

FEBRUARY 3, 1998

**THE ROLE OF BEHAVIORAL SURVEY METHODOLOGISTS IN
NATIONAL STATISTICAL AGENCIES**

by

Don A. Dillman¹
Washington State University
PO Box 644014
Pullman, Washington 99164-4014
e-mail: dillman@wsu.edu

A few years ago I learned of a new mail survey being designed by a national statistical agency for surveying millions of households. A lengthy memorandum had been written outlining more than 20 specific topics for research that should be done to decide how the survey would be designed. This survey was the first of its kind, and numerous design challenges had to be met. It involved measurement over time, mixed-mode data collection (telephone after several mail contacts), and numerous other measurement and nonresponse challenges. I read the list with much anticipation inasmuch as it had been portrayed to me as defining the potential research agenda. I expected there to be considerable emphasis on how to design the questionnaire, how to achieve good measurement of the concepts and combine data collected by mail and telephone, and how to assure high response rates from diverse segments of the population. Much to my surprise, none of these topics was on the list. In fact, there was not a single research topic on the list that a survey methodologist like myself would undertake. All of the topics concerned statistical issues, ranging from alternative sample designs to methods for imputing missing data. It was as though other kinds of methodological problems did not exist.

An employee of the national statistical agency in another country described to me a comparable experience. He had just been through a long and challenging experience of coauthoring a major publication about reducing nonresponse error on which he had collaborated with several social

¹Revision of keynote presentation to the International Association of Survey Statisticians sessions at the Joint Statistical Meetings, Anaheim, California, August 12, 1997. Don A. Dillman is Deputy Director for Research and Development at the Washington State University Social and Economic Sciences Research Center. From 1991-1995 he served as Senior Survey Methodologist in the Office of the Director at the U.S. Bureau of the Census. Appreciation is expressed to Cynthia Clark, David Binder, Robert Groves, Lars Lyberg, Elizabeth Martin, Clyde Tucker and Bryan Burke for comments on this paper. However, the opinions expressed in this paper are the author's and should not be attributed to these reviewers, the U.S. Bureau of the Census or Washington State University.

scientists. He commented about what he defined as the most important question to be addressed, i.e.,

How can we adjust for nonresponse?

He then noted that some of his co-authors defined the problem quite differently. They would ask first,

How can we reduce nonresponse?

There are three things that I find noteworthy about these different perspectives. First, they are both important perspectives on nonresponse error, and for each there are scientifically-based perspectives on how to carry them out.

Second, the two scientific perspectives seem to have rather little in common. The statistical adjustment perspective is an after-the-fact procedure. It takes as its starting point the available data set and uses the tools of statistical analysis and auxiliary information to input values and otherwise adjust the results. In contrast, the nonresponse reduction perspective takes a human behavior perspective on why people behave as they do (respond or not respond) and uses theories of behavior to design data collection approaches that will improve response. And, most important it decreases the need for nonresponse adjustment.

Third, I believe that most statistical agencies have done, and continue to do, a much better job of developing and using statistical principles associated with adjustment than they do in using human behavior theories to provide tools for reducing nonresponse. Statistical agencies have relatively few scientists whose primary training is in the social and behavioral sciences, and those who do have such training are frequently located apart from the divisions in which most survey design decisions are made.

My purpose here is to discuss the error reduction role of behavioral survey methodologists in statistical agencies. Specifically I'll describe three important roles these methodologists can be expected to fill in statistical agencies, and the organizational arrangements that will increase the likelihood of success in filling each role.

I define behavioral survey methodologists as individuals with advanced training in theories of human behavior (typically psychology or sociology) and who apply their scientific knowledge to the reduction of survey error. The antecedent term behavioral is used to distinguish these individuals from statistical survey methodologists who have no training in theories of human behavior and whose efforts to adjust for survey error tend to be based upon mathematical or statistical theories. My focus in this paper is entirely on the role of behavioral survey methodologists.

Survey Error and Theories of Human Behavior

The role of behavioral survey methodologists in statistical agencies is shaped by the particular scientific knowledge they bring to bear on survey design issues. Improving the quality of survey data by reducing survey error can be approached from many perspectives. In this regard I tend to follow Groves (1989), and his elaboration of four fundamental types of survey error:

Coverage Error: The result of all units of a population not having a known, nonzero probability of inclusion in the sample which is drawn to represent the population.

Sampling Error: The result of surveying a sample of the population rather than the entire population.

Measurement Error: The results of inaccurate response stemming from poor question wording, poor interviewing, survey mode effects and/or the respondent's behavior.

Nonresponse Error: The result of nonresponse from people in the sample who if they had responded, would have provided different answers than those who did respond to the survey.

Statistical survey methodologists and other statisticians are less likely than behavioral survey methodologists to deal with measurement and nonresponse error. In those cases where they do focus on it, they tend to deal with it from the perspective of data adjustment rather than error reduction, i.e., through improvement in response behaviors. This difference in approach to understanding and dealing with survey error undergirds the argument I will make about the role of survey methodologists in statistical agencies.

