Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior

Jack M. Feldman
School of Psychology
Georgia Institute of Technology

John G. Lynch, Jr.
Department of Marketing
University of Florida

Drawing from recent developments in social cognition, cognitive psychology, and behavioral decision theory, we analyzed when and how the act of measuring beliefs, attitudes, intentions, and behaviors affects observed correlations among them. Belief, attitude, or intention can be created by measurement if the measured constructs do not already exist in long-term memory. The responses thus created can have directive effects on answers to other questions that follow in the survey. But even when counterparts to the beliefs, attitudes, and intentions measured already exist in memory, the structure of the survey researcher's questionnaire can affect observed correlations among them. The respondent may use retrieved answers to earlier survey questions as inputs to response generation to later questions. We present a simple theory predicting that an earlier response will be used as a basis for another, subsequent response if the former is accessible and if it is perceived to be more diagnostic than other accessible inputs. We outline the factors that determine both the perceived diagnosticity of a potential input, the likelihood that it will be retrieved, and the likelihood that some alternative (and potentially more diagnostic) inputs will be retrieved.

This article examines the effects of measurement operations on revealed correlations among survey measures of belief, attitude, intention, and behavior. The potential reactivity of measurement has long been of concern in psychology. Methodologists (e.g., Campbell & Stanley, 1966; Cook & Campbell, 1979; Runkel & McGrath, 1972) warn of measurement-induced distortions relating to social desirability, evaluation apprehension, and sensitization to experimental treatments. Although it is true that the problem of reactivity of measurement affects work in both the social and physical sciences, in the physical sciences, measurement effects are expressed in terms of substantive theory. For instance, Heisenberg's uncertainty principle links basic physical processes to the consequences of measurement operations. Our goal is to develop a similar linkage in psychology, relating theories of the generation of judgment and behavior to the effects of measurement, and in the process take steps toward a theory of the instrument (Cook & Campbell, 1979).

Studies of social, organizational, and consumer behavior are often guided by reasoned action or expectancy-value theories, including both multi-attribute and structural equation models, as well as cognitive response approaches to attitude formation and change. These theories specify relations among (and therefore, require measurement of) beliefs, attitudes, intentions, and behavior—or some subset of these (e.g., Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975; A. Mitchell, 1986; T. Mitchell, 1974; Mobely, Griffeth, Hand, & Meglino, 1979; Naylor & Ilgen, 1984; Petty, Cacioppo, & Goldman, 1981).

We develop a rudimentary theory of when and how the timing, order, and method of measurement of belief, attitude, and intention affects the observed relations among them, and between them and behavior. We draw upon recent developments in the field of social cognition to develop our framework (Fazio, Chen, McDonel, & Sherman, 1981; Fazio, Powell, & Herr, 1983; Hastie & Park, 1986; Lichtenstein & Srull, 1985; Lingle & Ostrom, 1979; Sherman, Aghin, Berman, & Lynn, 1978). This work supports the general proposition that momentarily activated cognitions have disproportionate influence over judgments made about an object or on related behaviors performed shortly after their activation. Which cognitions receive momentary activation is a function of the environmental cues directing attention to some subset of the object's stimulus features, priming, retrieval factors, and individual differences.

The implication of this work for survey research on attitudes, intentions, and behaviors is that the process of measuring any one of these variables directs attention to some subset of the stimulus field and some subset of the respondent's prior knowledge. This increases the likelihood that either the actual response measured or its immediate cognitive antecedents will be retrieved for use in guiding subsequent judgment or behavior pertaining to the same object (Fazio et al., 1981). Moreover, by a process of output interference (Alba & Chattopadhyay, 1985a,
of making the first judgment may suppress the retrieval of cognitions that would have influenced the second judgment or behavior in the absence of prior measurement. If multiple measures of each construct are collected (e.g., to test some hypothesized causal flow via LISREL), the problem is exacerbated.

Note, however, that these processes are not peculiar to responses generated because a researcher has asked the subject to fill out a questionnaire (Schuman & Ludwig, 1983). Similar processes arise when events in everyday life (e.g., conversing with a coworker, reading a story in a magazine, seeing a product in a store) alter the salience of potential inputs to behavioral decisions. To say this, though, does not preclude the possibility that one’s research method makes salient a set of inputs that departs markedly from the set salient in some population of real-world settings to which one would like to generalize.

For instance, consider a researcher attempting to model relations among cognitions about an object (e.g., product, job, or social issue), attitude toward the object, behavioral intentions (e.g., purchase, turnover, or voting), and actual behavior. A serious question is whether each construct measured exists (or is spontaneously generated) in any form in the absence of the researcher’s query. If not, the act of measurement changes the phenomenon under study, producing the thought processes predicted by the theory being tested and quite possibly influencing behavior. We term this state of affairs self-generated validity. But even when these beliefs, affective responses, and intentions exist prior to measurement, the specific measurement methodology can profoundly affect the revealed structural relations among these constructs, and between them and subsequent behavior. We will develop a rudimentary theory of these effects.

Computational and Retrieval Processes

Reasoned Action Versus Constructed Action

We take a partially constructionist view of attitudes, beliefs, intentions, and behaviors. Our conceptualization may best be understood by contrasting it with a more traditional view—one that seems quite consistent with an extensive body of research in social, organizational, and consumer psychology (e.g., Fishbein & Ajzen, 1975; Ryan, 1982; Steele & Ovalle, 1984).

1. Measures of belief, attitude, intention, and so forth, serve only to indicate what is already present and highly robust.
2. Behavior is (except in trivial circumstances) under conscious control—that is, under the control of behavioral intention.
3. Beliefs, attitudes, intentions, and behavior are spontaneously activated and influence each other in a fixed causal hierarchy—for example, beliefs are integrated into some overall affective reactions, which influence intentions, which are the immediate antecedents of behavior (Fishbein & Ajzen, 1975; Lavidge & Steiner, 1961).
4. Implicitly, these four constructs (i.e., beliefs, attitudes, intentions, and behavior) are treated as context independent.

We concur with the view that hot and cold responses toward many objects may be developed over time, stored, and elicited very rapidly by appropriate stimuli. However, the available evidence leads us to believe the following:

1. Only some of the beliefs about objects, attitudes, and intentions that are typically measured in laboratory and survey research already exist in some memory storage location. We assume a principle of cognitive economy (e.g., Wyer & Srull, 1986). Resources are not used to develop judgments, beliefs, or plans of action unless some motive exists. Our arguments here are consistent with findings of nonattitudes in the survey research literature (e.g., Bishop, Oldendick, & Tuchfarber, 1984; Bishop, Oldendick, Tuchfarber, & Bennett, 1980; Hawkins & Coney, 1981; Schuman & Presser, 1980).
2. Of those responses already stored, and therefore available, only a small subset will be readily accessible at the time of measurement.
3. This presents no barrier to the subject in responding to our queries. If the desired answer cannot be directly retrieved from memory, an answer can be easily computed from other available inputs (Fischhoff, Slovic, & Lichtenstein, 1980).
4. Both retrieved and computed responses can be modified by the elicitation context—that is, by externally provided and internally generated retrieval cues. The elicitation context can influence both the inputs to any particular judgment and the nature of the process acting upon those inputs.

