Finally, the implementer of the naturalistic study must deal with several inevitable problems: managing problem/contract disjunctions, dealing with aspects of unfolding design, and managing field problems. The naturalist's lot is not an easy one. To suggest that persons engage in naturalistic inquiry because it is so much easier and less rigorous than conventional inquiry is to betray ignorance of what is actually involved.

11

Establishing Trustworthiness

Lincoln Y & Guba E (1985)
Establishing trustworthiness
In:
Naturalistic inquiry (Lincoln & Guba)
Newbury Park, CA, Sage, pp 289-327

How can I trust thee? Let me count the ways . . .
(with apologies to Elizabeth Barrett Browning)

THE ASSAILABILITY OF NATURALISTIC STUDIES

Probably no anthropologist is better known than Margaret Mead; her *Coming of Age in Samoa* is familiar to every literate American. The recent publication, therefore, of Derek Freeman's *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth* produced more than a ripple of surprised reaction. Could his charges that Mead was "astronomically wrong" about Samoa be true? Was it the case, as Freeman claims, that Mead had failed to acquire even a rudimentary acquaintance with Samoan culture before concentrating prematurely on her specialty, adolescent girls, thereby grossly mistaking the meaning of the observations she made? Was it true that Mead had come to her conclusions because she imposed, albeit unknowingly, her own ideology, emphasizing the "nurture" side of the "nature-nurture" controversy because that was what her mentor, Franz Boas, expected her to find? And finally, isn't the fact that Freeman, himself an experienced Samoan researcher, could arrive at such gross disagreements with Mead more than ample evidence of the untrustworthiness of such uncontrolled findings, irrespective of whether his judgments are right or wrong?

The naturalistic inquirer soon becomes accustomed to hearing charges that naturalistic studies are undisciplined; that he or she is guilty of "sloppy" research, engaging in "merely subjective" observations, responding indiscriminately to the "loudest bangs or brightest lights." Rigor, it is asserted, is not the hallmark of naturalism. Is the naturalist

inevitably defenseless against such charges? Worse, are they true? It is the purpose of this chapter to deny those allegations, and to provide means both for shoring up and for demonstrating the trustworthiness of inquiry guided by the naturalistic paradigm.

WHAT IS TRUSTWORTHINESS?

The basic issue in relation to trustworthiness is simple: How can an inquirer persuade his or her audiences (including self) that the findings of an inquiry are worth paying attention to, worth taking account of? What arguments can be mounted, what criteria invoked, what questions asked, that would be persuasive on this issue?

Conventionally, inquirers have found it useful to pose four questions to themselves:

(1) *"Truth value":* How can one establish confidence in the "truth" of the findings of a particular inquiry for the subjects (respondents) with which and the context in which the inquiry was carried out?
(2) *Applicability:* How can one determine the extent to which the findings of a particular inquiry have applicability in other contexts or with other subjects (respondents)?
(3) *Consistency:* How can one determine whether the findings of an inquiry would be repeated if the inquiry were replicated with the same (or similar) subjects (respondents) in the same (or similar) context?
(4) *Neutrality:* How can one establish the degree to which the findings of an inquiry are determined by the subjects (respondents) and conditions of the inquiry and not by the biases, motivations, interests, or perspectives of the inquirer?

Within the conventional paradigm, the criteria that have evolved in response to these questions are termed "internal validity," "external validity," "reliability," and "objectivity."

Internal validity may be defined in conventional terms as the extent to which variations in an outcome (dependent) variable can be attributed to controlled variation in an independent variable. A causal connection between independent and dependent variables is usually assumed. Thus Cook and Campbell (1979, p. 37) define internal validity as "the approximate validity [the best available approximation of the truth or falsity of a statement] with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of a cause." Since a variety of factors (plausible rival hypotheses) may influence the outcome, the purpose of design is either to control or to randomize those factors. Data analysis consists of testing the outcome variance against the variance of the randomized factors (error).

