



The Problem of Informant Accuracy: The Validity of Retrospective Data
Author(s): H. Russell Bernard, Peter Killworth, David Kronenfeld and Lee Sailer
Source: *Annual Review of Anthropology*, Vol. 13 (1984), pp. 495-517
Published by: [Annual Reviews](#)
Stable URL: <http://www.jstor.org/stable/2155679>
Accessed: 22/11/2013 11:24

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Annual Reviews is collaborating with JSTOR to digitize, preserve and extend access to *Annual Review of Anthropology*.

<http://www.jstor.org>

THE PROBLEM OF INFORMANT ACCURACY: The Validity of Retrospective Data

H. Russell Bernard

Department of Anthropology, University of Florida, Gainesville, Florida 32611

Peter Killworth

Department of Applied Mathematics and Theoretical Physics, University of Cambridge,
Cambridge, England

David Kronenfeld

Department of Anthropology, University of California, Riverside, California 92521

Lee Sailer

Department of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

"A measurement whose accuracy is completely unknown has no use whatever" [Wilson (107, p. 232)].

"A serious obstacle in the use of replications for increasing accuracy is the tendency to get closely agreeing repetitions for irrelevant reasons" [Wilson (107, p. 253)].

"My people don't lie to me" (Anonymous Anthropologist).

1. INTRODUCTION

This review focuses on the "fugitive problem" of informant accuracy in reporting past events, behavior, and circumstances. A simple example of the problem is this: if an informant says that she drives 6 miles to and from work, then does she? If she *really* drives 5.3 miles each way to work, then is her report close enough, and under what circumstances is it inadequate?

In one sense, her report is a valid operational measure of her commuting behavior if and only if she actually drives each day, and if she is accurate to an acceptable degree. However, our description of her behavior exists in the absence of an explicit formal theory of her behavior and in the absence of any explicit purpose for studying her behavior. Either theory or prediction would give us a standard by which we could decide whether her description was sufficiently accurate. A theory would also give us the basis for assessing the degree of validity of her description (for example, how much it matters whether she walks instead of drives).

In this paper we will be concerned with validity and accuracy in what we call their naive senses, because most anthropological theory is too inexplicit to permit more specific or subtle determinations. Note that a big improvement in anthropological description could come from the development of theoretical propositions detailed enough to allow us to (a) deduce judgments about the validity of proposed operational measures and (b) estimate just how accurate our measurements are, i.e. determine reasonable error bounds for our instruments (cf 40, pp. 21–31).¹

Naive questions of the type we have described here can be asked about much of our data in anthropology. We often ask informants to provide data on their behavior and on the behavior of others: Did you attend church last week? Did Ralph? We ask them to report on sequences of events: Where did you live after that? Does the tattooing come before or after the plantains are eaten? We ask them to report on environmental and economic conditions: When was the last time it rained around here? How many sheep do you own? In what follows, we survey the literature on informant accuracy, discuss some of the implications, and review some recent attempts to find a way out of the problem.

2. THE LITERATURE ON INFORMANT ACCURACY

In this section, we report on all of the relevant studies that we are aware of. We have found three substantive areas where the question of informant accuracy has been moderately well studied: 1. recall of child care behavior; 2. recall of health seeking behavior; and 3. recall of communication and social interactions. (A fourth major tradition, represented by the works of D'Andrade and Shweder and by Sudman and Bradburn, attempts to deal constructively with the problem of informant accuracy. Those studies are considered separately in Section 4, WHAT CAN BE DONE?) In addition, there have been many isolated reports on a variety of topics. Be warned that the sum of all these reports can be very depressing to the behavioral scientist who relies on recall

¹For a discussion of the difference between validity and reliability, and how both relate to instrument accuracy in the social sciences, see (29c, 68).

and report in lieu of more expensive forms of data collection such as participant observation or direct observation.

Child Care

Child development researchers (13, 15, 38, 39, 61, 64, 77, 78, 101, 104, 105, 111, 112) have tested the accuracy of mothers' recall of their children's scholastic abilities, health status, and so on, against both written records, for validity, and against prior interview data, for reliability. Consistently, mothers' recall was inaccurate by about a third to three quarters, usually as a result of underreporting.

Weisner et al (101, pp. 237–40) compared “trained field observers’ judgments that children were or were not in the role of caretaker of other children, or in charge of another child, with the reports of the children themselves.” Children and observers disagreed 69% of the time that no one was taking care of the children; they disagreed 75% of the time that a parent was caretaker; and 43% of the time that a sibling was responsible. By contrast, they disagreed only 22% of the time about whether or not the child was in a caretaker role. Overall, “when asked ‘who is taking care of you?’ (children) agree with the observer a little less than 50% of the time. But their accuracy is better in response to our other major question, ‘Are you caring for anyone else?’ In this instance the agreement with the observer is approximately 80%.”

Health Care

In a decade of research, Cannell (18) and his associates compared “data obtained from interviews with data obtained from objective records” on pediatric history and on various kinds of health behavior. In all cases, Cannell reports “discrepancies between the two sources.” Twenty-six percent of one-day hospitalizations were found to be unreported in interviews. Thirty-seven percent of accidents without personal injury were unreported if they occurred 9–12 months prior to the interview. Fifty-nine percent of chronic conditions went unreported if a year or more had passed between the time of the last visit to the doctor and the interview (pp. 1–16). Cannell’s 1977 survey monograph is the most complete source of studies on the causes of informant inaccuracy in reporting health events.

The causes include severity of event, time since event, and interaction between certain characteristics of the interviewer and respondent. For example, 100% of illnesses were reported if they occurred “last week,” irrespective of whether or not activity restraints or medical attendance were involved. Where such restraints were involved, only 86% were reported if the illness occurred as recently as “two weeks ago,” and where no such restraints or medical attendance was involved the figure dropped to only 48%.

If four months or more went by since the last visit to a clinic for a chronic condition, then 66% of the visits went unreported in interviews, even when informants were given a checklist of conditions to look at in order to jog their memories. (This is called “aided recall.”) Without checklists, the figure for unreported conditions rose to 84%. Cannell concluded that “the best documented phenomenon of underreporting of health events as well as of a wide variety of other types of events and behaviors, is the decrease in the reporting of events as time elapses” (18, p. 7). For related studies, see (17, 58, 59). For further work on the reliability, validity, and use of health diaries, see (3b, 52a, 95b, 98b).

Communications and Social Interactions

Bernard, Killworth, and Sailer (6–9, 50, 51), hereafter designated BKS, have reported on a series of experiments that tested informant accuracy in reporting of communication. Killworth & Bernard (50) reported an experiment in which an ongoing group of deaf teletype owners in the Washington, D.C. area was asked to “rank the members of the group in the order in which you communicate with them,” providing report data on whom they thought they communicated with. Twenty-five members of the group then logged the output from their teletypes for three weeks, providing a diary of their actual communication. The person most often communicated with was included in the top four ranks only 52% of the time. That experiment was replicated with 54 informants who were asked whom they believed they *might* communicate with in the next month. Then they logged their teletype contacts for a month. They were asked to select, from a list of 387 registered teletype owners, those with whom they had communicated within the past month, and to rank or to scale how much they had communicated with each one (rankers and scalars were assigned to two different groups). Again, their reports were correct about 50% of the time.

