



© Yan Yu

Stanley Presser

*Annual Review of Sociology*

The Role of Doubt in  
Conceiving Research:  
Reflections from a Career  
Shaped by a Dissertation

Stanley Presser

Department of Sociology and Joint Program in Survey Methodology, University of Maryland,  
College Park, Maryland, USA; email: stanleyp@umd.edu

ANNUAL  
REVIEWS **CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Sociol. 2022. 48:1-21

First published as a Review in Advance on  
February 15, 2022

The *Annual Review of Sociology* is online at  
[soc.annualreviews.org](http://soc.annualreviews.org)

<https://doi.org/10.1146/annurev-soc-080321-073017>

Copyright © 2022 by Annual Reviews.  
All rights reserved

**Keywords**

doctoral education, graduate training, failure, sociology of science, survey methodology, measurement error, questionnaire evaluation, nonresponse error, contingent valuation

**Abstract**

The doctoral dissertation often shapes the career that follows it, influencing both opportunities encountered and research conducted. This article describes the ways this has been true for me and then argues that, given the dissertation's importance, graduate programs do not focus sufficiently on strategies for conceiving research. As a result, many students flounder at the dissertation proposal stage. Drawing on the role of doubt in my career and in science more generally, I propose changes in doctoral programs to reduce the problem.

## INTRODUCTION

As this article—indeed, my career in survey research—would not have come about without Howard Schuman, I begin by noting Howard’s dissertation, “Social Structure and Personality Constriction in a Total Institution.” The title will surprise those familiar with Howard’s work because he never published from his thesis or in related areas. Thus, an influential research career may follow a dissertation that apparently leads to a dead end.

Yet such cases are atypical.<sup>1</sup> More commonly, the dissertation gives shape to a career. It is not unusual for scholars’ most widely cited publications to grow from their dissertations. Among well-known examples are Theda Skocpol’s *States and Social Revolutions* and Aldon Morris’s *The Origins of the Civil Rights Movement*.

How dissertations are conceived and the ways they shape careers are, regrettably, mostly unexplored subjects. The first part of this article recounts how my dissertation came about and the ways it influenced my later work. The second part of the article considers why many students have trouble devising a dissertation (which contributes to high rates of uncompleted degrees) and suggests changes in graduate programs to reduce the problem. The theme connecting the two parts of the article is the role of doubt (and error) in advancing knowledge.

## GRADUATE SCHOOL

My first year in the University of Michigan sociology doctoral program, I worked with William Gamson on research that culminated in his book *The Strategy of Social Protest*, a study of protest groups between 1800 and 1945. The project dovetailed with the interest in political sociology that led me to graduate school, and thus I was disappointed to learn Bill would be on sabbatical in Jerusalem the following year.

That next year I enrolled in a seminar Howard Schuman offered on public opinion, which seemed relevant to the fate of protest groups, an idea confirmed when we read Howard’s “Two Sources of Antiwar Sentiment in America” (Schuman 1972). Much of the course, however, dealt with theoretical and methodological issues. Howard’s discussion of the connections between method and substance resonated with claims about knowledge I had encountered in a philosophy of social science course my senior year in college.

At the seminar’s conclusion, Howard told me he had submitted a National Science Foundation proposal to study question wording in attitude surveys. Would I like to work with him and write my dissertation as part of the project? If the grant were awarded, it could pay my tuition and stipend. I don’t remember how I came to a decision, and speculation about why I agreed is apt

---

<sup>1</sup>The proportion of such cases among sociologists who became prominent survey methodologists, however, seems unusually high. In addition to Howard, Robert Groves and Mick Couper never published on the subjects of their dissertations (“Intra-Employer Status Mobility: The Role of the Firm in Occupational Achievement” and “Immigrant Adaptation in South Africa,” respectively). Similarly, Eleanor Singer and Nora Cate Schaeffer did not return to the subjects of their dissertations (“Birth Order, Educational Aspirations, and Educational Attainment” and “Distress in Major Adult Roles and Depression Among White Men and Women,” respectively) after publishing one article from them. The relatively recent emergence of survey methodology as a field of study may have played a role in these cases. (After reading this, Nora told me that survey methods “would not have been acceptable as a dissertation at Chicago,” and Bob said his dissertation was a way to establish legitimacy as a sociologist in case survey methodology didn’t “work out.”) Determining whether there are other clusters of such cases and, more generally, understanding the causes and consequences of abandoning one’s original research area in favor of a totally different one are, to my knowledge, problems unexamined in the sociology of science.

to be untrustworthy. But my account of the decision's consequences rests on sounder footing. Those consequences, which included devoting more time to research that did not appear in my dissertation than to research that did (making the dissertation experience inseparable from the experience of the larger project), suggest the advantages of doing a dissertation as part of a project of an experienced researcher who becomes one's mentor.

## QUESTIONS AND ANSWERS IN ATTITUDE SURVEYS

On joining Howard's project, I was a graduate student with almost no survey experience and had published a few articles in my high school newspaper. Howard was a full professor with a wealth of survey experience and had published three books and many articles in leading sociology journals. Yet these differences seemed of little concern to him. Howard treated me as though I were an equal partner. There were numerous design issues to be settled; countless choices about which questions to test and how; and innumerable decisions about conducting and then interpreting data analyses, which led to ideas for additional analyses and for further experiments. We discussed these matters extensively. In retrospect, it is obvious Howard contributed much more than I did, but he made it seem our contributions were inseparable: that although we may have had different initial thoughts, it was unclear who had come up with the idea or plan we eventually settled on.<sup>2</sup>

Likely an important reason for Howard's acting this way was his conception of research as a search for knowledge, as opposed to a method for proving what one suspects or believes. Our collaboration taught me what Howard knew: It was more rewarding to make a discovery than to demonstrate one's suspicion or belief was correct. And, as I learned from working with Howard, the chances of discovery often depend less on one's status and experience than on new ideas and not-taken-for-granted perspectives. In other words, I believe Howard was acting on the belief that discovery is promoted by seeking and encouraging the views of others that differ from your own—that making discoveries requires a paradoxical blend of self-confidence and humility.<sup>3</sup>

Over seven years we designed, conducted, analyzed, and wrote up more than 200 experiments administered in almost three dozen national surveys. The experiments examined formal item features, word choices, and question orderings. My dissertation, on the effects of middle alternatives and "don't know" options (a revision of which became two of our book's ten main chapters), was completed about halfway through the project.<sup>4</sup> After receiving my degree, I continued almost full-time on the project for another year in Ann Arbor. Then I left for an appointment at the University of North Carolina, which allowed me to work part-time (and a summer) on the project until our book (Schuman & Presser 1981) was published.

Three lessons from the research seem particularly important. First, although we found that question effects on marginals (univariate distributions) were common and frequently sizeable,

---

<sup>2</sup>During the first half of the project, Jean M. Converse was a key participant in many of these discussions. Her definitive history of survey research (Converse 1987) should be read by everyone interested in the field.

<sup>3</sup>Compare Robert M. Hauser's (2017, p. 13) reflection: "I did *not* teach students. I *worked* with them. They *taught* themselves—and me. That is what graduate training is about" (emphases in the original). I would amend only the last sentence to "That is what graduate training should be about," as frequently it is not.

<sup>4</sup>In addition to Howard and Jean, my dissertation acknowledgments mention a third project member, Jane Fountain (now Jane Barrett), who was both data manager and statistical consultant. Due to Jane's patiently showing me how I had gone astray in applying Leo Goodman's recently developed log linear analysis, I not only avoided reporting incorrect results but learned an invaluable lesson about the utility of mistakes. To understand my errors required delving more deeply into the analytic approach, which resulted in a considerably securer grasp of it than I otherwise would have had. I return to the value of error later in this article.

relations among variables (the focus of social science) were usually unaltered even in the face of large marginal shifts. Thus, what we called the assumption of form-resistant correlations, which undergirds most research, was often justified. Nonetheless there were notable exceptions where form or wording did alter associations, demonstrating this key assumption does not always hold.

Second, we showed wording or context effects were not simply artifacts of measurement; they had substantive meaning. Substance and method are interwoven, and this is clearer when method influences substantive conclusions.

Third, we found replication was critical. My favorite part of the book (though probably rarely read because it is an appendix) is titled “Mysteries of Replication and Non-Replication.” Replication failures led us to trust only results we obtained at least twice.