The discipline of statistics differs from those of psychology and sociology, in which survey methodologists are typically trained, with regard to their focus on human behavior and the motivations for it. Statisticians view human behavior as something to be described, and the theories and concepts of statistics offer no substantive explanations of why humans behave in particular ways. In contrast, psychology and sociology are explicitly about the reasons for human behavior, offering theories of why people behave in certain ways and not others.

Choosing to respond or not to respond and deciding what answer to give to a survey question are conscious human actions. Certain theories and concepts from sociology and psychology have been utilized and/or developed as a means of explaining these actions, a brief discussion of which follows.

Some Explanations for Measurement Error

Survey measurement error manifests itself in many different ways, e.g., question order effects, category order effects, item nonresponse, effects of question structure, i.e., open vs. closed, the use of scalar vs. nominal categories, and the choice of words for questions and answers alternatives. Research on understanding the causes of measurement error in surveys has been

very active during the last two decades, and many concepts have been invoked to describe and explain it. Those listed below illustrate the diversity of this literature but are by no means a complete listing of possible explanations for measurement error.

Social desirability effects: The tendency of respondents to provide answers which put them in a favorable light with the interviewer and/or perceived societal norms (e.g., DeMaio, 1984).

Acquiescence: The tendency to agree (rather than disagree) with statements because of cultural influences and/or the effects of hierarchical position (e.g., Schuman and Presser, 1981).

Norm-of-even-handedness: The tendency for the answer to a follow up question to be influenced by the answer given for the preceding question because respondents are trying to be fair. An example is the obtaining of less favorable answers to this question, "Should the U.S. be allowed to restrict imports from Japan," when it was preceded by, "Should Japan be allowed to restrict imports from the United States." (e.g., Bishop et al., 1988).

Satisficing: Tendency to give an answer that is deemed satisfactory but is not a full response (e.g., choosing from among the first answers of a list without reading all of the choices) (Krosnick et al. 1996).

Primacy/Recency: Tendency to choose from among the first answers on a self-administered survey, but from among last answers on an interview (Krosnick and Alwin, 1987).

Category range choices: Tendency to choose different answers depending upon what ranges of a behavior (e.g., hours students study or watch television each week) are provided (e.g., Rockwood et al., 1997).

Subtraction effects: Providing a different answer to the second of two related questions because the first question in a series changes the perceived meaning of a question (e.g., "How are economic conditions in your community?" "How are economic conditions in your state?") (e.g., Mason et al., 1994).

Addition effects: Providing a different answer to the second of two questions because of carryover from the first (e.g., more positive answer to question about life in general following question about quality of marriage) (e.g., Schuman and Presser, 1981).

Estimates without needed cognitive recall: Accuracy of estimates for given time period when asked out right (How many times have you visited a doctor in last six months?) Is less than when warm-up questions have established important events and dates in respondent's life during the previous six months prior to asking the recall question (e.g., Means et al., 1988).

Theories to account for such differences as these tend to come from one or the other of two overarching theoretical frameworks. One framework is from cognitive psychology and breaks into subparts the process of responding to a question--comprehension of the question, retrieval from memory of relevant information for providing an answer, making a judgment on what information to provide, and reporting an answer (Tourangeau and Rasinski, 1988). It has been used to explain phenomena such as primacy and recency effects (Krosnick and Alwin, 1987). The other framework is more sociological in nature and emphasizes complying with perceived norms or values of significant others, the effect of which is to produce biased answers. This framework has been relied upon to explain socially desirable answers as well as those consistent with the norm-of-even-handedness. In some cases these general frameworks, i.e., cognitive psychology and sociology, provide competing explanations for certain effects such as satisficing (Krosnick et al., 1996).

Reliance on mixed-mode surveys for which some data are collected by interview and other data by self-administered questionnaire, has increased in recent years. It has also become apparent that differences sometimes exist with regard to the answers respondents give to one type of survey vs. the other (Schwarz et al., 1991). At least seven different kinds of mode effects, have been reported in the literature (Dillman et al., 1996). Some (e.g. subtraction effects) appear to result from cognitive influences, and others from normative influences (social desirability and acquiescence) and still others may occur because of joint cognitive and normative effects (e.g. primacy/recency effects).

Recent research on survey mode effects has introduced additional concepts that contribute to our understanding of measurement error problems on self-administered questionnaires. Self-administered questionnaires consist of two types of information. The first type consists of words and the way they are arranged (information organization). The second is graphical layout and design which influences what respondents tend to see or not see, and the order in which they proceed through the survey questions (Jenkins and Dillman, 1995, 1997). Whether respondents follow the prescribed navigational path depends on visual perception concepts such as the figure-ground format, bottom-up and top-down processing of information, and pattern recognition. In addition, characteristics of the elements of the graphic layout, including size, brightness, color, familiarity, location, and contrast influence perception of elements on the page. What respondents comprehend is affected according to the Law of Pragnanz (figures with simplicity, regularity and symmetry are more easily perceived and remembered), law of proximity (similar figures located together are seen as belonging to the same group), and law of similarity (similar shapes are seen as belonging together). The importance of these concepts for a theory of self-administered questionnaire design is that they can be manipulated to increase the likelihood that respondents will understand the questions in a prescribed manner. Getting respondents to process the same words in the same sequence as done for interview questionnaires may diminish the likelihood of differential mode effects.