In all, BKS did a series of seven experiments, each testing various aspects of informant accuracy in recall of social network or communications contacts. They monitored a ham radio network for a month and then asked 44 users of the network to scale how much they had talked to each person in the network. They did behavior sampling in a university department, in a commercial office, and in a fraternity, each time asking people to report who they communicated with, and how much, over the time period of the experiment. Finally, they monitored a group of communicants on a computer-based electronic mail network. In this last case they were able to collect fully automated data about who talked to whom and how much. They were even able to automate the collection of reports of communication by asking informants for this information on the computer-based network itself.

Many questions about accuracy could be asked of the seven data sets. BKS concluded:

1. Informants who usually kept records of their behavior (e.g. ham radio operators) were *not* more accurate than those who did not.
2. Slightly obtrusive observation, such as occurs with direct observation, has no noticeable effect on informant accuracy.
3. Telling people in a group that they are expected to get more accurate in repeated experiments over time produces no significant improvement in accuracy of reporting communications.
4. Asking people “who do you like?” produces about the same answers as asking them “who do you talk to?”
5. Asking people about the significance or importance of their interactions with others is of little use since it produces no better results than simply asking them simply who they talked to.
6. In general, the more recent the time “window” over which informants report their communication (ie. Who did you talk to during the past 24 hours? as opposed to Who did you talk to last week?) the more accurate they tend to be. However, only 6% of informant accuracy can be accounted for by this factor. (As a matter of interest, questions about interactions two weeks previously were the least accurately reported, with better accuracy for shorter and longer times ago.)
7. Over all data sets, people can recall or predict less than half of their communications, measured either on amount or on frequency.

BKS (9, p. 31) concluded, on the basis of seven experiments, that what people say about their communications bears no useful resemblance to their behavior. Furthermore, individual differences in accuracy could not be accounted for by any of the usual characteristics of people or groups, such as sex, age, time in group, centrality in group, etc. They also concluded (8, p. 71) that “the error is so great that statistical and numerical techniques for washing data collected by recall instruments cannot solve the problem.” Recently, Romney and his associates (79, 80) have been working with BKS’s data and report significant progress in teasing out accurate signals from noisy data from informants. More on this in Section 4 when we discuss “WHAT CAN BE DONE?”

Hyett (41a, pp. 137–38) reports an effort by the British Post Office to find out more about the telephone use habits of its customers. They asked 2000 of their customers to keep diaries of all their calls for two weeks. “In order to test the validity of the diary entries, the calling behaviour of a sub-sample of 354 diarists was specially monitored at their respective exchanges.” In general, the people who made very few calls tended to overreport, and those who made a great many calls tended to underreport. In the analysis, Hyett compared grouped call data with grouped report data. That is, he compared 1–5 calls actually made, with 1–5 calls reported in the diaries, 6–10 actual calls, with

6–10 reported calls, and so on. Even with this generous grouping of data, only 71% of the diarists were “in the same category from both sources” of data.

In one of the few studies of informant accuracy by anthropologists, Polgar & Hammer (76) actually observed the speech interactions of 20 regular patrons of a doughnut shop in Manhattan and asked informants about their relations with one another. They reported a .69 correlation between the number of observed interactions and the number of reported interactions, accounting for 47% of the variance.

Some Isolated Studies

In addition to these three areas, other studies have dealt with informant accuracy in recall of sick leave taken, voting, petition signing, crime events, food intake, employment history, drug use, language use, births, household expenditures, and general daily activities.

Young & Young (114) asked a number of community leaders in rural Mexico questions like “when did you get electricity here?” and “what percentage of the people here customarily eat eggs?” They found that for publicly available information (about electricity, for example) there was both agreement and accuracy. However, for questions where the answers were not public (as in egg consumption), there was very little agreement, and hence high inaccuracy.

Gray (34) compared the reported sick leave of British Government employees with their actual record of leave days taken. Of the 433 informants tested, 205 took no leave time during the period tested (4.5 months). Of these, 192 correctly reported that they had taken no leave. Of the 228 who had taken some leave, 74 gave the correct answers.

Traugott & Katosh (96) compared informants’ reports of having registered and voted during the 1976 Presidential election with actual voter registration and voting records. Of 1417 persons who had registered and voted, only 55 said that they had not registered and only 27 claimed not to have voted. However, of 736 informants who had neither registered nor voted, 288 claimed to have registered and 217 reported having voted (p. 365). (See also 21, 66, 72 for studies of the validity of voting behavior reports.)

Pierce & Lovrich (74) validated informants’ reports of having signed a petition. In early 1979 they mailed questionnaires to 300 of 10,000 persons who had signed a petition in December of 1977. (The petition was to place a proposed Hydropower and Water Conservation Act on the Idaho ballot.) Only 56% of the known signers of the petition reported that they had signed it (74, p. 167).

According to a report of the National Research Council (69), 73% of crime incidents that occurred within three months before an interview were placed in the correct month by victims. This dropped to 60% for incidents that occurred

within six months prior to the interview, and 49% for incidents occurring up to 11 months previously (69, p. 38).

Greger & Entyre (36), tested the recall, over 24 hours, of food intake of female adolescents. They reported recall to be valid for intake of energy, protein, calcium, and zinc, but found that informants “were unable to recall their food intake with enough accuracy so that their intake of vitamins A and C, thiamin, riboflavin, niacin, and iron could be calculated within the range of two-thirds to four-thirds of their actual intake” (36, p. 72). For further work on informant inaccuracy in nutrition studies, see (4, 11b, 29b, 32, 54, 65, 113a, b).

Morgenstern & Barrett (67) compared reported unemployment during “the previous week” with reported unemployment “during the previous year.” They found that “women more than men, and youth of both sexes, all appear to understate significantly their unemployment when asked to recall it up to a year or more later compared with statements made for the previous week. For some groups this understatement is as high as 50 percent” (p. 357). By contrast, “Whites in high unemployment years, males 25 and over, and females 45 and over all evidence some tendency to overstate their unemployment” when they report on the past year as compared to their reports for the previous week (p. 357). They did not provide exact figures, however. [See also (3a) for further evidence of underreporting of employment history.]

Withey (108) examined the accuracy of recall of income by asking people to report in January 1948 their income for the year 1947 (i.e. their current income); to report in January 1949 their income for the year 1947 (i.e. their recalled income); and to report in January 1949 their income for the year 1948 (a further report of current income). He found that “a respondent’s recall of his previous income was not generally reliable and that the errors in recall were not random” (108, p. 197). “If, however, one is satisfied with rather gross measures, which” he concluded “would tolerate errors of around one thousand dollars, the error can, for recall periods of at least one year, be disregarded” (p. 204). This figure would have to be adjusted considerably today: 86% of Withey’s respondents had incomes of \$1000–\$7500. A thousand dollar error in 1947–48 was a substantial one. By contrast, Ferber & Birnbaum (30) found that “current self-reported wages differed from published ones on the average by 5.1 percent, while the comparable figure for retrospective data is 6.0” (p. 118). They attribute this partly to the fact that only 3% of their respondents failed to provide data on current salaries, while 29% of past salaries were not reported. “A plausible explanation” they said “is that respondents, in general, refused to make wild guesses and tended to report only salaries they remembered” (p. 116).

Bachman & O’Malley (2) note that “among high school seniors who report

any drug use during the past year, reported use during the past month is generally inconsistent with reported use during the past year; either the annual frequencies are too low, the monthly frequencies too high, or both" (p. 536). Although they did not have matched cognitive and behavioral data with which to test the alternatives, they conclude that "the discrepancies we have observed are not genuine, but rather reflect some considerable inaccuracies in reporting" (2, p. 545). They conclude, on the basis of Cannell and his associates' work (see above), that drug use is probably highly underreported, especially when informants are asked to scan long periods of time like a year.

