
Value Elicitation

Is There Anything in There?

Baruch Fischhoff

*Department of Social and Decision Sciences and
Department of Engineering and Public Policy,
Carnegie Mellon University*

Eliciting people's values is a central pursuit in many areas of the social sciences, including survey research, attitude research, economics, and behavior decision theory. These disciplines differ considerably in the core assumptions they make about the nature of the values that are available for elicitation. These assumptions lead to very different methodological concerns and interpretations, as well as to different risks of reading too much or too little into people's responses. The analysis here characterizes these assumptions and the research paradigms based on them. It also offers an account of how they arise, rooted in the psychological and sociological contexts within which different researchers function.

Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?

—National Opinion Research Center (NORC), 1978

Think about the last time during the past month that you were tired easily. Suppose that it had been possible to pay a sum of money to have eliminated being tired easily immediately that one time. What sum of money would you have been willing to pay?

—Dickie, Gerking, McClelland, & Schulze, 1987, p. 19 (Appendix 1)

In this task, you will be asked to choose between a certain loss and a gamble that exposes you to some chance of loss. Specifically, you must choose either: Situation A. One chance in 4 to lose \$200 (and 3 chances in 4 to lose nothing). OR Situation B. A certain loss of \$50. Of course, you'd probably prefer not to be in either of these situations, but, if forced to either play the gamble (A) or accept the certain loss (B), which would you prefer to do?

—Fischhoff, Slovic, & Lichtenstein, 1980, p. 127

600 people are ill from a serious disease. Physicians face the following choice among treatments: Treatment A will save 200 lives. Treatment B has 1 chance in 3 to save all 600 lives and 2 chances in 3 to save 0 lives. Which treatment would you choose, A or B?

—Tversky & Kahneman, 1981, p. 454

Problematic Preferences

A Continuum of Philosophies

A critical tenet for many students of other people's values is that "If we've got questions, then they've got answers." Perhaps the most ardent subscribers to this belief are ex-

perimental psychologists, survey researchers, and economists. Psychologists expect their "subjects" to behave reasonably with any clearly described task, even if it has been tortuously contrived in order to probe esoteric theoretical points. Survey researchers expect their "participants" to provide meaningful answers to items on any topic intriguing them (or their clients), assuming that the questions have been put into good English. Economists expect "actors" to pursue their own best interests, thereby making choices that reveal their values, in whatever decisions the marketplace poses (and economists choose to study).

This article examines this *philosophy of articulated values* both in its own right and by positioning it on a continuum of philosophies toward value formation and measurement. At the other end of this continuum lies what might be called the *philosophy of basic values*. It holds that people lack well-differentiated values for all but the most familiar of evaluation questions, about which they have had the chance, by trial, error, and rumination, to settle on stable values. In other cases, they must derive specific valuations from some basic values through an inferential process.

Perhaps the clearest example of this latter perspective might be found in the work of decision analysts (Raiffa, 1968; von Winterfeldt & Edwards, 1986; Watson & Buede, 1988). These consultants lead their clients to decompose complex evaluation problems into basic dimensions of concern, called *attributes*. Each attribute represents a reason why one might like or dislike the possible outcomes of a decision. For example, the options

Preparation of this article was supported by National Science Foundation Grant No. SES-8175564 and by the Carnegie Corporation of New York, Council on Adolescent Development. The views expressed are those of the author.

My special thanks go to Robert Abelson, who suggested the juxtaposition that is explored here, and to Charles Turner, who has stimulated concern for nonsampling error for many years. My thinking on these issues has benefited from discussions with many people, including Lita Furby, Robyn Dawes, Paul Slovic, Sarah Lichtenstein, Amos Tversky, Daniel Kahneman, Alan Randall, and Robin Gregory. I have also received valuable comments from participants in the National Research Council Panel on Survey Measure of Subjective Phenomena; the Russell Sage Foundation Conference, "Towards a Scientific Analysis of Values"; and the U.S. Forest Service Conference on Amenity Resource Valuation.

Correspondence concerning this article should be addressed to Baruch Fischhoff, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.

facing someone in the market for a car are different vehicles (including, perhaps, none at all), whose attributes might include cost, style, and reliability.

The relative attractiveness (or unattractiveness) of different amounts of each attribute is then captured in a *utility function*, defined over the range of possible consequences (e.g., Just how much worse is breaking down once a month than breaking down twice a year?). After evaluating the attributes in isolation, the decision maker must consider their relative importance (e.g., Just how much money is it worth to reduce the frequency of repairs from annual to biennial?). These tradeoffs are expressed in a multiattribute utility function. Having done all of this, the consequences associated with specific actions are then evaluated by mapping them into the space spanned by that function.

Between the philosophies of articulated values and basic values, lie intermediate positions. These hold that although people need not have answers to all questions, neither need they start from scratch each time an evaluative question arises. Rather, people have stable values of moderate complexity, which provide an advanced starting point for responding to questions of real-world complexity. Where a particular version of this perspective falls on the continuum defined by the two extreme philosophies depends on how well developed these partial perspectives are held to be.

Each of these philosophies directs the student of values to different sets of focal methodological concerns. For example, if people can answer any question, then an obvious concern is that they answer the right one. As a result, investigators adhering to the articulated values philosophy will worry about posing the question most germane to their theoretical interests and ensuring that it is understood as intended. On the other hand, if complex evaluations are to be derived from simple evaluative principles, then it is essential that the relevant principles be assembled and that the inferential process be conducted successfully. That process could fail if it required too much of an intellectual effort and, also, if the question were poorly formulated or inadequately understood. If people have thought some about the topic of an evaluation

question, then they have less far to go in order to produce a full answer. Yet, even if people hold such partial perspectives, there is still the risk that they will miss some nuances of the question and, as a result, overestimate how completely they have understood it and their values regarding the issues that it raises.

A Choice of Paradigms

The effort to deal with these different worries in a systematic fashion has led to distinct research paradigms (Kuhn, 1962). Each such paradigm offers a set of methods for dealing with its focal worries, along with empirical tests of success in doing so. Each has evolved some theory to substantiate its approach. As paradigms, each is better suited to answering problems within its frame of reference than to challenging that frame. Thus, for example, the articulated values paradigm is better at devising additional ways to improve the understanding of questions than at determining whether understanding is possible.

This is, of course, something of a caricature. Many investigators are capable of wearing more than one hat. For example, survey researchers have extensively studied the properties of the *don't know* response (T. Smith, 1984). Still, when one is trying to get a survey (or experiment or economic analysis) out the door, it is hard to address these issues at length for every question. It may be easier to take *no answer* for an answer in principle than in practice. At the other extreme, it may be unprofitable for a consulting decision analyst to deal with situations in which the answer to a complex evaluation question is there for the asking, without the rigamarole of multiattribute utility elicitation.

To the extent that studies are conducted primarily within a single paradigm, it becomes critical to choose the right one. Table 1 summarizes the costs of various mismatches between the assumed and actual states of people's values. Above the diagonal are cases in which more is expected of people than they are prepared to give. The risk here is misplaced precision, reading too much into poorly articulated responses and missing the opportunity to help people clarify their thinking. Below the diagonal are cases in which too little is expected of people.

