

Concordance Between Ethnographer and Folk Perspectives: Observed Performance and Self-Ascription of Sibling Caretaking Roles

THOMAS S. WEISNER, RONALD GALLIMORE, and ROLAND G. THARP

This paper compares field observers' perceptions of role performance with culture members' reports of what their roles are. Children's and field observers' judgments concerning child caretaking in Honolulu, Hawaii, were compared. The results show (1) statistically significant concordance between the two sets of reports; (2) fairly low (50%) agreement on who cares for the child; and (3) fairly high (80%) agreement on children's reports of being a caretaker. The concordance between children and caretakers is influenced by children's age, sex, and the social setting. Situational factors associated with sibling care (mother absence, distance from home, and numbers of children present) increased agreement on the caretaker role, but not on being cared for by others. Girls tend to overreport that they are caretakers and charges, and boys tend to underreport. There is a pattern of partial and systematically variable agreement with field observers' judgments. Concordance refers to folk and observer agreement on clusters of behaviors occurring in a social context; this differs from the related issues of reliability, veracity and accuracy of report, or validity.

ESTABLISHING THE DEGREE OF CONCORDANCE between folk and observer views is central for all kinds of ethnographic research. The issue of concordance and its implications for the cross-cultural comparability of field data has been debated primarily by those interested in comparative research using standardized measures of behaviors or beliefs. However, some form of concordance, negotiated between field observer and culture member, is central to every kind of field study. This problem is not limited to those with explicitly comparative goals or those using formal measures of cognition or frequency counts of behaviors. No field worker can assume that his or her views are identical to the cultural members' under study.

All fieldwork implicitly assumes that there has been or will be sufficient concordance between field worker and folk views that an account can be written. But what criteria are used to make the judgment that "sufficient" concordance has been achieved? Ethnologists usually have available only indirect evidence to support their assertions to have captured the folk

or cultural or emic view; and in many cases the criteria for making such assertions are not public and explicit. We believe that the monitoring of folk and observer agreements and disagreements would improve our understanding of this issue. Such a view is consistent with the idea that anthropology is a comparative science, public in its methods and in its ultimate intentions for the uses of field data.

One important domain for comparing observer and folk views concerns judgments about role performance. When are culture members acting out certain sets of norms and role expectations? To what degree are they aware of these roles and do they share the observers' views? It is this issue—comparing field observer perception of role performance with what culture members report their role to be—that is the focus of the study reported here.

Concordance judgments in fieldwork, as we define these, refer to context-specific agreements over social role behaviors. Concordance is not the same as reliability of recall, nor is it the same as accuracy of report, as these terms are usually used. Concordance is dependent on situational context and does not depend on agreement concerning reports of specific behaviors. We return to the role of concordance compared to other ways to assess reliability, validity, or veridicality in field research in the discussion.

This paper presents data collected as part of a larger study of sibling caretaking in Honolulu. We systematically 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. Results indicate statistically significant concordance between the two sets of reports; but also patterned variability in the degree of agreement. The extent of agreement was found to associate with situational variables (distance from home, maternal presence, and sex of child) and whether the child is being asked about the *caretaker* role (with substantial agreement) or the *charge* role (with lower agreement). (Throughout

Thomas S. Weisner is Associate Professor of Anthropology, Departments of Psychiatry and Anthropology, University of California, Los Angeles. Ronald Gallimore is Professor of Psychology, Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles. Roland G. Tharp is Professor of Psychology, Department of Psychology, University of Hawaii, Manoa, and Principal Investigator, Kamehameha Research Institute, Honolulu. This research was supported by the Kamehameha Early Education Project of the Kamehameha Schools and the Princess Bernice Pauahi Bishop Foundation, Honolulu. Additional support was provided by the Sociobehavioral Research Group, Mental Retardation Research Center, UCLA, and the Department of Psychiatry, UCLA. Computing was done by the Computer Resources Group, MRRC, UCLA. Our special appreciation goes to Linda Kobayashi and Sharon Omori.

the paper, a "caretaker" of another child is in the superordinate, responsible role; a "charge" is a child in the subordinate role, being cared for by someone else.)

We suggest that both results are essential for understanding the process of negotiating agreements between folk and observer views. Statistically significant overall agreement permits appropriate generalization. But understanding the patterning of agreements and constructing a plausible account of what appears to produce disagreements is equally important. Understanding disagreements is especially important in inferring contextual factors affecting judgments.