Campbell and Stanley (1963) suggest that there are eight "threats" to the internal validity of a study: *history*—the specific external events occurring between the first and second measurement other than the experimental variable(s); *maturation*—processes operating within the respondents as a function of the passage of time per se; *testing*—the effects of taking a test upon the scores of a second testing; *instrumentation*—changes in the calibration of a measurement instrument or changes in the observers or scores used; *statistical regression*—tendencies for movement toward the mean when comparison groups have been selected on the basis of initial extreme scores or positions; *differential selection*—effects of comparing essentially noncomparable groups; *experimental mortality*—the effects of differential loss of respondents from comparison groups, rendering them noncomparable; and *selection—maturation interaction*—an effect that in certain designs may be mistaken for the effect of the experimental variable. The rival hypotheses represented in these eight threats must be invalidated if a study is to have internal validity.

External validity may be defined, as do Cook and Campbell (1979, p. 37), as "the approximate validity with which we infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times." It is the purpose of randomized sampling from a given, defined population to make this criterion achievable. If a sample is selected in accordance with the rule that every element of the population has a known probability (not necessarily equal) of being included in the sample, then it is possible to assert, within given confidence limits, that the findings from the sample will hold for (be generalizable to) the population. It should be noted that the criteria of internal and external validity are placed in a trade-off situation by their definition. If, for the sake of control (internal validity), strenuous laboratory conditions are imposed, then the results are not generalizable to any contexts except those that approximate the original laboratory.

LeCompte and Goetz (1982) point out that, just as there are identifiable threats to internal validity, so are there to external validity. They identify four: *selection effects*—the fact that constructs being tested are specific to a single group, or that the inquirer mistakenly selects groups to study for which the constructs do not obtain; *setting effects*—the fact that results may be a function of the context under

investigation; *history effects*—the fact that unique historical experiences may militate against comparisons; and *construct effects*— the fact that constructs studied may be peculiar to the studied group.

Reliability is typically held to be, in the words of Kerlinger (1973, p. 422), synonymous with "dependability, stability, consistency, predictability, accuracy." Having described a "reliable man" as one whose behavior is consistant, dependable, and predictable, Kerlinger (1973, p. 443) goes on to say,

> So it is with psychological and educational measurements: they are more or less variable from occasion to occasion. They are stable and relatively predictable or they are unstable and relatively unpredictable; they are consistent or not consistent. If they are reliable, we can depend on them. If they are unreliable, we cannot depend on them.

It must be reasonable, as Ford (1975, p. 324) suggests, "to assume that each repetition of the application of the same, or supposedly equivalent, instruments to the same units will yield similar measurements."

Reliability is not prized for its own sake but as a precondition for validity; an unreliable measure cannot be valid, a fact illustrated by the well-known mental test theorem that the validity of a test cannot exceed the square root of its reliability (Gulliksen, 1950). Reliability is usually tested by replication (Ford's "repetition"), as, for example, the odd-even correlation of test items, or the test-retest or parallel-forms correlation. Reliability is threatened by any careless act in the measurement or assessment process, by instrumental decay, by assessments that are insufficiently long (or intense), by ambiguities of various sorts, and a host of other factors.

Objectivity is usually played off against subjectivity. In what Scriven (1971, p. 95) refers to as the "quantitative" contrast between these two, a contrast that is the one usually intended by conventionalists,

> "subjective" refers to what concerns or occurs to the *individual* subject and his experiences, qualities, and dispositions, while "objective" refers to what a *number* of subjects or judges experience—in short, to phenomena in the public domain.

In this sense, the usual criterion for objectivity is intersubjective agreement; if multiple observers can agree on a phenomenon their collective judgment can be said to be objective. Another conventional approach to the problem of establishing objectivity is through methodology; to use methods that by their character render the study beyond contamination by human foibles. Such a methodology is the experiment, as Campbell has observed:

> The experiment is meticulously designed to put questions to "Nature Itself" in such a way that neither the questions, not their colleagues, nor their superiors can affect the answer. (cited in Brewer & Collins, 1981, pp. 15-16)

Objectivity is threatened, then, by using imperfect methodologies that make it possible for inquirer values to refract the "natural" data—putting questions not directly to "Nature Itself" but through an intervening medium that "bends" the response; by engaging in inquiry with an openly ideological purpose; or by relying exclusively on the data provided by a single observer.

