory are being used to shape questionnaire design, findings from the cognitive psychology laboratory are being taken into the field in order to test their generalizability and thus enrich the academic field of cognitive psychology, as well as to ascertain their usefulness for the survey enterprise. Here is yet another instance of interaction between the academic world and the government. For example, a laboratory finding is that people recall visits to health care providers more easily and accurately if they begin with the earliest first (Fathi, Schooler and Loftus 1984). A recent investigation explores whether this advantage holds in the field situation of the pre-test of the NHIS (White and Berk 1987).

The movement to integrate methods from the cognitive sciences into the design of sample surveys is important for several reasons. First, it has brought a renewed scientific base to the problems of questionnaire design. Second, it has opened up the survey domain to the study of selected cognitive phenomena. But most important, it had brought new vigor to the survey enterprise and raised anew issues about the structure and format of the survey interview, going far beyond questionnaire design, that many statisticians thought were resolved in the 1940s and 1950s.

6. COMMENTS

Traditional reviews of the history of survey methods have focused on the role of probability sampling and its refinements, and occasionally on the study of non-sampling errors. Here we have attempted to set this methodological history in the context of the tradition of social science research that evolved over the nineteenth and early twentieth centuries and the institutions, in and outside of government, that facilitated and occasionally directly spawned the methodological developments. This perspective should help remind readers that factors other than the advance of statistical theory have helped to shape the survey domain as we know it today. It should also help them follow the evolution of survey theory and practice as it continues to be shaped by institutional change.

There is an additional facet of institutional shaping of the survey enterprise that we have not addressed heretofore. We wrote above about the permeability of the membranes separating the three sectors: government, market research (the private domain), and the universities and

other academic institutions. We believe that these membranes are becoming even more permeable with the increased presence of a fourth kind of institution, which we shall refer to as a "bridge". We saw earlier how the ASA-SSRC Committee on Government Statistics and Information Services, a bridge between academia and government, prepared the ground for federal statistical coordination. ASA and SSRC continue to provide bridging functions, but other such institutions also exist.

Some vivid examples of other bridges come to mind. For over 40 years the American Association for Public Opinion Research has been bringing together survey practitioners from all sectors in local chapters and in national conferences at which new findings are disseminated and issues of common concern are discussed. The National Science Foundation program on Measurement Methods and Data Improvement (MMDI), under the direction of Murray Aborn, has explicitly seen as part of its mandate the fostering of government/academic collaboration. The mission has been implemented, for example, through the funding of research by academics that both uses and improves government databases (the 1983 seminar on cognitive aspects of survey methodology was sponsored by MMDI) and the funding of an ASA-sponsored fellowship program. That fellowship program places academic researchers for a semester or a year in government statistical agencies to carry out their own research, bring new ideas to the agency, and return to their academic bases with new knowledge and contacts in the federal agencies and new awareness of government data bases and statistical concerns. The National Research Council, an arm of the National Academy of Sciences, maintains a Committee on National Statistics that brings statisticians from academia and the private sector together to interact with representatives of the government agencies. Here, in formal panel studies and informal interaction, individuals come to know one another and common problems are tackled.

While these and other bridges will surely not totally erase the boundaries between the sectors, we see their existence as a positive force for progress in the development of survey methodology. Developments in one sector move more quickly to others across these bridges, but perhaps more important, the bridges facilitate a process whereby problems faced by any sector become legitimate research questions in all sectors.

ACKNOWLEDGEMENTS

The preparation of this paper was supported in part by the National Science Foundation under Grant No. SES-8701606 to Carnegie Mellon University and Grant No. SES-8701816 to the State University of New York at Stony Brook. An earlier version appeared under the title "Some History of Survey Methods and Data Collection Technology," in the *Sesquicentennial Invited Paper Sessions* volume of the 1989 Proceedings of the American Statistical Association, 393-405.