Data and Methods

Two trained field observers visited eight Hawaiian-American children between the ages of five and nine, one boy and one girl randomly chosen from each grade level from kindergarten through grade 3, attending the Kamehameha Early Education Project research and demonstration school in Honolulu (Tharp and Gallimore 1979). The children and their families live in a low-income housing area near the school and are representative of many Hawaiian-American families in this community. Seven children were visited between 3:00 and 4:30 P.M. on 20 separate occasions, and one, 16 times. These visits were randomized and counterbalanced by observer, household, and time of the visit (early or later in the afternoon). Field observational techniques included a spot observation at the beginning of the visit (Rogoff 1978) assessing people present, activities, proximity to the home and to the mother, and affect. Field notes and ratings were made of language use, instruction or information exchange involving the child, and patterns of caretaking.

We utilized a repeated measures design for doing the field visits. Thus we have a relatively large N of 156 observations, but only eight children. The results of our analyses are all appropriately generalized to home settings in the afternoons—but not to Hawaiian boys and girls, or children of age five to nine. Pooling of field observations is appropriate for the questions raised in this study, since we are asking about concordance between child and observer perceptions of naturally occurring caretaking *situations*. These situations vary widely across the home visits. (Revisiting the same children can also produce order effects, and these are discussed below.)

The criteria for observers' judgments of child caretaking were drawn from previous ethnographic and interview studies on sibling caretaking in Hawaii. Fieldwork in a peri-urban community outside of Honolulu (Gallimore, Boggs, and Jordan 1974) established the importance of sibling care, and its basic characteristics. Further comparative work examined cross-cultural data on sibling care (Weisner and Gallimore 1977). A sibling care interview with mothers of children in the Kamehameha Early Education Project research and demonstration school explored folk views of various dimensions of such care (how structured; tasks associated with it; age and sex; appropriate timing, etc.) (Weisner, et al. n.d.; Weisner, Gallimore, and Tharp 1977). These earlier studies provided a blend of folk and observer criteria for defining sibling caretaking.

The two field observers in the present study were residents of Honolulu, familiar with Hawaiian families and child-rearing patterns, and worked during pilot field visits and

reliability training to establish consistent rules and definitions. The field observers were thus the beneficiaries of prior negotiated understandings between both cultural "insiders" and research "outsiders" of what constituted child caretaking in this community.

Observers' judgments of caretaking were based on carefully prepared criteria: clear indicators of cultural expectations of caretaking; direct observation of assistance; clear signs of responsibility for another child (e.g., warning a child concerning danger); helping a child do a chore or task; and/or requests for compliance with an indication that another child complied or recognized the appropriateness of the request. After the 20-30-minute observation ended, the observer asked the target child if he or she was taking care of another child, and if so, whom; or if another child was taking care of him or her, and if so, who that child was. The questioning of children was done informally and casually, in as natural a way as possible. The observers were well known to all the children from their school experience and their many visits. Nonetheless, to ask young Hawaiian-American children about child caretaking arrangements is indeed an unfamiliar thing to do. Not wanting to probe or ask extensive questions about child caretaking that might influence subsequent visits, we deliberately kept the discussions brief. Since we did 20 repeated visits with each child, the question and interview situation obviously became more familiar over time. This may have made children more comfortable answering, but it might also have altered their responses in some ways—such as increasing the number of "yes, I am caretaking" responses or producing more socially expected responses over time (girls saying "yes" more; boys saying "no" more; etc). To test for these kinds of effects, we did a runs test (Siegel 1956:52-58) across the 20 observation sessions, testing for change in the pattern of children's responses to the observers' questions as a function of the sequence of observations. There were no significant patterns.¹ Overall, although there are inevitable ambiguities in these short field interviews, we are confident that they were asked appropriately and understood by the children.

Reliability between the two observers was assessed using other families not in the sample of eight. Recalibration and reliability visits were done after every ten field visits. Overall agreement of parallel field visits was 76.3%. A series of post hoc MANOVA analyses were also run to test for possible observer, time of visit, or obtrusiveness effects on the various outcome measures (see also Sykes 1978). In general, outcome measures were unrelated to these possible sources of bias (Weisner et al. n.d.). We are confident that these observations were reliable across observers, time of day, and levels of obtrusiveness, and that the definitions and coding procedures used were consistent. How did they correspond to what the children themselves reported was going on?