Akers et al (1) report that juvenile reports of their smoking behavior were validated by a saliva test (presence of elevated level of thiocyanate). Their data show that thiocyanate levels "can be used to distinguish clearly the moderate to heavy smoker from the non-smoker and to estimate proportions of valid responses to self-reported smoking" (p. 241). The data cannot distinguish whether an informant really smokes a pack a day or half a pack a day. Nevertheless, Akers et al conclude that their data "add to the preponderance of evidence from other studies that self-reports of deviant behavior in adolescent populations are usually valid" (p. 248). Considering Bachman and O'Malley's findings (2), this conclusion seems overly optimistic.

Doreian (28, p. 159) asked Scottish Gaelic speakers who they spoke Gaelic with, and what they talked about. Since she had already done extensive ethnographic fieldwork, she was in a position to judge the accuracy of her informants' responses. She reports that "where the questions were posed in terms of interlocutors the inaccuracy was in the direction of *overestimating* the amount of Gaelic spoken." Questionnaire responses to the effect that respondents "always" or "usually" spoke Gaelic with spouses or children were highly inflated. On the other hand, "where the questions were posed in terms of verbal activities and topics, the inaccuracy was generally in the direction of *underestimating* the amount of Gaelic spoken." People whom she had heard discussing local affairs in Gaelic claimed never to do so.

Even questions about the number of births a woman has had or the number of siblings a person has are subject to substantial error (11a), as are questions regarding household and consumer expenditures (46, 47, 70, 71, 95a).

In psychology, there is a tradition of research based on "assessment of daily experiences" in which subjects provide data about the events in their daily lives. Stone & Neale (91) developed a method for gathering data which consisted of having both partners in a marriage record the daily life experiences of one partner. The idea was that this would increase the accuracy of the reporting. The average concordance of the recorded experience was 31%. They tried to increase the concordance by telephoning the participants during the day and reminding them to record the data, but this did not help. Finally they eliminated all instances where the wife, for example, did not know about an

event that her husband had recorded, and all events that the wife recorded but that the husband forgot to record. This raised concordance to 67%.

Summary of the Literature

The results of all of these studies leads to one overwhelming conclusion: on average, about half of what informants report is probably incorrect in some way.

One could argue that this is neither bad nor surprising. Social scientists are accustomed to low correlations *between variables representing different concepts*. The case here, however, is very different. The previous parade of studies documenting low correlations between self-report data and observation data concern a concept and its measurement. A weak relationship between two distinct concepts is to be expected. A weak relationship between a concept and the accurate measurement of the concept is unacceptable.

One could also argue that social scientists are not so naive as to assume that self-reports are genuine proxies for behavior—that they know their data only represent human cognition about an external reality. There are two things wrong with this argument. First, even if it were true, we would still be left with the annoying problem of trying to explain why we are interested in retrospective data about behavior in the first place. If we claim, in defense, that we are really interested in people's cognition about their behavior, then we should be making our contributions to the memory and cognition literature and not masquerading as behavioral scientists. And second, the argument just isn't true . . . that is, the behavior of social scientists of all disciplines is not consonant with the claim that they are aware of the lack of validity of retrospective data.

In a recent article in the *Archives of Sexual Behavior*, Catania & White (20) report a study of masturbation among aged persons. They say that "frequency of masturbation was assessed by asking the participant (in the study) to report the number of times he/she had masturbated to orgasm in the previous month" (p. 240). In another study in the same journal, five researchers (82) report the results of an investigation on the effects of testosterone replacement on the sexual behavior of hypogonadal men. "Each patient was given a diary and was asked to fill in daily the number of spontaneous erections and the frequency of masturbation, coitus, and other sexual activities" (p. 348).

It is easy to dismiss this kind of error on the grounds that the particular behavior being investigated is "sensitive" and that one would not expect people to report on it accurately. But of all the possible behaviors that we might be interested in, which would we expect to be reported most accurately? How about the personal networks of urban residents? Keefe (45), in her study of such networks, noted that "informants were asked to name the people with whom they had talked face to face in the week prior to the interview" (p. 65).

Christmas gift giving? Caplow (19) examined kin networks and the giving of Christmas gifts. Caplow notes that

the interview schedule had a separate form for each Christmas gathering attended by the respondent in 1978, which called for a list of the persons attending, the menus of the meals served, the circumstances under which pictures were taken, a description of all the gifts given and received by the respondent, the relationship of givers and receivers to the respondent, and if recall permitted, a description of gifts exchanged between other persons (19, p. 384).

Interviewing took place in February 1979, with the assumption that people can, after two months have passed, give a researcher these data. What effect would there be on the analytic results of the study had respondents been off by, say, 40% on the number of gatherings attended? Or by, say, 20% on the number of gifts they received? What if the figures were 60% and 50%? How accurate should these data be in the first place? Clearly, no one asked these questions—not in the design of the research, not in its conduct, not in its reporting, and not in its review for publication.²

In sum, despite the evidence, the basic fact of informant inaccuracy seems not to have penetrated either graduate training or professional social science research. Informant inaccuracy remains both a fugitive problem and a well-kept open secret.

²Unlike informant accuracy, other threats to data are well documented and have even become major areas of methodological research, commanding a great deal of time, money, and attention. Survey and interview data, for example, are threatened by “response effects,” which include such things as nonresponse (primarily for mailed questionnaires); cheating by respondents; reaction to the gender, or race, or socioeconomic class of either the respondent or the interviewer; and so on. Response effects in surveys have been studied by hundreds of scholars, and there are courses on the subject in many departments of sociology. Sudman & Bradburn (93) provide a thorough review of over 800 sources. (Also see the review article by Cannell (18) and the massive bibliography by Dalenius (23).

Direct observation data are threatened by reactivity, by expectations on the part of both observers and those observed, and even by outright cheating by hired observers who may try to cut down their work load. These problems have been examined in great detail, primarily by psychologists. They are discussed in graduate training and, whenever appropriate, in the methods sections of scientific reports. An excellent introduction to the problems associated with direct observation is provided by Kent & Foster (49). See D’Andrade (26) for a penetrating analysis of the frailties of human beings as “behaviorscopes.” On expectation biases, see the classic by Rosenthal (81) and the thorough review of the halo effect by Cooper (22). See Webb et al (99) for forceful statements on reactivity biases in observing human behavior.

There is a commonly held notion in anthropology that participant observation is a means of reducing the severity of reactivity and other biases in direct observation. Unfortunately, we do not know if this is true, and if so how *much* error is reduced by this kind of training. Nonetheless, there is a general consensus among anthropologists that personal biases do indeed color fieldwork, and that these biases should be discussed in the introduction to ethnographies. On participant observation see (5, 55, 60, 73, 84, and 90).

3. CONCEPTUAL VARIABLES: A WAY OUT?

Of course, if it were easy to collect accurate observational data about behavior, investigators would long ago have abandoned such a weak technique as asking informants about the past. Another technique for getting around the problem is to ask informants about conceptual variables like beliefs and attitudes. This is based on the assumption that internal states (beliefs and attitudes) generate behavior and must be highly correlated with it. Things like “machismo,” “alienation,” “authoritarianness,” “willingness to accept new technologies,” “support for abortion on demand,” “achievement motivation,” and “goal orientation” are examples of conceptual variables.