These lessons (the assumption of form- or method-resistant correlations is sometimes wrong, substance and method are intertwined, and replication is essential) shaped all my later work, partly by informing how I did research and partly by leading me to doubt the claims of some research (including my own), which stimulated questions and ideas for further research.

### MEASUREMENT ERROR FOR OBJECTIVE PHENOMENA

Having explored the measurement properties of attitudinal indicators, it seemed natural to turn next to the measurement properties of factual indicators—to ask whether survey respondents tell the truth. The first major attempt to answer this question dates to 1949, when the National Opinion Research Center (NORC) interviewed adults in Denver (where NORC was founded in 1941) and then checked records to assess respondents’ reports about matters including home ownership, voting, and Community Chest contribution. The estimates of misreporting varied from one-third (charitable contribution) to one-twentieth (home ownership).

In the next three decades, record-check studies for general population surveys analyzed other items and examined the relation of misreporting to respondent background characteristics. Yet none of this research investigated the degree to which misreporting was item-specific versus respondent-specific. The ubiquitous assumption of uncorrelated error terms had never been assessed in a survey of the general population.

Reanalyzing the Denver validity study (among the richest validation studies even today), I discovered that the true association between two measures (as gauged by the record-checks) predicted whether their measurement errors were related. This was both good news and bad. For 68 of the 70 pairs of items that tapped distinct constructs (e.g., contributing to charity and voting), the errors were unrelated. However, in the one instance of multiple items measuring the same construct, the assumption was violated. The errors among all seven electoral participation questions (voter registration and voting in three local elections and three national elections) were related. Moreover, the correlated errors affected the relationship of political interest to voting frequency, which was much stronger using respondent reports than records (Presser 1984a). Thus the assumption of form- or method-resistant correlations did not hold.

Similar results emerged from a conceptual replication that Michael Traugott and I conducted with the 1972–1974–1976 American National Election Survey (ANES) panel. According to record-checks, voting misreports were correlated across the three elections. Likewise, the measurement source affected an important conclusion about voting. Education was related to turnout over the three elections (net of political interest, political efficacy, and income) using respondent reports, but not the record-checks (Presser & Traugott 1992).

This analysis led to a collaboration with Santa Traugott, whose experience as ANES director of studies suggested the results of vote validation might be shaped by how the validation was conducted. We (Presser et al. 1990) constructed an index of record quality and access (e.g., whether

records were computerized and whether interviewers were able to access records or had to rely on an official to check them) and found respondent reports were more likely to be coded invalid where record quality/access was lower. Moreover, quality/access was related to respondent characteristics (for example, being Black and a large city resident), thus affecting results involving those variables in analyses based on records as opposed to survey responses. Others have continued this line of research (e.g., Berent et al. 2016), which I believe has led to a consensus that vote misreporting is often overestimated in record-check studies and that validated voting is not always more accurate than respondent reports.

After completing this work, I read Hadaway et al. (1993), who argued that US church attendance was greatly overreported in surveys. Their study drew on (a) self-reports of Ashtabula County, Ohio, Protestants and attendance counts from that county's Protestant churches and (b) estimates of the size of the Catholic population in twenty Catholic dioceses and officially reported attendance counts in those dioceses. This was a neat design, though (as the authors acknowledged) whether the results extended to America at large was unclear.

Validating reports of religious attendance in the nation as a whole seemed too difficult, so I wondered how to produce a national estimate not subject to social desirability bias. The answer that dawned on me was a time-use study conducted on Mondays. Probably not coincidentally, such a survey (conducted over two years) had recently been completed for the Environmental Protection Agency (EPA) by the University of Maryland Survey Research Center to assess exposure to harmful chemicals. As in most time-use surveys, respondents were asked to report everything they had done—and where—the day before the interview. Cases were distributed across weeks of the year and days of the week.

The estimate of Sunday religious service attendance from this 1992–1994 survey was 26%, which—after accounting for non-Sunday attendance—yielded a weekly attendance estimate of 29%. This was dramatically lower than that from both the General Social Survey (GSS) direct question (39% averaged across 1993–1994) and Gallup's direct questions (43% averaged across 1993–1994). Thus social desirability bias led to overestimating attendance by about 50% (Presser & Stinson 1998).<sup>5</sup>

Despite the marked shift in the marginals, there was—consistent with the assumption of form-resistant correlations—no effect on the relation of attendance to respondent background characteristics. The same conclusions about the links between religious attendance and variables such as age, sex, race, and education emerged from the time-use study as from the direct question studies.

By contrast, the assumption of form-resistant correlations did *not* hold for survey year. Both the GSS question (asked for almost a quarter-century) and the Gallup question (asked for a half-century) showed remarkable temporal stability, which had been interpreted as American exceptionalism because similar surveys in other Western nations showed substantial declines over time. But our analysis of earlier US time-use surveys revealed there had been a substantial decline in attendance between 1965 and 1975 followed by a smaller decline between 1975 and the 1990s. Hence measurement error had misled scholars about how religious attendance had changed in America.

Two final analyses, of practical importance given the much greater cost of time-use measures than direct questions, indicated that direct questions could avoid social desirability bias by eliminating the interviewer. A direct item in the mailed self-administered 1993–1994 Monitoring the Future follow-up of individuals who had graduated high school between 1984 and 1992

---

<sup>5</sup>Linda Stinson, then a graduate student, became involved in the project after asking whether she could play a role in research I was doing, which suggests the value of students being proactive in seeking research opportunities.

showed weekly attendance almost identical to that for the comparable group (19- to 28-year-olds who had graduated high school) in the 1992–1994 time-use survey and much lower than the 1993–1994 GSS and Gallup results for that group. In addition, the Gallup Youth Survey (conducted by interviewers) showed no change in weekly religious attendance between 1977 and 1993, whereas Monitoring the Future’s self-administered survey of high school seniors showed a decline of almost 20% over those years. These results are consistent with experimental demonstrations that self-administration reduces social desirability bias (e.g., Kreuter et al. 2008), which is a benefit of the recent extensive migration of surveys to the web.

## QUESTIONNAIRE EVALUATION METHODS

Because questions play a major role in determining measurement error, testing them is vital. For decades, however, testing consisted of what Converse and I (Converse & Presser 1986, p. 52) called “undeclared” pretests, in which respondents are not told they are participating in an evaluation of the questionnaire. As a result, conventional pretests only identify problems that interviewers note. In the 1970s and 80s, new testing methods were developed, including behavior codes (designed by Charles Cannell and his associates) and cognitive interviews (based on Anders Ericsson and Herbert Simon’s approach to studying problem solving). The latter is an example of what Converse and I called “participating” pretests, which provide richer information by asking individuals to think aloud when answering the questions or by questioning them about the questions.

By 1990, although the results of the new methods had been described, little was known about the degree to which they diagnosed different problems. Even less was known about the reliability of any of the methods, new or old. To address these issues, Johnny Blair and I compared conventional pretests, behavior codes, cognitive interviews, and expert review panels using a common questionnaire in repeated trials of each method. Among our findings was that behavior codes had the highest reliability, expert reviews identified the most problems, and there was substantial, though very far from complete, overlap in the problems identified by the different methods (Presser & Blair 1994).

To stimulate further work, I helped organize a conference on questionnaire evaluation (Presser et al. 2004). Although it succeeded in its aim, I came increasingly to doubt the assumption in most of this research (including my own) that problems identified by the methods would necessarily affect survey estimates. It seemed very likely that evaluation methods would not only miss important problems (produce false negatives) but also identify problems that would not alter estimates (false positives). Consequently, one method might identify more total problems than another but fewer that would change estimates. Methods with high false positive rates would misleadingly appear successful.

With this in mind, Aaron Maitland and I compared the different methods’ predictions of item characteristics known to affect estimates: invalidity, unreliability, and visible problems such as missing data. Our collaboration resembled my work with Howard. I suggested the idea to Aaron for his dissertation; we worked closely in designing the project; and, after Aaron received his degree, we reconsidered parts of the analysis which led to further results.

The validity study (Maitland & Presser 2019) compared the evaluations made by five methods of 44 factual questions whose level of accuracy (based on record-check studies) varied from 54% to 100%. The two automated methods (the Survey Quality Predictor and Question Understanding Aid) provided diagnoses unrelated to an item’s accuracy. The Questionnaire Appraisal System predicted item accuracy in the wrong direction, a result probably best treated as due to sampling error. Only cognitive interviews and expert reviews correctly predicted item inaccuracy, the combination of the two accounting for a third of the variation in accuracy.