Table 1
Risk of Misdiagnosis

Assumption made	Proper assumption		
	Articulated values	Partial perspectives	Basic values
Articulated values	—	Get incomplete values Inadvertently impose perspective	Get meaningless values Impose single perspective
Partial perspectives	Promote new perspectives Distract from sharpening	—	Impose multiple perspectives Exaggerate resolvability
Basic values	Shake confidence Distract from sharpening	Discourage Distract from reconciliation	—

Note. Above diagonal: misplaced precision, undue confidence in results, missed opportunity to help. Below diagonal: needless complication, neglect of basic methodology, induced confusion.

The risk here is misplaced imprecision, needlessly complicating the task and casting doubt on already clear thinking.

The choice of a paradigm ought to be driven by the perceived costs and likelihoods of these different mismatches. Thus, one might not hire a survey researcher to study how acutely ill individuals evaluate alternative medical procedures, nor might one hire a philosopher to lead consumers through the intricacies of evaluating alternative dentifrices. Evaluation professionals should, in turn, devote themselves to the problems most suited to their methods.

Yet, it is in the nature of paradigms that they provide clearer indications of relative than of absolute success. That is, they show which applications of the set of accepted methods work better, rather than whether the set as a whole is up to the job. After describing these paradigms in somewhat greater detail, I will consider some of the specific processes by which work within them can create an exaggerated feeling for the breadth of their applicability.

As a device for doing so, I will highlight how each paradigm might interpret several sets of potentially puzzling results, namely those produced by the studies posing

the four evaluation questions opening this article. In each case, two apparently equivalent ways of formulating the question produced rather different evaluations. Assuming that the studies were competently conducted, an articulated values perspective would hold that if the answers are different, then so must the questions have been. Any inconsistency is in the eye of the beholder, rather than in the answers of the respondents.

A basic values philosophy leads to quite a different interpretation: If their responses are buffeted by superficial changes in question formulation, then people must not know what they want. As a result, none of the evaluations should be taken seriously. At best, they reflect a gut level response to some very general issue. According to the intermediate, partial perspectives philosophy, each answer says something about respondents. However, neither should be taken as fully representing their values.

A Sample of Problems

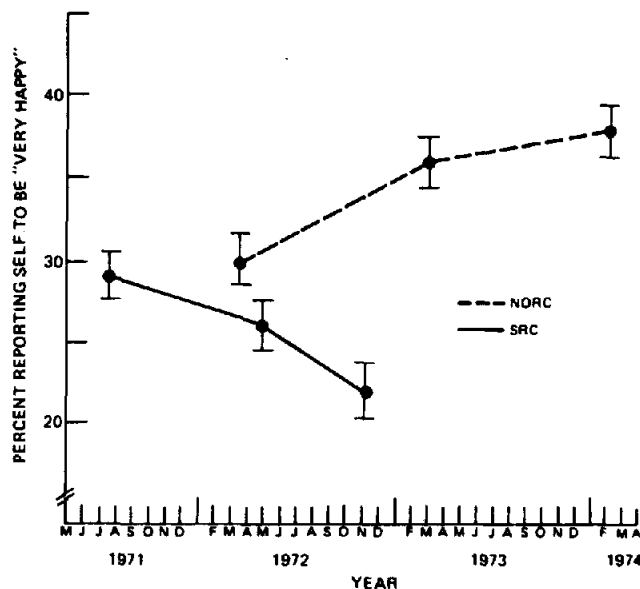
Happiness. Surveys sometimes include questions asking respondents to evaluate the overall state of their affairs. Answers to these questions might be used, for example, as barometers of public morale or as predictors of responses on other items (i.e., for statistical analyses removing individual mood as a covariate). In reviewing archival data, Turner and Krauss (1978) discovered the apparent inconsistency revealed in Figure 1. Two respected survey organizations, asking virtually identical happiness questions, produced substantially different proportions of respondents evaluating their situation as making them *very happy*. If the temptation of naive extrapolation is indulged, then quite different societies seem to be emerging from the two surveys (happinesswise, at least).¹

After a series of analyses carefully examining alternative hypotheses, Turner and Krauss (1978) concluded that the most likely source of the response pattern in Figure 1 was differences in the items preceding the happiness question. In the NORC survey, these items concerned family life; in the Survey Research Center (SRC) survey (Campbell, Converse, & Rodgers, 1976), they were items unrelated to that aspect of personal status.²

If respondents have fully articulated values, then different answers imply different questions. Inadvertently, the two surveys have created somewhat different happiness questions. Perhaps Happiness₁ (from the NORC survey) emphasizes the role of family life, whereas Happiness₂ (from the SRC survey) gives respondents more freedom in weighting the different facets of their lives.

From the opposing perspective, the same data tell

Figure 1
Trends in Self-Reported Happiness, 1971–1973



Note. Estimates are derived from sample surveys of noninstitutionalized population of the continental United States, aged 18 and over. Error bars demark ± 1 standard error around sample estimate. NORC = National Opinion Research Center; SRC = Survey Research Center. Questions were "Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?" (NORC); and "Taking all things together, how would you say things are these days—would you say you're very happy, pretty happy, or not too happy these days?" (SRC). From "Why Do Surveys Disagree? Some Preliminary Hypotheses and Some Disagreeable Examples" (p. 166) by C. F. Turner, 1984, in C. F. Turner and E. Martin, *Surveying Subjective Phenomena*, New York: Russell Sage Foundation. Copyright 1984 by the Russell Sage Foundation. Reprinted by permission.

¹ The two questions did differ slightly in their introductory phrase. One began "taken all together," the other "taking all things together." Only the bravest of theoreticians would try to trace the pattern in Figure 1 to this difference.

² Subsequent research (Turner, 1984; Turner & Martin, 1984) has shown a somewhat more complicated set of affairs—which may have changed further by the time this article is printed and read. Incorporating the most recent twists in this research would change the details but not the thrust of the discussion in the text.

quite a different story. If a few marginally related questions can have so great an impact, then how meaningful can the happiness question (and the responses to it) be? Conceivably, it is possible to take all things together and assess the happiness associated with them. However, as long as assessments depend on the mood induced by immediately preceding questions, that goal has yet to be achieved.

According to the partial perspectives philosophy, the two responses might be stable. However, neither should be interpreted as a thoughtful expression of respondents' happiness. Achieving that would require helping respondents generate and evaluate alternative perspectives on the problem, not just the one perspective that happens to have been presented to them.

Headache. According to Executive Order 12291 (Bentkover, Covello, & Mumpower, 1985), cost-benefit analyses must be conducted for all significant federal actions. Where those actions affect the environment, that often requires putting price tags on goods not customarily traded in any marketplace. For regulations governing ozone levels, one such good is a change in the rate of subclinical health effects, such as headaches and shortness of breath. In order to monetize these consequences, resource economists have conducted surveys asking questions like the second example in the set of quotations at the beginning of this article (Cummings, Brookshire, & Schulze, 1986; V. K. Smith & Desvousges, 1988).