REFERENCES

- ABOWD, J.M., and ZELLNER, A. (1985). Estimating gross labor force flows. *Journal of Business and Economic Statistics*, 3, 254-283.
- ANDERSON, M.J. (1988). *The American Census. A Social History*. New Haven: Yale University Press.
- AMERICAN ECONOMIC ASSOCIATION (1899). The Federal Census. Report of the committee on the Twelfth Census. *Publications of the American Economic Association*, New Series, No. 2, 1-7.

- BAILAR, B.A. (1990). Contributions to statistical methodology from the federal government. *Survey Methodology*, 16, 51-61.
- BOOTH, C. *et al.* (1889-1891) 1902-1903. *Life and Labours of the People in London*. London: Macmillan.
- BORUCH, R.F., and PEARSON, R.W. (1988). Assessing the quality of longitudinal surveys. *Evaluation Review*, 12, 3-58.
- BOURGUET, M.-N. (1988). Décrire, Compter, Calculer: The debate over statistics during the Napoleonic Period. In *The Probabilistic Revolution. Volume 1, Ideas in History* (Eds. L. Kruger, L.J. Daston and M. Heidelberger). Cambridge: MIT Press.
- BOWLEY, A.L. (1926). Measurement of the precision attained in sampling. *Bulletin of the International Statistical Institute*, 22, 6-62.
- BRADBURN, N.M., SUDMAN, S., and Associates (1979). *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- BROOKS, C.A., and BAILAR, B.A. (1978). *An Error Profile: Employment as Measured by the Current Population Survey*. Statistical Working Paper 3, Office of Federal Statistical Policy and Standards. Washington: U.S. Department of Commerce.
- CANTRIL, H. (1944) (1947). *Gauging Public Opinion*, Princeton: Princeton University Press.
- COCHRAN, W.G. (1939). The use of the analysis of variance in enumeration by sampling. *Journal of the American Statistical Association*, 34, 492-510.
- CONVERSE, J.M. (1987). *Survey Research in the United States: Roots and Emergence 1890-1960*. Berkeley: University of California Press.
- DEMING, W.E. (1944). On errors in surveys. *American Sociological Review*, 19, 359-369.
- DUBOIS, W.E.B. (1899) (1973). *The Philadelphia Negro: A Social Study; Together With a Special Report on Domestic Service by Isabel Eaton*. Millwood, N.Y.: Kraus Reprint.
- DUNCAN, J.W., and SHELTON, W.C. (1978). *Revolution in United States Government Statistics, 1926-1976*. U.S. Department of Commerce. Washington: U.S. Government Printing Office.
- FATHI, D., SCHOOLER, J., and LOFTUS, E. (1984). Moving survey problems into the cognitive psychology laboratory. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 19-21.
- FIENBERG, S.E. (1978). Victimization and the National Crime Survey: Problems of design and analysis. In *Survey Sampling and Measurement* (Ed. K. Namboodiri). New York: Academic.
- FIENBERG, S.E., and TANUR, J.M. (1983). Large scale social surveys: Perspectives, problems, and prospects. *Behavioral Science*, 28, 135-153.
- FIENBERG, S.E., and TANUR, J.M. (1986). The design and analysis of longitudinal surveys: Controversies and issues of cost and continuity. In *Survey Research Designs: Towards a Better Understanding of Their Costs and Benefits* (Eds. R.W. Pearson and R.F. Boruch). New York: Springer-Verlag.
- FIENBERG, S.E., and TANUR, J.M. (1987). Experimental and sampling structures: Parallels diverging and meeting. *Canadian Journal of Statistics*, 55, 75-96.
- FIENBERG, S.E., and TANUR, J.M. (1988). From the inside out and the outside in: Combining experimental and sampling structures. *Canadian Journal of Statistics*, 16, 135-151.
- FIENBERG, S.E., and TANUR, J.M. (1989). Combining cognitive and statistical approaches to survey design. *Science*, 242, 1017-1022.
- GINI, C. (1928). Une application de la méthode représentative aux matériaux du dernier recensement de la population Italienne (1er décembre 1921). *Bulletin of the International Statistical Institute*, 23, 198-215.
- GINI, C., and GALVANI, L. (1929). Di una applicazione del metodo rappresentativo all'ultimo censimento Italiano della popolazione (1 dicembre 1921) (In Italian). *Annali di Statistica*, Series 6, 4, 1-107.