Results: Children's and Observers' Perceptions of Caretaking Arrangements

Tables 1 and 2 show the direct comparisons between observers' and children's reports on child care. Both tables show a significant proportion of agreement between the children and the field observers, using the kappa statistic (Fleiss 1973:143-47).² Kappa for Table 1 is .194 ($p < .0001$)

TABLE 1. OBSERVER'S AND CHILD'S JUDGMENTS OF WHO IS TAKING CARE OF CHILD

Observer's judgments on who is taking care of child	Child's report ^a										
	No One			Mother or Father			Sibling			Total	Col. %
	n	Row %	Col. %	n	Row %	Col. %	n	Row %	Col. %		
No one	28	60.8	52.8	4	8.7	30.8	14	30.4	45.2	46	47.4
Mother or father	17	60.7	32.1	7	25.0	53.8	4	14.3	12.9	28	28.9
Siblings	8	34.8	15.1	2	8.7	15.4	13	56.5	41.9	23	23.7
Total	53			13			31			97	100.0
Row %	54.6			13.4			32.0				

kappa = .1935 (p < .0001)

^a In 57 observations the children were not asked because the mother or other adult was present in the same room with the child and observer (see text). (Two visits are missing a judgment.) In 44 of these 57 cases (77.2%), the observer judged mother was in charge, in 12 cases (21.1%), the siblings, and in one case (1.8%), no one. The mother or father data in this table are instances where the child was asked, and said that his parent(s) were caring for him.

and for Table 2, .499 (p < .0001). Field observers' judgments corresponded to the reports of the children themselves concerning the structure of caretaking. This result is encouraging for validating our coding categories and for inferring individual-level outcomes from social-setting-level judgments.

But this statistical agreement is only a first step in exploring why and when children and observers agreed or disagreed and in using our systematic contrast of folk and observer reports to help understand the variability in perceptions of caretaking. For instance, Table 1 also shows considerable disagreement between the child's verbal report on who is taking care of him or her (if anyone) and the observer's judgment. Observer and child could agree only 28 of 46 times (61% of observers' judgments) that no one was acting as caretaker for the child; only 7 of 28 times (25%) that the mother or father was, and 13 of 23 times (57%) that a sibling was responsible.

However, Table 2 shows a different result when the child was asked about being in the caretaker role: in 14 of 18 observations (78%) in which observers said that children were caretakers, the children also reported that they were caring for other children. Children's reports that they are in a caretaker role agree with observers' judgments more often than children's judgments of who is caring for them.

In addition to the agreement on caretaking, the "base rates" for attributing child-care responsibility also varied somewhat between children and observer, as reflected by the row and column total percentages in Tables 1 and 2. Thus, children were more likely to say that a sibling was caring for them (32%) than were the observers (23.7%). At the same time, children were less likely to report that their parents were caring for them (13.4%) than were the observers (28.9%; see Table 1). Furthermore, children generally said that they were acting as caretakers of other children more often than did the observers (28% versus 18%; see Table 2) and rarely said that they were not caring for other children when observers judged that they were.

One way to phrase the base rate difference is that children "overreport" that they are caretakers. The children were perhaps reporting on a combination of objective situational indicators, as well as a subjective, felt role responsibility for another child. But the observers could base their judgments only on role performance, that is, on a display of caretaking behaviors of some kind. The observers' criteria were more focused and limited than the children's were, since the child's felt sense of being in a responsible role was not always shown in observable role behaviors, and hence was not always accessible to the observer.

TABLE 2. OBSERVER'S AND CHILD'S JUDGMENTS OF WHETHER CHILD IS CARING FOR OTHER CHILDREN

Observer's judgment on whether child is caring for other children	Child's report ^a							
	No			Yes			Total	
	n	Row %	Col. %	n	Row %	Col. %	n	Col. %
No	68	82.9	94.4	14	17.1	50.0	82	82.0
Yes	4	22.2	5.6	14	77.8	50.0	18	18.0
Total	72			28			100	100.0
Row %	72.0			28.0				

kappa = .4989 (p < .0001)

^a In 55 cases the children were not asked because the mother or other adult was present. (One visit is missing a judgment.) No observer's judgment was made in these situations. This table has 3 additional cases because of changes in the field situation allowing observers to ask the "whether or not" question but not the "who" question in table 1.

Results: Situational Influences on Sibling Caretaking—Children Present, Maternal Absence, and Distance from Home

In another paper (Weisner et al. n.d.) and in earlier work (Weisner and Gallimore 1977), a variety of antecedent conditions were found to be associated with the likelihood of sibling care. Social-structural conditions such as work pressures on parents, the structure of the daily routine, and kinship and residence patterns interact with demographic circumstances (such as numbers of children available in the household and family size) to determine the likelihood and incidence on non-parental, child-child caretaking.