The reliability study used 53 items (Maitland & Presser 2016) and the study of visible problems such as missing data used 88 items (Maitland & Presser 2018)—in each case a mix of factual and attitudinal items from a nationwide panel survey. All five evaluation methods contributed independently to predicting the panel's estimates of reliability and of problems such as missing data, yet some methods did so for only one item type. This result underscores the importance of replication with other items. Moreover, because there are various ways to conduct the non-computer-based methods (and no consensus on which is best), future research also needs to test other ways of implementing the evaluation methods. (Ironically, survey methodologists frequently fail to appreciate that their most important insight—methods shape results—applies to their own work.) Such research should lead to improvements in the methods used to evaluate survey items and thus help us better assess whether items measure what we think they do.

## NONRESPONSE ERROR

Researchers often attend less to measurement error than to nonresponse error. This may be because indicators of measurement error are usually costly (and thus unavailable), whereas a nonresponse error indicator—the nonresponse rate—is relatively inexpensive and thus commonly available. Nonresponse error is a multiplicative function of the nonresponse rate and the difference between respondents and nonrespondents, and the conventional wisdom was that the error increases as the rate increases.

Research questioning this idea can be traced to 1997 when Andrew Kohut, founding director of the Pew Research Center, asked Robert Groves and me to join him and Scott Keeter (then a political science professor working part-time at Pew) to assess the consequences of the Pew surveys' high nonresponse rates (relative to those of academic surveys). Pew surveys (like most media and commercial polls) are generally fielded over a few days, whereas comparable academic surveys typically have field periods of a month or more. Accordingly, we designed an experiment in which a diverse set of items was asked of US adults drawn from a sample randomly assigned to a field period of either 5-days or 8-weeks.

As expected, the nonresponse rates differed greatly: 64% (short) versus 39% (long). Yet the two surveys produced strikingly similar substantive results; across almost 100 items the difference averaged only 2 percentage points (Keeter et al. 2000).

Demonstrating the power of conventional wisdom, I found this outcome baffling. More generally, it seemed essential to replicate the results before trusting them. Not having resources to collect new data, I settled on another approach: comparing estimates from a low nonresponse rate survey to those from the same survey restricted to the cases completed prior to various dates in the field period corresponding to increasingly higher nonresponse rates.

The Michigan Survey of Consumer Attitudes (SCA), which Howard and I used for many of our experiments, seemed ideal for this purpose. It had a relatively low nonresponse rate, retained call-record information, and—having been conducted for decades—enabled replications as well as estimates of change over time. I thought both Richard Curtin, the SCA's economist director, and Eleanor Singer, with whom I was studying cooperation on the decennial census (Presser et al. 2000), would be interested in collaborating. The result was Curtin et al. (2000), which, counter to my expectations, replicated the Pew findings: There was little difference in conclusions about the Index of Consumer Sentiment (including its change over time) after increasing the nonresponse rate by 5 to 50 percentage points!

Soon the Pew and SCA articles were widely cited as challenging the conventional wisdom that a high nonresponse rate signals high nonresponse bias. The central problem of the conventional wisdom is its conception of nonresponse bias as a survey attribute instead of an attribute of a survey statistic. In retrospect it is clearer that nonresponse bias can vary across a survey's

measures and therefore may not be predicted by the nonresponse rate, which—for any given survey—is a constant. To demonstrate substantial nonresponse bias (which now was important to me since some readers interpreted the Pew and SCA results to mean nonresponse could be ignored) requires a measure strongly related to the likelihood of becoming a respondent (versus a nonrespondent).

After the Pew and SCA articles were published, I read *Bowling Alone* (Putnam 2000), which argued that, although most indicators of social capital in the United States declined in the 25 years after 1975, volunteering increased substantially. Putnam's explanations for this anomaly were unpersuasive to me. I searched for recent volunteering estimates and found a 1996 Gallup poll (Putnam's source) reporting 49%, the 1996 National Household Education Survey (NHES) reporting 39%, and the 2002 Current Population Survey Volunteering Supplement (CPSVS) reporting 28%. This variation seemed clearly artifactual, reflecting differences in the surveys' methods. Recalling what I knew about the surveys suggested an explanation. The surveys had dramatically different nonresponse rates (CPSVS, 18%; NHES, 41%; and Gallup, over 60%), with higher nonresponse corresponding to higher volunteering estimates. Because nonresponse rates had skyrocketed between 1975 and 2000, the volunteering increase Putnam reported may have represented only rising survey nonresponse.<sup>6</sup>

A conversation with Katharine Abraham led to our collaborating to assess whether volunteering estimates were in fact subject to serious nonresponse bias. When she was Commissioner of the Bureau of Labor Statistics, Katharine was responsible for development of the American Time Use Survey (ATUS), which, as an unanticipated result of budget constraints, is well-designed to gauge nonresponse bias. To save money, the ATUS sample is drawn from households interviewed in the prior year's Current Population Survey (CPS) (and yields a much higher nonresponse rate than the CPS). Consequently, data are available from the CPS about both ATUS respondents and nonrespondents. We hypothesized that ATUS respondents would be much more likely than ATUS nonrespondents to have been volunteers. The results showed exactly that: 36% of ATUS respondents reported volunteering in the CPSVS versus 20% of ATUS nonrespondents. As the ATUS nonresponse rate was 47%, nonresponse bias inflated the volunteering estimate by 7 percentage points, producing a 24% overestimate ( $\frac{36}{36-7}$ ).<sup>7</sup>

Despite its substantial impact on marginals, nonresponse had no effect on cross-sectional models of volunteering. Conclusions about the impact of respondent characteristics (e.g., age, gender, education, and income) were unaffected by whether nonrespondents were included or excluded (Abraham et al. 2009). Thus the assumption of form- or method-resistant correlations held. Indeed, the overall pattern resembled that for religious attendance. Both measurement bias (for religious attendance) and nonmeasurement bias (for volunteering) affected marginals and estimates of change over time (misleading religion scholars in one instance and social capital scholars in the other) but not the associations between the measures and respondent background characteristics.

In our conclusion, we noted that volunteering's strong relationship to both response propensity and many survey variables suggested that adding volunteering to conventional nonresponse weights (using the annual CPSVS benchmark) would improve survey estimates. Subsequently, Peytchev et al. (2018) showed that incorporating volunteering in weights reduced the mean square error of GSS estimates (on average by 27%) compared with weights using only

---

<sup>6</sup>Between 1979 and 2003, SCA's nonresponse rate rose from 28% to 52% despite interviewer calls more than doubling (Curtin et al. 2005). As Gallup may not have increased effort to the same extent, its nonresponse may have increased even more.

<sup>7</sup>This result, which has a very small sampling error due to the very large  $n$ , was foreshadowed by both Groves et al. (2000) and the Pew study (for details, see Abraham et al. 2009, n. 4).

demographics—and argued that the utility of weights might often be enhanced by adding variables beyond demographic ones.

## ENVIRONMENTAL ECONOMICS

Economists have increasingly relied on surveys (Presser 1984b), and this is particularly true in environmental economics to measure the value of public goods or services (those not traded in markets). An approach often used is contingent valuation (CV), which, to oversimplify, asks respondents how much they would be willing to pay for a good or service. Environmental sociologist Robert Mitchell and economist Richard Carson published a state-of-the-art review of CV just before the 1989 Exxon Valdez oil spill. Shortly afterwards, the state of Alaska retained Mitchell, Carson, three other economists (W. Michael Hanemann, Raymond Kopp, and Paul Ruud), and me to design a CV study to estimate the dollar value of the spill's damages. Our study brought the measurement of nonuse value (in this case of the difference between Prince William Sound pre- and post-spill) to the attention of policy makers.

The study (later published as Carson et al. 2003) was criticized by scholars retained by Exxon, who argued CV could not provide trustworthy results. To assess the matter, the National Oceanic and Atmospheric Administration (NOAA) appointed a panel cochaired by Nobel laureates Kenneth Arrow and Robert Solow that included three other economists and Howard Schuman (whom I suggested to provide a survey perspective). The panel concluded that CV could measure nonuse value and proposed requirements for doing so.<sup>8</sup>

By then the Exxon litigation appeared to have settled, and our team (joined by psychologist Jon Krosnick and economist V. Kerry Smith) was retained by NOAA to conduct a CV of the damages (most notably the loss of bald eagles and peregrine falcons) arising from the world's largest DDT disposal site. In planning the new study, we responded to the NOAA panel's concerns that CV results might be sensitive to transitory media coverage and that CV studies ought to offer an explicit "don't know" option in the willingness-to-pay question. After media coverage of the Exxon spill had largely disappeared, we repeated the Alaska study and included an experiment on the effect of an explicit "don't know" option (reminiscent of my dissertation), which we had not offered in the original survey. The results showed the damages estimate was unchanged after two years—i.e., was highly reliable (Carson et al. 1997)—and including an explicit "don't know" did not alter estimates of willingness to pay or other important results (Carson et al. 1998).