Because of the different disciplines, concepts, and overarching theoretical frameworks relied on

by survey methodologists to examine measurement error, it is little wonder that they sometimes find themselves struggling to communicate with one another. This situation also helps to explain why a particular measurement effect is sometimes described in different ways using different concepts. It is complicated further by the fact that many of the concepts used to gain an understanding of measurement error have relatively little in common with the concepts and theories relevant to understanding nonresponse error.

Some Explanations for Nonresponse Error

Many factors may determine whether a sampled individual, household, business or other organization responds to a survey. For household surveys Groves and Couper (1996) have divided these factors into those which are under the surveyor's control (survey design and interviewer characteristics), and those outside the surveyor's control (social environment and householder characteristics). The decision to cooperate or refuse is described as the result of interaction between householder and interviewer. From this perspective and various theoretical orientations, numerous hypotheses are generated about the determinants of respondent cooperation. They include opportunity cost, (social) exchange, social connectedness, authority, topic saliency and fear of crime, to mention a few. Although developed as a general model for survey response, the elements obviously differ somewhat for self-administered questionnaires where the contact between respondent and surveyor is not mediated by an interviewer. Yet, this model has much similarity to the results of earlier research by Heberlein and Baumgartner (1978) on self-administered questionnaires which showed that respondent characteristics, as well as survey characteristics (from sponsor to number of contacts with respondents), both have a substantial influence on response rates.

As in the case of measurement error, some of the concepts called upon to account for response behavior come from a branch of psychology but not that of cognitive psychology, which has been so helpful for understanding measurement error. The relevant source of concepts about motivations comes from social psychology and suggests that many different characteristics of the response task can influence whether a person will comply with a response request. Groves et al. (1992), have applied six such concepts, as described by Cialdini (1984), to understanding the decision to respond to a survey. They include:

Reciprocation: The tendency to favor requests from those who have previously given something to you.

Commitment and Consistency: The tendency to behave in a similar way in situations that resemble one another.

Social Proof: The tendency to behave in ways similar to those like us.

Liking: The tendency to comply with requests from attractive others.

Authority: The tendency to comply with requests given by those in positions of power.

Scarcity: The tendency to attach greater value to rare opportunities to participate in something enjoyable or rewarding.

To enhance the likelihood of survey response, these concepts are used in conjunction with efforts to keep respondent-interviewer interaction going, thus allowing more opportunity to invoke these influences. In practice, this means doing callbacks rather than insisting that the interview be completed when contact is first made and tailoring interviewer comments to characteristics of respondents as learned through the extended interaction process (Groves et al., 1992).

For self-administered questionnaires, social-exchange theory has been used as a frame of reference to build a system of multiple contacts in which the concepts of increasing perceived rewards, reducing perceived costs, and trusting that the benefits of responding outweigh the costs of responding or not responding (Dillman, 1978; Goyder, 1987). This approach discourages viewing the questionnaire in isolation from the cover letter, or each of the contacts in isolation from one another, and has produced multiple hypotheses about the causes of high response rates (Dillman, 1991). Reciprocal obligation, a concept encompassed by the theory of social exchange as well as part of the taxonomy described by Cialdini, has been used to explain the effects of token financial incentives, which have been found far more effective than much larger post-payments for improving response to questionnaires (James and Bolstein, 1992).

Another theoretical approach used to improve response to surveys is the foot-in-the-door technique, based on the argument that if people respond first to a small request they are more likely to respond later to a larger request. This perspective has been used to design personal-delivery approaches for questionnaires that have proved to be highly effective (Dillman et al., 1995).

Can the Science Be Trusted and Who Can Provide it?

The literature on measurement and nonresponse errors is enormous, with thousands of articles having been written on each topic, much of which is not generally accessible to professionals without considerable social science training. In addition, the theoretical perspectives that have mostly guided research on measurement error are different from those that have guided research on nonresponse error. Even within each area there are often different perspectives that are competing to account for the same phenomenon. Further, there are considerable differences between how interview and self-administered questionnaires must be designed. In general, the scientific contributors to our understanding of how one method works are not among the chief research contributors to how the other works.

Frequently I have listened to leaders of statistical agencies describe their desire to hire a behavioral survey methodologist who could deal competently with all aspects of measurement and nonresponse error. My usual response to such questions is to ask the inquirer to specify what kind of survey methodologist is needed. Is it someone who knows the theories and

literature for conducting interviews, or is it someone whose speciality is self-administered surveys? Also, is the concern one wants addressed measurement or nonresponse error? And, within the area of measurement error, are the issues of concern cognitive or motivational? Do any of the issues concern graphical layout and design of questionnaires for interviewers or self-administration?

