

Brief Communications

Words, Deeds and Social Structure: A Preliminary Study of the Reliability of Informants

by LINTON C. FREEMAN AND A. KIMBALL ROMNEY

Linton Freeman is Professor in the School of Social Sciences, University of California, Irvine, CA 92717. A. Kimball Romney is Professor of Anthropology and Mathematical Social Science at the same institution. The material presented in this paper is based in part on work supported by the National Science Foundation under Grant No. SES84-8410209.

Key words: informant accuracy, observational data, method

In a recent series of articles, Bernard, Killworth and Sailer have raised a critically important issue for research in social science (Bernard and Killworth 1977; Bernard et al. 1980, 1981, 1982; Killworth and Bernard 1976, 1979). They observed particular events involving human social interaction and later questioned the individuals observed concerning that same interaction. They concluded that the data provided by the verbal reports were consistently inaccurate when matched against the criterion provided by the data based on observations. In a new paper that reviews their earlier work (Bernard et al. 1984:503) they conclude, "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."

These authors argue that this conclusion is far reaching in its implications for social scientific research. They stress the fact that social scientists often collect verbal recall data about actions or interactions and treat them as if they were accurate descriptions of those actions or interactions. Such research seems to embody the assumption that a verbal report of behavior can be taken as a proxy for having observed that behavior. To the degree that the two approaches generate markedly different results, however, a good deal of established social science research is thrown open to serious question.

Some Thoughts about Words and Deeds

Our position is that this work confronts an important issue. It is certainly true that acts and information about those acts based on recall of the actors are typically in something less than complete agreement. Moreover, it is also clear that to the degree that social scientists use recall data as if they had observed the remembered behavior, their results are open to

serious question. However, at this point we depart from the reasoning of Bernard and colleagues.

Bernard et al. (1984:503) suggest that the relatively weak correlations between these two kinds of data are distressing because they represent a poor correspondence between "a concept and the accurate measurement of that concept." This is an unfortunate phrase. They seem to be saying that observations of behavior are somehow a "concept," and that they provide an error-free criterion against which potentially inaccurate recall data must be matched. In passing, they touch on the possibility that self-reports might be treated as cognitive phenomena, but this notion they dismiss as psychological.

Leaving aside the philosophical difficulties of such notions as "true" criteria, we are uncomfortable with the authors' glib dismissal of an approach to these problems that is grounded in cognitive theory. It is our perspective that verbal recall data are by their very nature produced by perceptual and cognitive processes, and that, in principle, such data cannot be understood in any other terms. Indeed, in view of the fact that recall data necessarily are pre-processed in an individual's mind, we would be greatly surprised if we were faced with recall data that exactly matched records of observations! The problem is not one of the relation of a concept to its measurement, it is a theoretical problem having to do with the recall of interpersonal events. Recollections undoubtedly have some relation to observed "actual" events, but, as D'Andrade (1974) has pointed out, they also necessarily embody conditions and processes in the psychological space of the perceiving individual.

Tacit recognition of this fact is implied in the Bernard et al. distinction between error-free and fallible indices of interaction. Thus, these authors assumed that their verbal report data were overlaid with psychological processes, and characterized these data as fallible. Without such a psychological overlay, then, observed data could be assumed to be "true."

This perspective might be reasonable—even appropriate—in studies of legal testimony and the like where the nature of the problem requires an accurate reconstruction of actual events that occurred at some prior time. But as Hammer (1980) has indicated, it is not appropriate in scientific network studies of human social structure.

On the face of it, data that actually record human social interaction seem to be tapping into a domain that is traditionally associated with the concept of social structure. It is unlikely, however, that any particular set of recorded interactions itself captures anything that we would like to call social structure. It is more reasonable to assume that social

structure refers to a relatively prolonged and stable pattern of interpersonal relations. Indeed, if the expression "social structure" refers to anything at all, it refers to such regular recurrent patterns of interaction (Freeman 1986).

As students of social structure, then, we seek to uncover and to understand such regular long-term patterns of human interaction. Obviously, we cannot study them directly. To do so we should have to (unobtrusively) follow hordes of people from cradle to grave recording their every social act. So we compromise.

If we are well-funded, patient and as clever in research design as Bernard and his colleagues, we can record samples of observed social interactions over a relatively short run and for a relatively small set of individuals. Then, using more or less standard statistical procedures, we can make inferences about overall structural patterning on the basis of those samples. But this is still rather a time and labor intensive mode of research, and it truly does require a more than average amount of creativity in research design.