In Dickie et al.'s (1987) survey, people who reported having experienced being tired easily estimated that they would be willing to pay \$17, on average, to eliminate their last day of feeling tired easily. Later in the same survey, the interviewer computed the overall monthly cost of eliminating each respondent's three most serious ozone-related health effects. This was done by multiplying how much people reported being willing to pay to eliminate the last occurrence of each effect by the number of reported episodes per month, then summing those products across symptoms. Respondents were then asked, "On a monthly basis is [—] what you would be willing to pay to eliminate these three symptoms?" (p. 20, Appendix 1). If respondents recanted, they were then asked what monthly dollar amount they would pay for the package. The markedly reduced dollar amount that most subjects provided was then prorated over the individual health effects. By this computation, respondents were now willing to pay about \$2 to eliminate a day of being tired easily.

From a regulatory perspective, these strikingly different estimates indicate markedly different economic benefits from reducing ozone levels. (Indeed, the Office of Management and Budget [A. Carlin, personal communication, 1987] has seriously criticized the Dickie et al., 1987, study as a basis for revising regulations under the Clean Air Act.) From an articulated values perspective, they imply that the two questions must actually be different in some fundamental ways. For example, people might be willing to pay much more for a one-time special treatment of their last headache than for each routine treatment. From a basic values perspective, these results indicate that people know that symptomatic relief is worth

something, but have little idea how much (even after an hour of talking about health effects). As a result, respondents are knocked about by ephemeral aspects of the survey, such as the highly unusual challenge to their values embodied by the request to reconsider. The investigators in this study seem to have adopted a partial perspectives philosophy. They treat respondents' values seriously, but not seriously enough to believe that respondents have gotten it right the first time. Rather, respondents need the help provided by showing them the overall implications of their initial estimates (Furby & Fischhoff, 1989).

Gamble. In samples of people shown the third example (Fischhoff et al., 1980), most people have preferred the gamble to the sure loss. However, they reverse this preference when the sure loss is described as an insurance premium, protecting them against the potentially greater loss associated with the gamble (Fischhoff et al., 1980; Hershey & Schoemaker, 1980). This difference is sufficiently powerful that it can often be evoked within subject, in successively presented problems.

From an articulated values perspective, the appearance of equivalence in these two versions of the problem must be illusory. Observers who see inconsistency in these responses must simply have failed to realize the differences. Perhaps, as a matter of principle, people refuse both to accept sure losses and to decline insurance against downside risks. In that case, these seemingly superficial differences in description evoke meaningful differences in how people judge themselves and one another. People want both to preserve a fighting chance and to show due caution. How they would respond to a real-world analog of this problem would depend on how it was presented.³

From a basic values perspective, these results show that people know that they dislike losing money, but that is about it. They cannot make the sort of precise tradeoffs depicted in such analytical problems. As a result, they cling to superficial cues as ways to get through the task.

In this case, some subsidiary evidence seemingly supports the intermediate perspective. When both versions are presented to the same person, there is an asymmetrical transfer effect (Poulton, 1968, 1989). Specifically, there are fewer reversals of preference when the insurance version comes first than when it comes second. This suggests that viewing the sure loss as an insurance premium is a relevant perspective, but not one that is immediately available. By contrast, respondents do realize, at some level, that premiums are sure losses. Studies of insurance behavior show, in fact, some reluctance to accept that perspective. For example, people prefer policies with low deductibles, even though they are financially unattractive. Apparently, people like the higher probability of getting some reimbursement, so that their premium does not have to be viewed as a sure loss (Kunreuther et al., 1978).

Disease. About two thirds of the subjects responding to the fourth problem (Tversky & Kahneman, 1981)

³ Thus, these results would lead one to expect lower renewal rates on insurance policies were subscribers to receive periodic bills for sure losses, rather than for premiums.

have been found to prefer Treatment A, with its sure saving of 200 lives. On the other hand, about the same portion prefer the second treatment when the two alternatives are described in terms of the number of lives that will be lost. In this version, Treatment A now provides a sure loss of 400 lives, whereas Treatment B gives a chance of no lives lost at all.

Applying the alternative philosophies to interpreting these results is straightforward. One difference in this case is that there is not only some independent evidence but also some theory to direct such interpretations. The discrepancies associated with the three previous problems were discovered, more or less fortuitously, by comparing responses to questions that happened to have been posed in slightly different ways. In this case, the discrepancies were generated deliberately. Kahneman and Tversky (1979) produced the alternative wordings as demonstrations of their *prospect theory*, which predicts systematic differences in choices as a function of how options are described, or *framed*. The shift from gains (i.e., lives saved) to losses is one such framing difference.

Prospect theory embodies a partial perspectives philosophy. It views these differing preferences as representing stable derivations of intermediate complexity from a set of basic human values identified by the theory. The sources of these differences seem ephemeral, however, in the sense that people would be uncomfortable living with them. Adopting an articulated values philosophy here would require arguing that people regard the different frames as meaningfully different questions—and would continue to do so even after thoughtful reflection.

In the absence of a theoretical account (such as prospect theory) or converging evidence (such as the asymmetrical transfer effect with the sure-loss-premium questions), one's accounting of seemingly inconsistent preferences becomes a matter of opinion. Those opinions might reflect both the particulars of individual problems and the general orientation of a paradigm. The next section describes these paradigms. The following section considers how they could sustain such different views on the general state of human values.

The Paradigms

However the notion of *paradigm* is conceptualized (Lakatos & Musgrave, 1970), it is likely to involve (a) a focal set of methodological worries, (b) a corresponding set of accepted treatments, (c) a theoretical basis for justifying these treatments and directing their application, and (d) criteria for determining whether problems have been satisfactorily addressed. Table 2 characterizes the three paradigms in these terms. This section elaborates on some representative entries in that table.

Philosophy of Articulated Values

Investigators working within this paradigm have enormous respect for people's ability to articulate and express values on the most diverse topics. Indeed, so great is this respect that investigators' worrying often focuses on ensuring that evaluative questions are formulated and un-

derstood exactly as intended. Any slip could evoke a precise, thoughtful answer to the wrong question (Fischhoff & Furby, 1988; Mitchell & Carson, 1989; Sudman & Bradburn, 1982).

A hard-won lesson in this struggle involves recognizing the powerful influence that social pressures can exert on respondents (DeMaio, 1984). As a result, investigators take great pains to insulate the question-answer relationship from any extraneous influences, lest those become part of the question. To prevent such complications, interviewers and experimenters stick to tight scripts, which they try to administer impassively in settings protected from prying eyes and ears. Lacking the opportunity to impose such control, economists must argue that marketplace transactions fortuitously have these desirable properties, in order to justify interpreting purchase decisions as reflecting just the value of the good and not the influences, say, of advertising or peer pressure.

At first blush, this protectiveness might seem somewhat paradoxical. After all, if people have such well-articulated preferences, why do they need to be shielded so completely from stray influences? The answer is that the investigator cannot tell just which stray influence will trigger one of those preferences. Indeed, the more deeply rooted are individuals' values, the more sensitive they should be to the nuances of how an evaluation problem is posed.

For example, it is considered bad form if the demeanor of an interviewer (or the wording of a question) suggests what the investigator expects (or wants) to hear. Respondents might move in that direction (or the opposite) because they aim to please (or to frustrate). Or, they might be unmoved by such a hint because they are indifferent to the information or social pressure that it conveys. Because a hint becomes part of the evaluation question, its influence is confounded with that of the issues that interested the investigator in the first place.