In the absence of either a well-defined theory (cf 40, chap. 1; 35, chap. 10) or a clearly described dependent variable it becomes meaningless to discuss the validity or accuracy of such conceptual variables. We have no way of comparing our operational versions with any kind of external reality. In effect, such variables are simply “invented.” A successfully predicted dependent variable, in the absence of a real theory, speaks to the accuracy problem, but says nothing about the validity of either variable as a representative of some underlying reality. It is only in the context of a well-defined theory that we can talk seriously about validity, and can begin to separate our assessment of the validity of the operational measure from the more general question of the validity of the theory itself. The problem becomes like the measure of mass in physics—the theory contains sets of propositions which allow one to evaluate the adequacy of one’s operational data measures. But even here, the variable (concept) only has meaning in the context of the theory, and hence its operationalization can only have meaning if (to the degree that) the including theory is valid.

Most anthropological (indeed, most social scientific) variables are not rooted in such theory and so fall in the “invented” category. In addition, conceptual inventions lack even the kind of naive face validity that behavioral variables (e.g. did Smith feed Jones?) have by virtue of their amenability to direct observation. Intelligence tests, for example, are highly reliable, and with varying degrees of accuracy can be used to predict various kinds of academic performance. But everyone knows that their validity is open to serious question: are they measuring intelligence to begin with? By whose “inferred” definition? Accuracy of response is similarly irrelevant for such queries as: “What is your position on abortion?” or “Are ancestral spirits easily provoked?” Since there is no external referent against which to compare the answers to such questions, we simply take the answers as data and leave it at that.

Fair enough, too. Given that people tell the truth as they see it, and are not practicing active deceit in response to our queries (and in the absence of

appropriate theoretical assumptions), it is hard to imagine how to test the validity of attitudinal or other internal-state responses against anything. [But see (10) for an interesting discussion of patterned, culturally sanctioned dishonesty in reporting one's behavior and personal economic circumstances.] If someone says that they believe abortion is bad, or that the ancestral spirits are not easily provoked, then that's the end of it. Of course, one does want to know whether the data are reliable, whether they predict other attitudes, and whether they predict *some* kind of behavior.

Unfortunately, the case for attitudes as predictors of behavior is far from convincing. Attitude-behavior inconsistency was identified forcefully in a classic article by La Pierre (53) in 1934. Accompanied by a Chinese couple, La Pierre traveled by car a total of 10,000 miles, crossing the U.S. twice between 1930–32. The three travelers were served in 184 restaurants (refused in none), and were refused accommodation in only one out of 67 hotels. Later, La Pierre sent letters to all the establishments they had patronized, asking if they would "accept members of the Chinese race." Ninety-two percent replied "no."

Since that time, there have been hundreds of studies testing the relationship between attitudes and behavior. [For reviews, see (27, 48, 62, 63, 85, 106). For a review of the literature on the relationship between physiological and mental phenomena, see (44).] Some of the studies have been clever experiments: Freeman & Atavov (31) tested the amount of actual cheating behavior among students, compared with the students' reported attitudes about cheating. They concluded that "Since all correlations were insignificant" the results of their study "cast some doubt upon the validity of attitudinal survey questions for the assessment of overt behavior" (31, p. 46).

Some writers (27) have concluded that the situation is practically hopeless. Others have focused on making the best of a tough situation. Weigel & Newman (100), for example, found that they could increase attitude-behavior correspondence by increasing the scope of the behavioral referent. Since attitudes are general, while individual behaviors are specific, they observe that we should not expect to find much correspondence between attitudinal measures and individual behaviors. By amalgamating the behavioral measure into an index (rather than relying only on individual behaviors), they were able to increase attitude-behavior correspondence to .62 in an experimental situation. The optimistic view is that .62 is a high correlation. The pessimistic view is that, even by broadening the scope of the measure, the attitudinal index failed to correlate with about a third of the observed data (or 62% of the variance).

Kahle & Berman (42) offer evidence that attitudes actually cause behaviors, and Kahle et al (43) conclude from experimental evidence that attitudes are "nonspuriously antecedent to behaviors" (p. 402). They studied the relationship of survey items that were designed to measure "outgoing attitudes" and "outgoing behaviors" among high school boys. They found a high degree of

consistency in expressed attitudes and reports of behavior, and were careful to note that their results depended upon an assumption that “self-reports of behavior correspond to observations of behavior” (p. 406). They said that they “attempted to avoid faults of measurement that tend to decrease the validity of self-reports.” They found a correlation of 0.56 between observed behavior and self-reports, accounting for 31% of the variance.

So the question remains: Are we to be reassured by instruments that fail to correlate with a third to a half of the data, i.e. miss a third to two-thirds of the variance? From our perspective, both ways of looking at these results (correlation and variance accounted for) reflect an important fact: an instrument with this level of accuracy would be unacceptable in most fields of natural science. In fact, it is almost unthinkable that results without error bounds would be published in most natural science journals.

4. WHAT CAN BE DONE?

Scholars who have considered the problem have tried to elucidate the causes of inaccuracy, on the assumption that nine-tenths of cure is diagnosis. While no cure is in sight, the work of Sudman & Bradburn (92–94), D’Andrade (24–26), and Shweder & D’Andrade (88, 89) provide some valuable clues about what the cause might be.

Sudman & Bradburn (92) argue that “memory effects in surveys can be described by a <mathematical> function that is the product of effects due to omissions and telescoping” (p. 815). They offer a plausible three-parameter psychological model of two processes. Two parameters are necessary to describe simple omissions due to the fact that memory decays in time. The third parameter describes the fact (well known to psychologists) that informants tend to underestimate (i.e. telescope) the time dimension. Simply stated, events in the past are likely to be recalled as being more recent than they actually are. Sudman & Bradburn (92) find several naturally occurring circumstances where detailed records may be compared with memory reports. For example, they compare recollections of department store purchases with store credit card records.

They then use the model to compare a variety of data collection techniques such as “aided recall,” where the informant is presented with fixed alternatives (“Did you buy a coat? Any shirts? Any shoes?”) “record assisted recall,” where the informant is encouraged to keep an accurate daily diary; and “bounded recall,” where the informant is interviewed once to set a baseline, and then again later, after the desired time has passed, to guard against the problem of telescoping. They find that the memory model fits “pretty well,” and they conclude that the various interviewing techniques increase accuracy “by an average of about ten percent.” This report is typical of studies from the sample

survey field (94). They often show how simple rephrasings, clearer page layouts, reordering of questions, and so on, can increase the accuracy of responses by 5 to 10%.

Unlike Sudman and Bradburn, D'Andrade's and Shweder's work goes beyond time decay, or "memory drift," and concentrates on the idea that memory is subject to systematic distortion due to cultural training. Informants respond to questions by reporting cultural norms, or "what goes with what," rather than dredging up actual events, circumstances, behaviors, or personality traits. They reason that information about events and such must be stored in memory with lexical labels attached; culturally defined labels then become the mnemonics that enable us to retrieve the information that was stored. But these mnemonics are obviously mere abstractions—systematically distorted shorthand for an immense amount of data about behavior, personality characteristics and so on.

For example, Shweder & D'Andrade (89) point out that, given the following choice

M. G. has self-esteem. Therefore M. G. is probably not a leader.

and

M. G. has self-esteem. Therefore M. G. probably is a leader.

most informants endorse the second statement, despite the fact that, as Shweder (87, p. 643) has shown, people believe that most people with self-esteem are not leaders. Shweder and D'Andrade explain this as follows:

When informants tell us that self-esteem and leadership go together, or draw the inference that someone with self-esteem is likely to be a leader, they are not processing information about the correlation of two variables or the conditional probability of one given the other.