- GRAUNT, J. (1662) (1939). *Natural and Political Observations Made Upon the Bills of Mortality* (Edited and with an introduction by Walter F. Willcox). Baltimore: Johns Hopkins Press.
- GROVES, R.M., BIEMER, P.P., LYBERG, L.E., MASSEY, J.T., NICHOLLS, W.L. II, and WAKSBERG, J., eds. (1988). *Telephone Survey Methods*. New York: Wiley.
- GROVES, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.
- GROVES, R.M., and KAHN, R.L. (1979). *Surveys by Telephone*. New York: Academic Press.
- HANSEN, M.H., and HURWITZ, W.N. (1942). Relative efficiencies of various sampling units in population inquiries. *Journal of the American Statistical Association*, 37, 89-94.
- HANSEN, M.H., and HURWITZ, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-362.
- HANSEN, M.H., HURWITZ, W.N., and MADOW, W.G. (1953). *Sample Survey Methods and Theory*. New York: Wiley.
- HANSEN, M.H., HURWITZ, W.N., MARKS, E.S., and MAULDIN, W.P. (1951). Response errors in surveys. *Journal of the American Statistical Association*, 46, 147-190.
- HIPPLER, H.-J., SCHWARZ, N., and SUDMAN, S., eds. (1987). *Social Information Processing and Survey Methodology*. New York: Springer-Verlag.
- JABINE, T.B., STRAF, M., TANUR, J.M., and TOURANGEAU, R., eds. (1984). *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington: National Academy Press.
- JENSEN, A. (1926). Report on the representative method in statistics. *Bulletin of the International Statistical Institute*, 22, 359-380.
- KALTON, G., KASPRZYK, D., and DUNCAN, G.J., eds. (1989). *Panel Surveys*. New York: Wiley.
- KIAER, A.N. (1895-1896). Observations et expériences concernant des dénombrements représentatifs. *Bulletin of the International Statistical Institute*, 9, 176-183.
- KRUSKAL, W.H., and MOSTELLER, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895-1939. *International Statistical Review*, 48, 169-195.
- LAZARSFELD, P.F., BERELSON, B., and GAUDET, H. (1944). *The People's Choice: How the Voter Makes up his Mind in a Presidential Campaign*. New York: Columbia University Press.
- LECUYER, B., and OBERSCHALL, A. (1978). Social research, the early history of. In *International Encyclopedia of Statistics* (Eds. W.H. Kruskal and J.M. Tanur). New York: Macmillan and the Free Press.
- LEVENSTEIN, A. (1912). *Die Arbeitsfrage mit besonderer Berücksichtigung der sozialpsychologischen Seite des modernen Grossbetriebes und der psychophysischen Einwirkungen auf die Arbeiter*. (In German) Munich: Reinhardt.
- MADANSKY, A. (1986). On biblical censuses. *Journal of Official Statistics*, 2, 561-569.
- MASSEY, J.T. (1988). An overview of telephone coverage. In *Telephone Survey Methods* (Eds. R.M. Groves et al.). New York: Wiley.
- MOSTELLER, F. (1978). Errors: I. Nonsampling errors. In *International Encyclopedia of Statistics* (Eds. W.H. Kruskal and J.M. Tanur). New York: Macmillan and the Free Press.
- MOSTELLER, F., HYMAN, H., MCCARTHY, P.J., MARKS, E.S., and TRUMAN, D.B. (1949). *The Pre-election Polls of 1948*. Bulletin 60. New York: Social Science Research Council.
- NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-625.
- NEYMAN, J. (1938). *Lectures and Conferences on Mathematical Statistics and Probability*. Washington, DC: Graduate School, U.S. Department of Agriculture.
- NICHOLLS, W. L., II (1988). Computer-assisted telephone interviewing: A general introduction. In *Telephone Survey Methods* (Eds. R.M. Groves et al.). New York: Wiley.