Unfortunately, the logical consistency of this position can border on tautology, inferring that a change is significant from respondents' sensitivity to it and inferring that respondents have articulated values from their responses to changes in questions now known to be significant. Conversely, responding the same way to two versions of a task means that the differences between them are not irrelevant and that people know their own minds well enough not to be swayed by meaningless variations.

The potential circularity of such claims can be disrupted either by data or by argument. At the one extreme, investigators can demonstrate empirically that people have well-founded beliefs on the specific questions that they receive. At the other extreme, they can offer theoretical reasons why such beliefs ought to be in place (bolstered, perhaps, by empirical demonstrations in other investigations). Developing these data and arguments in their general form has helped to stimulate basic research into nonverbal communication, interviewer effects, and even the psycholinguistics of question interpretations (e.g., Jabine, Straf, Tanur, & Tourangeau, 1984; Rosenthal & Rosnow, 1969; Turner & Martin, 1984).

Table 2
Three Paradigms for Eliciting Values

Worry	Treatment	Theoretical base	Test of success
Assumption: People know what they want about all possible questions (to some degree of precision)			
Inappropriate default assumptions (for unstated part of question)	Examine interpretation, specify more, manipulate expectations	Nonverbal communication, experimenter-interviewer effects, psycholinguistics	Full specification, empathy with subjects
Inappropriate interpretation of stated question	Use good English, consensual terms	Survey technique, linguistics	Sensible answers, consensual interpretation of terms
Difficulty in expressing values	Choose correct response mode	Psychometrics, measurement theory	Consistency (reliability of representation)
Strategic response	Proper incentives, neutral context	Microeconomics, demand characteristics	Sensible answers, nonresponse to "irrelevant" changes
Assumption: People have stable but incoherent perspectives, causing divergent responses to formally equivalent forms			
Deep consistency in methods across studies (failing to reveal problem)	"Looking for trouble": multiple methods in different studies; "asking for trouble": open-ended questions	Framing theory, new psychophysics, multiple disciplines, anthropology	Nonresponse to irrelevant changes
Eliciting values incompletely (within study)	Multiple methods within study, open ended	Same as above, counseling skills	Inability to elicit more
Inability to reconcile perspectives	Talking through implications	Normative analysis, counseling skills	Unpressured consistent response to new perspectives
Assumption: People lack articulated values on specific topic (but have pertinent basic values)			
Pressure to respond	Measure intensity, allow no response, alternative modes of expression	Survey research, social psychology	Satisfaction, stability among remainder
Instability over time	Accelerate experience	Attitude formation, behavioral decision theory	Stable convergence
Inability to relate	Client-centered process	Normative (re)analysis	Full characterization
Undetected insensitivity	Ask formally different questions	Normative analysis	Proper sensitivity

Within this paradigm, the test of success is getting the question specified exactly the way that one wants and verifying that it has been so understood. A vital service that professional survey houses offer is being able to render the questions of diverse clients into good English using consensual terms. This very diversity, however, ensures that there cannot be specific theory and data for every question that they ask. As a result, the test of success is often an intuitive appeal to how sensible answers seem to be. The risks of circularity here, too, are obvious.⁴

Assuming that respondents have understood the question, they still need to be able to express their (ready) answer in terms acceptable to the investigator. The great edifice of psychometric theory has evolved to manage potential problems here by providing elicitation methods compatible with respondents' thought processes and investigators' needs (Coombs, 1964; Nunnally, 1968). The associated tests of success are, in part, external—the abil-

ity to predict responses to other tasks—and, in part, internal—the consistency of responses to related stimuli. The risk in the former case is that the theoretical tie between measures is flawed. The risk in the latter case is that respondents have found some internally consistent way to respond to questions asked within a common format and varying in obvious ways (Poulton, 1989).

Perhaps surprisingly, the main concern of early contingent valuation investigators was not that respondents would have difficulty expressing their values in dollar terms. On the contrary, they feared that subjects would be able to use the response mode all too well. Knowing

⁴ One is reminded of the finding that undetected computational errors tend to favor investigators' hypotheses. A nonmotivational explanation of this trend is that one is more likely to double-check all aspects of procedure, including calculations, when results are surprising (Rosenthal & Rosnow, 1969).

just what they want (and how to get it), subjects might engage in strategic behavior, misrepresenting their values in order to shift to others the burden of paying for goods that they value (Samuelson, 1954). In response, investigators developed sophisticated tasks and statistical analyses. Applications of these methods seem to have allayed the fears of many practitioners (Brookshire, Ives, & Schulze, 1976).⁵

Philosophy of Basic Values

From the perspective of the philosophy of basic values, people's time is very limited, whereas the set of possible evaluative questions is very, very large. As a result, people cannot be expected to have articulated opinions on more than a small set of issues of immediate concern. Indeed, some theorists have argued that one way to control people is by forcing them to consider an impossibly diverse range of issues (e.g., through the nightly news). People who think that they can have some opinion on every issue find that they do not have thoughtful opinions on any issues (Ellul, 1963). The only way to have informed opinions on complex issues is by deriving them carefully from deeply held values on more general and fundamental issues (Rokeach, 1973).

Taking the headache question as an example, a meaningful answer is much more plausible from someone who has invested time and money in seeking symptomatic relief, which can serve as a firm point of reference for evaluating that special treatment. (Economists sometimes call these *averting behaviors* [Dickie et al., 1987].) Otherwise, the question seems patently unanswerable—and the wild discrepancies found in the research provide clear evidence of respondents' grasping at straws.

From the perspective of this paradigm, the existence of such documented discrepancies means that not all responses can be taken seriously. As a result, investigators adhering to it worry about any aspects of their methodology that might pressure respondents to produce unthoughtful evaluations. In this regard, an inherent difficulty with most surveys and experiments is that there is little cost for misrepresenting one's values, including pretending that one has them. By contrast, offering no response may seem like an admission of incompetence. Why would a question have been posed if the (prestigious?) individuals who created it did not believe that one ought to have an answer? With surveys, silence may carry the additional burden of disenfranchising oneself by not contributing a vote to public opinion. With psychological experiments, it may be awkward to get out, or to get payment, until one has responded in a way that is acceptable to the experimenter.

One indication of the level of perfunctory responses in surveys may be seen in the repeated finding (Schuman & Presser, 1981) that explicitly offering a *don't know* option greatly increases the likelihood of subjects offering no opinion (e.g., from 5% to 25%). Yet, even that option is a rather crude measure. Respondents must determine how intense a degree of ignorance or indifference *don't know* implies (e.g., Does it mean absolutely, positively

having no idea?). Investigators must, then, guess at how respondents have interpreted the option.

Hoping to say something more about the intensity of reported beliefs, survey researchers have conducted a lively debate over alternative statistical analyses of seemingly inconsistent attitudes (e.g., Achen, 1975; Converse, 1964). Its resolution is complicated by the difficulty of simultaneously evaluating questions and answers (Schuman & Presser, 1981; T. Smith, 1984). For example, one potential measure of value articulation is the stability of responses over time. When people say different things at different times, they might just be responding randomly. However, they might also have changed their underlying beliefs or settled on different interpretations of poorly worded questions. Changes in underlying opinions may themselves reflect exogenous changes in the issues addressed by the question (e.g., "My headaches are worse now than the last time I was asked") or endogenous changes in one's thinking (e.g., "I finally came to realize that it's crazy to be squirreling money away in the bank rather than using it to make myself less miserable").