* * *

It should be evident that these formulations of criteria intended to respond to the four basic questions are themselves dependent for their meaning on the conventional axioms, such as naive realism and linear causality. We shall have more to say about that later, but for the moment the point to be made is that criteria defined from one perspective may not be appropriate for judging actions taken from another perspective, just as, for example, it is not appropriate to judge Catholic dogma as wrong from the perspective of say, Lutheran presuppositions.

Gareth Morgan (1983a) has made the same point in relation to his management of the project reported in the recent volume, *Beyond Method*. He set himself the task of presenting a variety of research perspectives (each written by an author committed to it) to illustrate the point that each has its own assumptions and provides a separate option for an investigator to consider. But as the project developed unforeseen issues began to emerge:

> For example, there was the question as to how the reader could come to some conclusion regarding the contrary nature, significance, and claims of the different perspectives. Using the work of Gödel (1962) as a metaphor for framing this issue, I realized that there was a major problem here: There was no obvious point of reference outside the system of thought represented in the volume from which the different perspectives could be described and evaluated. As Gödel has shown in relation to mathematics, there is a fallacy in the idea that the propositions of a system can be proved, disproved, or evaluated

> on the basis of axioms within that system. Translated into terms relevant to the present project, this means that it is not possible to judge the validity or contribution of different research perspectives in terms of the ground assumptions of any one set of perspectives, since the process is self-justifying. Hence the attempts in much social science debate to judge the utility of different research strategies in terms of universal criteria based on the importance of generalizability, predictability and control, explanation of variance, meaningful understanding, or whatever *are inevitably flawed: These criteria inevitably favor research strategies consistent with the assumptions that generate such criteria as meaningful guidelines for the evaluation of research.* It is simply inadequate to attempt to justify a particular style of research in terms of assumptions that give rise to that style of research.... Different research perspectives make different kinds of knowledge claims, and the criteria as to what counts as significant knowledge vary from one to another. (Morgan, 1983a, pp. 14-15; emphasis added)

Or, in the vernacular of the streets, "different strokes for different folks." Different basic beliefs lead to different knowledge claims and different criteria.

THE CRITERIA APPROPRIATE TO THE NATURALISTIC PARADIGM

Just what is it that makes the conventional criteria inappropriate to the naturalistic paradigm? If they are inappropriate, what shall we substitute in their place? There is no question that the naturalist is at least as concerned with trustworthiness as is the conventional inquirer. We say "at least" because it is precisely on the point of trustworthiness that the naturalistic investigator is most often attacked, as we tried to show in the opening paragraphs of this chapter. It therefore becomes of utmost importance that (1) the inappropriateness of the conventional criteria be well demonstrated, and (2) acceptable alternative criteria be proposed and their use defended. We may consider the four criterion areas one at a time.

(1) "Truth value." On the assumption of a single, tangible reality that an investigation is intended to unearth and display, the ultimate test of internal validity for the conventional inquirer is the extent to which the findings of an inquiry display an isomorphism (a one-to-one relationship) with that reality. But the determination of such isomorphism is *in principle* impossible, for, in order to make it, the inquirer would need to know the nature of that ultimate tangible reality

a priori. But it is precisely the nature of that reality that is at issue; if one already "knew" it there would be no need to mount an inquiry to determine it.

The conventional inquirer must therefore fall back on a less compelling test; thus the statement by Cook and Campbell cited earlier that internal validity is the "approximate validity with which we infer that a relationship between two variables is causal." The game is played by *postulating* a relationship and then *testing* it against nature (thereby preserving the naive realist posture)—putting the question to "Nature Itself." The hypothesis cannot of course be *proved* (the underdetermination problem) but it can be *falsified* (Popper, 1959).