In a prefatory note to our Alaska CV article, the journal's coeditors wrote, "the shift in magnitude in potential liability, associated with the use of CV in the Exxon Valdez case had the outcome an economist would have expected; injuries from oil spills in the United States have substantially decreased while there has been little change from the previous trend elsewhere in the world" (Turner & Bateman 2003, p. 255). I don't know how this judgment held up after the 2010 BP blowout in the Gulf of Mexico. But even before the BP well was sealed, most of the team reassembled and, joined by new members, spent nearly five years preparing for NOAA a monetary estimate of the resulting damages (Bishop et al. 2017). As was true of the earlier cases, the BP litigation settled before trial. Although never presented in court, our research in these cases—dealing with difficult measurement issues, the solutions to which were made possible by the extraordinary resources available to us because of the billions of dollars at stake—improved the CV method used across the world.<sup>9</sup>

<sup>8</sup>Fourcade (2011) provides a sociological analysis of CV.

<sup>9</sup>With funds awarded from litigation following a Shell Oil spill in California, we also designed a CV study to serve as a reference for future oil spills in that state (Carson et al. 2004).

## RELATED ACTIVITIES

In addition to enabling my contributions to valuing the natural environment, becoming a survey methodologist led to my involvement in other policy-related litigation in which surveys played a key role. These included large class-action lawsuits about discrimination against women in employment, deceptive advertising of lite cigarettes, and the attempt to add a citizenship question to the 2020 Census. Likewise, I have applied my survey experience to governmental matters in nonjudicial settings, for example, designing a survey of employers for the congressionally mandated evaluation of the Immigration Reform and Control Act of 1986 and providing advice as a member of National Academy of Sciences' panels on issues including "Measuring Dimensions of Social Capital to Inform Policy" and "Rationales and Approaches for a U.S. Program of Human Space Exploration."

The challenges of this applied research have often been as great as—if not greater than—those of basic research, and they deepened my understanding of research in various ways. For instance, the requirement in most litigation not to identify the survey's sponsor nor purpose and, in some cases, not to destroy respondent identifiers enhanced my appreciation of the meaning of confidentiality and informed consent (Presser 1994). Because the consequences of these applied inquiries usually dwarfed those of basic research (generally leading to more rigorous scrutiny than scientific peer review), I became more careful in all my work. The research frequently involved collaborating with scholars from other disciplines, which exposed me to new perspectives and heightened my sense that familiarity with other fields can be an invaluable ingredient in solving sociological problems. The work expanded my horizons in other ways as well, provided additional income, and—by bringing high quality data and analysis to bear on important issues—contributed to more informed public policy decisions.

My administrative activities that stemmed from being a survey methodologist were similarly rewarding, sometimes in unexpected ways. Editing *Public Opinion Quarterly* (a major forum for survey methods research) resembled teaching in the sense of helping authors improve their papers while learning from them. Serving on the American Association for Public Opinion Research council probably contributed to Andy Kohut's inviting me (shortly after he was president and I was past president) to join the Pew nonresponse project. Directing the University of Maryland Survey Research Center when it conducted the EPA survey almost certainly inspired my realization that time-use studies would measure religious attendance without social desirability bias. Finally, and perhaps most importantly, joining Bob Groves and Graham Kalton to create the Joint Program in Survey Methodology led to my collaborations with many who became its faculty and students.

## TROUBLE CONCEIVING A DISSERTATION

That my career has been shaped by my dissertation is not unusual, though the connection was strengthened because the dissertation involved working with Howard on the larger project. Such an arrangement has various advantages, but numerous students will not have the opportunity to join professors' projects or be interested in those available.<sup>10</sup> Although many students who work on their own succeed in devising a dissertation proposal, many others have little sense of what to propose, find it hard to choose among possibilities, or focus on one possibility yet have trouble developing it (sometimes due to resource constraints). High attrition and lengthy time-to-degree—according to the best estimate, only a little more than half of US doctoral students complete their

---

<sup>10</sup>Noy & Ray's (2012) study suggests some students (e.g., women of color) may be less likely to be offered this kind of mentoring.

degree 10 years after enrolling (Sowell et al. 2008)—are due partly to the difficulties many students experience at the proposal stage.<sup>11</sup> In turn, these difficulties likely stem in part from a problem in doctoral education.

## DOCTORAL EDUCATION'S FLAW

Although one might expect education for a doctoral degree to be very different from that for a bachelor's, in key respects this is not true in US universities. Many doctoral courses are similar to those for upper-level undergraduates. Both feature recent research—commonly university press books and articles from leading journals—and consider methods and theories.<sup>12</sup> Likewise, the comprehensive examination at the end of doctoral coursework often involves summaries and critiques of subfields comparable to those required on exams in advanced undergraduate courses.

The doctoral task that differs most from undergraduate tasks is proposing research to make a new contribution. Remarkably, however, coming up with such a proposal is not something for which most doctoral programs provide formal guidance. Students read numerous studies that contribute to knowledge and are trained in the methods for conducting them. But they are typically given little insight into where the ideas or problems for such studies came from or what considerations led researchers to choose one idea or problem over another. This helps explain why many students struggle in fashioning a dissertation proposal.

The absence of training for coming up with research proposals dovetails with how little has been written about it. Although there is a large literature about the dissertation and a modest one on the proposal, this work usually assumes the student has already developed a problem or gives cursory attention to how to do so. *Writing a Proposal for Your Dissertation: Guidelines and Examples* by Steven Terrell (2016) is typical. Only 4 of its almost 300 pages deal with devising a problem. Terrell advises reading the suggestions for future research at the end of publications, consulting with experts, and attending conferences. These are sensible recommendations, though often insufficient.<sup>13</sup>

Why is so much less attention paid to formulating research proposals than to conducting research? Hans Reichenbach's (1938) distinction between the context of discovery and the context of justification suggests an answer. As Sprague & Zimmerman (1983, p. 259) note,

The context of discovery refers to the origin of research questions. The context of justification refers to the set of processes through which research questions are tested. The context of justification [follows] the rule of the scientific method. . . the context of discovery [involves] the seemingly idiosyncratic and mysterious process through which research questions and creative, novel hypotheses emerge.

Although most philosophers no longer see a sharp distinction between the two contexts (Schickore 2018), teaching how to come up with ideas is far more difficult than teaching how

---

<sup>11</sup>Though not completely comparable because of differences in study designs, estimates of completion in Canada (Elgar 2003), the United Kingdom (High. Educ. Funding Counc. Engl. 2005), and Australia (Martin et al. 2001) are similar to those in the United States. For reflections on the Australian system, see Connell (1985), and for a more general analysis of US attrition, see Lovitts (2001).

<sup>12</sup>The assumption that upper-level undergraduates should read the same work as graduate students is conveyed in book reviews. Two recent examples are Ryan Light's (2019, p. 855) judgment in the *American Journal of Sociology*, "[This book is] appropriate for advanced undergraduate and graduate classes on social networks or computational social science," and Elizabeth Armstrong's endorsement on the back cover of Martin (2017), "I look forward to assigning [this book] in both undergraduate and graduate methods courses."

<sup>13</sup>The most detailed and helpful suggestions I have found are those of Booth et al. (2016, chapters 3–4).

to test ideas. Nonetheless, much can be done to reduce the “idiosyncratic and mysterious process” of formulating research questions and thus aid students in conceiving proposals.

If one’s objective is to make an original contribution to knowledge, it helps to understand the nature of science and how it advances. Paradoxically, however, understanding scientific progress—contributions “fruitful of further experimentation [or] observations” (Conant 1951, p. 25)—means understanding scientific failure, because doubt about what science has concluded often drives progress. Failure and doubt suggest lessons for how to imagine research proposals.