Although some individuals make Herculean attempts to ~~cover~~ both interview and self-administered questionnaire issues, it is becoming increasingly difficult to do so well. The title of behavioral survey methodologist no longer describes a common specialty as much as it describes a cluster of specialties, training for which may be provided by quite different disciplines. The quality of professional one needs for designing the details of a particular survey is more than those that can now be mustered by a methodology generalist. This increasing specialization of methodological skills also underscores the need for multiple professionals with complimentary specializations within a single agency.

My suggestion to leaders of some statistical agencies that highly trained specialists in various aspects of survey methodology should be added to their staffs in order to do a better job of reducing survey error rather than simply trying to adjust for it afterwards has often been countered by a general criticism of survey methodology. That criticism is that the principles espoused by survey methodologists, of any type, are more opinion than established fact, whereas statistics is a more established science with formulae that have been proven. Why, the question is posed, add specialized survey methodologists to an agency staff if there is disagreement on the principles they espouse?

Nonresponse and measurement error problems are unlikely to be resolved by focusing mostly on adjustment and without the benefit of insights from theories of human behavior and also working on the reduction side of the problem. Although the behavioral theories about some aspects of human behavior are not yet fully articulated and tested, I don't think we'll find solutions by looking in an area devoid of behavior theories, where the answers are unlikely to be found.

When I contemplate the varied issues addressed by methodologists concerned with reducing measurement and nonresponse error, the variety of theoretical perspectives being employed in this research and the frequently differing perspectives on these matters, I come away quite impressed with the progress being made in gaining an understanding of these phenomena. I also view competing theories as a sign that our understanding of phenomenon as complex as measurement and nonresponse is advancing.

A useful science does not develop without concepts, hypotheses, and integrating theories. Competing perspectives are desirable because they stimulate the research necessary for separating useful design principles from speculative hypotheses. I also view the existing knowledge as a very useful frame of reference for thinking through methodological challenges in situations and on problems that have not been addressed in previous research. What some might view as a glass half empty, I tend to view as a glass half full and available for consumption.

Three Roles for Behavioral Survey Methodologists in Statistical Agencies

It seems important that behavioral survey methodologists perform effectively in three roles, each of which is critical for reducing errors in surveys. Performance of these roles has implications for hiring practices and organizational management of our national statistical agencies.

Role 1: Bring an error reduction perspective to bear in an influential way for all aspects of designing and implementing agency surveys.

The first role is to bring an error reduction perspective to bear in an influential way on the design and implementation of all statistical agency surveys. It was precisely this role that was not represented in the research issues memorandum discussed in the introduction to this paper; behavioral survey methodologists were not involved in discussing and defining the issues to be researched during the design phase of this important national survey. In some agencies I sense that this omission may be the rule rather than the exception. The tendency has been to borrow fairly standard questionnaire design and implementation methods from previous surveys, rather than to consider whether they are subject to error and need to be improved.

In a previous paper I wrote about the difficulty of achieving innovation in government surveys, especially on nonresponse and measurement issues (Dillman, 1995). I argued that three important characteristics of the organizational structure coalesced to inhibit innovation in government agencies. They were: (1) the co-existence of research and operations cultures separated by divisional lines; (2) major differences in the dominant value systems of those cultures; and (3) the difficulty of resolving those differences in hierarchically-oriented organizations (Dillman, 1995).

There is no absence of willingness on the part of agency professionals without training in survey methodology being willing to make survey questionnaire and implementation design decisions which influence nonresponse and measurement error. Decisions on all aspects of survey design that influence error reduction--number of contacts, interviewing protocols and call-back instructions, content of cover letters, the wording of survey questions, and the ordering of questions into a questionnaire, are all being accomplished by these personnel on a routine basis. They include professionals within the existing operations culture who have relatively little understanding of the theories and principles important to survey error reduction, just as behavioral survey methodologists typically have relatively little understanding of sampling and adjustment methodologies.

However, the writing of questions, layout of questionnaires, and implementation methods are seen as something which everyone, regardless of background, has considerable knowledge. In the paper just mentioned, I commented:

When a report (to work groups) is made on sampling design, questions are seldomly raised. When they are, formulas are discussed and the participation is limited to those with strong statistical backgrounds, especially people from the research culture. However, when discussion shifts to the way questions should be written, the form of the questionnaire or the procedures that should be used to improve response, virtually everyone enters the discussion.

In various project work groups where I have observed votes being taken on matters of questionnaire page layout or design, all participants, regardless of training or background, seem to view it as appropriate to contribute to such decisions. The reason for this is that matters of question wording, questionnaire design, content of cover letters, and sequencing of contacts are seen as within the everyday experience world. Personal experience and ideas about how I would respond to this postcard reminder vs. another one are viewed equally as having as much validity as perspectives based upon theory and research.