What they are doing is judging the extent to which two events co-occur by the extent to which the events affiliate in their minds or have strong verbal associative connections (89, p. 51).

And, as D'Andrade points out (26, p. 177)

With this type of memory error, any attempt to discover how human behavior is organized into multibehavior units—such as dimensions or clusters—which is based on data consisting of long-term memory judgments will result in conclusions which primarily reflect the cognitive structure of the raters.³

Other researchers have sought to explain the existence of informant inaccuracy with similar arguments. Recall that Weisner et al (101), studying children's

³D'Andrade also points out the fundamental idea of systematic distortion was adumbrated by Newcomb in 1931. A comparison of judges' ratings for the behavior of a group of boys showed a correlation of .49. This was not much to crow about, but a correlation of the judges' ratings with records of behavior was much worse (.14). Newcomb concluded that the "close relation between the intra-trait behaviors which is evident in the ratings may, therefore, be presumed to spring from logical presuppositions in the minds of the raters, rather than from actual behaviors" (p. 228; quoted in 26, p. 177).

perceptions of their roles as either caretakers or wards, found that up to 50% of their informants' answers were inaccurate. They argue that they are not really studying specific behaviors, but rather "perceptions of a felt role performance" (p. 242). They say "we believe that this kind of inaccuracy occurs because people do not count their behavior . . . Rote recording has limited functions in daily living; it is not adaptive to store uncoded . . . bits of information—albeit information vital to the social scientists whose methods are predicated on accurate reporting" (p. 241).

Where does this leave us? Informants *are* inaccurate; memory *does* decay exponentially with time. [See (29a) for the earliest work on this phenomenon, and (86) for a review of the literature on memory and cognition; see (57b) for a review of work on recognition memory; see (37) for a review of memory decay in everyday life.] And on top of all this there appears to be systematic distortion in how informants recall just about everything. Furthermore, recall may be effected by the subject of the study, by whether informants are aided in their recall in some way during the interview (e.g. giving them checklists rather than open-ended questions), by whether they keep diaries, by conditions of the interview, or by a variety of cultural factors. There has hardly been any research at all on any of these things. It is not enough to point out that there are reasons for the fact that informants are poor instruments for studying human behavior. We must begin a systematic effort to discover, for each kind of data that we want, what the error bounds are for the instruments that we use to collect those data.

Are there some general rules that we can use to make a reasonable guess about the efficacy of our instruments? Is informant accuracy a function of the questions we use? Or is it a function of informant personality and/or socioeconomic characteristics? Do people in some cultures report things in general with greater accuracy than people in others? Do people in some cultures report environmental events, or personal property ownership (or whatever) better than people in others? Or is the level of informant accuracy a random event? How can informants be made more accurate? How *much* more accurate can they be? How can the most accurate informants be identified without laborious checks on their reports? Are some areas of behavior easier for informants to recall than others? Some of these questions have been addressed by several researchers, but the surface is still shiny. Kronenfeld et al (52b), for instance, report an experiment in which informants leaving restaurants were asked to report on what waiters and waitresses were wearing. Informants showed much higher agreement about what waiters were wearing than about what waitresses were wearing, despite the fact that none of the restaurants in question had waiters. Informants provided greater detail about the kind of music that was playing in restaurants that, in fact, did not have music than they provided for restaurants that did have music. Kronenfeld's presumption was that, in the absence of

specific memories about things that informants had seen or heard, informants accepted the presuppositions (buried in the questions) that there had been waiters and/or music and they turned to cultural norms for descriptions of what must have been there.

Informants who had noticed something (about waitresses or music that *was* present) were like the blind men describing the elephant. They reported various components of the actual events or circumstances, because that was all they could have experienced at one time. (This is clearly the case for eyewitnesses, where informant inaccuracy can have devastating effects on the lives of wrongly convicted individuals; for reviews see 14, 33, 56, 57a, 103, 109, 110). This explains the thinness and variation in their descriptions. Those who started from whole cultural cloth were able to provide rich descriptions, unencumbered by partial memories and working from complex normative wholes, based on many experiences over a lifetime. The design of the experiment was not sufficiently complex to make these statements more than just suggestions. Clearly, however, this is an area ripe for research by anthropologists working in the field in many different cultural environments.

Perhaps informants answer the culturally appropriate questions nearest to the ones we ask them, rather than the questions we ask them? To test this more directly, Kronenfeld and Armstrong (Unpublished) questioned a sample of people who had recently been to a talk in a relatively unfamiliar room. The informants were asked specifically about the color of the chalkboard—which was, in fact, blue, and hence unique on campus. Most informants answered green (the color of all other campus chalkboards) and some answered blue. Regardless of which answer they gave, the experimenter badgered the informants, asking if in fact they were not mistaken. Those who said blue stuck by their answers, while most of those who said green quickly and easily caved in. The blues, of course, had really noticed the color and were not open to doubt. The greens were responding according to cultural norms and were thus open to alternatives from the investigator. The fact that those informants working off of a cultural norm tended strongly to back down while those who worked off of actual experience did not provide a way for a researcher to approach the problem of assessing the correctness of independent informants concerning some shared experiential fact for which the researcher lacks independent information. The answer that shows more resistance to badgering is more likely to be accurate. This work is consistent with Cancian's finding (16a) that deviations from truth (in Zinacanteco informants' descriptions of the *cargo* careers of others) tended to be in the direction of cultural norms. [For an extended discussion of how beliefs may operate in relation to behavior, see (16b; see also 41b).]

There is some evidence that groups are better able to recall details of events than are individuals. Leann Martin (personal communication) told some Bush-

men tales to her classes and had the students tell the stories back to her from memory. She varied the size of the groups in her experiment and found that the bigger the group the more accurate was the tell-back. Apparently, each student noticed different things, and none noticed very much. The tell-back of individuals consisted of the remembered elements, combined with enough elements from their own cultural models to make what seemed to them to be a reasonable story. Putting students together in larger and larger groups does not yield more of a hodge podge; group renderings were both fuller and more accurate than were individual ones. This finding, along with the chalkboard example, would seem to imply that at some level people do know the difference between what they actually heard and what they only presumed on the basis of norms.

Tversky & Kahneman (97, 98a) have contributed another piece to the puzzle. They show how subconscious decision making is systematic and patterned, but completely illogical. For example, people will go to great lengths to save five dollars on a twenty dollar item, but not on a hundred dollar item. Logically, it is the same five dollars, but people calculate in percentages. Tversky and Kahneman have shown, in many experiments, how minute rewording or restaging of situations *radically* affects the way people will respond. People grossly overestimate small probabilities and underestimate large ones. These studies are relevant to informant accuracy because ultimately the answer to a question posed by a social scientist is the result of a large number of subconscious decisions. “What does he want to know? Does he already know that? What word should I use here? Did he understand? What should I say next? Is he going to use this information in a bad way? Am I likely to see him again? Is he likely to tell others?” and so on.