- NICHOLLS, W.L., II, and GROVES, R.M. (1986). The status of computer-assisted telephone interviewing: Part I - Introduction and impact on cost and timeliness of survey data. *Journal of Official Statistics*, 2, 93-115.
- NORWOOD, J.L., and EARLY, J.F. (1984). A century of methodological progress at the U.S. Bureau of Labor Statistics. *Journal of the American Statistical Association*, 79, 748-761.
- OLKIN, I. (1987). A conversation with Morris Hansen. *Statistical Science*, 2, 162-179.
- PAYNE, S. L. (1951). *The Art of Asking Questions*. Princeton: Princeton University Press.
- PORTER, T.M. (1986). *The Rise of Statistical Thinking, 1820-1900*. Princeton: Princeton University Press.
- RAO, J.N.K., and BELLHOUSE, D.R. (1990). The history and development of the theoretical foundations of survey based estimation and statistical analysis. *Survey Methodology*, 16, 3-29.
- RICE, S. (1928). *Quantitative Methods in Politics*. New York: Knopf.
- RUGG, D., and CANTRIL, H. (1944) (1947). The wording of questions. In *Gauging Public Opinion* (Ed. H. Cantril). Princeton: Princeton University Press.
- SCHWARZ, N. (1987). Cognitive aspects of labor surveys in a multinational context. Paper prepared for the Working Party on Labor Statistics, OECD. Paris, April 1987.
- SCHUMAN, H., and PRESSER, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic.
- SMITH, T.W. (1975). Social change and the General Social Survey: An annotated bibliography. *Social Indicators Research*, 2, 9-38.
- STASNY, E.A. (1988). Modelling nonignorable nonresponse in categorical panel data with an example in estimating gross labor force flows. *Journal of Business and Economic Statistics*, 6, 207-219.
- STIGLER, S.M. (1986). *The History of Statistics. The Measurement of Uncertainty Before 1900*. Cambridge: Harvard University Press.
- SUDMAN, S. (1987). *Reducing the Costs of Surveys*. Chicago: Aldine.
- SUDMAN, S., and BRADBURN, N.M. (1974). *Response Errors in Surveys: A Review and Synthesis*. Chicago: Aldine.
- TAEUBER, C. (1978). Census. In *International Encyclopedia of Statistics* (Eds. W.H. Kruskal and J.M. Tanur). New York: Macmillan and the Free Press.
- THORNBERRY, O.T. Jr., and MASSEY, J.T. (1988). Trends in United States telephone coverage across time and subgroups. In *Telephone Survey Methods* (Eds. R.M. Groves et al.). New York: Wiley.
- TREWIN, D., and LEE, G. (1988). International comparisons of telephone coverage. In *Telephone Survey Methods* (Eds. R.M. Groves et al.). New York: Wiley.
- VAN KLEECK, M. (1930). The Federal Unemployment Census of 1930. *Proceedings of the American Statistical Association*, 189-200.
- WHITE, A.A., and BERK, M.L. (1987). Recall strategies in personal interviewing: moving results from the laboratory to the field. *Proceedings of the Social Statistics Section, American Statistical Association*, 66-71.
- WILLCOX, W.F. (1930). Census. In *Encyclopaedia of Social Sciences* (Eds. E.R.A. Seligman and A. Johnson). New York: Macmillan.

COMMENT

ROBERT M. GROVES¹

The writing of histories of the development and use of survey methods signals a certain maturation of the field. Currently, we are seeing the fiftieth anniversary of several important survey innovations – Neyman’s breakthrough papers in stratification, the start of the U.S. Current Population Survey, and the greater visibility of election polling. With our attention called to such developments it is natural to review the intervening years, seeking to find some theme for events affecting the field. Professors Fienberg and Tanur have completed such an exercise in their paper.

My comments will review key parts of the work, offering comments as I proceed, and then note some errors of nonobservation, misplaced emphases, and other minor quibbles.