In the present case, immediate situational characteristics that varied included the number of children present in and around the home, mother's presence in the home, and the distance children were from their houses during the afternoon. Our results indicate that sibling care was more likely to occur when more children were present around the target child in the settings, the child was further from the home, and the daily schedule found the mother away from the home. These situational conditions clearly influenced both children's and observers' reports in the expected direction: over 80% of observers' and children's judgments of sibling caretaking agreed when high-likelihood sibling-care situations occurred.

However, this finding is stronger when children reported being caretakers than when they reported being charges. In circumstances facilitating sibling care and in settings where there was a minimum of situational ambiguity, children still tended to report that they were being cared for by other children when observers said that no one was caring for them. Reduction in situational ambiguity thus reduces observer-child differences when the child is in the caretaker role but not when the child is in the role of charge.

Results: Effects of Children's Sex and Age on Agreement

Situational factors interacted with the children's age and sex in influencing both the likelihood of being involved in sibling care, and in observer-child agreement in reports. Older children are more likely to be in situations where sibling care occurs (away from home, mother absent, several children present), and older girls are more likely to be in such situations than older boys. Thus, observers and children alike reported more involvement in sibling care as charge and caretaker by older girls than for older boys. For younger children, there is no strong difference by sex. Two age and sex effects are associated with this pattern of differential likelihood of being involved in sibling-care situations (and reported sibling care): (1) boys "underreport" that they are caretakers or charges compared to the observer, while girls "overreport" being charges or caretakers; and (2) girls agree more closely with the observer's reports of sibling care than do boys. Thus, girls agreed with the observer's judgment when the observer said that sibling care was occurring and disagreed when the observer said it was not occurring. Boys agreed with the observer more often when the observer reported that sibling care was *not* occurring and disagreed more when the observer said yes.

The sex difference in sibling-care incidence and in children's agreement suggests that children are using some version of a socially appropriate sex-role model of child-care responsibility in their report (Gallimore et al. 1974). Hawaiian girls are expected to carry out the child-care role, and to be more sensitive to nurturance and familial responsibility than boys. Girls may well have been using this pattern of social expectation in answering observers' questions. Even if the more objective-behavioral criteria needed by observers to infer the child-care role did not occur, girls may well have used a broader, generalized expectation of implied caretaking responsibility. Boys may well have been operating under the converse expectation—independence from "domestic" responsibility and self-reliance in peer situations.

Discussion: What Do Our Concordance Data Imply

Are the children accurate reporters? In one instance, they are not: when asked "who is taking care of you?" they 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%. So there is no simple answer to the question of whether the children could be depended on to be accurate³ reporters; it is a function of the question and the context. Our data showed pockets of concordance, which indicated that the children were offering an interpretation of their behavior in context. Settings influencing judgments were typically those where distance from home or the personnel present made caretaking roles ambiguous and where cultural norms influenced reports (e.g., girls responding positively to the caretaker query). Our conclusion from the statistical results, as well as from our qualitative knowledge of the field observations, is this: it is appropriate to ask children if they are caring for other children regardless of age, sex, or setting. However, considerable caution is required if children are asked who is taking care of them, since age, sex, and setting all influence levels of agreement. We also do not think it likely that there are general "error bounds" on these questions that would provide general levels of accuracy, since the social-situational factors that it would be necessary to correct for would change.

In any event, we believe that our data speak to more than the question of accuracy, and that is why we have used the term concordance rather than accuracy. In our usage, concordance refers to the degree of agreement between an observer and a research participant concerning the participant's social role attributions. Concordance judgments refer to context-specific folk and observer reports of social performances and participation. These are not judgments of discrete behaviors but rather of clusters of behaviors which occur in a meaningful setting. Concordance does not assume that observers and participants necessarily are basing their judgments on the same information, as is the case in accuracy of report assessment. Indeed, we take for granted that a participant may base his or her report on felt role-assumption or other qualities that are not manifest and not available to the observer, who must rely on observable behavior. For example, the participant may base a role attribution on felt responsibility without reference to behavior, while the observer may

attribute caretaking roles on the basis of situation-specific behavior. This possibility of a differential data base is one condition that distinguishes concordance from accuracy of report. Concordance judgments include the participants' and the observers' impressions or interpretations of a social setting and the subjective experience of the event. Concordance agreements may be low if the expectation of discrepant perceptions is high. Assessments of accuracy, on the other hand, ask whether discrete events did or did not happen; disagreement implies inaccuracy by one or the other observer or raises doubts that the event happened. Disagreement between participants and observers in concordance analyses may be high and yet produce *understanding* of the differing perceptions through analyses of the patterns of agreement and disagreement. The primary issue is not over whether the events did or did not happen; the analysis is over *how* two observers interpreted what did happen.