In order to provide some (persuasive if not compelling) evidence in favor of the claim that the hypothesis is true, it is necessary to eliminate the possibility that plausible rival hypotheses could be at work. "True" experimental designs (in the sense of Campbell & Stanley, 1963) are "true" precisely because they (putatively) unambiguously rule out all such plausible rivals. But, as Campbell and Stanley note, it is not often possible to mount such "true" designs in practice. Perforce one falls back on "quasi-experimental" designs that, while better than mere guesswork, may yield inauthentic results because they are exposed to the "threats" of certain common plausible rivals: history, maturation, and the other factors reviewed briefly above. "True" designs depend for their authenticity on the ability of the investigator to mount suitable controls and/or to randomize; quasi-designs are "imperfect" in one or more ways related to control or randomization.

To score naturalistic inquiry as nontrustworthy on the grounds that controls and/or randomization were not effected is to miss the point that, at bottom, those techniques are appropriate *only insofar as one can buy into the assumption of naive realism.* If that assumption is rejected or altered, then the rational argument summarized above is cut off at the root. When naive realism is replaced by the assumption of multiple constructed realities, there is no ultimate benchmark to which one can turn for justification—whether in principle or by a technical adjustment via the falsification principle. "Reality" is now a multiple set of mental constructions. But, we may note, those constructions are made by humans; their constructions are in their minds, and they are, in the main, accessible to the humans who make them (excepting, let us say, repressed constructions—but even those may become accessible via hypnotism or psychoanalysis). The test of isomorphism, in principle impossible to apply within the conventional paradigm, becomes *the method of choice* for the naturalist. In order

to demonstrate "truth value," the naturalist must show that he or she has *represented those multiple constructions adequately,* that is, that the *reconstructions* (for the findings and interpretations are also constructions, it should never be forgotten) that have been arrived at via the inquiry are *credible to the constructors of the original multiple realities.*

The operational word is *credible.* The implementation of the credibility criterion—the naturalist's substitute for the conventionalist's internal validity—becomes a twofold task: first, to carry out the inquiry in such a way that the probability that the findings will be found to be credible is enhanced and, second, to demonstrate the credibility of the findings by having them approved by the constructors of the multiple realities being studied. We shall in a subsequent section suggest techniques for accomplishing these goals.

We may note, finally, that even if Campbell and Stanley's criteria were to be taken seriously by the naturalist (and we are not arguing that they should be), naturalistic designs would probably score at least as well as the typical quasi-experimental design. Recall that it is Campbell and Stanley's point that *all eight* factors are potential "threats" to quasi-designs; is that also the case with naturalistic designs? Some of the threats can be read as equally applicable to both types; thus differential selection, differential mortality, history, and testing would affect both kinds of outcomes in about the same way. Score: 0-0. One of the threats is probably more likely in naturalistic studies—instrumentation, since changes can and do occur in human instruments and probably to a greater extent than is typical of paper-and-pencil or brass instruments. Score: quasi-designs 1, naturalistic designs 0. But one of these threats—statistical regression—does not apply at all unless quantitative methods are used, and their use is relatively rare in naturalistic studies. Score: quasi-designs 1, naturalistic designs 1. Finally, naturalistic approaches seem particularly useful in *overcoming* two of the threats—maturation and maturation/selection interaction—because naturalistic studies usually involve long-term and continuing interactions with respondents and hence facilitate the assessment of such effects. Final score: quasi-designs 1, naturalistic designs 3. The claim that naturalistic approaches score at least as well as conventional ones on Campbell and Stanley's criteria does not seem to be exaggerated.

(2) Applicability. The criterion of external validity has proved to be troublesome within the conventional framework, for, as we have already suggested, it is in a trade-off situation with internal validity.

The very controls instituted to ensure internal validity militate against clean generalizations. In the final analysis, results that are acquired in that epitome of the controlled situation—the laboratory—are discovered to be applicable only in other laboratores. In that connection we have already cited Urie Bronfenbrenner (1977) on the field of developmental psychology (Chapter 8).

For the naturalist, however, the difficulty with the concept of external validity is not simply that its achievement conflicts with the achievement of internal validity, but that it is based on a conventional axiom that is rejected by the naturalist paradigm. Indeed, naturalists make the opposite assumption: that at best only working hypotheses may be abstracted, the *transferability* of which is an empirical matter, depending on the degree of similarity between sending and receiving contexts. In the classic paradigm all that is necessary to ensure transferability is to know something with high internal validity about Sample A, and to know that A is representative of the population to which the generalization is to apply. The generalization will apply to *all* contexts within that same population.