Under these conditions the most powerful influence on decision-making tends to be what has been done previously, the cost and processing needs of other divisions, and perspectives that are not at all based in the available science. This tendency is amplified by divisional structures which separate survey methodologists from operations divisions where the final decisions are made, and which separate other functions such as question content, mailing services, data entry, and data analysis; each of which represents interests that have a direct, and sometimes deleterious influence on measurement and/or nonresponse.

In some cases professionals whose primary training is in statistics are called upon to make methodological design decisions that have significant impacts on the ability of experiments to produce interpretable results. In one agency I observed an important experiment being designed that required a very large sample size because of an attempt to identify differences within tenths of a percent. However, the design required a telephone interview to determine whether information previously provided on a self-administered questionnaire was correct. That immediately raised the specter of obtaining different answers by telephone simply because of the different mode being used and the likelihood of obtaining socially desirable answers by interview. In addition, the telephone questions were structured and worded differently, taking people through a different mental or navigational sequence than was the case for the self-administered questionnaire. Under these conditions obtaining a sampling error precision of less than one percent would still not allow policy decisions to be made with confidence. The potential for unknown measurement error differences between survey modes was likely to be much greater than the sampling precision which had become the central focus of the experiment.

Few would argue that any of the four major sources of survey error--sampling, coverage, measurement and nonresponse--are unimportant for designing quality surveys. In practice, however, less is done than is desirable for many government agency surveys with regard to reducing measurement and nonresponse error at the survey design stage. Thus, a critical role of behavioral survey methodologists is to participate in the design of all government surveys and

perform that role effectively.

Role 2: Bring theoretical efficiency and effectiveness to experimental tests of alternative questionnaire designs and implementation procedures through the use of theories, concepts, pretests and findings of past behavioral science research.

Field experiments represent an important means of identifying ways to improve the design of government surveys. A second important role of behavioral survey methodologists is to improve the efficiency and effectiveness of these tests by bringing behavioral science theories, concepts, and past research to bear on their design.

Funding constraints inevitably require that alternative experimental panels be limited in most tests. In such cases the selection of panels may become a tug-of-war among many legitimate agency interests as each seeks to have its concern tested. Theoretical knowledge and results of past research can be used as a criterion to select certain experimental treatments and forgo others in ways that increase the likelihood that results will be meaningful and cumulate across experiments and populations.

Frequently I have read research reports on experiments about survey implementation procedures and wondered if the appropriate role for a behavioral survey methodologist might have been simply to discourage their being undertaken. Some of the most common are experiments on sending payments to people for completing questionnaires vs. prepaid token financial incentives. Research experiments have consistently shown that post payments (an economic exchange), even large ones, do not work nearly as well as prepaid token incentives (a social exchange).

Similarly, I was asked a few years ago to comment on an experimental test that had been conducted on the effects of a prenotice letter. It was tested for use on a proposed national mail survey on health issues being developed by one of the nation's health care agencies. The request attracted my attention because this variable has been tested many times, and in the vast majority of studies had produced improved response, while this study had shown it to be ineffective. A quick examination of the procedures revealed the likely reason. The letter was more than a page long, consisting of many disclosures about the study that put it in a negative light without the ability for recipients to simultaneously see the questionnaire and how threatening the questions were. Such information is better saved for the cover letter. In addition, it was sent nearly a month prior to the mailing of the survey. The literature strongly suggests that a brief, anticipation building preletter sent just a few days ahead of the questionnaire mailing would have been more effective. The negative results with this preletter were hardly a surprise. Involvement of a behavioral survey methodologist earlier in the research process would have contributed in one of two ways--either to keep the experiment from being done, or to get it redesigned so that the tested method had a reasonable chance of working.

A personal experience at the U.S. Census Bureau also illustrates some of the complexities of this role. One of my responsibilities while there was to contribute to the design of experiments aimed at improving response rates for the next Census. In past Censuses two contacts, the initial

questionnaire mailing plus a postal patron addressed postcard follow up indicating there is still time to respond, a format which I believe was quite ineffective, were used to elicit response. A series of four experiments, consisting of 27 treatments, was developed in order to test the effects of 13 different variables on response rates. Results of these experiments showed that only five of these variables contributed significantly to improved response rates: prenotice letter (5-7%); postcard reminder (4-8%); replacement questionnaire (6-11%); respondent-friendly design following principles discussed earlier in this paper (3-4%); and a message indicating response was required by U.S. law (9-11%). Together these variables improved response rate from about 42% for one mailing of the questionnaire to about 78%, with each making an independent contribution of the magnitude listed here in parentheses (Dillman et al., 1984). The overall response rate was about 13% higher than achieved in the 1990 Census when the existence of a national Census climate also contributed to higher response.