Killworth and Bernard (1976) and Bernard, Killworth, and Salier (1981) provide an excellent example of a contrasting study of accuracy. Bernard et al. questioned users of a computer conferencing network with whom they had communicated over varying intervals of elapsed time between the question and the actual act of communication. Thus they asked, "Did you conference with anyone within the last 20 minutes; Whom? Three weeks ago?" Accuracy was defined as agreement between observer and informant on the occurrence of a given act of communication. Accuracy is thus the extent to which the participant's recall for behavior *matches an objective recording of discrete behavior*. Their study provides convincing evidence the network users were highly inaccurate, even with respect to whom they had conferenced with as recently as 20 minutes before the question. They properly conclude that people are not accurate reporters.

Unlike the Bernard study, we did not ask the Hawaiian children in our study to report on discrete behaviors taken out of the context of social role performance. Thus we did not ask them "Did you help your younger brother tie his shoe in the last ten minutes?" We did not ask the children the specific number of times that they spoke with others in the setting, nor with whom they spoke. If we had done so, we probably would not have found the children to be very accurate reporters of their own behavior. And they would have been hopelessly inaccurate if we had asked them about such details of a week or a month before. But it is precisely these kinds of data that are gathered in accuracy-of-report studies.

Why would our Hawaiian children/informants have been inaccurate? Why were Bernard and Killworth's computer scientists inaccurate reporters of their communication patterns? We do not believe that the children and the scientists are so cognitively disorganized and unobservant that they have no coherent scheme for remembering. We believe that this kind of inaccuracy occurs because people do not count their behavior; they rarely focus on it in the discrete bits that are required for objective accuracy-of-report studies. It is more consistent with contemporary theory and research to regard people as complex information processors; storing, recoding, and retrieving information in a variety of ways, not merely as rote recorders. Rote recording has limited functions in daily living; it is not adaptive to store uncoded or related bits of information—albeit information vital to the social scientists whose methods are predicated on accurate reporting.

Information sought by researchers is stored and retrieved via the same cognitive processes that an individual otherwise employs; and it is thus subject to the same influences and can be regarded as another case of cognitive operation. Ericsson and Simon (1980) argued that recent advances in cognitive research should be taken into account in our use of research methods that rely on human information processing. We should incorporate into our methods the ways in which the features of human cognitive processes influence verbal reports.

For example, what we know about short- and long-term memory processes can be used to predict effects on verbal reports, as a function of the latency between an experience and the request for a report. Also, important effects can be expected depending on whether a report is a "direct articulation or explication of the stored information" or one in which "the stored information is input to intermediate processes, such as abstraction and inference" (Ericsson and Simon 1980:223).

One class of intermediate processes influencing storage and retrieval is cultural categories, or classifications. What we term concordance analysis allows for the effects described by Ericsson and Simon (1980), including cultural biases introduced by culturally determined ways of classifying events, persons, and things. Thus concordance analysis allows us to assume that there are coherent, discoverable patterns influencing participant reports of social settings and role performances. Such an analysis examines concordances and nonconcordances to identify what might be affecting observer/participant agreement/disagreement; for example, setting effects that may be interpretable with available ethnographic information.

Such patterns are examples of how cognitive categories, built and sustained through culture, can structure perception and memory.⁵ Such biases certainly affect the Hawaiian children and perhaps account for the distressing inaccuracy of Killworth and Bernard's network communicators (1976:282). The nature and function of these biases have been examined by Shweder and D'Andrade (1980): as events are encoded in memory, lexical labels are attached—words and cultural categories are used as mnemonic devices to structure and store information and to retrieve it. But words in our language do not always go together in the same ways as do events in the world. Encoding of perceptions and experiences in natural language thus produces "systematic distortions" in how we store information and how we recall later. For example, there is evidence that personality traits are subject to this sort of systematic distortion: if one is perceived as cooperative, then instances of antagonism are misperceived, recategorized, attributed to transient factors, etc., in order that the cognitive map of the person fits our cultural bias that cooperativeness and antagonism are mutually exclusive traits.

Taken together, these considerations suggest an interesting perspective on our concordance data and accuracy-of-report studies. Our observers were trained to use Hawaiian concepts of child care to make a contextualized series of judgments about *role behavior* which were compared with global attribution of *role performance* by informant children. The disagreements and agreements seemed to reflect predictable biases, given the factors that earlier research has suggested influence caretaking role performance. However, achieving con-

cordance does not imply that child-tenders would necessarily *accurately report* an objective count of specific behaviors.