The naturalist rejects this formulation on several grounds. First, as we saw in Chapter 8, the concept of "population" is itself suspect. As every sampling statistician knows, inferences about populations can be made with greater and greater precision to the extent to which the population is divided into homogeneous strata. But of course such stratification amounts to the formation of subunits that are more and more contextually alike. If one wishes to know, under those circumstances, whether something found out about a stratum of Chicago residents also applies (is generalizable to), say, a stratum of New York residents, the two strata will have to be compared on those factors that define them. That is to say, in order to be sure (within some confidence limits) of one's inference, one will need to know about *both* sending and receiving contexts. We move then from a question of generalizability to a question of *transferability.* Transferability inferences cannot be made by an investigator who knows *only* the sending context.

The condition of representativeness is absolutely basic to the conventional axiom of generalizability. And that axiom in turn seems to depend upon the axiom of naive realism. If there are to be generalizations that are, in Kaplan's sense (see Chapter 5), nomic and nomological, that is, time and context free, there must be some basic rules of nature that govern situations under all circumstances. These basic rules cannot be mere inventions of the mind (constructions); they

must be "real," characteristics of Nature Itself, out there waiting to be discovered. Again the naturalist finds him- or herself in a fundamental propositional disagreement.

It should be clear from the above that if there is to be transferability, the burden of proof lies less with the original investigator than with the person seeking to make an application elsewhere. The original inquirer cannot know the sites to which transferability might be sought, but the appliers can and do. The best advice to give to anyone seeking to make a transfer is to accumulate *empirical* evidence about contextual similarity; the responsibility of the original investigator ends in providing sufficient descriptive data to make such similarity judgments possible. Even if the applier believes on the basis of the empirical evidence that sending and receiving contexts are sufficiently similar to allow one to entertain the possibility of transfer, he or she is nevertheless well advised to carry out a small verifying study to be certain.

Finally, we may note, as in the case of internal validity, naturalistic studies seem to be at least as impervious to the "threats" to external validity as are conventional ones. We noted earlier that LeCompte and Goetz (1982) have specified four threats. Selection effects are threats if the constructs being tested are specific to a single group, but this is precisely what the naturalist believes obtains in every instance unless there is evidence to the contrary, that is, evidence that would show that another group is sufficiently similar to warrant ignoring this possibility. Setting effects are threats because the results may be a function of the context under investigation. But the naturalist sees this state of affairs not as a threat but as the normal circumstance confronting investigators. History effects are threats because unique historical experiences may militate against comparisons. The naturalist expects that to happen. Construct effects are threats because the constructs studies may be peculiar to the studied group. Of course, says the naturalist. The naturalist sees these four states of affairs not as threats but as affirmations of the greater validity of the naturalist axioms. The axioms take these matters into account; they are seen not as effects that undermine external validity but as factors that have to be accounted for in making judgments of transferability.

(3) Consistency. As we have seen, the key concepts undergirding the conventional definition of reliability are those of stability, consistency, and predictability. Within conventional studies reliability is typically demonstrated by replication—if two or more repetitions of essentially similar inquiry processes under essentially similar conditions yield essentially similar findings, the reliability of the inquiry is indisputably established.

But replicability depends, again, upon an assumption of naive realism. There must be something tangible and unchanging "out there" that can serve as a benchmark if the idea of replication is to make sense. If the thing "out there" is ephemeral and changing, noted instabilities cannot be simply charged off to the inquiry procedure; they are at least as much a function of what is being studied as of the process of studying. The quotation from Ford (1975) that requires that the repetitions be applied "to the same units" is telling; it is precisely that condition that can never be met, just as one can never cross the *same* stream twice (if it is indeed possible to cross the *same* stream even once!). Replicability in the traditional sense can be determined only within a given framework—and that framework is itself a construction, not an inevitable and unchanging part of "reality."