On the other hand, if we are poor, impatient or unimaginative, we can always collect quick and dirty data on the structural patterns inherent in interaction simply by interviewing the actors involved in that interaction. Then again, we must make inferences—this time from the recollections of our informants. And, in this latter case, we cannot make simple statistical inferences. *We must explicitly take our informants' cognitive processes into account.*

Thus, from our perspective this is in no way a problem of the accuracy of a fallible verbal index as it strains toward approximating a "true" observed criterion. Instead, the problem is one of estimating the parameters of an unknown long-term structure of social relations by using one or both of a pair of indices based on two different operational definitions of that unknown structure. We can collect observational data on samples of relevant interaction, or we can collect verbal recollections of interaction from the participants. Both of these measures, we expect, will err as indices of social structure, but they can be expected to err in different ways. Together, they may tell us a good deal about the underlying patterns that are our real focus of interest.

Given this two-index perspective, any particular observed set of interactions is simply a sample from the unobserved underlying pattern. As such, it is, of course, subject to all the usual perturbations associated with sampling variation. Just how well any single sample of interactions represents the underlying structure is necessarily a function of the variance in interaction frequencies from time to time. To the degree that the variance departs from zero, individual sample data will be increasingly ineffective in predicting overall structure.

The potential relationship between social structure and an index of structure based on verbal responses to a question asking: "With whom did you interact in way x in the period from time t to t*?" is much more difficult to untangle *a priori*. Requests for detailed information based on the recall of particular events seems to mobilize a complex cognitive process. A particular event is apparently not stored in memory in the form of a direct replica of that event. So it turns out that any attempt to recall the factual details of some particular event immediately produces two kinds of errors. First, some facts are lost. But second, those facts that are remembered are apparently supplemented with constructed pseudo-facts. As

Loftus and Loftus (1976:117) put it, "To make up for the gaps in memory, the witness may unconsciously add bits and pieces to his recollection, even if those bits and pieces are not truly remembered."

Such pseudo-facts, however, are not constructed at random. Rather, as Clifford and Bull (1978:42) indicate, "Eye-witnesses report known population norms when they are unsure but being pressed to say something." Thus, if a person is asked to recall who attended a party last Saturday night, he or she will include not only names of those who are actually remembered, but names of those who were absent but who "always" go to parties of that sort. What will be remembered, then, is a mix between actual events and statistical norms based on expectations about the kinds of situations in which the remembered events took place.

All this suggests that, as Bernard et al. have stressed, people's recall of particular events cannot be trusted. Recall is biased, but it is probably biased in the direction of making it more like other "similar" interactions in the past—more like the norm. Thus, it is reasonable to expect that the recall of a particular period of interaction will be "laundered" to make it resemble a typical period—the norm—for interaction of the kind in question.

When stored in memory, then, particular events seem to be bound into "bundles" along with other "similar" events. Humans, it seems, are not simple memory devices; they are organizers and processors of information. And their mode of organization seeks to establish the normal, usual or typical patterns of events at the expense of total recall of individual occurrences.

To some extent, therefore, the verbal recall data should, at least in part, be tapping into the subject's recollections of the same kinds of regular patterns of interaction that we have characterized as social structure. Subjects will be systematically biased, but they will be biased in the direction of providing accurate descriptions of the underlying structure that is our true focus of interest.

Recent research by Romney and Weller (1984) bears on this problem. They have re-analyzed some data from Bernard et al. where subjects were asked to recall the others with whom they interacted during a defined time period. The responses of some subjects were quite similar—they produced essentially the same list of others with whom they remembered interacting. Moreover, the others who turned up on this consensual list were those persons with the highest observed overall interaction rates during the period of observation. These respondents, it seems, were like the others in being unable to give "accurate" lists of those with whom they interacted, but they did produce lists that captured the general overall activity patterns of others. To the degree that overall activity rates are stable through time, it could be argued that these subjects were displaying a bias toward recalling the norm in the long-range structural patterns of interaction in this group.

The general problem, then, is to discover the ways in which observations and verbal responses can be used to recover general patterns of social structure. It is reasonable to assume that extensive long-term observations can reveal a great deal about structure, but it is not obvious that brief short-run glimpses of interaction can be used to estimate such structure. This is a problem for study.

It is also not clear that verbal responses to questions about interaction can necessarily provide a useful estimate of social structure. We hypothesize, however, that as a consequence of the kinds of cognitive processes involved in recall, such verbal data may be biased in such a way as to make them particularly effective in estimating social structure.