A striking aspect of many contingent valuation studies is the high rate of refusals to provide acceptable responses among individuals who have already agreed to participate in the study (Cummings et al., 1986; Mitchell & Carson, 1989; Tolley et al., 1986). These protest responses take several forms: simply refusing to answer the evaluation question, offering to pay \$0 for a good that one has admitted to be worth something, and offering to pay what seems to be an unreasonably high amount (e.g., more than 10% of disposable income for relieving a headache). For investigators under contract to monetize environmental goods, these responses are quite troublesome.⁶ For investigators who have the leisure to entertain alternative perspectives, these responses provide some insight into how respondents having only basic values cope with pressure to produce more. It is perhaps a testimony to the coerciveness of interview situations how rarely participants say *don't know*, much less try to bolt (as they have in these contingent valuation studies).

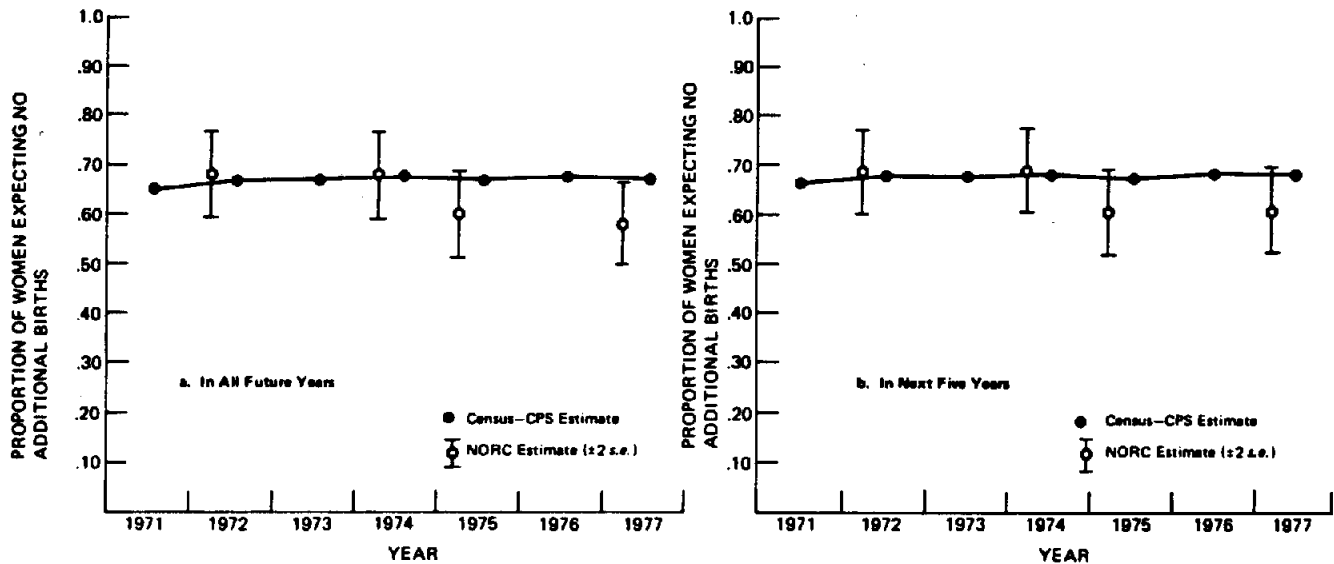
The term *protest response* implies hostility toward the investigator. Some of that emotion may constitute displaced frustration with one's own lack of articulated values. The investigator's "crime" is forcing one to confront not knowing exactly what an important good is worth. Perhaps a more legitimate complaint is that investigators force that confrontation without providing any help in its resolution.

As mentioned, investigators within the articulated values paradigm provide no help as a matter of principle. Elicited values are intended to be entirely those of the

⁵ The processes by which these fears were allayed might be usefully compared with the processes by which psychology convinced itself that it knew how to manage the effects of experimenter expectations (Rosenthal, 1967).

⁶ In actual studies, investigators sometimes just throw out protest responses. At times, they adjust them to more reasonable values (e.g., reducing high values to 10% of disposable income).

Figure 2
Estimates of Fertility Expectations of American Women: Proportion of Women Expecting No Further Children in (a) All Future Years, and (b) the Next Five Years.



Note. Samples included only married women aged 18–39; sample sizes in each year were approximately 4,000 (Census-CPS) and 220 (NORC). CPS = Current Population Survey; NORC = National Opinion Research Center. From "Why Do Surveys Disagree? Some Preliminary Hypotheses and Some Disagreeable Examples" (p. 192) by C. F. Turner, 1984, in C. F. Turner and E. Martin, *Surveying Subjective Phenomena*, New York: Russell Sage Foundation. Copyright 1984 by the Russell Sage Foundation. Reprinted by permission.

respondent, without any hint from the questioner. This stance might also be appropriate to investigators in the basic values paradigm in cases in which they want to know what is in there to begin with when an issue is first raised. However, basic values investigators might also be interested in prompting the inferential process of deriving specific values from general ones. That might be done nondirectively by leaving respondents to their own devices after posing an evaluative question and promising to come back later for an answer. In the interim, respondents can do whatever they usually do, such as ruminate, ask friends, listen to music, review Scripture, or experiment. Such surveys might be thought of as accelerating natural experiences, guided by descriptive research into how people do converge on values in their everyday life.

Alternatively, investigators can adopt a multiply directive approach. They can suggest alternative positions, helping respondents to think through how those positions might or might not be consistent with their basic values. Doing so requires a normative analysis of alternative positions that might merit adoption. That might require adding professions like economics or philosophy to the research team. Surveys that present multiple perspectives are, in effect, respondent centered, more akin to decision analysis than to traditional question-centered social research, with its impassive interviewers bouncing stimuli off objectified respondents. Studies that propose alternative perspectives incur a greater risk of sins of commission, in the sense of inadvertently pushing subjects in one of the suggested directions, and a reduced risk of sins

of omission, in the sense of letting respondents mislead themselves by incompletely understanding the implications of the questions that they answer.

As shown in the discussion of the questions opening this article, a clear hint that people have only basic values to offer is when they show undue sensitivity to changes in irrelevant features of a question. It can also be suggested by undue insensitivity to relevant features. Figure 2 shows the proportion of women who reported that they expect no additional births, either in all future years (left side) or in the next five years (right side).⁷ In each panel, there was considerable agreement between responses elicited by two respected survey houses. So, here is a case in which all of the irrelevant differences in procedures (e.g., interviewers, sampling, preceding questions) had no aggregate effect on responses. Across panels, however, there is a disturbing lack of difference. If there are women who intend to give birth after the next five years, then the curves should be lower in the left panel than in the right one. Although the investigators took care to specify time period, respondents either did not notice or could not make use of that critical detail.