The naturalist is willing to concede what might be called "instrumental" unreliability. Conventional theory tells us about unreliabilities of paper-and-pencil or brass instruments, and surely the human instrument displays its equivalents. Humans do become careless; there is "instrumental decay" such as fatigue; the human mind is tentative and groping and it makes mistakes. But the naturalist is not willing to have charged off to his or her "unreliability" changes that occur because of changes in the entity being studied (a construction, remember) or because of changes in the emergent design as insights grow and working hypotheses appear.

The naturalist sees reliability as part of a larger set of factors that are associated with observed changes. In order to demonstrate what may be taken as a substitute criterion for reliability—*dependability*—the naturalist seeks means for taking into account both factors of instability *and* factors of phenomenal or design induced change. It can be argued that this naturalist view is broader than the conventional, since it accounts for everything that is normally included in the concept of reliability plus some additional factors. We shall return later to the question of how this can be accomplished operationally.

(4) Neutrality. The conventional concept of objectivity may be viewed from three perspectives:

(a) Objectivity exists when there is an isomorphism between the data of a study and reality—when the questions are put to "Nature Itself" and it is "Nature Itself" that answers. One might term this the ontological definition, based on a correspondence notion, and it founders,

as must by now be evident, on the naive realist axiom. In all events it would never be possible to test objectivity if it were defined in this way.

(b) Objectivity exists when an appropriate methodology is employed that maintains an adequate distance between observer and observed. One might term this the "epistemological" definition, based on the notion that it is possible for an observer to be neither disturbing nor disturbed (a kind of naive positivism), and it founders on the axiom of subject-object dualism.

(c) Objectivity exists when inquiry is value-free. One might term this the "axiological" definition, based on the notion that is possible to allow Nature to "speak for itself" without impact from the values of the inquirer or any of his or her cohorts. It founders on the axiom of value-dependence.

As we have seen, and as Scriven (1971) points out, the typical criterion that is invoked to judge objectivity is that of intersubjective agreement. What a number of individuals experience is objective and what a single individual experiences is subjective; Scriven refers to this as the "quantitative" sense of objectivity. But, he argues, there is also a qualitative sense in which the subjective/objective distinction may be made. In this sense,

> there is a reference to the *quality* of the testimony or the report or the (putative) evidence, and so I call this the "qualitative" sense. Here, "subjective" means unreliable, biased or probably biased, a matter of opinion, and "objective" means reliable, factual, confirmable or confirmed, and so forth. (Scriven, 1971, pp. 95-96; emphasis in original)

Now the naturalist much prefers this second, qualitative (in Scriven's sense) definition of objectivity. This definition removes the emphasis from the investigator (it is no longer his or her objectivity that is at stake) and places it where, as it seems to the naturalist, it ought more logically to be: on the data themselves. The issue is no longer the investigator's characteristics but the characteristics of the data: Are they or are they not *confirmable?* The naturalist prefers this concept to that of objectivity; again, techniques for assessing confirmability will be discussed below.

The four terms "credibility," "transferability," "dependability," and "confirmability" are, then, the naturalist's equivalents for the conventional terms "internal validity," "external validity," "reliability," and "objectivity." These terms are introduced not simply to add to

naturalism's mystique or to provide it with its fair share of arcane concepts, but to make clear the inappropriateness of the conventional terms when applied to naturalism and to provide alternatives that stand in a more logical and derivative relation to the naturalistic axioms. If it is true, as Gareth Morgan asserts, that different paradigms make different knowledge claims, with the result that criteria for what counts as significant knowledge vary from paradigm to paradigm, then it is essential that the naturalistic paradigm be graced with its own, more appropriate set. We offer these four for consideration.

HOW CAN THE NATURALIST MEET THESE TRUSTWORTHINESS CRITERIA?

We turn now to a consideration of means whereby the naturalist's alternative trustworthiness criteria may be operationalized, dealing with each in turn.