Flexibility of Response Generation Processes

We should emphasize that this distinction between retrieval and computational processes is not peculiar to responses to surveys—it is one of the most fundamental in cognitive psychology (Lingle, Altom, & Medin, 1984). Lichtenstein and Srull (1985) illustrated this distinction with an example in which a consumer is asked, “Is the Buick Regal a luxury automobile?” If the respondent has already determined the answer to this question at some prior time, it can simply be retrieved from memory to respond to the query. If the consumer has never before considered this question (or if he or she cannot retrieve the previously made judgment), the answer would have to be computed on the spot. The computational process would involve retrieval of relevant information about the Buick Regal and the comparison of this information to one’s referent for luxury automobiles. Once the answer has been computed, it can be retrieved to answer subsequent questions—for example, “Are power windows standard equipment on the Buick Regal?”

It is important to our thesis that the process by which a respondent answers a given question may differ markedly as a function of whether a preceding question has been answered. In the example just given, the respondent may answer the second question by a retrieval and generalization process as follows: “The Buick Regal is a luxury automobile, and power windows are standard on most luxury automobiles.” The same respondent, if he or she were not asked the preceding question, might determine whether power windows were standard by first retrieving instances of Buick Regals seen previously, and then computing whether none, some, or all of them had power windows.

The argument could as easily be phrased in organizational terms. For instance, if a respondent in a job-satisfaction study is asked questions pertaining to unfamiliar constructs such as autonomy or growth needs, it is not unlikely that he or she would be required to compute a response. We believe that computa-
tional and retrieval processes of the sort outlined earlier may be involved in the generation of answers to questions about job characteristics (e.g., Thomas & Griffin, 1983), leadership behavior (e.g., Lord, Binning, Rush, & Thomas, 1978), or work groups (e.g., Staw, 1975).

A study by Fazio et al. (1984) illustrated the point that a researcher's request to make a judgment can prompt the respondent to create the construct being measured. Fazio et al. used an ingenious methodology that is generally applicable to the problem of discerning whether subjects are retrieving pre-existing judgments or are computing them for the researcher's benefit. Subjects were asked to play with a set of novel puzzles and, subsequently, to answer evaluative questions about each puzzle ("Letter series: Interesting?") by pushing one of two buttons marked yes and no. The key dependent variable was reaction time to respond to these questions (controlling for individual differences affecting speed of response).

One group of subjects had been asked to complete a paper and pencil attitude measure after playing with the puzzles but before the response time data were collected. Their response times were significantly faster than those of a second group that was not asked to fill out paper and pencil measures. This was interpreted as evidence that neither group had precomputed attitudes prior to questioning, but that the paper and pencil measures had stimulated the formation of retrievable attitudes. The response times of subjects who answer paper and pencil measures were no faster than those of other groups of subjects given cues that would cause them to form attitudes spontaneously (telling subjects to anticipate either a subsequent interview or a choice of which puzzle to play in the future). These results show that attitudes are sometimes formed spontaneously in the absence of a researcher's attempt to measure them—but only if it is functional to do so. (See also Kardes, 1986b.)

**Directive Effects of Measurement-Induced Cognitive States**

The theories of behavior that have most influenced decision research over the past decade have emphasized conscious choice from among alternative courses of action. Attitude research inspired by Fishbein and Ajzen's (1975) theory has this character:

> Human beings are usually quite rational, and make systematic use of the information available to them. We do not subscribe to the view that human social behavior is controlled by unconscious motives... nor do we believe that it can be described as capricious or thoughtless. (Ajzen & Fishbein, 1980, p. 5)

In behavioral decision theory, even the strategies by which decisions are made are often hypothesized to be deliberately chosen. Cost–benefit metatheories of situational effects on decision processes (E. Johnson & Payne, 1986; Payne, 1982) assumed that the decision maker chooses the heuristic offering the best combination of costs (effort) and benefits (accuracy).

We take issue with these views on at least two dimensions: the role of consciousness and the role of choice. A substantial body of evidence leads many investigators to conclude that much information processing, as well as action, occurs outside of awareness and without volition (e.g., Bargh, 1984; Kimble & Perlmutter, 1970; Posner, 1982). Lachman, Lachman, and Butterfield (1979) stated that "It is the exception, not the rule, when thinking is conscious" (p. 207). We view behavior as being at least partially "mindless" (Langer, 1978; Thorngate, 1976) in the sense that much day to day action occurs without deliberate intent and without awareness of controlling factors. The extent to which one consciously chooses to attend work or school on a given day, deliberately chooses to stay or quit one's workplace under other than unusual circumstances, or considers alternative brands of common consumer goods is certainly open to debate. Habit undoubtedly plays a large part in some of these behaviors.

A significant characteristic of these mindless behaviors is that no meaningful choice occurs. A script theoretic description (e.g., Langer & Abelson, 1972) might be more apt. The—probably unfalsifiable—claim can be made that some similar conscious choice must have been made at some point prior to habit formation, but this misses an important point. In performing the current behavior, only one possible course of action comes to mind. Moreover, which particular course of action comes to mind can vary markedly as a function of the particular retrieval cues active at the time, although the respondent has no awareness of their effects. See, for instance, Cialdini's (1984) account of the automatic character of much social influence.

Naturally occurring circumstances may, however, force conscious attention to normally unmonitored behaviors. Workers introduced to an incentive payment plan may consciously determine their behavior, just as an antismoking therapy might force attention to the number of cigarettes consumed. The foregoing discussion suggests a process of behavior generation in which volition or thoughtfulness is periodically activated rather than constant, and occurs only with respect to parts of a person's total behavior.

We contend that the typical survey of beliefs, attitudes, intentions, and behavior can function to bring normally automatic behavior under conscious control. Consider workers suddenly faced with a job satisfaction/turnover survey. Prior to that time, they may have been going to work without consciously choosing to do so. The possibility of doing otherwise is not salient to most of them. The researcher's survey causes them to focus attention on outcomes as described in the questionnaire, to form overall evaluations of alternative courses of action, and to state a behavioral intention. It is not unlikely that the entire sequence of questioning serves as the basis for attitude formation, which then serves as the most salient basis for the development of an intention. That these appear to be causally related when measures are analyzed is not disputed here. However, the antecedent cognitive states were not ones that would have occurred in the absence of measurement.

This is a measurement-produced counterpart of the effects documented by Fazio et al. (1981); Fazio et al. (1983); Fazio and Zanna (1981); and Snyder (1982). In those studies, manipulations heightening the accessibility of a newly computed attitude produced attitude-consistent behavior. Similarly, Sherman (1980) has shown that stating one's intention to perform a behavior increases the likelihood of performing that behavior.