A Preliminary Study of Words and Deeds

Even if we confine our observations to records of a single kind of interpersonal relation, the kind of long-range observational data discussed above are almost impossible to gather. One simply cannot follow people around day and night over a period of months in order to record all instances of even one kind of social interaction. We can evade this problem, at least in part, by limiting our observations to some *situation specific* form of interpersonal behavior.

We have elected to observe interactions in a situation that involves people who naturally come together regularly in a specific context. For this first look, we have collected data on a group of faculty and graduate students that participate weekly in a university colloquium series. In this context, interaction is defined simply as coattendance at the conference. Such a definition is not entirely inappropriate inasmuch as the typical colloquium consists of a speaker and 10 to 20 listeners who are seated more or less in a circle around the room. General group discussion typically takes place, and it is reasonable to suppose that everyone present is more or less aware of his or her fellow participants.

Here, we shall focus in particular on one colloquium session that was attended by 16 participants in the spring of 1983. An investigator was present and constructed a seating map that recorded who sat where and the exact time each participant entered and left the room. This map provides the EVENT data—an explicit record of actual attendance at one colloquium.

To provide information on regularities of interaction—NORM data—long-term attendance records are needed. Persons who attend, then, can be assigned values proportional to their probabilities of attending any particular meeting. Such values can be taken as an index of the social structure of that set of people in that specific context. Unfortunately, in the present preliminary study, such long range attendance data were impossible to collect. Here we had to settle for attendance records for the two meetings that took place immediately prior to the one that was the main focus of this study.

Given this kind of limitation, we could construct only a crude estimate of the attendance norm for the event series as a whole. We decided simply to classify persons as members of the NORM set if they had been present at both previous meetings and as non-members otherwise. The names of 14 persons were thus included in the NORM set. A week after the end of the mapped meeting, each of the 16 persons who attended was seated at a micro-computer that was programmed to conduct an interview. The computer described the colloquium session, named the speaker and his topic as well as the date, and asked the subject to recall all the people who were present. Subjects were encouraged to take the task seriously and to remember everyone they could. Further, they

were told that if they remembered an attendee but could not recall his or her name, a detailed description would do. The computer then recorded each name or description listed along with the exact time it was entered. This was done to permit the study of both the order of recall and the phenomenon of differential delay in remembering different names.

Results

If we examine the data from the "accuracy" perspective of Bernard and colleagues, our results are consistent with theirs. Subjects remembered from 6 to 16 names. They made errors of both types. In all, there were 90 errors of omission and 19 errors of false recall. In the extreme cases, one individual who was not present was named as present by 12 of the 16 respondents and another who was present was named by only three.

From our own perspective, involving the notion of a structural norm, these results begin to take shape. Among the 90 errors of omission, only 33 were members of the NORM set, while 57 were names of persons who attended but who were not in the NORM set. Moreover, all 19 cases of false recall were members of the NORM set—people who had attended the previous two sessions. The two extreme cases illustrate this point. According to a consensus established in a discussion of the results among the people involved, and according to the reports of the persons themselves, the over-named absentee was a person who had never missed a session before, while the person who was present but got only three nominations had never attended before.

Thus, while these results seem to show the usual inaccuracy from the Bernard, Killworth and Sailer perspective, from our viewpoint they seem to be exhibiting a bias toward the norm. As a matter of fact, this bias may make the verbal responses better predictors of the norm than the observations on attendance.

We can begin to examine this question by examining the multidimensional scaling (MDS) plot shown in Figure 1. This is a plot of the respondents' similarities in naming those who were present at the colloquium session studied. There are 16 respondents, plus two "dummy" respondents: the EVENT—the list of those observed to have been present, and the NORM—the list of those who were present at the two earlier sessions. The EVENT is shown by a cross-hatched point and the NORM by a solid black point.

The thing to note about the plot in Figure 1 is the fact that the human verbal responses are centered, not around the EVENT, but rather around the NORM. Since the MDS representation captures all of the interpoint distances, the center of the plot is a kind of "balance point" around which all of the observations are arrayed. In this case, since the NORM falls almost exactly at the intersection of the axes in the plot, the NORM is some sort of average of everyone's perceptions. In contrast, the EVENT is located in the upper left quadrant and is not a center in the same sense.