Credibility

We shall suggest five major techniques: activities that make it more likely that credible findings and interpretations will be produced (prolonged engagement, persistent observation, and triangulation); an activity that provides an external check on the inquiry process (peer debriefing); an activity aimed at refining working hypotheses as more and more information becomes available (negative case analysis); an activity that makes possible checking preliminary findings and interpretations against archived "raw data" (referential adequacy); and an activity providing for the direct test of findings and interpretations with the human sources from which they have come—the constructors of the multiple realities being studied (member checking).

(1) Activities increasing the probability that credible findings will be produced. There are three such activities: prolonged engagement, persistent observation, and triangulation. The first, *prolonged engagement,* is the investment of sufficient time to achieve certain purposes: learning the "culture," testing for misinformation introduced by distortions either of the self or of the respondents, and building trust. We saw in the opening paragraphs of this chapter that a major criticism leveled by Freeman (1983) against Margaret Mead was that she spent virtually no time learning about Samoan culture before she focused intensively on the special area she had carved out for herself: adolescent girls. But the meaning of adolescence presumably cannot be appreciated except in terms of larger cultural parameters. Similarly, one

might suggest, it is not possible to understand *any* phenomenon without reference to the context in which it is embedded; indeed, Schwartz and Ogilvy (1979) argue that objects and behaviors take not only their meaning but their very existence from their contexts. It is imperative, therefore, that the naturalist spend enough time in becoming oriented to the situation, "soaking in the culture through his or her pores," to be certain that the context is thoroughly appreciated and understood. Just how long is that? The answer to that question is of course relative to the context's scope and sophistication, but at a minimum it must be: "Long enough to be able to survive without challenge while existing in that culture."

Prolonged engagement also requires that the investigator be involved with a site sufficiently long to detect and take account of distortions that might otherwise creep into the data. First and foremost the investigator must deal with personal distortions. The mere fact of being "a stranger in a strange land" draws undue attention to the inquirer, with its attendent overreaction. It seems likely that unless the inquirer began as an accepted member of the group or agency being studied, distortions can never be overcome; Philip Jackson (1968) points out that in his yearlong study of a California classroom—one in which he sat virtually every day—even his sneezes continued to draw attention until the end of the year, although no one attended to the sneezes of any of the "regular" members of the class. But the investigator also introduces distortions based on his or her own a priori values and constructions. No one enters a site in a mindless fashion; there are always prior formulations, as attested to by the fact that it is always possible to write out ahead of time what one expects to find there. Fortunately this possibility also provides the basis for a test: If the investigator produces field notes and makes interpretations that are continuously predictable from the original formulation, then that investigator has either not spent enough time on site or has persisted against all logic in his or her ethnocentric posture.

There are also distortions introduced by the respondents. Many of these are *unintended;* so, for example, Bilmes (1975) describes a series of sources of "misinformation," including *perceptual distortions and selective perception* (Bilmes admittedly operates from a correspondence view of reality, so the naturalist would want to take this category with a grain of salt); *retrospective distortion and selectivity; misconstruction of investigator's questions*—and hence of the answers given to them; and *situated motives,* such as wanting to please the investigator, saying normatively appropriate things, or simply not being motivated to address the investigator's concern fully. But some distortions are *intended* to deceive or confuse; Douglas (1976) is particularly articulate about the lies, fronts, and deceptions that may be practiced by informants. Indeed, he argues that the cooperative posture that characterizes most inquiry is a case of misplaced confidence; that everyone has something to hide; and that investigators are well advised to adopt an investigative posture. Whether one wishes to be as cynical as Douglas must remain an open question, but there are surely times and places in which the techniques he suggests are useful. During the period of prolonged engagement the investigator must decide whether he or she has risen above his or her own preconceptions, whether misinformation has been forthcoming and whether that misinformation is deliberate or unintended, and what posture to take to combat that problem.