It is not our contention that the beliefs, attitudes, and intentions measured in typical surveys necessarily are created by the process of measurement. Later in this article, we discuss how
prior knowledge and involvement moderate the sensitivity of responses to the context created by questioning. Research on the boundary conditions of expectancy theories of work behavior (Duchler & Mobley, 1973; Graen, 1969; Kopelman & Thompson, 1976; T. Mitchell, 1974) shows that strong contingencies between behavior (or performance) and outcomes must exist before the predictions of the model are supported. These contingencies (and the feedback and/or reinforcement supporting them) have the effect of creating an environment demanding a conscious choice in a context in which certain outcomes are highly salient (see T. Mitchell, 1974, p. 1970). In effect, the organizational environment creates the same circumstances as measurement may establish. These circumstances maximize the likelihood of spontaneous intention formation and the influence of intentions on behavior, reducing the power of automatic retrieval processes of the sort described earlier. However, the efficacy of goal setting, behavior modification, and incentive programs in organizations shows that such conditions are the exception rather than the rule.

Other events in everyday life influence the likelihood that respondents will already possess stored answers to the questions we might ask them. For instance, Fazio and Zanna (1978, 1981), Fazio et al. (1981), and Smith and Swinyard (1983) have presented evidence suggesting that people are more likely to form an attitude toward an object spontaneously if they have direct experience with it than if they simply read information about it. If, following direct experience or receipt of information, measures of attitude are collected, attitude–behavior correlations are stronger for those with direct experience. Our interpretation of these results is that persons with direct experience have already formed and rehearsed overall attitude judgments prior to attitude measurement. Both responses to the attitude scales and subsequent behavior are based on a retrieval of the same inputs, maximizing correlations.

But even if the process of measurement does not create the beliefs, attitudes, and intentions held to be the distal and proximate causes of behavior, the context of questioning can have profound influence over the structural relations observed among measures of these constructs. The existence of these cognitions in respondents' memories does not guarantee that they will be activated to determine related responses in hierarchies postulated by various extant theories. By shaping the nature of the computational and retrieval processes by which answers to survey questions are generated, the measurement context can increase or decrease revealed correlations among these constructs. In the next section of this article we develop an elementary theory of factors that influence the likelihood that a respondent will answer one question in a series by retrieving his or her response to some previous question. In general, such retrieval processes will increase correlations among the two measures involved.

Lability of Observed Relations Among Measures of Beliefs, Attitudes, Intentions, and Behaviors

In social, organizational, and consumer psychology, it has been popular to postulate hierarchies among cognition, affect, and behavior. In the classic high-involvement hierarchy (Fishbein & Ajzen, 1975; Lavidge & Steiner, 1961; M. Ray, 1982), cognitions about an object are followed by the formation of integrated affect toward the object, which then results in behavioral intention or behavior. In low-involvement hierarchies proposed in the field of consumer research (Kassarjian, 1978; Krugman, 1965; M. Ray, 1982), very low-level cognitions are formed on the basis of mere exposure to the object (e.g., through advertising). Only after purchasing and using the product is anything resembling beliefs and attitudes formed (Smith & Swinyard, 1983).

We regard these hierarchies to be highly labile, rather than fixed. In general, if memories of beliefs, attitudes, intentions, or past behaviors exist, cues directing activation of any one of these can cause it to be the direct determinant of a judgment or behavior.

We believe that any of 16 possible causal flows (i.e., belief causing related belief, belief causing attitude, belief causing intention, belief causing behavior, . . . , behavior causing subsequent behavioral intention, behavior causing subsequent behavior) can be demonstrated, especially if the two elements in the pair are measured in immediate temporal proximity. From a methodological standpoint, this is significant because it implies that—except under certain conditions to be discussed later—one cannot rule out the possibility that measurement procedures might have inflated or deflated the observed correlation between two measures merely by showing that the correlation is unaffected by the order in which the two measures are collected.

Some of the 16 causal flows will be observed more frequently than others. We present a theory of which causal flows will be observed in any given instance, with special focus on how measurement operations affect these flows. In our theory, we consider the effects of three (somewhat interrelated) factors—the perceived diagnosticity of the first judgment for the second, the accessibility of the first judgment in memory, and the accessibility of alternative inputs to the second judgment—on the likelihood that a subject's response to a measure of one construct will be used as a basis for a response to a subsequently measured construct.

Diagnosticity

The perceived diagnosticity of the first judgment or decision for a second (later) one is the degree to which the respondent perceives that the answer to the first question correctly identifies how the second should be answered. A respondent's propensity to base an answer to the second question in a series on his or her answer to the first is a positive function of the perceived diagnosticity of the first for the second.

Several conceptual points might be made about diagnosticity. First, the specific factors governing the perceived diagnosticity of a particular prior response will depend on the nature of the second question and the goals it engenders. We should not expect to be able to specify some set of determinants of diagnosticity common to all judgment and choice tasks. For example, Einhorn and Hogarth (1986) have argued that the diagnosticity of observations in each of the four cells of a $2 \times 2$ (Presumed Cause Present vs. Absent $\times$ Effect Present vs. Absent) contingency table differs as a function of whether respondents are asked to judge the likelihood (a) that the cause will produce
the effect or (b) that the effect was produced by that particular cause.¹

Second, the diagnosticity of one’s response to Judgment A for Judgment B might differ from the diagnosticity of B for A. For instance, one’s response to “Are politicians honest?” might be more diagnostic for “Is Ronald Reagan honest?” than one’s answer about Ronald Reagan would be for judging the honesty of politicians in general. Third, the diagnosticity of A for B depends on the response to A. Thus, a response of “No” to “Are any politicians honest?” may be more diagnostic for “Is Ronald Reagan honest?” than is an answer of “Yes” to the former question. This example highlights a fourth point. Diagnosticity is itself the result of a judgment process and, thus, depends on knowledge structures that are chronically accessible or are temporarily accessible (Bargh, 1984) at the time of the second judgment. Thus, one’s recollection that one answered “Are any politicians honest?” negatively is highly diagnostic for “Is Ronald Reagan honest?” if the response considers that Ronald Reagan is a politician and applies some logical or quasi-logical inference rule (Wyer & Hartwick, 1980) or implicit theory (Murphy & Medin, 1985). If one lacked this knowledge or did not retrieve it in answering the question about Ronald Reagan, one’s earlier answer about any politicians might be perceived to be considerably less diagnostic.

An illustration of the effects of diagnosticity on judgment-retrieval processes comes from a fascinating article by Schuman and Ludwig (1983). They reviewed four published studies in which survey respondents were asked to judge their approval of certain actions taken by each of two competing parties—for example, “Do you think a Communist country like Russia should let American newspaper reporters come in and send back American news as they see it?” and “Do you think the United States should let Communist newspaper reporters from other countries come in here and send back the news as they see it?” The order of the two questions was counterbalanced. When the question about American reporters was asked in the first position, 90% of respondents answered in the affirmative. When the question about Communist reporters was asked first, only 36% agreed. However, when each question was asked in the second position, the action by the favored group became less popular (66% for American reporters), and the same action by the less favored group became more popular (73% for Communist reporters). Similar results were obtained in the other studies that Schuman and Ludwig reviewed. Regardless of whether the first question asked was one to which predominant responses were yes or no, subjects tended to give the same response to the following question about the action by the other party.