## FAILURE IN SCIENCE

The most publicized failure of science is “the replication crisis.” Ioannidis (2005) famously argued that common scientific practices—especially misuse of statistical significance measures—mean most published medical research (genetics to epidemiology) is wrong. Although his conclusion was based on logic and simulations, later replications of published work showed high rates of failure to reproduce the original results. A National Academies report (Natl. Acad. Sci. Eng. Med. 2019a, p. 358) found that “issues of reproducibility occur in all domains of science,” including the social sciences (Chang & Li 2015, Open Sci. Collab. 2015).<sup>14</sup>

The replication crisis has led to proposals for changing scientific practices (e.g., Simmons et al. 2011, Natl. Acad. Sci. Eng. Med. 2019b). Although some of the remedies are controversial, others are not and are beginning to be incorporated into graduate training. Yet none of the remedies will eliminate widespread, serious error because error in science arises from a source more fundamental than those that produced the replication crisis.

The more basic issue, characterizing both quantitative and qualitative research, is science’s tentative nature. George Polya (1954, p. v; italics in original) observed:

all our knowledge outside mathematics and demonstrative logic (which is, in fact, a branch of mathematics) consists of conjectures. . . . We secure our mathematical knowledge by *demonstrative reasoning* but we support our conjectures by *plausible reasoning*. . . . Demonstrative reasoning is safe, beyond controversy, and final. Plausible reasoning is hazardous, controversial, and provisional.

Since science has an “eternally provisional nature” (Polanyi 1946, p. 43), its results will often need more than qualification. Discovery, Ludwig Fleck (1979, p. 30) argued, is “inextricably interwoven with what is known as error. To recognize a certain relation, many other relation[s] must be misunderstood, or overlooked.”

Because all scientific knowledge is uncertain, Richard P. Feynman (1998, pp. 27–28) argued that

it is of paramount importance, in order to make progress, [to] recognize. . .doubt. Because we have the doubt, we then propose looking in new directions for new ideas. . . . [I]f we did not have a doubt or recognize ignorance, we would not get any new ideas. There would be nothing worth checking, because we would know what is true. . . . Doubt is . . . of very great value.

More than 50 years later, Gian Giudice, head of the theory department at the European Organization for Nuclear Research, expressed this idea somewhat differently (Overbye 2017, p. D1):

---

<sup>14</sup>Freese & Peterson (2017, p. 148) note that sociologists Raymond Mack and Arnold Rose emphasized the importance of replication in the early 1950s. In 1970, Denton Morrison and Ramon Henkel brought together a series of papers by statisticians, sociologists, and psychologists in *The Significance Test Controversy* (Morrison & Henkel 1970), which likewise anticipated recent concerns. Psychologist Paul Meehl’s critique (e.g., Meehl 1967, 1990) was particularly trenchant. Concerns were also expressed by scholars in other disciplines [e.g., the economist D.N. McCloskey (1985)]. Why these concerns failed to resonate until recently is a question ripe for investigation by the sociology of science (an inquiry for which Freese and Peterson’s review provides clues).

“It’s a high point of research when we have confusion. . . . Confusion means an opportunity for new ideas.” Just as doubt and confusion lead to new discoveries, new discoveries lead to doubt and confusion. This is conveyed in paleoanthropologist Susan Antón’s reaction to the unearthing of a new species of ancient humans (Wade 2019, p. 108): “we know a lot less [about human evolution] than we thought we did.” Thus, in science, as Max Weber (1946, p. 138) noted,

each of us knows that what [s]he has accomplished will be antiquated in ten, twenty, fifty years. That is the fate to which science is subjected; it is the very meaning of scientific work. . . . [which] asks to be “surpassed” and outdated.

An awareness of the limits of what we know, an awareness that what we think we know may not be so, and—maybe most fundamentally—an awareness of the importance of embracing not knowing are central not only to science but to other modes of knowing. Observations like those of Feynman, Giudice, and Antón are common in the arts, humanities, and business, illustrating Polanyi’s (1966, p. 82) claim that the foundations of science “broadly hold also for all other creative systems of the modern mind.” Consider just a few examples. Novelist Siri Hustvedt (2016, pp. 149–50) recognized:

Doubt is fertile because it opens a thinker to foreign thoughts. Doubt is a question generator. . . . One of the few universals when it comes to ideas is that questions are normally better than answers.

T.S. Eliot (1971, p. 127) reflected:

In order to arrive at what you do not know  
You must go by the way which is the way of ignorance.

General Motors CEO Alfred P. Sloan (Drucker 1974, p. 472) observed:

I take it we are all in complete agreement on the decision here. . . . Then I propose we postpone further discussion of this matter until our next meeting to give ourselves time to develop disagreement and perhaps gain some understanding of what the decision is all about.

Nobel literature laureate Wislawa Szymborska (1996) noted:

Whatever inspiration is, it’s born from a continuous “I don’t know.”

A quote sometimes attributed to the Canadian poet Bliss Carman declares:

What are facts but compromises? A fact merely marks the point where we have agreed to let investigation cease. Investigate further and your fact disappears.<sup>15</sup>

---

<sup>15</sup>This quote (evoking Goethe’s maxim “every fact is already a theory”) appeared in “On Having Known A Poet” (in the May 1906 issue of the *Atlantic Monthly*) attributed to the unnamed poet of the title referred to as a friend by the article’s author, who—in turn—was not named by the *Atlantic*. Hence, the quote is sometimes attributed to Anonymous (e.g., Sullivan 2019, p. 100), but also—without a stated rationale—to Carman (e.g., Bohle 1967, p. 90). Sorfleet (1990, p. 198), again without a stated rationale, attributes the *Atlantic* article (not the quote) to Peter McArthur, who, according to *The Dictionary of Canadian Biography*, was Carman’s friend. This is consistent with the claim of James Cappon, who also knew Carman (Bentley 2004, p. 100; Campbell 2013). Cappon (1929, pp. 656–57) refers to the *Atlantic* article as “McArthur’s sketch” of Carman and observes about the “facts as compromises” quote: “The report is not perhaps meant to be verbally accurate but only

Given the central roles of error and ignorance in human inquiry, it is astonishing how little of this is conveyed to doctoral students. The emphasis in graduate training is mainly on what is known, with limited attention to the fragility of that knowledge. To prepare for a research career, students should understand the ubiquity and importance of error as well as what microbiologist Martin Schwartz (2008, p. 1771) calls “the importance of stupidity in scientific research.” Yet this, Schwartz notes, “can be difficult for students who are accustomed to getting the answers right.” Thus he observes “education might do more to ease what is a very big transition: from learning what other people once discovered to making your own discoveries.”

Schwartz’s (2008, p. 1771) one-page essay was not designed to offer suggestions for improving the training of students, but it provides a useful illustration:

Preliminary and thesis exams have the right idea when the faculty committee pushes until the student starts getting the answers wrong or gives up and says, “I don’t know.” The point of the exam isn’t to see if the student gets all the answers right. If they do, it’s the faculty who failed the exam. The point is to identify the student’s weaknesses, partly to see where they need to invest some effort and partly to see whether the student’s knowledge fails at a sufficiently high level that they are ready to take on a research project.

To convey the fundamental importance of both error and not knowing and therefore help more students come up with productive research proposals, changes in doctoral programs are needed.

## IMPROVING DOCTORAL EDUCATION

Courses might be revised in various ways.<sup>16</sup> Some could compare extensively cited and rarely cited articles from similar years, topics, and journals, which would stimulate discussion about the character of influential research. Others could examine earlier state-of-the-art research whose conclusions are no longer accepted and analyze how and why the consensus changed, providing a foundation for diagnosing problems in current research. Discussion of current work’s problems would likewise be promoted by assigning readings at odds with the prevailing consensus. The organization of courses might be informed by the insight of cosmologist John D. Barrow (1998, p. 115):

It is in the process of first learning the subject that you are most likely to have new ideas about it. Once you allow that process to be entrained by an influential standard approach invented by someone else, you are relinquishing an opportunity to see it in a new way.<sup>17</sup>

Courses might also adopt the suggestion of physicist Edwin T. Jaynes (1993, p. 262) to cover

only a few problems, but in . . . real depth. It doesn’t even matter very much what those few problems are; once a student knows what it feels like to analyze something in depth, [s]he can do it for h[er]self on whatever other problems may come h[er] way. Equally important, [s]he can recognize in the work of others the distinction between a superficial study and one that is deep enough to be capable of finding new things.

---

an echo of many such utterances. Peter is not quite a first class reporter in that line, and in his sketch of the poet has mixed in traits from Roberts and D.C. Scott [two other Canadian poets of the period] with the idea of making it more interesting.” Thus the attribution to Carman exemplifies the quote. (For a wide-ranging history of the meaning of fact, see Poovey 1998.)