Some promising work comes from Romney and his associates (79, 80). As noted earlier, they have reanalyzed several of the datasets from BKS (6–9, 50, 51), and have achieved encouraging results. In their final paper in the informant accuracy series, BKS (9) noted that while individuals in a group could not recall whom they communicated with, there was evidence that the “group at large” did “know” who the most popular (ie. communicated-with) persons were. Romney & Faust (79) found that 1. the more similarly two people judge the communication pattern of others, the more they interact with each other; and 2. the more two people share accurate knowledge of others, the more they interact with each other. Romney & Weller (80) followed these findings at the aggregate level by trying to predict individual differences in accuracy from the patterns of recall among informants. They offer a theory of individual differences in informant accuracy—a theory which may generalize eventually to other types of data. Romney and Weller’s work allows one to distinguish *relatively* accurate informants from *relatively* inaccurate ones without direct information about the actual behavior. However, it does not allow one to

determine exactly how accurate the particular recall (i.e. cognitive) data set might be. In other words, for many purposes they would still have to depend on having matched data sets for both informant cognition about some behavior and the behavior itself. There is no way to tell, a priori, how accurate a particular cognitive data set might be.

If informants agree with one another in information that they independently provide, then there must be a basis for this. The two likely explanations are norms and actual experience. In situations that lack a strong specific norm, agreement should imply shared experience. In situations where a specific norm is clear and agreement could be on either basis, then, if one gets two clusters of answers, Kronenfeld's badgering technique provides a way of approaching the problem. Much empirical work is needed concerning the attributes of eliciting situations and of tasks that skew informant responses toward norms or towards experience.

However, there is an even more fundamental point that needs to be made. It is worthwhile and meaningful in any science to account for the variation of accuracy both between and within instruments. If a previously accurate thermometer refuses to give correct readings somewhere, then perhaps we can explain this by the drastic reduction in air pressure in that region. But that is not the end of the story, only the beginning. By *allowing* for the pressure changes, can we restore the accuracy of the thermometer? If the answer is "yes," then we have 1. acquired a reliable instrument and 2. learned some more facts about the physical world. On the other hand, if the answer is "no," then we have only achieved (2) and the thermometer is *still* unreliable and can only be used under tightly controlled circumstances.

What matters, then, is what level of accuracy can we achieve in the instrument we use *after making empirical and/or theoretical corrections*? Knowing which informants are likely to be least accurate is surely important; but knowing whether *any* informant is sufficiently accurate to rely upon his or her responses is more important still.

As the lead quotes from Wilson (107) at the top of this article make clear, interviewing many inaccurate informants will not solve the accuracy/validity problem. Of course, each of us in anthropology believes that fieldwork (going out and staying out and collecting our own data "on the ground," as they say) lessens or even nullifies the effects of informant inaccuracy. *Our* informants are accurate; and long-term participant observation ensures that we *know* which informants are accurate and which are not. Our discipline teaches us to seek out "good" informants (i.e. people who know a lot and who can report accurately on the culture we are studying), and our mythology teaches that we can, in fact, do it through long-term participant observation.

Our mythology may be correct. Surely, some people *are* better informants than others (cf 12, 75, 83, 102), and perhaps fieldwork does sharpen our ability

to choose such informants over less capable ones [though Young & Young (114) found otherwise]. It is time we determined whether key informants are really better than a representative survey, and if so by how much and under what circumstances.

From our perspective, then, the evidence of informant inaccuracy ought not to lead to complaints or to despair. It ought to lead instead to a rich, relatively unexplored arena of research. Surely our informants are not to blame for being inaccurate. It is not even their problem. People everywhere get along quite well without being able to dredge up accurately the sort of information that social scientists ask them for. If we have a great deal of inaccuracy in our data, then we have only ourselves to blame for using the instruments of our craft—that is, the questions that we use to tap our informants' memories—so uncritically.

If every field researcher were to test the accuracy (that is, the error bounds) of just *one* instrument, one query; and if all such tests were published in very brief notes in a central place; then in a few years we would know a *lot* more than we do today. There is an unfortunate tendency in our discipline to talk about methodological studies as "just methodological." In the natural sciences, good instruments are highly prized and the search for such instruments is a noble endeavor. Understanding informant accuracy is the interface, we believe, between method and theory. We hope that the search for such an understanding will soon become a major research focus in anthropology.

Literature Cited

1. Akers, R. L., Massey, J., Clarke, W., Lauer, R. M. 1983. Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Soc. Forces* 62:234–50
2. Bachman, J. G., O'Malley, P. M. 1981. When four months equal a year: Inconsistencies in student reports of drug use. *Public Opin. Q.* 45:536–48
- 3a. Bancroft, G. 1940. Consistency of information from records and interviews. *J. Am. Stat. Assoc.* 35:377–81
- 3b. Baranowski, T., Dworkin, R. J., Cieslik, C. J., Hooks, P., Clearman, D. R., et al. 1984. Reliability and validity of self report of aerobic activity: family health project. *Research Quarterly for Exercise and Sport*. (Forthcoming)
4. Beaton, G. H., Milner, J., Corey, P., McGuire, V., Cousins, M., et al. 1979. Sources of variance in 24-hour dietary recall data: Implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* 32:2546–59
5. Becker, H., Geer, B. 1957. Participant observation and interviewing: a comparison. *Hum. Organ.* 17:28–32
6. Bernard, H. R., Killworth, P. D. 1977. Informant accuracy in social network data, II. *Hum. Commun. Res.* 4:3–18
7. Bernard, H. R., Killworth, P. D., Sailer, L. 1980. Informant accuracy in social network data, IV. A comparison of clique-level structure in behavioral and cognitive data. *Soc. Networks* 2:191–218
8. Bernard, H. R., Killworth, P. D., Sailer, L. 1981. A review of informant accuracy in social network data. In *Modelle für Ausbreitungsprozesse in sozialen Strukturen*, ed. H. J. Hummel, W. Sodeur. Duisberg: Sozialwissenschaftlichen Kooperative
9. Bernard, H. R., Killworth, P. D., Sailer, L. 1982. Informant accuracy in social network data, V. An experimental attempt to predict actual communication from recall data. *Soc. Sci. Res.* 11:30–66
10. Bilmes, J. 1975. Misinformation and ambiguity in verbal interaction: a northern Thai example. *Int. J. Soc. Lang.* 5:63–75
- 11a. Blacker, J., Brass, W. 1979. Experience of retrospective demographic inquiries to determine vital rates. In *The Recall Method in Social Surveys*, ed. L.