Schuman and Ludwig (1983) argued that these results could not be explained in terms of some consistency principle. As evidence, they cited a study by Schuman and Presser (1980) in which subjects were more likely to agree that lawyers were mercenaries (72%) than to agree that doctors were mercenaries (50%), but agreement rates were not affected by order of the questions. Schuman and Ludwig maintained that the effects of sequence result from the implications of a norm of even-handedness that is made salient when questions about favored and disfavored parties are asked contiguously. The two parties must be perceived to be in competition with each other for the context effects to be observed—a condition that is not met in the lawyer versus doctor example.

We interpreted these studies somewhat more broadly, in terms of diagnosticity. When one’s answer to the first question is diagnostic of what the answer to the second might be, it is unnecessary to compute an answer to the second question or to engage in an alternative retrieval strategy involving effortful search of long-term memory. In our view, this suggests measurement effects on responses to the second question of each pair, with greater distorting effects of measurement when the first question in the series was relatively high in diagnosticity.

In the Schuman and Ludwig (1983) article, the evidence of measurement effects on the structural determinants of responses came from changes in mean level of responses as a function of the order of questions. But changes in mean level are not a necessary consequence of such structural changes. Our view suggests a second type of test. Suppose that one developed a pool of pairs of questions, A(i) and B(i), and roughly measured diagnosticity by the following method. For each pair of questions, a group of pretest subjects would first be asked to answer Question A(i) in each pair, then to indicate what answer to B(i) would be suggested on the basis of the answer to A(i) alone, then to rate their confidence that this answer to Question B(i) is correct or acceptable. A second randomly assigned group of pretest subjects would follow the same procedure, but would first judge B(i), its implication for A(i), and their confidence that the implied answer to A(i) is correct. In a separate experiment, subjects will simply be asked to answer the questions in each pair, either in the order A(i), B(i), or in the order B(i), A(i).

Our theory predicts that the correlations among the measures of A(i) and B(i) will be unaffected by the order of measurement for those pairs of items that are symmetrically diagnostic (i.e., the diagnosticity of A for B equals the diagnosticity of B for A), as measured by the confidence ratings of the pretest subjects. However, for pairs of items that are asymmetrically diagnostic, the correlation between A(i) and B(i) should be greater in the order A(i), B(i) than in the reverse order if A(i) is more diagnostic for B(i) than B(i) is for A(i).

A study of consumer information processing by Lichtenstein and Srull (1985, Experiment 2) bears on the issue of diagnosticity and the probability that previously computed judgments will be retrieved in response to a subsequent question. It will be discussed in some detail because it illustrates an innovative method for the separation of computational and retrieval processes. (For other methods pertinent to the problem, see also Fazio et al., 1984; Lichtenstein & Srull, 1987; Lynch, 1981.)

Lichtenstein and Srull (1985) presented subjects with complex print advertisements for two products, giving the subjects one of two tasks. Some subjects were asked to read the ads to form an evaluation of each product so they would later be able to judge its desirability in comparison with competing brands. Other subjects were asked to read the ads in order to judge them

¹ The discussion of similarity judgments by Tversky (1977) and Tversky and Gati (1978) closely parallels our conception of the factors influencing diagnosticity. This raises the possibility that either diagnosticity may be mediated in some fashion by similarity or that the two constructs may have their origins in common processes. Unfortunately, space limitations prevent a full exploration of these issues.
on stylistic dimensions (e.g., language, art work). Crossed with this Processing Objective factor was the nature of an intervening task given to subjects 1 hr after they read the ads; they either (a) recalled as much of the information as possible, (b) evaluated the products along the same scale as described, or (c) did nothing. After a 48-hr delay, all of the subjects were asked to recall as much information as possible from the advertisements.

These recall scores showed that, regardless of processing objectives, subjects who had been asked to recall the information 1 hr after they read the ads; they either (a) recalled as much of the information as possible, (b) evaluated the products along the same scale as described, or (c) did nothing. After a 48-hr delay, all of the subjects were asked to recall as much information as possible from the advertisements. This simply verified the paradigmatic assumption that the prior attempt to recall the information strengthened memory for that information.

The most interesting results, though, were for subjects whose intervening task was to make evaluative judgments of the products. Two days later, these subjects remembered much of the specific information in the advertisement as did subjects given the intervening task of recalling that same information. However, subjects who had first judged the product and then judged it again 1 hr later could recall no more of the specific information 2 days later than did the control group that had performed no intervening task.

These data suggest that subjects in the style-judgment condition spontaneously retrieved the original ad information to compute a second (product) judgment, and thus had later recall scores equivalent to those of the direct-recall group. Those who formed an evaluative judgment initially needed only to retrieve the judgment, and so lacked the rehearsal responsible for the other groups' greater recall of ad information.

The data are also consistent with the hypothesized effects of diagnosticity and similarity in judgment retrieval. Subjects making style judgments could have retrieved this information for use in making product judgments; MacKenzie, Lutz, and Belch (1986) presented evidence that judgments of this type are possible. In fact, this did not occur; only the subjects possessing a relevant, diagnostic prior judgment retrieved it for later use. In this example, it is clear that initial evaluative judgments are more diagnostic of (identical) evaluative judgments made 1 hr later than are initial style judgments. It is equally clear, however, that further theoretical and empirical work is necessary to specify the causes of diagnosticity. This is made more difficult by the task-specific nature of these causes. (See Lynch, Marmorstein, & Weigold, 1987, for results bearing on the diagnosticity of recalled attributes and prior evaluations in choice tasks.)

**Memory Accessibility of Potential Inputs to Judgments**

In our theory of observed relations among measures of beliefs, attitudes, intentions, and behaviors, the second factor that influences whether a prior cognition will be used as an immediate input to a later related judgment is its accessibility in memory. Memory accessibility, in turn, will be a function of the following conditions: (a) the time since most recent activation of that cognition (Wyer & Srull, 1986), (b) the amount of interfering material encountered in the same general content domain (Keller, 1987; Lynch et al., 1987), (c) elaboration and rehearsal of the original information (Ross, Lepper, Strack, & Steinmetz, 1977; Sherman, 1980; Sherman et al., 1978), (d) characteristics of the information itself that determine the rate of decay in the respondent's ability to retrieve it, such as vividness (Reyes, Thompson, & Bower, 1980) or abstraction and summarizing power (Chattopadhyay, 1986; Lingle, Geva, Ostrom, Leippe, & Baumgardner, 1979), (e) motivation and processing goals at the time of initial encoding of the information (Biehal & Chakravarti, 1983; Loken & Hoverstadt, 1985), and (f) retrieval cues, whether internally generated by virtue of prior knowledge or externally provided by priming, similarity of contextual cues at encoding and retrieval, and so forth (Bettman & Sujan, in press).

These conditions affect the likelihood that any previously formed cognition will be used as an input to a judgment—not just one activated in responding to previous questions in a survey. We now turn to a detailed analysis of how the first four of these conditions would operate to influence observed correlations among measures of beliefs, attitudes, intentions, and behaviors.

**Effects of time and intervening material.** The probability that any particular cognition will be retrieved as an input to some judgment or behavior is inversely related to the amount of time since its most recent activation (Wyer & Srull, 1986) and the amount of material in the same content domain encountered in the interim (Keller, 1987).