<sup>16</sup>Some of these approaches are not new, though I believe few courses use more than one or two of them and many use none.

<sup>17</sup>Thus Zuckerman & Merton (1972, p. 313) observe that “naïveté and focused ignorance evidently have their functions in science.”

Perhaps most importantly, courses ought to search for and focus on puzzles (within and across readings), because puzzles, highlighting what we don't know, stimulate productive inquiry.<sup>18</sup>

Many students would also benefit from reading more critically. It is no surprise when students accept what they read at face value as most of their readings have undergone extensive review (and have been judged suitable to assign). Having worked to understand the conclusions and how they were arrived at, students are often tempted to assume they have mastered the material. But, it is at this point that the harder work of evaluation—motivated by the assumption that all research is flawed or limited—should begin.

Telling students to read more critically is generally insufficient, however, as it is easier to recommend than to accomplish. To make critical reading less difficult, I have suggested that students first read only the statement of the problem(s) or question(s) the study sets out. Next, not knowing what was actually done to examine the problem(s) or answer the question(s), they should develop an ideal approach to meet the study's objectives. At that point, they should read the rest of the study. Since the study's approach will probably differ in major ways from the imagined ideal, one then has a useful perspective from which to think about the study's limitations or flaws.

Another way to improve students' critical thinking is to ask them to reproduce a table from a publication of their choice that analyzed archived data. Over many years of assigning this exercise, my students have often been unable to produce the published numbers, typically because authors were unclear about what they did. In addition to revealing the importance of documentation, the exercise reveals the myriad decisions (none of which has a single correct choice) required to conduct research and how each may shape results. The lesson is frequently underscored by the second part of the exercise, which is to develop, and then test, a hypothesis about how the original finding will be affected by changing one or more decisions.<sup>19</sup>

Given the importance of understanding that all results are shaped by the procedures used to produce them, it is troubling how often this is not highlighted in methodology courses. The problem is partly that social scientists typically conceive of methodology in terms of deriving information from data (i.e., applied statistics), as opposed to deriving data from the world (i.e., data collection). Yet even instruction about the latter is bedeviled by the misleading nature of the metaphor "data collection," since data are not collected but instead produced or manufactured in a social process. Methodology courses need to stress this latter idea, which, when understood, engenders a critical view of the results of all data analyses.<sup>20</sup>

Similarly, theory courses should convey the view of standpoint theories (e.g., Collins 1986, Harding 1986) that all knowledge is socially situated—both the language we use to think about research and the questions we ask (which influence what we learn) are shaped by our social positions and the social structures of which those positions are part. This perspective facilitates both evaluating research critically and identifying important questions and problems that research has not yet addressed (or even raised). It also underscores the need for greater

---

<sup>18</sup>Merton (1948, p. 506) offers related observations; Sandberg & Alvesson (2011) and Gustafsson & Hagstrom (2018) consider "gap-spotting" and puzzle construction as ways to develop research proposals; and Zuckerman (2017) provides an in-depth analysis of the framing of research puzzles.

<sup>19</sup>Lucas et al. (2013, p. 230) note the implication for theory of this methodological lesson: "Strict methodological replications can produce evidence that the results of a prior study were not due to chance, but they do not help resolve the issue of whether empirical indicators are accurate reflections of theoretical concepts. Only tests with diverse indicators provide this opportunity to advance theory."

<sup>20</sup>Robert Groves (personal communication) aptly expressed an implication of this point: Until you try to develop measures of a construct you are susceptible to the misbelief you understand it.

efforts not just to recruit students from underrepresented groups, but to promote their success. As chemist Joseph DeSimone (quoted by Malcolm 2019, p. 1; italics added) argues,

There is no more fertile ground for innovation than a diversity of experience [which]... arises from a difference of cultures, ethnicities, and life backgrounds. A successful scientific endeavor is one that attracts a diversity of experience, draws upon the breadth and depth of that experience, *and cultivates those differences, acknowledging the creativity they spark.*<sup>21</sup>

In addition, programs ought to emphasize the connections between success and failure. Consider psychologist Jan Van Bavel, who, in his lab meetings (Kafka 2018)

regularly celebrat[ed] successes... they'd revel over a publication accepted, a dissertation defended, a student award. He... realized that might give younger researchers, particularly, a false sense of what life as a researcher is really like. So now... grad students, postdocs, and undergrads also share their defeats. "I decided to flip the script," Van Bavel says. "We're also going to announce the failures, because they are far more common and... also part of the process... I wanted... to make it transparent how common failure is... Those first few years are full of failure and criticism... Opening up the conversation normalized the process and created an instant... support session. It also sparked a conversation about how we all deal with rejection."

In this example, focusing on failure is a way to help students deal with their inevitable setbacks.<sup>22</sup> Failure should also be presented as essential to scientific progress. Instead of being disappointed when their hypotheses are not confirmed, for example, students would benefit from understanding that disconfirmed expectations are more likely than those that are confirmed to lead to puzzles, thereby increasing the odds of discovery.<sup>23</sup> To convey the roles of failure and error in science, readings in the history, philosophy, and sociology of science (including work on the replication crisis) ought to be central to introductory pro-seminars.

Pro-seminars might likewise introduce students to the autobiographical articles in the *Annual Review* volumes for the various social sciences and to the essay collections portraying how and why social scientists developed research problems (e.g., Hammond 1964, Riley 1988, Glassner & Hertz 2003, Khan & Fisher 2014, Hunter 2018), which often address students' concerns about connecting research to their lives. Similarly, instructors might look for opportunities in their classes to describe how they devised problems and why they pursued some and abandoned others.

---

<sup>21</sup>Lawrence Bobo (personal communication) noted, "[In] focus group discussions with white participants, I have never heard a view expressed that isn't available in the published survey based literature [but]... just the opposite is true for minority focus group participants whether Black, Latinx, or Asian [from whom] something NOT already widely examined in the literature has almost always emerged." Other approaches to thinking critically about research are described by Becker (1998, p. 7), who offers "ways of expanding the reach of our thinking, of seeing what else we could be thinking and asking," and by Martin (2017).

<sup>22</sup>As a graduate student, I learned Otis Dudley Duncan—just elected to the National Academy of Sciences—had a paper rejected by *Public Opinion Quarterly* (POQ). When I later met the same fate at POQ, instead of wondering if I were meant to be a researcher, I felt (albeit for the first and last time) in the same league as Duncan. The importance of knowing of others' failures is suggested by Crittenden & Wiley's (1980) finding that making external (versus internal) attributions in response to a journal rejection was related to making further efforts to publish the manuscript.

<sup>23</sup>"There is no surer way to screw up an experiment than to be certain of its outcome" is how a related idea was expressed by Stuart Firestein (2012, p. 17), the author of two excellent books (Firestein 2012, 2016) on the integral roles of ignorance and failure in science. More generally, "a theory is confirmed by observing things which it predicts *that are otherwise unexpected*" (Jaynes 1993, p. 274; italics in original). The aim is "discovering something [you] hadn't known in advance" (Smith 2003, p. 155), an idea mined by Fleck (1979, pp. 85–86).

Despite the fact that students frequently learn a great deal from each other, many aspects of doctoral programs (e.g., course assignments done individually and conference travel funds awarded only to sole-authored papers) discourage collaborative work among students. Programs could provide more opportunities and support for such work. This seems especially important now that the Internet (not to mention COVID-19) has dramatically reduced the likelihood of people working in their university offices, thereby decreasing the informal interactions that often play a central role in doctoral education.

Finally, students can be directly helped in formulating dissertation proposals. Advisors ought to discuss dissertations with their advisees no later than the end of their first year. Dissertation ideas could be explored in many courses. Successfully defended dissertation proposals should be made easily accessible to students. And seminars to develop dissertation proposals might be offered—possibly jointly by two or more departments, as participants from different fields are apt to assist students to see their problems in new ways.<sup>24</sup> Even two instructors from one department will often disagree in ways that benefit students.<sup>25</sup> Structuring such seminars so that students critique each other's work might also lead to the formation of dissertation support groups, thereby reducing the isolation many students feel when writing their dissertation as well as yielding better and more quickly completed dissertations.

## CONCLUSION

Donald Rumsfeld famously observed:

There are known knowns; these are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. . . . It is the latter category that tend to be the difficult ones.