When I shared these findings to a group consisting mostly of survey methodologists conversant with the mail survey literature, several people registered their surprise that resources were spent to test some of these variables. After all, the most enduring finding of past mail survey research is that multiple contacts is the most powerful means to improving rates. Reporting that the addition of three contacts--prenotice, reminder, and replacement questionnaire--improved response was hardly revolutionary. Why, they asked, was it necessary to reprove the most obvious and consistent findings of past research?

It was also pointed out in this discussion that the mandatory disclosure notice adheres to the authority principle espoused by Cialdini (1984), and the usefulness of emphasizing government sponsorship is supported by a 1978 meta-analysis by Heberlein and Baumgartner that showed that government sponsorship tended to improve response rates. That left the experimental evidence on respondent-friendly design as potentially the only truly new contribution to the literature.

My take on the importance of these findings is somewhat different. It was possible to design these experimental tests and individual panels in a way that the results would accumulate across experiments (Dillman et al., 1984). Specifically, experimental panels from earlier experiments were carried over to later ones in a manner that would allow us to measure the marginal gain of adding individual elements. Four of the elements found significant were added at least in at least one of the experiments so that the amount of value contributed by each element could be determined for each of the first four elements, a finding far more important than simply whether each made a significant difference in an experiment. In addition the Cialdini and the Heberlein and Baumgartner concepts had not to our knowledge ever been tested experimentally.

Apart from these concerns, reproving what has been proven elsewhere may sometimes be essential in order to bring about change. The admonition was offered many times that when attempting to change a survey as large and as entrenched in a particular methodology as the Decennial Census had been (similar questionnaire and mail-out procedures for three decades), the experiments were essential in order to change beliefs about what was possible. Theory and past research, although important, were not enough!

A variety of skills and perspectives must be brought to bear effectively on statistical agency research. Some of those skills are from practically oriented operations staff, some are from sampling statisticians. Each in their own way harbors certain perspectives and knowledge that the other lacks but which contributes to efficiency and effectiveness of experimental tests. Defining motivational elements to be tested, and experimental designs that will cumulate across experiments is an important contribution to be made by survey methodologists

Role 3: Contribute to the Expanding Science of Survey Methodology.

The third important role for behavioral survey methodologists is to contribute intentionally and explicitly to the expanding science of survey methodology. I have sensed among many government agency administrators a concern that their agency is not authorized to do basic research. Thus, it is reasoned that the scientific value of research experiments should not be a major consideration in the design of agency experiments. Such research should presumably be left to university scientists. To do that would leave much needed methodological research undone and invite allowing the hunch and personal experience to continue to override the available science.

Relatively little money is available for basic methodology research, and the average funding level per grant is generally quite small in the United States. As a result, the vast majority of university experiments are done on small sample sizes. Also, other modes of testing theoretical ideas, for example meta-analyses, often get favored over experimentation as the only practical alternative. Much of the important work to be done in survey methodology concerns relatively small effects (say, 1-3%) which requires fairly large numbers of experimental panels with large sample sizes.

In addition, the equipment needed for research on some methodological issues is getting more complex and remains expensive. For example, I expect that optical scanning and imaging will soon become the dominant mode of data entry for most government surveys. There are two approaches to improving scanning rates--better equipment that can read and decipher handwriting or better questionnaire layout and design that improves the way people write and mark answers so that less demand is placed on software development (Dillman and Miller, In Press). To conduct such research many basic questions need to be answered about how people use questionnaire pages and how they are influenced to navigate pages in the desired manner. The kind of experimentation and testing that needs to be done involves laboratory work on eye movements in response to graphic design alternatives, cognitive testing laboratories, the building of questionnaires with color and symbols that require graphical design skills and equipment, and large samples for field tests. The chances of these resources being assembled within a large government agency seem much better to me than the prospects of putting them together within a university setting.

As I look at the kinds of methodological issues facing survey methodologists as we turn the century, I am concerned about the complexity of research that needs to be done but also

optimistic that we have a marvelous opportunity to develop a unified methodology for designing and implementing surveys. However, I'm not at all optimistic that this will happen unless national statistical agencies provide leadership for the conduct of basic methodological research.

Effective performance of this role will also address in an important way the disturbing tendency to sometimes view behavioral survey methodologists as one-time consultants. For example, when a measurement or nonresponse problem is recognized by the agency they are called in to provide a quick and inexpensive fix. The most prevalent example of this that I have observed in several agencies is to ask cognitive psychologists to do a few cognitive interviews and fix a question that is not working. In many respects these requests represent a significant advancement over the past, when questionnaires were finalized around the conference table, and put directly into the field. And, qualitative methods have repeatedly proved their importance in statistical agencies by identifying questionnaire problems. However, conducting a few cognitive interviews with volunteers is a very limited base for changing questionnaire designs for major surveys; the same laws of probability apply to cognitive interviews as to other types of survey interviews and the science of doing them needs to continue to develop. Such requests too often get made as a means of trying to avoid experimental tests. It is important that the contribution of survey methodologist not be limited to a small-scale fix-it role that inhibits the building of the larger science of survey design.