The clearest implication of findings such as these for survey research on relations among beliefs, attitudes, intentions, and behaviors is that responses to a question, i, asked as part of a series, are more likely to be retrieved as a basis for a subsequent response, i + n, the closer together the two items appear in the survey.

However, we hypothesized that the effects of time and number of intervening items depends on the similarity among the items in two respects. First, assume that questionnaire items i and i + n pertain to the same job dimension—say, a supervisor's behavior. The properties of retroactive interference make it less likely that Response i will be retrieved as a basis for i + n if the intervening items are similar in content to Item i (e.g., the frequency of other supervisory behaviors) than if they are dissimilar. Second, the decay rate in propensity to use Response i as a basis for i + n (as a function of the number of intervening items) is also sensitive to the similarity of those two items because similarity of two items enhances the likelihood that the first response will be retrieved at the time of the second judgment and positively influences the diagnosticity of the first response for the second. Thus, the probability of basing Response i + n on Response i (§P^i+n) can be formulated as follows:

\[
P^i+n = f[^\text{Similarity}(i, i+n)/n].
\]

(1)

If two highly similar items (e.g., two items in a multi-item scale measuring the same construct) are separated by n unrelated ones (that is, items not eliciting the same schema or processing strategy) in a questionnaire, it may be quite likely that the earlier response would be retrieved and used as an input to the latter, with this likelihood relatively insensitive to the size of n.

If two mildly similar items (e.g., belief about an attribute of an object and overall evaluation of the same object) are separated by the same n unrelated items, one's propensity to retrieve i as a basis for i + n may be lowered substantially if n is moderately large. If Items i and i + n are contiguous (or nearly contigu-
ous), however, judgment retrieval may be invoked even if the two items seem only moderately similar (e.g., Lingle & Ostrom, 1979). This occurs because the capacity of working memory is limited. Wyer and Srull (1986) hypothesized that in making a judgment, subjects first search working memory. If a sufficient basis for making the judgment is found, search terminates.

Intervening items make it less likely that one's answer to Question \( i \) will be in working memory when Question \( i + n \) is encountered. Thus, subjects must either compute a new response or engage in effortful search of long-term memory.

These arguments imply that the problem of artificial influence of early judgments on later judgments should be greatest when multiple rating scales are used to measure the same construct, as in (a) the assessment of reliability, (b) some applications of multitrait, multimethod analyses, and (c) causal modeling when each construct is measured by multiple indicators. In these cases, even if the multiple indicators are interspersed throughout a longer questionnaire, it is likely that responses to later items will be based on answers to earlier ones. In such cases, high levels of internal consistency can scarcely be interpreted to mean that measurement error accounts for only a small proportion of variance in the multi-item scale. The memory dynamics have produced an effect similar to that obtained by sampling only from a small part of the domain of potential items. The same points apply to the assessment of test–retest reliability.

The retrieval principles involved here suggest an interesting interpretation of some forms of method variance. Measuring two or more constructs by the same method increases the surface similarity between items. In light of the arguments developed earlier, this enhances the probability that cognitions generated in answering one question will be retrieved to answer subsequent questions. Shared method variance in multiple verbal report methods will occur to the extent that the question formats are perceived to be similar by respondents. However, under the assumption that the resultant similarity among items is only moderate, it should be possible to reduce method variance and shared method variance by increasing the time interval or amount of unrelated material separating items sharing a common method. As a practical matter, though, it is often convenient to ask all of the questions that share a common method or format contiguously, as when ratings of a series of objects on the same scale are to be collected. This increases method variance.

Parenthetically, the hypothesized interactive effects of similarity and time interval between two measures provides an attractive explanation of factors affecting the observed correlations between measures of attitude or behavioral intention with actual behavior. Attitude and intention correlate more closely with behavior if the two measures correlated match in their levels of specificity with respect to the action itself, time, context, and target, and the smaller the time interval between collection of the attitude or intention measures and the measure of behavior (Davidson & Jaccard, 1979; Fishbein & Ajzen, 1975). Matching in terms of levels of specificity increases the likelihood that the attitude or intention will be retrieved as an input to the subsequent behavior, as does shortening the time interval. Because, in most studies, the time interval is substantial, similarity of the behavior measure to its assumed antecedent is particularly important.

Effects of elaboration and rehearsal. A number of studies in social cognition support the conclusions that elaboration or rehearsal of an information item increases the likelihood that it will be retrieved for use in making a subsequent judgment.

1. Sherman (1980) showed that the statement of intention judgments can cause behavior consistent with stated intentions. Research on goal setting (Locke, Shaw, Saari, & Latham, 1981) indicated that goals may influence behavior over substantial time periods. Garland (1985) has shown that internal goals function as standards for people's judgments of their own performance, and this could occur only if the goal (or some procedural representation; see Read, 1987) is frequently accessed.

2. Sherman et al. (1978) demonstrated that subjects who recalled their ratings of the importance of recycling behaved consistently with their ratings, whereas those who were not induced to elaborate their ratings did not.

3. Fazio et al. (1983, Experiment 2) showed that attitude toward one object, if primed, can influence beliefs formed about a related object.

4. Fazio et al. (1981) provided evidence that repeated association of an attitude object with subjects' attitudinal evaluations enhanced both attitude accessibility and attitude–behavior correlations. Fazio, Sanbonmatsu, Powell, and Kardes (1986) provided further evidence that the strength of object–evaluation associations affects the probability that these evaluations will be spontaneously activated by the presence of the object.

5. Higgins and Rholes (1978) showed that causing subjects to elaborate selectively favorable, as opposed to unfavorable, information about a stimulus person caused subsequent evaluations of that stimulus person to be either positive or negative, respectively.

There are four immediate implications of results like these for survey research studies. First, it is possible to make the retrieved memory of any prior expression of belief, attitude, intention, or behavior the immediate antecedent of any other belief, attitude, intention, or behavior by appropriate elaboration of the former.

Second, when any particular construct is measured in a questionnaire prior to another construct, multiple measures of the former will increase correlations with the latter. Psychometric reasons for this effect are well known. But the available evidence in social cognition suggests that the increased rehearsal provided by multiple measurements should increase the accessibility of memories of responses to the presumed antecedent construct, increasing correlations above and beyond the effects of increased measurement reliability.

Third, different methods of measuring any particular construct differ in the degree to which they cause subjects to elaborate the implications of their responses, altering the likelihood
that these responses will be retrieved to guide future behavior. For example, subjects' responses to open-ended questions (e.g., cognitive-response protocols) should be more retrievable than similar responses to structured scales (Slamecka & Graf, 1978).

Wright and Kriewall (1980) examined the effects of utility measurement procedures on the accuracy with which high school seniors' subsequent college choices could be predicted. Predictive accuracy was greater when subjects were instructed to engage in deliberations prior to the utility measurement interview than when they were not. Accuracy was also greater when utility functions were elicited under instructions to imagine that the actual college choice was to be made immediately rather than in 9 months (as was actually the case). Finally, accuracy of decompositional (conjoint) utility functions was less than that of compositional utility functions (based on subjects' ratings of the attractiveness of the individual levels of each attribute used in the decompositional task).