- Moss, H. Goldstein, pp. 48–60. London: Univ. London Inst. Educ., Stud. Educ. (new ser.) No. 9
- 11b. Block, K. G. 1982. A review of dietary assessment methods. *Am. J. Epidemiol.* 115:492–505
 12. Boster, J. S. 1983. Requiem for the omniscient informant: There's life in the old girl yet. In *Directions in Cognitive Anthropology*, ed. J. Dougherty. Urbana: Univ. Ill. Press
 13. Brekstad, A. 1966. Factors influencing the reliability of anamnestic recall. *Child Dev.* 37:603–12
 14. Buckhout, R. 1974. Eyewitness testimony. *Sci. Am.* 231 (6):23–31
 15. Burton, R. V. 1970. Validity of retrospective reports assessed by the multi-trait-multimethod analysis. *Dev. Psychol. Monogr.* Vol. 3, No. 3, part 2
 - 16a. Cancian, Frank. 1963. Informant error and native prestige ranking in Zinacantan. *Am. Anthropol.* 65:1068–75
 - 16b. Cancian, Francesca 1975. *What are Norms?* New York: Cambridge Univ. Press
 17. Cannell, C. F. 1969. International comparisons of health care utilization: A feasibility study. *Vital and Health Statistics*, US Public Health Serv. Publi. No. 1000-ser. 2, No. 33. Natl. Cent. for Health Stat., USGPO
 18. Cannell, C. F. 1977. A summary of research studies of interviewing methodology, 1959–1970. *Vital and Health Statistics*, Ser. 2, data evaluation and methods research: No. 69. DHEW Publ. No. HRA-77-1343. Rockville, Md: Public Health Serv., Health Resour. Admin., Natl. Cent. Health Stat.
 19. Caplow, T. 1982. Christmas gifts and kin networks. *Am. Sociol. Rev.* 47:383–92
 20. Catania, J. A., White, C. B. 1982. Sexuality in an aged sample: Cognitive determinants of masturbation. *Arch. Sex. Behav.* 11:237–44
 21. Clausen, A. 1968. Response validity: vote report. *Public Opin. Q.* 32:588–606
 22. Cooper, W. H. 1981. Ubiquitous halo. *Psychol. Bull.* 90:218–44
 23. Dalenius, T. 1977. Bibliography on non-sampling errors in surveys, I, II, III. *Int. Stat. R.* 45:71–88, 181–97, 303–18
 24. D'Andrade, R. G. 1965. Trait psychology and componential analysis. *Am. Anthropol.* 67(5):215–28
 25. D'Andrade, R. G. 1973. Cultural constructions of reality. In *Cultural Illness and Health*, ed. L. Nader, T. W. Maretzky, pp. 115–27. Washington, DC: Am. Anthropol. Assoc.
 26. D'Andrade, R. G. 1974. Memory and the assessment of behavior. In *Measurement in the Social Sciences*, ed. H. M. Blalock Jr., pp. 159–86. Chicago: Aldine
 27. Deutscher, I. 1973. *What We Say/What We Do*. Glenview, Ill: Scott, Foresman
 28. Doreian, N. C. 1981. *Language Death: The Life Cycle of a Scottish Gaelic Dialect*. Philadelphia: Univ. Penn. Press
 - 29a. Ebbinghaus, H. 1964. *Memory*. New York: Dover (orig. publ. 1885)
 - 29b. Emmons, L., Hayes, M. 1973. Accuracy of 24-hour recalls of young children. *J. Am. Diet. Assoc.* 62:409–15
 - 29c. Featherman, D. L. 1979–80. Retrospective longitudinal research: methodological considerations. *J. Econ. Bus. (Temple Univ. Sch. Bus. Adm.)* 32:152–69
 30. Ferber, M. A., Birbaum, B. 1979. Retrospective earnings data: some solutions for old problems. *Public Opin. Q.* 43:112–18
 31. Freeman, L. C., Atavov, T. 1960. Invalidity of indirect and direct measures of attitude toward cheating. *J. Pers.* 28:443–47
 32. Garn, S. M., Larkin, F. A., Cole, P. 1978. The real problem with 1 day records. *Am. J. Clin. Nutr.* 31:1114
 33. Goldstein, A. G. 1977. The fallibility of the eyewitness: psychological evidence. In *Psychology in the Legal Process*, ed. B. D. Sales. New York: Spectrum
 34. Gray, P. G. 1955. The memory factor in social surveys. *J. Am. Stat. Assoc.* 50:344–63
 35. Greenberg, J. 1968. *Anthropological Linguistics*. New York: Random House
 36. Greger, J. L., Entyre, G. M. 1978. Validity of 24-hour dietary recalls by adolescent females. *Am. J. Public Health* 68:70–72
 37. Grunenberg, M. M., Morris, P. E., Sykes, R. N., eds. 1978. *Practical Aspects of Memory*. London: Academic
 38. Haggard, E. A., Brekstad, A., Skard, A. G. 1960. On the reliability of the anamnestic interview. *J. Abnorm. Soc. Psychol.* 61:311–18
 39. Hefner, L. T., Mednick, S. A. 1969. Reliability of developmental histories. *Pediatric Dig.* 11:28–39
 40. Homans, G. 1967. *The Nature of Science*. New York: Harcourt Brace
 - 41a. Hyett, G. P. 1979. Validation of diary records of telephone calling behavior. See Ref. 11, pp. 136–38
 - 41b. Johnson, A. 1974. Ethnoecology and planting practices in a swidden agricultural system. *Am. Ethnol.* 1:87–101
 42. Kahle, L. R., Berman, J. J. 1979. Attitudes cause behaviors: A cross-lagged panel analysis. *J. Pers. Soc. Psychol.* 37:315–21
 43. Kahle, L. R., Klingel, D. M., Kulka, R.

- A. 1981. A longitudinal study of adolescents' attitude-behavior consistency. *Public Opin. Q.* 45:402-14
44. Kallman, W. M., Feuerstein, M. 1977. Psychophysiological procedures. In *Handbook of Behavioral Assessment*, ed. A. R. Ciminero, K. S. Calhoun, H. E. Adams, pp. 329-64. New York: Wiley
45. Keefe, S. E. 1980. Personal communities in the city: Support networks among Mexican-American and Anglo-Americans. *Urban Anthropol.* 9:51-74
46. Kemsley, W. F. F. 1979. Collecting data on economic flow variables using interviews and record keeping. See Ref. 11, pp. 115-32
47. Kemsley, W. F. F., Nicholson, J. L. 1960. Some experiments in conducting family expenditure surveys. *J. R. Stat. Soc. Ser. A* 123:307-28
48. Kendler, H. H., Kendler, T. S. 1949. A methodological analysis of the research area of inconsistent behavior. *J. Soc. Issues* 5:27-31
49. Kent, R. N., Foster, S. N. 1977. Direct observational procedures: Methodological issues in naturalistic settings. See Ref. 44, pp. 279-328
50. Killworth, P. D., Bernard, H. R. 1976. Informant accuracy in social network data. *Hum. Organ.* 35:269-86
51. Killworth, P. D., Bernard, H. R. 1979. Informant accuracy in social network data, III. A Comparison of triadic structures in behavioral and cognitive data. *Soc. Networks* 2:19-46
52. Kronenfeld, D. B., Kronenfeld, J., Kronenfeld, J. E. 1972. Toward a science of design for successful food service. *Institutions and Volume Feeding Management* 70:38-44
- 57b. Mandler, G. 1980. Recognizing: the judgment of previous occurrence. *Psychol. Rev.* 87:252-71
58. Marquis, K. H., Cannell, C. F. 1971. Effect of some experimental interviewing techniques on reporting in the Health Interview Survey. See Ref. 17, No. 41
59. Marquis, K. H., Cannell, C. F., Laurent, A. 1972. Reporting health events in household interviews: Effects of reinforcement, question length and reinterviews. See Ref. 17, No. 45
60. McCall, G. J., Simmons, J. L., eds. 1978. *Identities and Interactions: An Examination of Human Associations in Everyday Life*. New York: Free Press
61. McGraw, M., Molloy, L. B. 1941. The pediatric anamnesis: Inaccuracies in eliciting developmental data. *Child Dev.* 12:255-65
62. McGuire, W. J. 1975. The concepts of attitudes and their relations to behaviors. In *Perspectives on Attitude Assessment: Surveys and Their Alternatives*, ed. H. W. Sinaiko, L. A. Broedling, pp. 16-40. Washington DC: Manpower and Advisory Serv., Smithsonian Inst.
63. McNemar, Q. 1946. Opinion-attitude methodology. *Psychol. Bull.* 43:289-374
64. Mednick, S. A., Schaffer, J. B. P. 1963. Mothers' retrospective reports in child rearing research. *Am. J. Orthop.* 33:457-61
65. Meredith, A., Mathews, A., Zichefoose, M., Weagley, E., Wayave, M., Broom, E. G. 1951. How well do school children recall what they have eaten?" *J. Am. Diet. Assoc.* 27:749-51
66. Miller, M. 1952. The Waukegan study of voter turnout prediction. *Public Opin. Q.* 16:381-98
67. Morgenstern, R. D., Barrett, N. S. 1974. The retrospective bias in unemployment reporting by sex, race and age. *J. Am. Stat. Assoc.* 68:805-15
68. Muir, D. E. 1977. Disentangling instrument reliability, validity, and accuracy in the social and behavioral sciences. *Sociol. Symp.* No. 20, Fall:61-72
69. National Academy of Sciences, National Research Council. 1976. *Surveying Crime: Panel for the Evaluation of Crime Surveys*. Washington DC: Natl. Acad. Sci.
70. Neter, J. 1970. Measurement error in reports of consumer expenditure. *J. Market. Res.* 7:11-25
71. Neter, J., Waksberg, R. 1964. A study of response errors in expenditures data from household interviews. *J. Market.* 59:18-55
72. Parry, H. J., Crossley, H. M. 1950. Validity of responses to survey questions. *Public Opin. Q.* 14:61-80
73. Pelto, P., Pelto, G. 1978. *Anthropological Research: The Structure of Inquiry*. New York: Cambridge Univ. Press. 2nd ed.
74. Pierce, J. C., Lovrich, N. P. 1982. Survey measurement of political participation: Selective effects of recall in petition signing. *Soc. Sci. Q.* 63:164-71
75. Poggie, J. J. Jr. 1972. Toward quality control in key informant data. *Hum. Organ.* 31:23-30
76. Polgar, S. K., Hammer, M. 1966. A group's picture of itself: A comparison of observed and reported data. Presented at Ann. Meet. Am. Anthropol. Assoc., 65th, Pittsburgh
77. Pyles, M. K., Stolz, H. R., MacFarlane, J. W. 1935. The accuracy of mothers' reports on birth and developmental data. *Child Dev.* 6:165-76