This is clear when we contrast our study with Killworth and Bernard's. They were unable to find much to encourage use of behavior recall. No pattern of accuracy was found; no factor was identified that influenced accuracy: "In short, everything we have measured seems to be related to inaccuracy in a reasonable way. The problem is that nothing seems to matter very much" (Bernard et al. 1981:24). In contrast, the pockets of concordance in our data showed patterns that depended on child and setting characteristics, for example, distance from home, personnel present, and cultural norms.

We also identified many disagreements between observers and children. But we regard disagreements in this instance to be more than mere errors of measurement.⁴ Disagreements and agreements form a pattern that reflects cultural constraints and, in this sense, can be predictable. Thus, girls are subject to normative, cultural pressures toward responsibility, and so they perceive more child caretaking; or nearness to home and mother involves a cultural view of the caretaking hierarchy that would reduce self-attribution of the caretaking role. One might well *expect* children's and observers' reports to differ, depending on whether the child was reporting his role as caretaker or charge. The caretaker role is more clearly defined, and self attributions of control and authority over others may be more likely than reports of being under the control of others (e.g., Sarbin and Allen 1968:503-06).

Killworth and Bernard did not do a concordance study; they did not discover any pattern to the way their computer scientists *did* perceive the computer network and remember with whom they talked and when. We did not do a study of accuracy of report in Hawaii; we did not systematically match discrete behaviors under the general category "caretaker" or "charge" that were elicited from children and compare these with the identical behaviors reported by observers. Each kind of study has value and extends understanding of the data; they are not mutually exclusive in our view. In naturalistic field observational studies, however, patterns of concordance between folk participants and participant observers is the primary research situation. It is in this setting that anthropologists or any social scientists can improve their understanding of and confidence in ethnographic inferences through systematically analyzing the concordance between folk and observer views.

But what of the relation between concordance of this kind and counts of discrete behaviors? Noncontextualized counts of behavior frequency will overlap with observer or participant reports to the extent that the discrete behaviors overlap with the social roles being enacted in that setting. It is also likely that certain domains (e.g., child caretaking) may show more such overlap than others (e.g., contact within a computer network). We believe that discrete behavior frequencies will be related to more global folk and observer judgments under some circumstances. But neither we nor Bernard et al. have data that speak directly to this topic.

Where is the line drawn between "discrete behavior" categories and "clusters," such as in the definition of a role performance? There are many ways to tie another child's shoe, and there are social categories for children's roles that subsume "caretaker of another child." There is no absolute line; even behaviors not consciously perceived can influence

behavior and be culturally regulated (e.g., body language). Our distinction between concordance and accuracy/discrete behavior studies depends on two dimensions: the fact that we are comparing perceptions of a felt role-performance, not a specific behavior; and the fact that the level of behavioral description is culturally appropriate and ordinary—that is, the descriptive level would not be considered unusual in normal discourse. In contrast, asking children to list specific instances of child-care acts is not culturally ordinary.

Practical Applications of Concordance Estimates in Fieldwork

We are hesitant to suggest that concordance assessment should become a standard practice in fieldwork. There are already many exhortations in ethnographic texts to do fieldwork in more systematic and careful ways, with long lists of conditions to control for or watch for in field research (Bogdan and Taylor 1975; Golde 1979; Johnson 1975; Levine et al. 1980; Lofland 1976; Longabaugh 1980; Schwartz and Jacobs 1979; Spradley 1979; Wax 1971). We are sometimes dismayed by all the items on such lists, because in practice it is not possible to control for them all or keep them all in mind when doing participant-observation and writing field notes. (Nor is it possible to control for all the known sources of possible influence in experimental and quasi-experimental research done under controlled conditions [Cronbach 1975].) Are we adding more new strictures on fieldwork, on top of an already substantial array?

We think it more useful to identify a limited number of fieldwork conditions and designs where assessing concordance between folk and observer views would be most desirable. We suggest three general conditions:

1. Does the study require inferences across levels of data, for example, between situational influences and individual cognitive states, or between a social *situation* (like the existence of a caretaking hierarchy) and a presumed *felt* sentiment or personality disposition (like nurturance or responsibility)? If field workers want to be sure that whatever situational influences or perceptions of events they think important are in fact *perceived* similarly by actors themselves, then some systematic attention to comparing folk and observer views would be indicated.