An analogous result in contingent valuation research was Tolley et al.'s (1986) finding that people were willing to pay as much for 10 days worth as for 180 days worth of a fixed improvement in atmospheric visibility. Even

⁷ This is a question of prediction, rather than of evaluation, except in the sense that intentions to have children reflect the perceived value of having them.

more dramatic is Kahneman and Knetsch's (Kahneman, 1986) finding that respondents to a phone survey were willing to pay equal amounts to preserve the fisheries in one Ontario lake, in several Ontario lakes, and in all of the lakes in Ontario. These results could, of course, reflect articulated values based on utility functions that flattened out abruptly after 10 days and one lake. More likely, they reflect a vague willingness to pay a little money for a little good.

Philosophy of Partial Perspectives

By adopting an intermediate position, individuals working within the partial perspectives paradigm must worry about the problems concerning both extremes. On the one hand, they face the risk of inadequately formulated and understood questions, preventing respondents from accessing those partially articulated perspectives that they do have. On the other hand, investigators must worry about reading too much into expressions of value produced under pressure to say something.

These worries may, however, take on a somewhat different face. In particular, the existence of partial perspectives may give a deceptive robustness to expressions of value. Thus, investigators using a single method may routinely elicit similar responses without realizing the extent to which their success depends on the method's ability to evoke a common perspective. That fact may be obscured further when a family of related methods produces similar consistency. It takes considerable self-reflection for investigators to discern the structural communalities in methods that seem to them rather different. Speculative examples might include a tendency for surveys to emphasize hedonic rather than social values by asking respondents for their personal opinions, or for experimental gambles to encourage risk taking because participants cannot leave with less than they went in with,⁸ or to discourage emotional involvement because the scientific setting seems to call for a particularly calculating approach. Discovering the perspectives that it inadvertently imposes on itself is part of the continuing renewal process for any scientific discipline. In the social sciences, these perspectives may also be imposed on the people being studied, whose unruly behavior may, in turn, serve as a clue to disciplinary blinders (e.g., Furby, 1986; Gergen, 1973; Gilligan, 1982; Wagenaar, 1989).

Research methods may create consistent response sets, as well as evoke existing ones (Tune, 1964). When asked a series of obviously related questions on a common topic, respondents may devise a response strategy to cope with the experiment. The resulting responses may be consistent with one another, but not with responses in other settings. Indeed, those investigations most concerned about testing for consistency may also be the most vulnerable to generating what they are seeking. Think, for example, of an experiment eliciting evaluations for stimuli representing all cells of a factorial design in which each factor is a different outcome attribute. Why not come up with some simple rule for getting through the task?

For example, Poulton (1968, 1989) has conducted

detailed secondary analyses of the quantitative estimates elicited in psychophysics experiments in an effort to capture the subjective intensity of physical stimuli (e.g., sweetness, loudness). He argued that the remarkable internal consistency of estimates across stimulus dimensions (Stevens, 1975) reflects the stability of investigators' conventions in setting up the details of their experiments. Although subjects have no fixed orientation to such unfamiliar forms of evaluation, they do respond similarly to structuring cues such as the kind of numbers to be used (e.g., integers vs. decimals) and the place of the standard stimulus in the range of possibilities.⁹ The (nontrivial) antidotes are what might be called *looking for trouble* and *asking for trouble*—eliciting values in significantly different ways and using sufficiently open-ended methods to allow latent incoherence to emerge.

Economists hope to reduce these problems by discerning people's values from the preferences revealed in market behavior. Such actions ought to be relatively free of pressures to respond. After all, you don't have to buy. Or do you? Even if choices are voluntary, they can only be made between options that are on offer and with whatever information respondents happen to have. For example, you may hate ranch style homes but have little choice other than to choose one that makes the best of a bad situation in some locales. In that case, the preferences thereby revealed are highly conditional. Furthermore, even if the choice sets are relatively open and well understood, they may be presented in ways that evoke only a limited subset of people's values. By some accounts, evoking partial perspectives is the main mission of advertising (by other accounts, it is just to provide information). Some critics have argued that some perspectives (e.g., the value of possessing material goods) are emphasized so effectively that they change from being imposed perspectives to becoming endorsed ones.¹⁰

If one wants to predict how people will behave in situations presenting a particular perspective, then one should elicit their values in ways evoking that perspective.¹¹ If one wants to get at all their potentially relevant perspectives, then more diverse probing is needed. This is the work of many counselors and consultants. Although some try to construct their clients' subjective problem representation from basic values (along the lines of decision analysis), others try to match clients with general

⁸ The need to protect human subjects poses this constraint. Even without it, there would be problems getting people to risk their own money in a gamble contrived by some, possibly mistrusted, scientist.

⁹ Many contingent valuation studies have elicited values by asking subjects questions such as "Would you pay \$1, \$2, \$3, . . . ?" until they say no. One might compare the implicit structuring of this series of questions with that achieved by "Would you pay \$10, \$20, \$30, . . . ?" or by moving down from \$100 in \$1 increments.

¹⁰ This is just the tip of the iceberg regarding the methodological difficulties of inferring values from observed market behavior (Campen, 1986; Fischhoff & Cox, 1985; Peterson, Driver, & Gregory, 1988). In many cases, technical difficulties make inferring values from behavior an engaging fiction.

¹¹ Fischhoff (1983) considered some of the difficulties of predicting which frames are evoked by naturally occurring situations.

diagnostic categories. Each category then carries prognoses and recommendations. As mentioned, the counselor stance is unusual in social research. Like any direct interaction, it carries the risk of suggesting and imposing the counselor's favored perspective. Presumably, there is a limit to how quickly people can absorb new outlooks. At some point, they may lose cognitive control of the issue, wondering perhaps, "Whose problem is it, anyway?"

How Could They Think This Way?

Described in its own right, any paradigm sounds like something of a caricature. Could proponents really believe that one size fits all when it comes to methodology? Surely, decision analysts realize that some values are already so well articulated that their decomposition procedures will only induce confusion. Surely, survey researchers realize that some value issues are so important and so unfamiliar in their details that respondents will be unable to resist giving uninformed answers to poorly understood questions. Surely they do. Yet, equally surely, there is strong temptation to stretch the envelope of applications for one's favored tools.

Some reasons for exaggerating the applicability of one's own discipline are common to all disciplines. Anyone can exaggerate the extent to which they are ready for a challenge. Each discipline has an intact critique of its competitors. People who ask questions know what they mean and also know how they would answer. What might be called anthropology's great truth is that we underestimate how and by how much others see the world differently than we do. Paradigms train one to soldier on and solve problems, rather than to reflect on the whole enterprise.

The inconsistent responses opening this article present an interesting challenge for that soldiering. As shown in the discussion of those results, each paradigm has a way to accommodate them. Yet, investigators in the basic values paradigm seem much more comfortable with such accommodation. They seem more ready to accept them as real (i.e., produced from sound, replicable studies) and much more ready to see them as common. Basic values investigators sometimes seem to revel in such discrepancies (e.g., Hogarth, 1982; Nisbett & Ross, 1980), whereas articulated values investigators seem to view them as *bona fide*, but still sporadic, problems (e.g., Schuman & Presser, 1981).¹² Insight into these discrepant views about discrepancies can be gained by examining the institutional and methodological practices of these paradigms.