Wright and Kriewall (1980) interpreted these effects as showing the superiority of utility measurement procedures that capture the true state of mind of the decision maker at the time of choice. An alternative interpretation is that the methods that forecasted more accurately actually were more effective in shaping choices, by causing greater elaboration of certain values. These elaborated values were then more likely to be retrieved as inputs to the actual college choice. For example, the finding that decompositional techniques forecasted less accurately than compositional ones parallels Carlston's (1980) finding that memory for the evaluative implications of implicit judgments of a person's traits decayed more over time than did memory for explicit judgments of the same traits.

Fourth, as argued earlier, surveys (like other events in life) can force conscious attention to cognitions pertinent to behaviors that are normally automatic. Smith and Lerner (1986) discussed the development of automatism of information processing in person perception tasks, presenting evidence that memory performance is better for stimuli that are processed in a controlled (vs. automatic) fashion. Presumably, this is a function of the increased elaboration of stimuli processed consciously. Arguably, then, when surveys bring normally automatic processes under conscious control, they increase the likelihood that the response will be retrieved as input to some later judgment.

This speculation is testable, at least in principle. Cacioppo, Petty, and their colleagues (Cacioppo & Petty, 1981; Cacioppo, Petty, Losch, & Kim, 1986) have shown that facial electromyographic (EMG) activity can be used to measure both valence and intensity of affective reactions spontaneously generated on the mere presentation of an attitude object. Suppose that attitudes could be measured for some subjects by standard survey methods and for another group by facial EMG activity. Following either a brief or a long delay, behavior toward the same object could be measured. Our theory suggests that after a brief delay, measures of attitude collected by the two methods might correlate about equally with behavior. After a long delay, however, the correlation of the EMG measures with behavior are predicted to decline markedly, whereas the correlation between the paper and pencil measures of attitude and behavior are not. Because perioral EMG reflects the degree of conscious processing, elaboration and rehearsal in response to survey measures can be assessed in a manner complementary to verbal protocol and thought-listing techniques (Cacioppo & Petty, 1981).

**Information characteristics affecting decay in retrievability**

Not all information is equally memorable. Some types decay more rapidly than others, in the sense that the probability of their retrieval declines faster with the passage of time. This may be due at least in part to differences among types of information in the degree of elaboration and integration received at encoding. For instance, Reyes, Thompson, and Bower (1980) provided evidence that the vividness of information affected its likelihood of being recalled and used as an input to a decision made 48 hr after receiving the information, but not when the same decisions were made immediately after receiving the information.

Perhaps more pertinent to issues of correlations among measures of beliefs, attitudes, intentions, and behaviors is research that suggests that memory for overall evaluations and other summary judgments decays less rapidly than does memory for the original information on which these were based (Carlston, 1980; Chattopadhyay, 1986; Lingle et al., 1979). A potential implication of findings such as these is that attitude and behavioral intention—being relatively integrative and abstract in nature—might be more memorable than beliefs about attributes or consequences and their associated evaluations. If so, this might make it more likely that behavior will be guided by attitudes and intentions than by beliefs and evaluations.

Consider the implication of this speculation for correlations of behavior with attitude as measured by the sum of Belief X Evaluation products versus attitude as measured by overall evaluative ratings. If behavior is measured immediately after each attitude measure, the two methods might predict equally well. If behavior is measured after a delay, however, the overall evaluation measures should exhibit greater correlation.

**Accessibility of Alternative Inputs**

The preceding section detailed the evidence that the probability of using any particular belief, attitude, intention, or behavior as an input to a researcher's subsequent measure of any one of these is a positive function of the accessibility of the former in memory. In the present section, we outline the case that this same probability is inversely related to the memorability of alternative inputs. We develop three main themes:

1. The increased accessibility of an input produced by its elaboration simultaneously reduces the likelihood that other inputs will be retrieved from memory because of output interference effects.
2. Inputs that are perceived to be only moderately diagnostic may be ignored when more diagnostic ones can be retrieved, but used when they cannot.
3. Important individual differences exist that affect one's ability to generate alternative inputs and, hence, affect sensitivity to the context produced by questioning. These include expertise, central life interests, affective polarization, and some forms of involvement.

**Output interference.** In their influential chapter on value liability, Fishhoff et al. (1980) offered the following hypothesis: "when conflicting values are relevant to a particular issue, the priming or evocation of one will tend to suppress the accessibil-
ity of its counterpart" (p. 127). Such evocation effects are general phenomena of social behavior, and are not limited to survey responding. For example, Regan (1971) found that willingness to help another depended on the subjects' liking for the recipient of their aid. When, however, the recipient had previously performed a favor for the subjects, a reciprocity norm determined their behavior and the attitude–behavior correlation disappeared.

Recent cognitive theory supports an accessibility explanation for such effects. For instance, Wyer and Srull (1986, p. 331) postulated that

No more information is retrieved for use in attaining a processing objective than is sufficient to allow the objective to be attained. When this minimal amount has been retrieved, the search terminates. [Moreover,] When information relevant to a processing objective is required, the contents of the Work Space are searched first.

Together, these retrieval dynamics produce a situation in which the most accessible cognition sufficient to determine a response is used, whether it be a recently activated behavioral norm or an answer to a prior survey question. This would occur even if other cognitions or previously formed responses (e.g., intentions) were present in long-term memory, as long as they were momentarily less accessible.

Hoch (1984) provided a somewhat different account of the retrieval dynamics underlying these effects, and offered empirical evidence of their role in explaining a robust finding in the social judgment literature: Asking subjects to imagine or explain a hypothetical future event increases subsequent estimates of the probability that that event will actually occur (e.g., Carroll, 1978; Ross et al., 1977). Hoch interpreted these results in terms of associative interference effects, as shown in the verbal learning literature (Roediger, 1978; Rundus, 1973). Recall of information is construed as a process in which one node activates one of a set of connected nodes. Which node is activated is a probabilistic function of the relative strengths of pathways connecting these nodes. Generating one reason why an event might occur strengthens the association between the event and that information, increasing the likelihood that thinking about the event will resample the same information. The probability of recalling a new item of information is a decreasing function of the amount of previously recalled material.

To test this process in a judgment setting, Hoch (1984) conducted three experiments. Results showed that the order in which reasons for the occurrence or nonoccurrence of an event were elicited influenced accessibility of those reasons. He also demonstrated that the accessibility of the reasons subjects generated mediated their judgments of the probabilities of the event in question.

Related work interpretable in terms of output interference effects on memory and judgments has been reported by Fishhoff, Slovic, and Lichtenstein (1978) in their research on fault trees, and in Alba and Chattopadhyay's (1985a, 1985b, 1986) experiments on part-list cuing of attributes and brands of consumer products. In short, questioning procedures influence accessibility, which in turn influences judgment.

Similar interference effects may underlie Tesser's (1978) finding of self-generated attitude polarization. Tesser and his coworkers found that time spent thinking about an object or an issue renders schema (category)-consistent beliefs more accessible, and the effect presumably associated with these beliefs produces a more extreme attitude response. (See Tesser, 1978, and Chaiken & Yates, 1985, for alternative theoretical accounts.)