Implications for Government Agencies and Survey Methodologists

The Necessity of Hiring Specialists

The time is past when individual behavioral survey methodologist could be expected to be knowledgeable about all aspects of survey methodology--interviews, self-administered questionnaires, measurement, visual layout of questionnaires, and response rate improvement. Not only are the literatures on each topic enormous, but the intellectual heritages of work being done on these topics differs. The fact that a person is very good on one topic may mean, increasingly, that they aren't very knowledgeable about other topics.

Thus, the first implication for agencies is the need to include on their staffs different kinds of behavioral survey methodologists, consistent with their survey needs, and not expect individuals to cover all areas. It simply is no longer possible, assuming it ever was, to be a methodological expert in everything. Nor is it possible for statisticians to learn a little bit of social science and thus become a methodological specialist in reducing survey measurement and nonresponse error.

This issue is probably not an easy one for existing government behavioral survey methodologists to deal with. If a staff consists mostly of individuals trained to deal with interview methods, it may be difficult for them to accept the need for professionals whose primary training and experience is with issues related to self-administered questionnaires or vice versa. The time has come for existing methodological staffs to consider explicitly the breadth of methodological skills needed for researching and contributing to the design of 21st century survey methods. The more

that staff members resemble one another in training and background, the less adequate that staff is likely to be for achieving that goal. This issue also underscores the need for larger staffs of survey methodologists in many statistical agencies so that the diverse needs of survey design can be competently handled.

It's been suggested that one of the impediments to hiring various types of behavioral survey methodologists in statistical agencies is that the best and brightest are not likely to be attracted to positions in which they are unable to interact with a team of other scientists like them. In my view this is much less of a barrier than it might have been in the past. The Internet connections of scientists with one another and the development of greater specialization means that behavioral scientists are no longer confined to working locally with professionals having similar interests and skills. Being viewed as a "chore person" to provide quick methodological fixes and not being able to do fundamental research, as I have suggested under role three, strikes me as a much greater inhibitor to the hiring of such individuals. Access to equipment and the ability to contribute to the design of large scale experiments strike me as more important attractants than the disincentive of possible professional isolation.

An Organizational Dilemma

Where survey methodologists are located in statistical agencies needs to be rethought. Herein lies a basic dilemma. On the one hand it is argued that methodologists need to be located together so that they can learn from and help one another respond to varied agency problems. Being located in a single unit helps develop a research culture orientation that encourages the production of good science and in the long run better quality work. It promotes the development of high methodological standards that become shared.

On the other hand, when survey methodologists are clustered in one place under one authority, they are sometimes placed in the role of a temporary consultant to operational divisions where there is a tendency to be on the margins of decision-making. It is much easier to ignore the advice of methodologists when they are in a consulting role. When located in one division apart from where the survey work is being implemented, there is also a tendency for methodologists to get left out of the routine communication. Consequently, changes with major implications for measurement or nonresponse error are likely to get made without the survey methodologist being able to provide appropriate assistance or advice. Most importantly the leadership of some divisions may simply decide they have no reason to involve a methodologist in the design of their surveys and leave it to the operations part of their division to work out all of those procedural things.

An opposite point of view is the concern that assignment to divisions would result in less qualified personnel being assigned to research methodology positions (an operations person being promoted) and the person would then be co-opted by the division and that a consulting relationship would therefore be better.

In the previous paper I argued for matrix management as a potential solution for this problem, whereby methodologists have dual responsibilities, one of which is to a division where their expertise is evaluated, and the other being where their work is done. In the absence of significant increases in well-trained behavioral survey methodologists, I believe matrix management would be the most desirable approach and needs to receive serious consideration. Dual accountability to a **A**methods@division for scientific knowledge and a **A**program@division for solving the problems of particular surveys has been successful in other contexts and should be considered by government statistical agencies.

Conclusion

I have briefly argued in this paper that statistical agencies do a far better job of dealing with survey error adjustment than with error reduction. I have also argued that the skills for reducing survey error, especially in the areas of nonresponse and measurement, are skills that come from advanced training in the social sciences--psychology and sociology in particular. If government statistical agencies are to deal effectively with error reduction there needs to be a greater presence of such persons in those agencies.

I have also argued that the scientific knowledge that undergirds survey methodology is increasingly complex, and individual professionals cannot effectively cover all areas of that knowledge. It's important that this state of affairs be realized by the leaders of government statistical agencies and existing survey methodologists.

Staffing statistical agencies with survey methodologists to meet the challenges of the 21st century requires creating a mix of theoretical and conceptual skills within methodology staffs that have not been well recognized in the past, but which is essential for designing high quality surveys that reduce error rather than only adjusting for it. Doing so poses many challenges for the current and future leadership of these agencies.

References

Bishop, G. F., Hippler, H., Schwarz, N., and Strack, F. 1988. **AA** Comparison of Response Effects in Self-Administered and Telephone Surveys.@ In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, and J. Waksberg (eds.), *Telephone Survey Methodology*. New York: Wiley.