Fienberg and Tanur express the purpose of their paper in two ways “to note that technical developments in surveys can be understood only in the context of institutions within which they occur” (p. 31) or at another point to note that “factors other than the advance of statistical theory have shaped the survey domain” (p. 42). Consistent with this they note:

1. the role of ruling, governing institutions which perceive a need for information on the population’s welfare or its reaction to taxation;
2. later, the role of academics in the social sciences in framing central statistical and measurement issues in surveys;
3. the role of mass media use of surveys for election and current events monitoring; and
4. still later, the use of surveys by commercial entities in the market economy.

They document the resolution of controversies in the government sector about use of probability sampling.

Along the way we learn some interesting facts – for example, that for 12 U.S. censuses (120 years) there was no permanent organization for the Census Bureau; that the Department of Agriculture data collection began with need for information about food supplies in the Civil war; that another boost for surveys occurred in the New Deal’s creation for government programs. There seems to be a recurring theme here that governments emphasizing services for the welfare of the populace demand more information about their societies than do those pursuing other goals. In addition, we see that governments most sensitive to public opinion demand more measures of that opinion (that reminds this reader of Gallup’s early metaphor of the survey as a voting analog).

The focus in the paper on the role that institutions played in the development is convincing only for parts of the review. For example, the institutional focus is appealing in describing Lazarsfeld’s evangelical efforts to bring commercial survey and academic inquiry together. The role of the Bureau of Applied Social Research at Columbia University in his partial success at that is enlightening. So too the move of Likert and others from a government agency home (U.S. Department of Agriculture) to academia in order to spread the method to new domains is largely a story of groups of people and organizations which make them effective.

However, the identification of organizations or institutions as the focus can be misunderstood as the stimulus to developments. Nothing I read in the paper changes my opinion that the survey field at its origins attracted broad, creative thinkers. Many were intelligent and charismatic; they led by ideas and mobilized others to work diligently at the definition

¹ Robert M. Groves, The University of Michigan and U.S. Bureau of the Census.

of the new field. Institutions permitted this to happen. They didn't produce the developments. They were homes for the best and brightest.

Within the focus on the institutional, I wished the emphasis of the paper might have been placed more on two related points:

1. Different tasks were more easily accomplished in the different domains. For example, government agencies were by their nature restricted to questions of monitoring social welfare, the commercial, to newsworthy or dollar worthy interests, and the academic to longer term, more basic social issues. Those involved in early developments shaped their agenda to the goals of the organization.
2. Stories of the early days of survey research, as told by those who lived them are filled with the excitement of a new field. I missed in the paper sufficient acknowledgement that the young researchers involved in the work shared an evangelical mission – spreading the gospel of probability sampling, inventing new methods of interviewing because nothing existed. The institutional focus misses the human drama of those days.

Fienberg and Tanur also note “the membranes separating the institutions are extremely permeable”. That is, researchers move back and forth between the institutions, contributing to each of them, and transferring knowledge as they move. The evidence the authors cite is the experience of Lazarsfeld addressing basic design issues while conducting radio audience research within an academic setting and of Likert moving from the insurance industry to the Department of Agriculture to the University of Michigan. These moves seem the exception rather than the rule. I have not conducted the appropriate careerline research to demonstrate this, but my impression is that the fences between the sectors have been and remain high and painful to transgress. Further, movement among academic government, commercial is asymmetric. Rarely is there movement from the commercial or government sector to the academic sector (current demands on publication history prevent this). The government-commercial interchange is larger.

The result of this insularity is the development of techniques not shared across the different sectors (edit and imputations schemes, nonresponse reduction techniques). The three sectors to some extent have developed their own language to describe their work (*e.g.*, “stem and banners”, “tabs” versus “contingency tables”).

The membrane metaphor also fails to observe the large differences in the centrality of surveys to organizations in the three sectors. Academic survey research is not central to any university in the world. It was not central early in the history of the method (*viz.* the inability of the Likert group to obtain university parking stickers because of their nonfaculty status). Even now it is often viewed as a haven of technicians (several steps below the chemistry laboratory staff) currently on many campuses. In contrast there are government and commercial organizations fully devoted to survey design, collection, and analysis. These have decision-making hierarchies constantly monitoring cost and error structures of surveys without the ongoing debate about the relative worth of the enterprise.