There is a clear implication of this work for survey research on relations between beliefs, attitudes, intentions, and behavior. First, research seeking to establish the causal prominence of a particular type of attribute (e.g., intrinsic outcomes of work) may, by the nature of the questioning procedure, cause an increase in the accessibility of cognitions of the very type that are held to be causal and a decrease in alternative cognitions. For instance, social cue and bogus feedback effects (Adler, Skov, & Salvemini, 1985; Thomas & Griffin, 1983) can be explained by appealing to similar concepts. Similarly, it is easy to see the possibility that the researcher's measures of modal salient beliefs in a typical expectancy-value study can suppress the retrieval of idiosyncratic beliefs. If (a) attitude or intention is measured shortly after modal salient beliefs and their associated evaluations and (b) the attitudes or intentions are constructed rather than retrieved directly from memory, the observed correlations tend to overstate the degree to which attitude or intention depends on this specific subset of cognitions in other (non-research) settings.

Effects of ability to retrieve more diagnostic inputs on the use of less diagnostic ones. In answering a researcher's question about belief, attitude, intention, or behavior, the respondent will use the most diagnostic of some small set of easily accessible potential inputs. Therefore, inputs of modest diagnosticity may be ignored if more diagnostic ones are accessible, but may be used if this is not the case. For instance, consider Bem's (1972) construal of the process by which subjects generate answers to questions about their attitudes. If attitudes are nonexistent or not easily retrieved, subjects will use memories of their past behavior and the external circumstances in which it occurred as inputs. But the existence of an easily retrievable prior attitude removes any need to rely on these less diagnostic inputs to construct answers to the researcher's measures.

Our earlier discussion of information characteristics affecting decay in memory accessibility is particularly relevant to this discussion. The accumulated evidence suggests that abstractions, overall evaluations, and other higher level encodings, once formed, can be retrieved to make subsequent memory-based judgments without retrieving the specific information on which they were originally based. Moreover, memory for these higher level encodings decays very slowly over time, in comparison with memory for the raw details on which they are based (e.g., Carlston, 1980; Higgins & King, 1981; Lingle et al., 1979). If the framing of a question, context of a decision, or the activation of a previously learned heuristic makes the overall evaluation relatively less diagnostic or relevant than the original information, that original information may well be retrieved and used (e.g., Alba & Hasher, 1983; M. Johnson & Raye, 1981) even when the summary judgment is easily recalled. If time or interfering material makes the original information inaccessible, however, these same higher level encodings might come to dominate decisions (Wyer & Srull, 1986).

The evidence on this point is mixed (cf. Carlston, 1980; Kardes, 1986a), but its clearest implication for self-generated
Individual differences in ability to generate alternative inputs. In previous sections we have described the way in which the context of measurement and other extraneous factors, (e.g., social cues) may cause the creation or recreation of beliefs, attitudes, and intentions, influencing observed correlations among these constructs and between them and observed behavior. Important individual differences exist, however, in sensitivity to the context produced by measurement. In particular, intense preexisting affect toward some object or category and expert-level category systems—both of which are associated with highly articulated values or central life interests—are thought to reduce or eliminate what we have called self-generated validity.

This point was forcefully made by Bettman and Su jan (in press). These researchers faced expert and novice photographers with a choice either between two cameras (comparable alternatives) or between a camera and a personal computer (noncomparable alternatives). In each case, one option was described partly in terms of specific attributes related to its potential to be used creatively, whereas the other option was described partly in terms of specific attributes related to reliability. Prior to the choice task, all of the subjects had completed an ostensibly unrelated memory experiment that primed either creativity or reliability or no primed attribute.

Interest centered on the effects of this priming manipulation on choice processes. Expert photographers choosing between two cameras showed no priming effects. These subjects already had well-formed and easily retrievable evaluative criteria. Novice consumers making either type of choice focused more on the primed attribute than on the unprimed one. These subjects tended to rate whichever attribute had been primed as more important, to evaluate the alternative characterized in terms of the primed attribute more positively than the one characterized by the unprimed attribute, and to choose accordingly. The absence of previously formed evaluative criteria made these subjects highly sensitive to the priming manipulation. We found it interesting that expert photographers—even those with expertise in computers as well—were similarly sensitive to the priming manipulation when they were given a choice between a camera and a computer. These subjects had no retrievable criteria for comparing such dissimilar alternatives. (For further evidence that the weighting schemes of novices but not experts are sensitive to context effects produced by questioning, see Chakravarti & Lynch, 1983; Hutchinson, 1983; and Kahneman & Miller, 1986.)

Both central life interests and what we term trait involvement may produce similar effects. Involvement can be defined as personal relevance (Chaiken, 1980; Petty & Cacioppo, 1979; Petty et al., 1981) or as a state of activation directed to some portion of the psychological field (Cohen, 1983). Our use of the term here is more congruent with the former. Personal relevance may be temporarily created by telling people that an issue will impact their lives (e.g., Petty et al. 1981). Alternatively, relevance may be chronic, resulting from a high level of affect associated with objects of a particular type. We will term these two types state and trait involvement (see also Houston & Rothschild, 1977). Most organizational research has focused on involvement of the trait type (e.g., Kanungo, 1979, 1982; Rabinowitz & Hall, 1977), and we will confine our analysis to the effects of these relatively stable individual differences.

Trait involvement is associated with enduring value systems and is triggered by value-relevant issues and objects. We take the position that value systems become well rehearsed and highly articulated if they are relevant to those parts of one's environment that mediate strong, frequent, positive or negative outcomes. Highly differentiated, expertlike category structures are developed in response to the same contingencies. The categories thus developed are highly affectively polarized, frequently rehearsed, and independent of context. With experience, affective responses and category-based perceptions become habitual and automatic, as do schemata for generating appropriate behaviors (Barsalou, 1987; Barsalou & Ross, 1986a, 1986b; see also Markus, 1977; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). This is clearly consistent with previous discussions of job and political involvement and commitment (Fiske & Kinder, 1981; Kanungo, 1979; Rabinowitz & Hall, 1977).

It has long been known that categories associated with a high level of affect are chronically more accessible than those less polarized (e.g., Tajfel, 1969), with effects on cognitive processing paralleling those documented in category accessibility studies (Bargh, 1984; Bargh, Bond, Lombardi, & Tota, 1986). Affect-laden categories are likely to be frequently used, and the associations that form them frequently (and recently) rehearsed, both of which would tend to increase their accessibility and context independence (Barsalou, 1987; Higgins, Bar gh, & Lombardi, 1985).

Because respondents high in trait involvement have accessible, affectively polarized categories, the operation of framing, priming, and context effects of measurement is mitigated. Strong affective responses have already been computed and need only to be retrieved. Indeed, they may occur automatically on stimulus presentation. Thus, these subjects would be less likely than those low in trait involvement to respond to a question about attitude toward an object by retrieving responses to earlier questions in the researcher's survey.