Interest in Discrepancies

Basic values investigators would like to believe that there are many robust discrepancies "out there in the world" because they serve a vital purpose for this kind of science. Discovering a peculiar pattern of unexpected results has been the starting point for many theories (Kahneman & Tversky, 1982). McGuire (1969) has gone so far as to describe the history of experimental psychology as the

history of turning artifacts into main effects. For example, increased awareness of experimenter effects (Rosenthal, 1967) stimulated studies of nonverbal communication (e.g., Ekman, 1985). In fact, some critics have argued that psychology is so much driven by anomalies that it tends to exaggerate their importance and generality (Berkeley & Humphreys, 1982). Anomalies make such a good story that it is hard to keep them in focus, relative to the sometimes unquirky processes that produce them (Fischhoff, 1988).

Interest in Order

On the other hand, articulated values investigators are more interested in *what* people think than in *how* they think. For those purposes, all these quirks are a major headache. They mean that every question may require a substantial development effort before it can be asked responsibly, with elaborate pretesting of alternative presentations. The possibility of anomalies also raises the risk that respondents cannot answer the questions that interest the investigators—at least without the sort of interactive or directive elicitation that is an anathema within this paradigm.

Of course, investigators in this paradigm are concerned about these issues. Some of the most careful studies of artifacts have come from survey researchers (e.g., Schuman & Presser, 1981). Classic examples of the effort needed to tie down the subjective interpretation of even seemingly simple questions may be found in the U.S. Department of Commerce studies of how to ask about employment status (Bailar & Rothwell, 1984). However, every research program with resource constraints is limited in its ability to pursue methodological nuances. When those nuances could represent fatal problems, then it is natural to want to believe that they are rare.

For survey research houses, these constraints are magnified by the commercial pressures to keep the shop open and running at a reasonable price. To some extent, clients go to quality and will pay for it. However, there is a limit to the methodological skepticism that even sophisticated clients will tolerate. They need assurance that investigators have the general skill needed to create workable items out of their questions. Clients might know, at some level, that "different questions might have produced different answers" (according to the strange wording that quality newspapers sometimes append to survey results). However, they still need some fiction of tractability.

Ability to Experiment

A further constraint on articulated values scientists is their theoretical commitment to representative sampling. The expense of such samples means that very few tests of alternative wording can be conducted. Conversely, it means

¹² This observation was sharply drawn by Professor Robert Abelson at a meeting of the National Research Council Panel on Survey Measure of Subjective Phenomena (Turner & Martin, 1981). This section of my article is, in large part, an attempt to work up the pattern that he highlighted.

that many discrepancies (like the happiness questions) are only discovered in secondary analyses of studies conducted for other purposes. As a result, there are typically confounding differences in method that blur the comparison between questions.

By contrast, basic values scientists are typically willing to work with convenience samples of subjects. As a result, they can run many tightly controlled experiments, increasing their chances of finding discrepancies. Multiple testing also increases the chances of finding differences by chance. If they are conscientious, these scientists should be able to deal with this risk through replications (which are, in turn, relatively easy to conduct). This indifference to sampling might reflect a self-serving and cavalier attitude. On the other hand, it may be the case that *how* people think might be relatively invariant with respect to demographic features that are known to make a big difference in *what* they think.

Precision of Search

The theories that basic values scientists derive to account for discrepancies are not always correct. When they are, however, they allow investigators to produce inconsistent responses almost at will. Much of experimental psychology is directed at determining the precise operation of known effects. For example, at the core of prospect theory is a set of framing operations designed to produce inconsistencies. The prevalence of phenomena under laboratory conditions has, of course, no necessary relationship to their prevalence elsewhere. Some extrapolation of prevalence rates from the lab to the world would, however, be only natural (Tversky & Kahneman, 1973). Furthermore, continuing absorption with a phenomenon should sharpen one's eagerness and ability to spot examples. Investigators who want and expect to see a phenomenon are likely to find it more often than investigators who do not. It would be only natural if the confirmation offered by such anecdotal evidence were overestimated (Chapman & Chapman, 1969).

The theoretical tools for seeking nuisance effects in an articulated values study would likely be more poorly defined. For example, the question might be posed as generally as "How common are order effects?" Given the enormous diversity of questions whose order might be reversed, the answer is, doubtless, "very low" in the domain of all possible questions. However, with questions of related content, order effects might be much more common (Poulton & Freeman, 1966). Moreover, questions are more likely to appear in surveys with somewhat related ones, rather than with completely related ones—even in amalgam surveys pooling items from different customers. Without a theory of relatedness, researchers are in a bind. Failure to find an order effect can just be taken as proof that the items were not related.¹³

Criterion of Interest

Surveys are often conducted in order to resolve practical questions, such as which candidate to support in an election or which product to introduce on the market. As a

Table 3

Conditions Favorable to Articulated Values

Personally familiar (time to think)
Personally consequential (motivation to think)
Publicly discussed (opportunity to hear, share views)
Uncontroversial (stable tastes, no need to justify)

Few consequences (simplicity)
Similar consequences (commensurability)
Experienced consequences (meaningfulness)
Certain consequences (comprehensibility)

Single or compatible roles (absence of conflict)
Diverse appearances (multiple perspectives)
Direct relation to action (concreteness)
Unbundled topic (considered in isolation)

Familiar formulation

result, the magnitude of an effect provides the critical test of whether it is worthy of notice. Unless it can be shown to make a difference, who cares? Laboratory results come from out of this world. If they cannot be mapped clearly onto practical problems, then they are likely to seem like curiosities. The psychologists' criterion of statistical significance carries little weight here. Survey researchers, with their large samples, know that even small absolute differences can reach statistical significance.

On the other hand, not all survey questions have that direct a relationship to action. One can assess the effects of being off by 5% in a pre-election poll or a product evaluation, as a result of phrasing differences. However, in many other cases, surveys solicit general attitudes and beliefs. These are widely known to be weak predictors of behavior (Ajzen & Fishbein, 1980). As a result, it may be relatively easy to shrug off occasional anomalies as tolerable. Discrepancies should become more important and, perhaps, seem more common as the questions driving research become sharper. The discrepancies associated with contingent valuation studies have come under great scrutiny recently because of their enormous, economic consequences. Changes in wording can, in principle, mean the difference between success and failure for entire companies or industries.

Thinking About Liability

How common are artifacts? is an ill-formed question, insofar as there is no clear universe over which the relative frequency of instances can be defined. Nonetheless, investigators' intuitive feeling for overall frequency must determine their commitment to their paradigms and their

¹³ A related example—for which I have unfortunately misplaced the reference and must rely on memory—is the finding that people respond more consistently to items on a common topic when those are grouped in a survey than when the questions are scattered. An (expensive) attempt to replicate this finding took as its common topic attitudes toward shop stewards, and found nothing. That could mean that the first result was a fluke or that *shop stewards* is not a meaningful concept of the sort that could induce consistent attitudes when brought to people's attention.

ability to soldier on in the absence of definitive data. Understanding the nature and source of one's own disciplinary prejudices is essential for paradigms to be used wisely and to evolve. Understanding other disciplines' (more and less legitimate) prejudices is necessary for collaboration. Implicit assumptions about the nature of human values seem to create a substantial divide among the social sciences. If they were to work together, the focal question might shift from how well articulated are values to where are they well articulated. Table 3 offers one possible set of conditions favorable to articulated values. Turner (1981) offered another. It might be informative to review the evidentiary record of discrepancies and nondiscrepancies in the light of such schemes.