The paper ends with a discussion of three developments since 1960 that are important to understanding surveys. At this point, the institutional context is dropped as the organizing principle of the paper and innovations are the focus. Three developments are highlighted: a) the use of the telephone as a data collection medium and later developments in computer assisted telephone interviewing (CATI); b) the use of longitudinal surveys to study micro-level change over time; and c) the application of cognitive psychological concepts to survey methods.

The authors note the movement of mode of data collection from face to face to phone and development of CATI, but they fail to note that this is largely a US phenomenon in the academic and government sectors (the commercial side had done it years ago). Indeed, it is an example of distinctive methodologies pursued by the three sectors. I share their belief that the merits of longitudinal surveys are increasingly being recognized and note that the 1980's is seeing this spread internationally. The Fienberg and Tanur team was instrumental in launching the U.S. effort to apply cognitive psychological concepts to survey measurement, and we are in their debt for this.

The paper does not make it clear whether the authors believe the CATI, longitudinal surveys, and the effort to "cognitize" survey methodological research are the most important three developments in surveys, but they clearly omit several other important ones. We can all choose our three most important developments since 1960; here are some other candidates:

1. Development of Generalized Statistical Software Packages

This development greatly expanded the number of researchers who could directly pose and answer questions using survey data. In the statistical and social sciences at this writing, it is common for undergraduates to perform analyses of survey data whose complexity would have prevented their being done 25 years ago.

2. Existence of Survey Data Archives

The archiving of survey data on computer media was a further democratizing force in survey analysis. With those developments replication and extension of analysis, a key component of the structure of scientific advance, became trivial. Unfortunately, there were also deleterious effects. Analysts of survey data could do their work in complete ignorance of the survey design, of the interviewer training and supervision guidelines, of nonresponse rates, and of a host of other design features known by those conducting the survey.

3. Growth of Commercial and Nonprofit Industry to do Government Surveys

The U.S. is distinctive in its reliance on academic and commercial groups to conduct surveys on behalf of government agencies. Some of this exists in many Western countries, but to a much smaller degree. This suggests that a cross-cultural strain in the paper might be interesting – to identify unique histories of survey research in various societies.

4. 1960 as Beginning of Widespread Acceptance in Academic Circles of the Social Psychological Model of the Interview

This typically describes survey interviews as "conversations with a purpose" and focuses the researcher's attention on the role of the two actors in the errors produced during measurement.

5. Ubiquity of Surveys

Survey measurement is now a way of life for most large corporations (prior to the breakup of ATT in the U.S. the corporation conducted over 7 million customer satisfaction interviews annually). Surveys are viewed as irreplaceable sources of information about customers, suppliers, and the general society.

6. Nonresponse and the Growing Reluctance of the Population to be Measured

This is certainly a phenomenon of great import to survey researchers in most Western countries. With statistical inference to large populations one of the key virtues of surveys versus other data collection schemes, this issue strikes at the heart of the tool. Again, a cross-national theme to the paper would have highlighted these issues.

We can apply the superpopulation metaphor to any historical account – that is, any series of events (which later we call history) is but one realization of an infinite set of possible series which defines the universe of possible realities. This fits the set of questions that remain unanswered.

1. Why after almost a century hasn't survey research fully evolved into a profession (with specified standards and training criteria)?
2. Why is there so little formal educational structure for survey researchers to get their knowledge base? Why are there departments of communications, operations research, naval architecture but none of survey research (teaching sampling, questionnaire design, data analysis)?
3. Would public education about surveys and statistics (like the ASA/NSF program in quantitative literacy) have made an impact on acceptance of surveys?

We are indebted to the Fienberg/Tanur team for reviewing our collective past. They have helped chronicle the birth and first 50 years or so of what is now an important component in most societies of the world. I do hope that the year 2040 will see the need to ask Fienberg and Tanur to update their paper for that occasion. I hope they will be able to report innovation during those 50 years that made a difference in survey methods.