This same tendency can be observed by calculating some conditional probabilities. The probability that a person who was present at the EVENT is in the NORM set is .62, while the probability that a person who was named as present by a respondent is in the NORM set is .80. This suggests that,

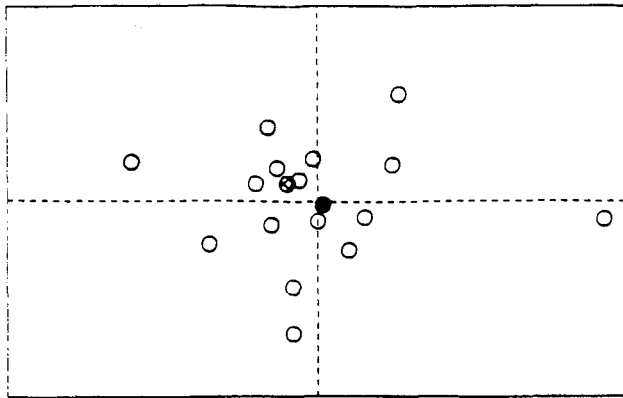


FIGURE 1. MDS OF AGREEMENT IN NAMING ATTENDEES

in their attempts to recall the event, the respondents were biasing their recall toward the usual—the normal—attendance patterns.

We can learn more about the recall process from the MDS plot in Figure 2. Here the points are the persons named as present, and the input data were their average temporal proximities on the respondents' lists, transformed to z-scores. Thus, this plot may be viewed as a kind of collective map—a picture of how the respondents dredged up the names of those they believed to have been present. In this case, students are shown as circles and faculty as filled circles. Those in the NORM set are marked with a carat (^), and those who were not present—those not in the EVENT set—have an asterisk (*).

Figure 2 shows several rather clear clusters. There is a faculty cluster and a student cluster. Moreover, the unexpected attendees (the non-NORM set) are clustered together, as are those who were expected but absent. There are, however, two anomalies—both faculty. One was the faculty member who received 12 nominations. He was expected to be present, but wasn't there. And on the MDS plot, he is placed at some distance from his faculty colleagues and grouped with the students who were both in the NORM set and present. The fact is his name was given later than other faculty on the respondents' lists. The other anomalous faculty member was the one who was not expected but who showed up. He is grouped along with the students who were expected to attend, but were absent. Again, he is way out of place in terms of the overall pattern.

All this seems to suggest that, in ordering their lists, respondents do respond to actual attendance—to the EVENT. But they seem also to respond to their expectations of attendance—to the NORM. And, at least for this group, there also seems to be a response to status, in this case to faculty versus student identification. Response to the faculty seems somehow to be more volatile; it results in striking relocations of the violators of expectations. But, in any case, the overall pattern revealed in Figure 2 suggests that the cognitive processes involved in the subjects' generation of lists involves the interaction of three factors: attendance, the norms of attendance, and status.

To check this last notion, we ran a three-way analysis of variance where the dependent variable was the number of

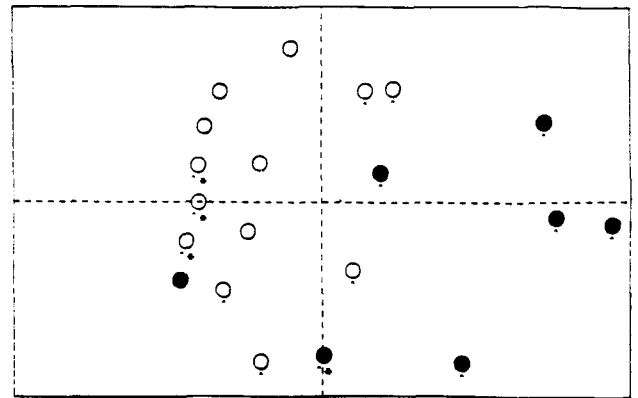


FIGURE 2. MDS OF PROXIMITY IN BEING NAMED

times a person was named as having been present and the three predictors were the dichotomies: NORM vs. non-NORM, EVENT vs. non-EVENT and faculty vs. non-faculty (F was huge, and the associated R was .92). Indeed, it seems that the naming process is intimately bound up with these three factors.

Summary and Discussion

Every face-to-face group develops an enduring pattern of interpersonal relations that may be referred to as the "social structure" of that particular group. This paper has examined the relative contributions of observed data and recalled data for understanding the social structure of the group. We presented the results of a simple experiment to illustrate some of the various factors that need to be considered in evaluating verbal recall and observed data as indicators of social structure.

Like previous investigators we found that there is a rather large discrepancy between observed and recalled data. The major new finding in the present research is our observation that the errors in the recall data are biased heavily in the direction of the social structure. Contrary to the conclusions of others who argue that observed data is superior to recall data, we believe that the bias in the recall data can be turned to advantage, and, in fact, make such recall data superior to observations for the study of a normative social structure.