REFERENCES

- Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review*, 69, 1218-1231.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Bailar, B. A., & Rothwell, N. D. (1984). Measuring employment and unemployment. In C. F. Turner & E. Martin (Eds.), *Survey measure of subjective phenomena* (pp. 129-142). New York: Russell Sage Foundation.
- Bentkover, J., Covello, V., & Mumpower, J. (Eds.). (1985). *Benefits assessment: The state of the art*. Amsterdam: Reidel.
- Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the "bias" heuristic. *Acta Psychologica*, 50, 201-250.
- Brookshire, D. S., Ives, C. C., & Schulze, W. D. (1976). The valuation of aesthetic preferences. *Journal of Environmental Economics and Management*, 3, 325-346.
- Campbell, A., Converse, P., & Rodgers, W. (1976). *The quality of American life: Perceptions, evaluations, and satisfaction*. New York: Russell Sage Foundation.
- Campan, J. T. (1986). *Benefit, cost, and beyond*. Cambridge, MA: Ballinger.
- Chapman, L. J., & Chapman, J. P. (1969). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 74, 271-280.
- Converse, P. E. (1964). The nature of belief systems in mass politics. In D. E. Apter (Ed.), *Ideology and discontent*. Glencoe, NY: Free Press.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Cummings, R. D., Brookshire, D. S., & Schulze, W. D. (Eds.). (1986). *Valuing environmental goods: An assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman & Allenheld.
- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Survey measure of subjective phenomena* (pp. 257-282). New York: Russell Sage Foundation.
- Dickie, M., Gerking, S., McClelland, G., & Schulze, W. (1987). *Improving accuracy and reducing costs of environmental benefit assessments: Vol. 1. Valuing morbidity: An overview and state of the art assessment* (USEPA Cooperative Agreement No. CR812954-01-2). Washington, DC: U.S. Environmental Protection Agency.
- Ekman, P. (1985). *Telling lies*. New York: Norton.
- Ellul, J. (1963). *Propaganda*. New York: Knopf.
- Fischhoff, B. (1983). Predicting frames. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 103-116.
- Fischhoff, B. (1988). Judgment and decision making. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 153-187). New York: Wiley.
- Fischhoff, B., & Cox, L. A., Jr. (1985). Conceptual foundation for benefit assessment. In J. D. Bentkover, V. T. Covello, & J. Mumpower (Eds.), *Benefits assessment: The state of the art* (pp. 51-84). Amsterdam: Reidel.
- Fischhoff, B., & Furby, L. (1988). Measuring values: A conceptual framework for interpreting transactions with special reference to contingent valuation of visibility. *Journal of Risk and Uncertainty*, 1, 147-184.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 117-141). Hillsdale, NJ: Erlbaum.
- Furby, L. (1986). Psychology and justice. In R. L. Cohen (Ed.), *Justice: Views from the social sciences* (pp. 153-203). New York: Plenum.
- Furby, L., & Fischhoff, B. (1989). *Specifying subjective evaluations: A critique of Dickie et al.'s interpretation of their contingent valuation results for reduced minor health symptoms* (U.S. Environmental Protection Agency Cooperative Agreement No. CR814655-01-0). Eugene, OR: Eugene Research Institute.
- Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, 26, 309-320.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Hershey, J. R., & Schoemaker, P. J. H. (1980). Risk taking and problem context in the domain of losses: An expected utility analysis. *Journal of Risk and Insurance*, 47, 111-132.
- Hogarth, R. M. (Ed.). (1982). *New directions for methodology of the social sciences: Question framing and response consistency*. San Francisco: Jossey-Bass.
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.
- Kahneman, D. (1986). Comment. In R. D. Cummings, D. S. Brookshire, & W. D. Schulze (Eds.), *Valuing environmental goods: An assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman & Allenheld.
- Kahneman, D., & Tversky, A. (1979). Prospect theory. *Econometrica*, 47, 263-292.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11, 123-141.
- Kuhn, T. S. (1962). *The structure of scientific revolution*. Chicago: University of Chicago Press.
- Kunreuther, H., Ginsberg, R., Miller, L., Sagi, P., Slovic, P., Borkin, B., & Katz, N. (1978). *Disaster insurance protection: Public policy lessons*. New York: Wiley.
- Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of scientific knowledge*. Cambridge, England: Cambridge University Press.
- McGuire, W. J. (1969). Suspiciousness of experimenter's intent. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research*. San Diego, CA: Academic Press.
- Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.
- National Opinion Research Center. (1978). *General Social Surveys, 1972-1978: Cumulative codebook*. Chicago: Author.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Nunnally, J. C. (1968). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Peterson, G. L., Driver, B. L., & Gregory, R. (Eds.). (1988). *Amenity resource valuation: Integrating economics with other disciplines*. State College, PA: Venture.
- Poulton, E. C. (1968). The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 69, 1-19.
- Poulton, E. C. (1989). *Bias in quantifying judgments*. London: Erlbaum.
- Poulton, E. C., & Freeman, P. R. (1966). Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 66, 1-8.
- Raiffa, H. (1968). *Decision analysis*. Reading, MA: Addison-Wesley.
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Rosenthal, R. (1967). Covert communication in the psychological experiment. *Psychological Bulletin*, 67, 356-367.
- Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research*. San Diego, CA: Academic Press.
- Samuelson, P. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, 36, 387-389.
- Schuman, H., & Presser, S. (1981). *Questions and answers*. San Diego, CA: Academic Press.
- Smith, T. (1984). Nonattitudes: A review and evaluation. In C. F. Turner

- & E. Martin (Eds.), *Survey measure of subjective phenomena* (pp. 215-256). New York: Russell Sage Foundation.
- Smith, V. K., & Desvousges, W. H. (1988). *Measuring water quality benefits*. Boston: Kluwer-Nijhoff.
- Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Tolley, G. et al. (1986). *Establishing and valuing the effects of improved visibility in the eastern United States* (USEPA Grant No. 807768-01-0). Washington, DC: U.S. Environmental Protection Agency.
- Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin*, *61*, 286-302.
- Turner, C. F. (1981). Surveys of subjective phenomena: A working paper. In D. Johnson (Ed.), *Measurement of subjective phenomena*. Washington, DC: U.S. Government Printing Office.
- Turner, C. F. (1984). Why do surveys disagree? Some preliminary hypotheses and some disagreeable examples. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (pp. 159-214). New York: Russell Sage Foundation.
- Turner, C. F., & Krauss, E. (1978). Fallible indicators of the subjective state of the nation. *American Psychologist*, *33*, 456-470.
- Turner, C. F., & Martin, E. (Eds.). (1981). *Surveys of subjective phenomena*. Washington, DC: National Academy Press.
- Turner, C. F., & Martin, E. (Eds.). (1984). *Surveying subjective phenomena*. New York: Russell Sage Foundation.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.
- von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. New York: Cambridge University Press.
- Wagenaar, W. A. (1989). *Paradoxes of gambling behavior*. London: Erlbaum.
- Watson, S., & Buede, D. (1988). *Decision synthesis*. New York: Cambridge University Press.