78. Robbins, L. C. 1963. The accuracy of parental recall of aspects of child development and of child rearing practices. *J. Abnorm. Soc. Psychol.* 66:261-70
79. Romney, A. K., Faust, K. 1982. Predicting the structure of a communications network from recalled data. *Soc. Networks* 4:285-304
80. Romney, A. K., Weller, S. 1984. Predicting informant accuracy from patterns of recall among individuals. *Soc. Networks*. In press
81. Rosenthal, R. 1966. *Experimenter Effects in Behavioral Research*. New York: Appleton-Century-Crofts
82. Salmimies, P., Kockott, G., Pirke, K. M., Vogt, H. J., Schill, W. B. 1982. Effects of testosterone replacement on sexual behavior in hypogonadal men. *Arch. Sex. Behav.* 11:345-52
83. Sankoff, G. 1971. Quantitative analysis of sharing and variability in a cognitive model. *Ethnology* 10:389-408
84. Schatzman, L., Strauss, A. L. 1973. *Field Research: Strategies for a Natural Sociology*. Englewood Cliffs: Prentice-Hall
85. Schuman, H., Johnson, M. P. 1976. Attitudes and behavior. *Ann. Rev. Sociol.* 2:161-207
86. Seamon, J. G. 1980. *Memory and Cognition*. New York: Oxford Univ. Press
87. Shweder, R. A. 1977. Likelihood and likelihood in everyday thought: Magical thinking in judgments about personality. *Curr. Anthropol.* 18:637-48
88. Shweder, R. A., D'Andrade, R. G. 1979. Rethinking culture and personality theory, part I: A critical examination of two classical postulates. *Ethos* 7:255-78
89. Shweder, R. A., D'Andrade, R. G. 1980. The systematic distortion hypothesis. In *New Directions for Methodology of Behavioral Science*, 4:37-58
90. Smith, H. W. 1982. Improving the quality of field research training. *Hum. Relat.* 35:605-19
91. Stone, A., Neale, J. M. 1980. *Development of a Methodology for Assessing Daily Experiences*. Tech. Rep. 80-1, Contract No. N00014-77-C-0693. Arlington, Va: Off. Naval Res.
92. Sudman, S., Bradburn, N. M. 1973. Effects of time and memory factors on response in surveys. *J. Am. Stat. Assoc.* 68:805-15
93. Sudman, S., Bradburn, N. M. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine
94. Sudman, S., Bradburn, N. M. 1982. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass
- 95a. Sudman, S., Ferber, R. 1979. *Consumer Panels*. Chicago: Am. Market. Assoc.
- 95b. Sudman, S., Lannom, L. B. 1980. *Health Care Surveys Using Diaries*. Washington DC: U.S. Dep. Health Human Serv. Natl. Cent. Health Serv. Res.
96. Traugott, M. W., Katosh, J. P. 1979. Response validity in surveys of voting behavior. *Public Opin. Q.* 43:359-77
97. Tversky, A., Kahneman, D. 1974. Judgment under uncertainty. *Science* 185:1124-31
- 98a. Tversky, A., Kahneman, D. 1981. The framing of decisions in the psychology of choice. *Science* 211:453-58
- 98b. Verbrugge, L. M. 1980. Health diaries. *Med. Care* 18:73-95
99. Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally
100. Weigel, R. H., Newman, L. S. 1976. Increasing attitude-behavior correspondence by broadening the scope of the behavioral measure. *J. Pers. Soc. Psychol.* 33:793-802
101. Weisner, T. S., Gallimore, R., Tharp, R. G. 1982. Concordance between ethnographer and folk perspectives: Observed performance and self-ascription of sibling caretaking roles. *Hum. Organ* 44:237-44
102. Weller, S. 1984. Consistency and consensus among informants: Disease concepts in a rural Mexican village. *Am. Anthropol.* In press
103. Wells, G. L. 1978. Applied eyewitness testimony research: System variables and estimator variables. *J. Pers. Soc. Psychol.* 36:1546-57
104. Wenar, C. 1963. The reliability of developmental histories. *Psychosom. Med.* 25:505-9
105. Wenar, C., Coulter, J. B. A. 1962. A reliability study of developmental histories. *Child Dev.* 33:453-62
106. Wicker, A. W. 1969. Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *J. Soc. Issues* 25:41-78
107. Wilson, E. B. 1952. *An Introduction to Scientific Research*. New York: McGraw-Hill
108. Withey, S. B. 1954. Reliability of recall of income. *Public Opin. Q.* 18:197-204
109. Woocher, F. D. 1977. Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness identification. *Stanford Law Rev.* 29:969-1030
110. Yarmey, A. D. 1979. *The Psychology of Eyewitness Testimony*. New York: Free Press

111. Yarrow, M. R., Campbell, J. D., Burton, R. V. 1964. Reliability of maternal retrospection: A preliminary report. *Fam. Process* 3:207-18
112. Yarrow, M. R., Campbell, J. D., Burton, R. V. 1970. *Recollections of Childhood: A Study of the Retrospective Method*. Monogr. Soc. Res. Child Dev., Vol. 35, No. 5, ser. no. 138. Chicago: Univ. Chicago Press
- 113a. Young, C. M. 1981. *Dietary Methodology in Assessing Changing Food Consumption Patterns*. Washington DC: Natl. Acad. Press
- 113b. Young, C. M., Hagan, G. C., Tucker, R. E., Foster, W. D. 1952. A comparison of dietary study methods. Dietary history vs. seven day record vs. 24 hour recall. *J. Am. Diet. Assoc.* 28:218
114. Young, F., Young, A. 1961. Key informant reliability in rural Mexican villages. *Hum. Organ.* 20:141-48