But equally difficult may be the known knowns that are wrong or misleading. These may be as common in science as in other areas of life. My research has often dealt with such knowledge—for instance, that US weekly religious attendance was stable over the half-century following World War II, that Americans' volunteering increased markedly in the last quarter of the twentieth century, and that increasing survey nonresponse rates signal increasing nonresponse bias. Doubting the known knowns—not only in the work of others but in one's own work as well—is central to the scientific enterprise. By emphasizing the importance of doubt in the development of knowledge, graduate programs might enable more students to complete their degrees, increase the likelihood that they see ways to advance their fields later in their careers, and help them train succeeding generations of students to do likewise.

---

<sup>24</sup>More programs should also consider requiring doctoral students to take courses in related fields. Because of the University of Michigan's graduate school cognate requirement, I took a cultural anthropology course taught by Marshall Sahlins and a cognitive psychology course taught by Robert Zajonc. They deepened my understanding of social science and made me more open to interdisciplinary collaboration, which has often been key to advancing knowledge.

<sup>25</sup>Reeve Vanneman (personal communication) noted, "Exposing students to respectful disagreements is one of the most important lessons we can leave them." On the idea of a co-taught seminar to design research, Michael Hout (personal communication) observed that before he moved from Berkeley to NYU, Neil Fligstein and Jeff Manza concluded that Princeton students were excelling on the job market because of Princeton's research and writing workshop and thus both departments adopted the idea: "It's expensive. Two colleagues team-teach a two semester—second semester of second year plus first semester of third year—course that emphasizes research design in the first and writing up results in the second. But it makes a huge difference, shortening time to proposal, to first publication, and to degree."

## DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

I thank Karen Cook and Douglas Massey for inviting me to prepare this article and Katharine Abraham, Christopher Antoun, Lawrence Bobo, Amy Corning, Dawn Dow, Robert Fay, Robert Groves, Michael Hanemann, Michael Hout, Meyer Kestnbaum, Jon Krosnick, Rashawn Ray, Nora Cate Schaeffer, and Reeve Vanneman for helpful comments on earlier versions. Preparation of the article was aided by a University of Maryland sabbatical, during which I was warmly hosted by Darren Pennay and his colleagues at Australia's Social Research Centre. For that experience living abroad (as well as for so much else that has enriched my life) I am grateful to Yan Yu.

## LITERATURE CITED

- Abraham KG, Helms S, Presser S. 2009. How social processes distort measurement: the impact of survey nonresponse on estimates of volunteer work in the United States. *Am. J. Sociol.* 114:1129–65
- Barrow JD. 1998. *Impossibility: The Limits of Science and the Science of Limits*. Oxford, UK: Oxford Univ. Press
- Becker HS. 1998. *Tricks of the Trade: How to Think about Your Research While You Are Doing It*. Chicago: Univ. Chicago Press
- Bentley DMR. 2004. *The Confederation Group of Canadian Poets, 1880–1897*. Toronto: Univ. Toronto Press
- Berent MK, Krosnick JA, Lupia A. 2016. Measuring voter registration and turnout in surveys. *Public Opin. Q.* 64:413–42
- Bishop RC, Boyle KJ, Carson RT, Chapman D, Hanemann WM, et al. 2017. Putting a value on injuries to natural assets: the BP oil spill. *Science* 356:253–54
- Bohle B, ed. 1967. *The Home Book of American Quotations*. New York: Dodd Mead
- Booth WC, Colomb GC, Williams JC, Bizup J, Fitzgerald WT. 2016. *The Craft of Research*. Chicago: Univ. Chicago Press. 4th ed.
- Campbell S. 2013. *Both Hands: A Life of Lorne Pierce of Ryerson Press*. Montreal and Kingston, ON: McGill-Queen's Univ. Press
- Cappon J. 1929. Bliss Carman's beginnings. *Queen's Q.* 36:637–65
- Carson RT, Conaway MB, Hanemann WM, Krosnick JA, Mitchell RC, Presser S. 2004. *Valuing Oil Spill Prevention*. Philadelphia: Kluwer
- Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, et al. 1997. Temporal reliability of estimates from contingent valuation. *Land Econ.* 73:151–63
- Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, et al. 1998. Referendum design and contingent valuation. *Rev. Econ. Stat.* 80:484–87
- Carson RT, Mitchell RC, Hanemann WM, Kopp RJ, Presser S, Ruud P. 2003. Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. *Environ. Resour. Econ.* 25:257–86
- Chang AC, Li P. 2015. *Is economics research replicable? Sixty published papers from thirteen journals say "Usually not."* Work. Pap., Finance Econ. Discuss. Ser. 2015–083, Federal Reserve Syst., Washington, DC
- Collins PH. 1986. Learning from the outsider within: the sociological significance of Black feminist thought. *Soc. Problems* 33:S14–32
- Conant JB. 1951. *Science and Common Sense*. New Haven, CT: Yale Univ. Press
- Connell RW. 1985. How to supervise a PhD. *Vestes Aust. Univ. Rev.* 28(2):38–41
- Converse JM. 1987. *Survey Research in the United States: Roots and Emergence 1890–1960*. Berkeley: Univ. Calif. Press
- Converse JM, Presser S. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire*. Thousand Oaks, CA: SAGE
- Crittenden KS, Wiley MG. 1980. Causal attribution and behavioral response to failure. *Soc. Psychol. Q.* 43:353–58

- Curtin R, Presser S, Singer E. 2000. Effects of response rate changes on the Index of Consumer Sentiment. *Public Opin. Q.* 64:413–28
- Curtin R, Presser S, Singer E. 2005. Changes in telephone nonresponse over the past quarter century. *Public Opin. Q.* 69:87–98
- Drucker PF. 1974. *Management: Tasks, Responsibilities, Practices*. Portsmouth, NH: Heinemann
- Elgar FJ. 2003. *Ph.D. completion in Canadian universities: final report*. Rep., Grad. Stud. Assoc. Can. and Dalhousie Univ., Halifax, NS
- Eliot TS. 1971. *The Complete Poems and Plays*. New York: Harcourt, Brace & World
- Feynman RP. 1998. *The Meaning of It All*. Boston: Addison Wesley
- Firestein S. 2012. *Ignorance: How It Drives Science*. Oxford, UK: Oxford Univ. Press
- Firestein S. 2016. *Failure: Why Science is So Successful*. Oxford, UK: Oxford Univ. Press
- Fleck L. 1979. *Genesis and Development of a Scientific Fact*, transl. F Bradley, TJ Trenn. Chicago: Univ. Chicago Press
- Fourcade M. 2011. Cents and sensibility: economic valuation and the nature of “nature.” *Am. J. Sociol.* 116:1721–77
- Freese J, Peterson D. 2017. Replication in social science. *Annu. Rev. Sociol.* 43:147–65
- Glassner B, Hertz R, eds. 2003. *Our Studies, Ourselves*. Oxford, UK: Oxford Univ. Press
- Groves RM, Singer E, Corning A. 2000. Leverage-salience theory of survey participation. *Public Opin. Q.* 64:299–308
- Gustafsson K, Hagstrom L. 2018. What is the point? Teaching graduate students how to construct political science research puzzles. *Eur. Political Sci.* 17:634–48
- Hadaway K, Marler P, Chaves M. 1993. What the polls don’t show: a closer look at U.S. church attendance. *Am. Sociol. Rev.* 58:741–52
- Hammond PE, ed. 1964. *Sociologists at Work*. New York: Basic
- Harding S. 1986. *The Science Question in Feminism*. Ithaca, NY: Cornell Univ. Press
- Hauser RM. 2017. A life in sociology. *Annu. Rev. Sociol.* 43:1–18
- High. Educ. Funding Counc. Engl. 2005. *PhD research degrees: entry and completion*. Issues Pap., High. Educ. Funding Counc. Engl., Bristol, UK
- Hunter MA, ed. 2018. *The New Black Sociologists: Historical and Contemporary Perspectives*. London: Routledge
- Hustvedt S. 2016. *A Woman Looking at Men Looking at Women: Essays on Art, Sex, and the Mind*. New York: Simon and Schuster
- Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Med.* 2:696–701
- Jaynes ET. 1993. A backward look into the future. In *Physics and Probability: Essays in Honor of Edwin T. Jaynes*, ed. WT Grandy Jr., PW Milonni, pp. 261–76. Cambridge, UK: Cambridge Univ. Press
- Kafka AC. 2018. In the lab, failure is part of the job description. *Chronicle of Higher Education*, June 15. <https://www.chronicle.com/article/in-the-lab-failure-is-part-of-the-job-description/>
- Keeter S, Miller C, Kohut A, Groves RM, Presser S. 2000. Consequences of reducing nonresponse in a national telephone survey. *Public Opin. Q.* 64:125–48
- Khan S, Fisher DR, eds. 2014. *The Practice of Research: How Social Scientists Answer Their Questions*. Oxford, UK: Oxford Univ. Press
- Kreuter F, Presser S, Tourangeau R. 2008. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opin. Q.* 72:847–65
- Light R. 2019. Book review. *Am. J. Sociol.* 125:853–55
- Lovitts B. 2001. *Leaving the Ivory Tower: The Causes and Consequences of Departure From Doctoral Study*. Lanham, MD: Rowman and Littlefield
- Lucas JW, Morrell K, Posard M. 2013. Considerations of the “replication problem” in sociology. *Am. Sociologist* 44:217–32
- Maitland A, Presser S. 2016. How accurately do different evaluation methods predict the reliability of survey questions? *J. Surv. Stat. Methodol.* 4:362–81
- Maitland A, Presser S. 2018. How do question evaluation methods compare in predicting problems observed in typical survey conditions? *J. Surv. Stat. Methodol.* 6:465–90