Cialdini, Robert. 1984. *Influence: The New Psychology of Modern Persuasion*. Quill: New York.

DeMaio, T. J. 1984. **ASocial** Desirability and Survey Measurement: a Review.@ In C. F. Turner and E. Martin (eds.), *Surveying Subjective Phenomena*. Vol. 2. Russell Sage: New York..

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. John Wiley

Co.: New York.

Dillman, Don A. 1991. "The Design and Administration of Mail Surveys." *Annual Review of Sociology* 17:225-249.

Dillman, Don A. And Kent J. Miller. In Press. "Optical Scanning Technologies, Response Rates, and Data Quality in Mail Surveys by Small Survey Organizations," In Couper, Mick et.al. *Computer-Assisted Survey Information Collection*. New York: Wiley-Interscience.

Dillman, Don A., Jon R. Clark and James Treat. 1994. "The Influence of 13 Design Factors on Response Rates to Census Surveys." Annual Research Conference Proceedings, U.S. Bureau of the Census. Washington, D./C. Pp. 137-159.

Dillman, Don A., Dana E. Dolsen, and Gary E. Machlis. 1995. "Increasing Response to Personally-Delivered Mail-Back Questionnaires by Combining Foot-in-the-Door and Social Exchange Methods." *Journal of Official Statistics* 11(2):129-139.

Dillman, Don A., Roberta L. Sangster, John Tarnai, and Todd Rockwood. 1996. "Understanding Differences in People's Answers to Telephone and Mail Surveys." In Braverman, Marc T. and Jana Kay Slater (eds.), *Current Issues in Survey Research*, New Directions for Program Evaluation Series, Chapter 4:45-62. Jossey-Bass: San Francisco.

Goyder, John. 1987. *The Silent Minority; Nonrespondents on Sample Surveys*. Westview Press: Westview Press.

Groves, Robert. 1989. *Survey Errors and Survey Costs*. John Wiley Co.: New York.

Groves, Robert, Robert Cialdini and Mick Couper. 1992. "Understanding the Decision to Participate in a Survey." *Public Opinion Quarterly*, 56(4):475-495.

Groves, Robert and Mick Couper. 1996. "Household-Level Determinants of Survey Nonresponse." In Braverman, Marc T. and Jana Kay Slater (eds.), *Current Issues in Survey Research*, New Directions for Evaluation Series, Chapter 5:63-79. Jossey-Bass: San Francisco.

Heberlein, Tom A. and Robert Baumgartner. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review*, 43:447-462.

James, Jeannine M. and Richard Bolstein. 1992. "Large Monetary Incentives and Their Effect on Mail Survey Response Rates." *Public Opinion Quarterly*, 56(4):442-453.

Jenkins, Cleo and Don A. Dillman. 1995. "The Language of Self-Administered Questionnaires as Seen Through the Eyes of Respondents," in *Volume 3, Statistical Policy Working Paper 23: New Directions in Statistical Methodology*, pp. 470 to 516. U.S. Office of Management and Budget.

Jenkins, Cleo and Don A. Dillman. 1997. AChapter 7: Towards a Theory of Self-Administered Questionnaire Design.@in Lyberg, Lars, et al. *Survey Measurement and Process Quality* New York: Wiley-Interscience, pp. 165-196.

Krosnick, Jon A. and Duane Alwin. 1987. AAn Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement.@ *Public Opinion Quarterly*, 51:201-219.

Krosnick, Jon A., Sowmya Narayan, and Wendy R. Smith. 1996. ASatisficing in Surveys: Initial Evidence.@ In Braverman, Marc T. and Jana Kay Slater (eds.), *Current Issues in Survey Research*, New Directions for Evaluation Series, Chapter 3:29-44. Jossey-Bass: San Francisco.

Mason, Robert, John E. Carlson, and Roger Tourangeau. 1994. AContrast and Subtraction in Part-Whole Items.@ *Public Opinion Quarterly*, 58(4): 569-578.

Means, B., D. J. Mingay, A. Nigamand, and M. Zarrow. 1988. AA Cognitive Approach to Enhancing Health Survey Reports of Medical Visits.@ In M. Grunberg, P. Morris, and R. Sykes (eds.), *Practical Aspects of Memory: Current Research*. John Wiley: Chichester.

Rockwood, Todd H, Roberta L. Sangster, and Don A. Dillman. 1997. AThe Effect of Response Categories on Questionnaire Answers: Context and Mode Effects.@ *Social Methods & Research*, 26(1):118-140.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Academic Press: San Diego.

Schwarz, Norbert, Fritz Strack, Hans J. Hippler, and George Bishop. 1991. AThe Impact of Administration of Mode on Response Effects in Survey Research.@ *Applied Cognitive Psychology*, 5:193-212.

Tourangeau, R. and K. A. Rasinski. 1988. ACognitive Processes Underlying Context Effects in Attitude Measurement,@*Psychological Bulletin*, 103: 299-314.