2. Is there likely to be internal cultural variability in how culture members feel about the relevant situations or in how they would label or classify them? Is it important for answering research questions to know which individuals differ and in what situations? In such cases, field observations that systematically tested for the folk views of a variety of culture members and compared these to observer judgments would be valuable. This is a more focused strategy than simply searching for intracultural variability in general. Before going to the trouble of doing a concordance study, one should probably have a specific set of hypotheses that would depend for their confirmation on finding patterned agreement and disagreement between culture members compared to field-worker judgments. For instance, our hypotheses about differences in child caretaking responsibility pressures on boys and girls are strengthened by the results of this study.

3. Some social situations have very ill defined boundaries

that are permeable and flexible. Some cultural categories are similarly fuzzy, ambiguous, or complex. Systematic attention to folk and observer concordance would be indicated in studies of such situations or cultural domains. In our own study, for instance, we need to know more about the ethnography of the situations where cultural members (siblings among themselves, or mothers and their children) disagree regarding assignment or caretaking responsibility. Our analysis points to the kinds of circumstances where a more intensive study of how caretaking roles are assigned, self-ascribed, and denied would be most fruitful.

NOTES

¹ Although this result is consistent with the argument that repeated visits and questioning of the children did not alter their reports or observer's perceptions, it in fact does not demonstrate it. Possible secular trends in caretaking frequencies over the five months of field visits could produce clusters of similar judgments (either of no or frequent child caretaking), yet no observer bias need be present. At the same time, such patterns of observers' and children's reports might be masking possible bias by working counter to its effects. In any case, we do not believe that this influence was powerful, even if present. Visits were well spaced, counterbalanced, and varied. The consistent patterning of observers' and children's reports by other conditions (settings, sex, age, etc.) also indicates that bias confounded with the sequence of visits is unlikely to be of any substantial importance.

² Fleiss (1973:143-47) points out that chi-square measures degree of association, not degree of agreement between (in this case) two judges (observer and child). The kappa statistic assesses proportion of agreement between judges, independent of overall level of association, and takes into account the expected proportion of agreement due to chance alone.

³ Accuracy of behavior recall is not equivalent to *reliability* of recall. As Bernard, Killworth, and Sailer (1981) point out, participants may be reliably reporting an inaccurate account of their communication behavior and network, in the sense that if asked repeatedly they would tend to give the same report. The problem is this: Bernard et al.'s data reveal that what people might recall reliably is almost certainly inaccurate.

⁴ This is somewhat analogous to generalizability theory, which holds that there may be many sources of variance in "test" scores, in contrast to classical psychometric theory which distinguishes only true score and error variance (Mitchell 1979).

⁵ Both the cultural-normative influence of role perception and the Shweder and D'Andrade semantic consistency model of perception and recall of events presume that certain psychological and cognition mechanisms were involved in shaping both observer and folk classifications, perceptions, memory, and recall of events. Does this mean that we must demonstrate these cognitive and other mechanisms in operation in the natural environment in order to infer that they were present and influencing the concordant views between children and observer? It is not necessary to know all that went into the making of a child's judgments in order to use those judgments to explore the implications of child's and observer's perceptions of an event. Were this criterion necessary, no ethnographic interview data would be valid, since only rarely are we able to reconstruct clearly the steps used by actors in reasoning and in generating their perception of settings. We know only the results of that process.

Indeed, requiring process equivalence as well as outcome equivalence would paralyze most research efforts in the social sciences. In cognitive research, for instance, elaborate experimental designs are necessary to infer the operation of process variables that are presumed to underlie differential learning rates (Butterfield 1978); for example, elaborate procedures are used to ensure that subjects

have engaged in covert rehearsal of material to be remembered during the course of experimental analyses of mnemonic storage and retrieval.

Inferences about decision-making processes are equally complex in ethnographic fieldwork (e.g., Quinn 1978) and in studies in ethnoscience (e.g., Tyler 1969:343-432). Finding concordance in observable behaviors and situational characteristics between participants and field observers, and inferring possible similarities in process, seems the only attainable goal at present.