We think that it is also important to note that both observed data and recall data are subject to sampling variability. In our example the sampling variability of the "NORM" would be rather large since it is based on observing only two previous colloquia. In future studies a longer run of observations and a more detailed consideration of sampling variability should be included.

One topic that we did not consider that should be considered in the future is that of individual variability in ability to report on the norms of the group. Individuals, of course, will differ both in their ability to recall particular events and their ability to express the norms. Many of these differences are doubtless associated with personal traits of the perceiving individuals. Clifford and Bull (1978:Chapter 7) present a review of trait-based differences and stress such factors as cognitive style, need for approval, and the like. Another topic

not addressed in this paper is how to maximize the fit between the verbal report and the observed behavior. If such a task had been our main interest we would have explored in greater depth the approach to accuracy suggested in Romney and Weller (1984). In a reanalysis of the Bernard, Killworth and Sailer data they found that reliability, defined as the correspondence between an individual's recall responses and aggregated group recall responses, was a better predictor of accuracy than was interaction.

In the context of this paper we are more interested in an individual centered variable that ties social structure and individual cognition together. We can draw on the work of social psychologists (Bavelas 1948; Newcomb 1961, 1968) to grapple with the problem of the effect of an individual's social position on his or her perceptions of and reports about social interaction. At the very least, a person's position in a structural pattern of interaction should influence that person's access to information. Romney and Faust (1982), for example, show that persons in more "central" positions of interaction are better informed about overall interaction patterns. In the simplest case, those persons who are frequently involved in interaction can be expected to have more information on recurrent patterns of interaction than those who are rarely involved. In future studies we hope to use this kind of information to "weight" the answers of some subjects more than others in making inferences about social structure.

REFERENCES CITED

- Bavelas, Alex
1948 A Mathematical Model for Group Structures. *Human Organization* 7:16-30.
- Bernard, H. Russell, and Peter D. Killworth
1977 Informant Accuracy in Social Network Data, II. *Human Communication Research* 4:3-18.
- Bernard, H. Russell, Peter D. Killworth, and Lee Sailer
1980 Informant Accuracy in Social Network Data, IV. A Comparison of Clique-level Structure in Behavioral and Cognitive Data. *Social Networks* 2:191-218.
- 1981 A Review of Informant Accuracy in Social Network Data. *In Modelle für Ausbreitungsprozesse in Sozialen Strukturen*. H. J. Hummell and W. Sodeur, eds. Duisberg: Sozialwissenschaftlichen Kooperative.
- 1982 Informant Accuracy in Social Network Data, V. An Experimental Attempt to Predict Actual Communication from Recall Data. *Social Science Research* 11:30-66.
- Bernard, H. Russell, Peter D. Killworth, David Kronenfeld, and Lee Sailer
1984 The Problem of Informant Accuracy: The Validity of Retrospective Data. *Annual Review of Anthropology* 13:495-517.
- Clifford, Brian R., and Ray Bull
1978 *The Psychology of Person Identification*. London: Routledge and Kegan Paul.
- D'Andrade, Roy
1974 Memory and the Assessment of Behavior. *In Measurement in the Social Sciences*. H. M. Blalock, Jr., ed. Pp. 159-186. Chicago: Aldine.
- Freeman, Linton C.
1986 Social Networks and Social Structure. *In Research Methods in Social Network Analysis*. L. C. Freeman, D. R. White and A. K. Romney, eds. Fairfax, VA: George Mason University Press.

- Hammer, Muriel
1980 Reply to Killworth and Bernard. *Connections* 3:14-15.
- Killworth, Peter D., and H. Russell Bernard
1976 Informant Accuracy in Social Network Data. *Human Organization* 35:269-286.
- 1979 Informant Accuracy in Social Network Data, III. A Comparison of Triadic Structures in Behavioral and Cognitive Data. *Social Networks* 2:19-46.
- Loftus, Geoffrey R., and Elizabeth F. Loftus
1976 *Human Memory: The Processing of Information*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Newcomb, Theodore
1961 *The Acquaintance Process*. New York: Holt, Rinehart and Winston.
- 1968 Interpersonal Balance. *In Theories of Cognitive Consistency: A Sourcebook*. R. P. Abelson et al., eds. Pp. 28-51. Chicago: Rand McNally.
- Romney, A. Kimball, and Katherine Faust
1982 Predicting the Structure of a Communications Network from Recalled Data. *Social Networks* 4:285-304.
- Romney, A. Kimball, and Susan Weller
1984 Predicting Informant Accuracy from Patterns of Recall among Individuals. *Social Networks* 6:59-78.