- Maitland A, Presser S. 2019. A comparison of five question evaluation methods in predicting the validity of respondent answers to factual items. In *Advances in Questionnaire Design, Development, Evaluation, and Testing*, ed. PC Beatty, D Collins, L Kaye, JL Padilla, GB Willis, A Wilmot, pp. 75–90. New York: Wiley
- Malcolm S. 2019. *Written testimony of Dr. Shirley Malcom, American Association for the Advancement of Science*. Presented to the Committee on Science, Space and Technology, U.S. House of Representatives, May 9
- Martin JL. 2017. *Thinking Through Methods*. Chicago: Univ. Chicago Press
- Martin YM, Maclachlan M, Karmel T. 2001. *How is Australia's Higher Education Performing? An Analysis of Completion Rates of a Cohort of Australian Post Graduate Research Students in the 1990s*. Canberra, ACT: Dep. Educ. Train. Youth
- McCloskey DN. 1985. The loss function has been mislaid: the rhetoric of significance tests. *Am. Econ. Rev.* 75:201–5
- Meehl PE. 1967. Theory-testing in psychology and physics: a methodological paradox. *Philos. Sci.* 34:103–15
- Meehl PE. 1990. Why summaries of research on psychological theories are often uninterpretable. *Psychol. Rep.* 66:195–244
- Merton RK. 1948. The bearing of empirical research upon the development of sociological theory. *Am. Sociol. Rev.* 13:505–15
- Morrison DE, Henkel RE, eds. 1970. *The Significance Test Controversy*. Chicago: Aldine
- Natl. Acad. Sci. Eng. Med. 2019a. *A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis*. Washington, DC: Natl. Acad. Press
- Natl. Acad. Sci. Eng. Med. 2019b. *Reproducibility and Replicability in Science*. Washington, DC: Natl. Acad. Press
- Noy S, Ray R. 2012. Graduate students' perceptions of their advisors: Is there systematic disadvantage in mentorship? *J. High. Educ.* 83:876–912
- Open Sci. Collab. 2015. Estimating the reproducibility of psychological science. *Science* 349(6251):aac4716
- Overbye D. 2017. Yearning for new physics at CERN, in a post-Higgs way. *New York Times*, June 20, p. D1
- Peytchev A, Presser S, Zhang MM. 2018. Improving traditional nonresponse bias adjustments: combining statistical properties with social theory. *J. Surv. Stat. Methodol.* 6:491–515
- Polanyi M. 1946. *Science, Faith and Society*. Chicago: Univ. Chicago Press
- Polanyi M. 1966. *The Tacit Dimension*. Chicago: Univ. Chicago Press
- Polya G. 1954. *Mathematics and Plausible Reasoning, Vol. I: Induction and Analogy in Mathematics*. Princeton, NJ: Princeton Univ. Press
- Poovey M. 1998. *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: Univ. Chicago Press
- Presser S. 1984a. Is inaccuracy on factual survey items item-specific or respondent-specific? *Public Opin. Q.* 48:344–55
- Presser S. 1984b. The use of survey data in basic research in the social sciences. In *Surveying Subjective Phenomena*, Vol. 2, ed. CF Turner, E Martin, pp. 93–114. New York: Russell Sage Found.
- Presser S. 1994. Informed consent and confidentiality in survey research. *Public Opin. Q.* 58:446–59
- Presser S, Blair J. 1994. Survey pretesting: Do different methods produce different results? *Sociol. Methodol.* 1994:73–104
- Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, et al. 2004. *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley
- Presser S, Singer E, Van Hoewyk J. 2000. Knowing versus feeling as factors in willingness to report information to the census. *Soc. Sci. Res.* 29:140–47
- Presser S, Stinson L. 1998. Data collection mode and social desirability bias in self-reported religious attendance. *Am. Sociol. Rev.* 63:137–45
- Presser S, Traugott M. 1992. Little white lies and social science models: correlated response errors in a panel study of voting. *Public Opin. Q.* 56:77–86
- Presser S, Traugott M, Traugott S. 1990. *Vote "over" reporting in surveys: the records or the respondents?* ANES Tech. Rep. Ser. No. 010157, ANES, Stanford Univ., Stanford, CA; and Univ. Mich., Ann Arbor, Mich.
- Putnam RD. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster
- Reichenbach H. 1938. *Experience and Prediction: An Analysis of the Foundations and Structure of Knowledge*. Chicago: Univ. Chicago Press

- Riley MW, ed. 1988. *Social Change and the Life Course*, Vol. 2: *Sociological Lives*. Thousand Oaks, CA: SAGE
- Sandberg J, Alvesson M. 2011. Ways of constructing research questions: gap-spotting or problematization? *Organization* 18:23–44
- Schickore J. 2018. Scientific discovery. In *The Stanford Encyclopedia of Philosophy*, ed. E Zalta. Stanford, CA: Stanford Metaphys. Res. Lab. <https://plato.stanford.edu/archives/sum2018/entries/scientific-discovery/>
- Schuman H. 1972. Two sources of antiwar sentiment in America. *Am. J. Sociol.* 78:513–36
- Schuman H, Presser S. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic
- Schwartz MA. 2008. The importance of stupidity in scientific research. *J. Cell Sci.* 121:1771
- Simmons JP, Nelson LD, Simonsohn U. 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22:1359–66
- Smith DE. 2003. Resisting institutional capture as a research practice. In *Our Studies, Our Selves*, ed. B Glassner, R Hertz, pp. 150–61. Oxford, UK: Oxford Univ. Press
- Sorfleet JR. 1990. Secondary bibliography of Bliss Carman's work. In *Bliss Carman: A Reappraisal*, ed. G. Lynch, pp. 197–204. Ottawa, ON: Univ. Ottawa Press
- Sowell R, Zhang T, Redd K. 2008. *Ph.D. completion and attrition: analysis of baseline program data from the Ph.D. Completion Project*. Rep., Counc. Grad. Sch., Washington, DC
- Sprague J, Zimmerman MK. 1983. Overcoming dualisms: a feminist agenda for sociological methodology. In *Theory on Gender/Feminism on Theory*, ed. P England, pp. 255–80. Chicago: Aldine
- Sullivan SP. 2019. Challenges for comparative fact-finding. *Int. J. Evid. Proof* 23:100–6
- Szyborska W. 1996. *The poet and the world*, transl. S Baranczak, C Cavanagh. Nobel Lecture, Stockholm, Dec. 6
- Terrell S. 2016. *Writing a Proposal for Your Dissertation*. New York: Guilford
- Turner RK, Bateman IJ. 2003. Editors' note. *Environ. Resour. Econ.* 25:255–56
- Wade L. 2019. New species of ancient human unearthed. *Science* 364(6436):108
- Weber M. 1946. Science as a vocation. In *From Max Weber: Essays in Sociology*, ed. H Gerth, CW Mills, pp. 129–56. Oxford, UK: Oxford Univ. Press
- Zuckerman EW. 2017. On genre: a few more tips to article-writers. *Confessions of a Supply-Side Liberal Blog*, May 27. <https://blog.supplysideliberal.com/post/2019/5/27/ezra-w-zuckermanon-genre-a-few-more-tips-to-article-writers>
- Zuckerman H, Merton RK. 1972. Age, aging, and age structure in science. In *Aging and Society, Volume 3: A Sociology of Age Stratification*, ed. MW Riley, M Johnson, A Foner, pp. 292–356. New York: Russell Sage Found.