REFERENCES CITED

- Bernard, H. R., P. D. Killworth, and L. Sailer
1981 Informant Accuracy in Social Network Data V: An Experimental Attempt to Predict Actual Communication From Recall Data. *Social Science Research*. (In press.)
- Bogdan, Robert, and Steven Taylor
1975 Introduction to Qualitative Research Methods. A Phenomenological Approach to the Social Sciences. New York: Wiley.
- Butterfield, Earl C.
1978 Observing Cognitive Development. *In* *Observing Behavior: Theory and Applications in Mental Retardation*, Vol. 1. Gene Sackett, ed. Pp. 137-76. Baltimore: University Park Press.
- Cronbach, L. J.
1975 Beyond the Two Disciplines of Scientific Psychology. *American Psychologist* 30:116-34.
- Ericsson, K. Anders and Herbert A. Simon
1980 Verbal Reports as Data. *Psychological Review*. 87: 215-51.
- Fleiss, Joseph L.
1973 *Statistical Methods for Rates and Proportions*. New York: Wiley.
- Gallimore, Ronald, J. Boggs, and C. Jordan
1974 *Culture, Behavior and Education. A Study of Hawaiian-Americans*. Beverly Hills: Sage.
- Golde, Peggy, ed.
1970 *Women in the Field*. Chicago: Aldine.
- Johnson, John M.
1975 *Doing Field Research*. New York: The Free Press.
- Killworth, P. D., and H. R. Bernard
1976 Informant Accuracy in Social Network Data. *Human Organization* 35:269-86.
- Levine, Harold, et al.
1980 Teaching Participant-Observation Research Methods: A Skills-Building Approach. *Anthropology and Education Quarterly* 11:38-54.
- Lofland, Lynn
1976 *Doing Social Life: The Qualitative Study of Human Interaction in Natural Settings*. New York: Wiley.
- Longabaugh, Richard
1980 The Systematic Observation of Behavior. *In* *Handbook of Cross-Cultural Psychology*, Vol. 2. Harry C. Triandis and John W. Berry, eds. Pp. 57-126. Boston: Allyn and Bacon.
- Mitchell, Sandra K.
1979 Interobserver Agreement, Reliability, and Generalizability of Data Collected in Observational Studies. *Psychological Bulletin* 86:376-390.
- Quinn, Naomi
1978 Do Mfantse Fish Sellers Estimate Probabilities in Their Heads? *American Ethnologist* 5:206-26.
- Rogoff, Barbara
1978 Spot Observation: An Introduction and Examination. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition* 2:22-26.
- Sarbin, Theodore R., and Vernon L. Allen
1968 Role Theory. *In* *The Handbook of Social Psychology*,

- Vol. 1. Gardner Lindzey and Elliot Aronson, eds. Pp. 488-567. Reading, Mass.: Addison-Wesley.
- Schwartz, Howard, and Jerry Jacobs
1979 *Qualitative Sociology. A Method to the Madness*. New York: The Free Press.
- Shweder, Richard A., and Roy G. D'Andrade
1980 The Systematic Distortion Hypothesis. *In* *New Directions for Methodology of Behavioral Science: Fallible Judgment in Behavioral Research*. Richard A. Shweder, ed. San Francisco: Jossey-Bass.
- Siegel, Sidney
1956 *Nonparametric Statistics for the Behavioral Sciences*. Pp. 52-58. New York: McGraw-Hill.
- Spradley, James
1979 *Participant Observation*. New York: Holt, Rinehart.
- Sykes, R. E.
1978 Toward a Theory of Observer Effect in Systematic Field Observation. *Human Organization* 37:148-56.
- Tharp, Roland, and Ronald Gallimore
1979 The Ecology of Program Research and Evaluation: A Model of Evaluation Succession. *In* *Evaluation Studies Review Annual*, Vol. 4. L. B. Sechrest, ed. Pp. 39-60. Beverly Hills: Sage.
- Tyler, Stephen A., ed.
1969 *Cognitive Anthropology*. New York: Holt, Rinehart.
- Wax, Rosalie, ed.
1971 *Doing Fieldwork: Warnings and Advice*. Chicago: University of Chicago Press.
- Weisner, Thomas and Ronald Gallimore
1977 My Brother's Keeper: Child and Sibling Caretaking. *Current Anthropology* 18:169-91.
- Weisner, Thomas, et al.
n.d. Cultural Homogeneity, Familial Variability, and Educational Adaptability: Sibling Caretaking among Urban Hawaiians. Unpublished ms.
- Weisner, Thomas, Ronald Gallimore, and Roland G. Tharp
1977b Sibling Caretaking in an Urban Hawaiian Setting. Paper presented at the Society for Cross-Cultural Research, East Lansing, Michigan.
- Whiting, John W. M., Irvin L. Child, and William W. Lambert
1963 *Field Guide for the Study of Socialization*. New York: Wiley.
- Whiting, John W. M., and Beatrice B. Whiting
1975 *Children of Six Cultures. A Psycho-Cultural Analysis*. Cambridge, Mass.: Harvard University Press.