Given that expertise and trait involvement via articulated value systems lessen the sensitivity of measures of belief, attitude, intention, and behavior to the sorts of self-generated validity effects we have discussed, how does one assess the likelihood of these distorting effects in any given study? It could be argued that survey researchers rarely ask people about opinions, attitudes, and behaviors about which they have little knowledge. Certainly in consumer research, however, it is clear that multiattribute attitude models have been applied to every mundane packaged good imaginable. Organizational researchers reading this article might maintain that issues they study (reenlistment, turnover, job satisfaction) are sufficiently central to the respondents' lives that self-generated validity effects might be minimal.

We believe, however, that situations conducive to self-generated validity effects are the rule, not the exception. Barsalou
(1987) has shown that even well-known taxonomic categories may be influenced by context. Given the degree of practice people have with such categories, it seems unlikely that our beliefs, attitudes, and so forth are more consistent. Fischhoff et al. (1980) have maintained that people have a very small number of central values and that responses to measures of any specific value are generated by a computational rather than a retrieval process. In the political realm, Converse (1964, 1980; see also Bishop, Oldendick & Tuchfarber, 1984; Bishop, Oldendick, Tuchfarber, & Bennett, 1980; Schuman & Presser, 1980) has proposed that the mass public has no well-articulated political ideology, but rather has opinions on specific issues of personal relevance and virtually no opinion on others. Responses to questions on which no opinion is held are highly susceptible to context effects. Apparently only highly educated persons or those with very strong political involvement have well differentiated and integrated ideologies guiding their responses to questions about domestic or international policy (but see Judd, Krosnick, & Milburn, 1981; Judd & Milburn, 1980, for a contrary opinion).

One of the oldest controversies in psychology is also related to the degree of elaboration characterizing an individual’s knowledge structure. Cronbach (1946, 1950) pointed out that a number of alternative factors, generally called response sets or styles, influence answers to items in measures of ability, personality, and attitude when test content is difficult or ambiguous—that is, when no prior structure exists that is sufficient to give specific meaning to the stimulus. When stimulus information and existing cognitive structures are insufficient to guide responding, default options might then be used; these might take the form of dependence on contexts created by extraneous cues (Kreitler & Kreitler, 1981), cues internal to the question format (e.g., Aiken, 1983), and other dispositional or environmental factors. This argument would lead to the prediction that response sets are easily altered and depend on prior procedural knowledge, which seems true (Cronbach, 1950), and that they are not general across content domains (e.g., J. Ray, 1984; Winkler, Kanouse, & Ware, 1982). It does seem difficult to predict just what style or set will be used by an individual on a given occasion (Rorer, 1965), possibly because so many different default options are available. A model such as that proposed by Link (1982) may be used to establish whether any consistent theme underlies a given set of responses; experimental manipulations of context, category priming, question structure, and so forth, combined with diagnostic measures (e.g., personal construct elicitation) may tell which of many is being used at any given time. The point is, simply, that differences in the degree and organization of knowledge do exist, and that when both are low, other factors influence question responding in systematic ways.

Conclusions

Getzels (1982) quoted Gertrude Stein to close his discussion of problem formulation in creativity and social behavior: “Suppose no one asked a question? What would the answer be?” (p. 48). This article has followed very much in the spirit of the work of Getzels and others who have studied the problem of question framing and response consistency (e.g., Fischhoff et al., 1980; Hogarth, 1982; Sherman, 1980). Our argument has been that all real-world experiences with an object have the potential to shape subsequent responses to it.

Our intent was to make two contributions to this literature. First, we have focused our analysis more narrowly than have previous authors, both in terms of the statistical indexes held to be affected by measurement and in terms of the content domain in which measures are collected. Our analysis shows how observed correlations among beliefs, attitudes, intentions, and behaviors may be affected by the process of measurement.

Second, we have set forth a reasonably comprehensive theoretical framework for analyzing when and how the context produced by measurement will influence these observed correlations. Measurement can alter correlations among beliefs, attitudes, intentions, and behavior in two ways. If the researcher asks a question about attitude (or belief, intention, etc.), and no such cognition already exists, the measure creates one. In answering subsequent questions in a survey, the newly created attitude may then be retrieved and used in generating answers to later questions about the same attitude object—creating, for example, a correlation between attitude and subsequently measured behavior. Had attitude not been measured prior to behavior, behavior would have been generated using different (nonattitudinal) inputs.

Alternatively, consider those cases in which answers to all of the researcher’s questions about belief, attitude, and so forth, already exist in long-term memory. The content of preceding questions in the survey, their ordering, and other aspects of measurement can still affect observed relations among constructs. Instead of directly retrieving the appropriate response from long-term memory, subjects may retrieve some related response(s) made earlier in the survey. These earlier responses may be used directly to answer the question at hand, or integrated with other inputs to recompose an attitude or belief that already existed.

Briefly, our theory predicts that an earlier response will be used as an input to a subsequent response if the former is accessible and if it is perceived to be more diagnostic than other accessible inputs. We outlined the factors that determine both the perceived diagnosticity of a potential input, the likelihood that it will be retrieved, and the likelihood that some alternative (and potentially more diagnostic) inputs will be retrieved. The same computational and retrieval mechanisms postulated to explain effects of measurement can be used to explain the effects of other events that occur in everyday life.

Our framework allows us to generate some fairly specific hypotheses about how correlations between two measures will be affected by controllable aspects of measurement operations: number and content of intervening items appearing between any two items in a questionnaire, ordering of items, temporal spacing of measures, format of questions, and so forth. Moreover, we have detailed methods that are capable of diagnosing the nature of the (computational and retrieval) cognitive processes hypothesized to underlie the predicted effects.

The reader may be bothered by questions of what can be done to avoid the distorting effects of measurement that we have outlined. We and others (L. James, personal communication, April 1987) can suggest steps to mitigate the problem, but these are
likely to be costly and time consuming to implement. They include the following:

1. Extensive pretesting of the subject population, including both interviews and behavioral observation, to help determine the beliefs, attitudes, and intentions that are spontaneously formed or pre-exist.

2. Determining the terms and constructs naturally used by respondents to express both cognition and affect, and the ecological form of behavior. This is a standard procedure in cross-cultural research.


4. Conducting field experiments that systematically assess the degree to which, for example, question placement and form influence obtained reliability and validity. In the latter case, we refer to both trait and nomological validity (Cook & Campbell, 1979).

5. Determining the extent to which different subgroups of the population differ in susceptibility to measurement effects of the sort we have discussed—for example, on the basis of individual differences in prior knowledge and involvement. Like Cronbach (1975) and Lynch (1982), we imply that our ability to avoid invalidity is necessarily modest because of the limited nature of our understanding of the phenomena we research. But our message should not be interpreted as one of despair. Researchers' measures are not unique in prompting responses that then shape subsequent judgments and decisions. Therefore, all research yields information about the true structural determinants among beliefs, attitudes, intentions, and behavior—under some conditions. We hope that our work will allow researchers to gauge more precisely the range of real-world instances to which they might expect their results to generalize.

Our discussion has necessarily been speculative because little of the research that has provided the theoretical and empirical base for this article has been concerned with effects of measurement operations per se. We hope, though, that by providing a theoretical framework, testable hypotheses, and methods for testing such hypotheses, our article will have heuristic value.

References


Received August 4, 1986
Revision received November 9, 1987
Accepted November 20, 1987