Finally, the period of prolonged engagement is intended to provide the investigator an opportunity to build trust. Now, building trust, as Johnson (1975) has eloquently pointed out, is not a matter of applying techniques that guarantee it. Moreover, trust is not a matter of the personal characteristics of the investigator: a "nice guy" to whom respondents will instinctively confide their innermost secrets. Rather, it is a *developmental* process to be engaged in daily: to demonstrate to the respondents that their confidences will not be used against them; that pledges of anonymity will be honored; that hidden agendas, whether those of the investigator or of other local figures to whom the investigator may be beholden, are not being served; that the interests of the respondents will be honored as much as those of the investigator; and that the respondents will have input into, and actually influence, the inquiry process. Building trust is a time-consuming process; moreover, trust can be destroyed in an instant and then take even more time to rebuild. Prolonged engagement is a must if adequate trust and rapport are to emerge.

Before leaving the topic of prolonged engagement, we wish to add a caveat against the danger of what anthropologists have sometimes referred to as "going native." Lincoln and Guba (1981, p. 4) describe this phenomenon as follows:

> When an anthropologist has become so like the group he is studying that he ceases to consider himself a part of the profession—or ceases to consider either his cultural or professional subgroup as his dominant reference group—he is contributing to the research and begins a "performance-understanding" role (Kolaja, 1956, p. 161) within the studied group. Paul, in a discussion of this problem, named Frank Cushing as an example of an anthropologist who simply refused to continue publishing the results of his field studies. Identification with

the "natives," or co-optation, as a persistent problem of inquirer identification, has been a part of the "warnings and advice" given to new participant observers for several decades. Gold (1969) suggests that going native is almost always the result of naivete, and happens as an unfortunate accident. In the process of attempting to gain *Verstehen,* he asserts, ". . . the field worker may overidentify with the informant and start to lose his research perspective by 'going native' " (p. 36). Moreover, "prolonged direct participation entails the risk that the researcher will lose his *detached wonder* and fail to discover certain phenomena that the relatively uninvolved researcher would discover" (p. 63-64, latter italics added).

It seems clear that any tendencies to "go native" will be abetted by prolonged engagement. The longer the investigator is in the field, the more accepted he or she becomes, the more appreciative of local culture, the greater the likelihood that professional judgments will be influenced. There are no techniques that will provide a guarantee against such influence either unconsciously or consciously; awareness is, however, a great step toward prevention.

The technique of *persistent observation* adds the dimension of *salience* to what might otherwise appear to be little more than a mindless immersion. If the purpose of prolonged engagement is to render the inquirer open to the multiple influences—the mutual shapers and contextual factors—that impinge upon the phenomenon being studied, the purpose of persistent observation is to identify those characteristics and elements in the situation that are most relevant to the problem or issue being pursued and focusing on them in detail. If prolonged engagement provides scope, persistent observation provides depth.

The inquirer must sooner or later come to terms with what Eisner (1975) has termed the "pervasive qualities" involved—those things that really count. That focusing also implies sorting out irrelevancies—the things that do *not* count. But rather than taking the view that the atypical is de facto also the "intrinsically uninteresting," the naturalist must be able to recognize when the atypical may have importance. These goals require that the naturalist continuously engage in tentative labeling of what are taken as salient factors and then exploring them in detail, to the point where either the initial assessment is seen to be erroneous, or the factors are understood in a nonsuperficial way. To satisfy this criterion of trustworthiness, the naturalist must be able to describe in detail just how this process of tentative identification and detailed exploration was carried out.

Persistent observation also has its pitfall, paralleling that of "going native" in relation to prolonged engagement. In this case the danger is that of premature closure. Pressed by demands of clients or funders, and perhaps subject to the intolerance of ambiguity so characteristic of the human species, the naturalistic inquirer may come to a focus too soon—as in the case of Margaret Mead (if Freeman's charge is to be credited). This problem is especially serious in those situations in which lies, fronts, or other deceptions are being practiced, for early closure makes it especially easy to bring off such deceits. The proper practice of persistent observation calls for an aura of skepticism surrounding an intention to come to those terms called for by the situation.

The technique of *triangulation* is the third mode of improving the probability that findings and interpretations will be found credible. It seems likely that the term "triangulation" had its origins in the metaphor of *radio* triangulation, that is, determining the point of origin of a radio broadcast by using directional antennas set up at the two ends of a known baseline. By measuring the angle at which each of the antennas receives the most powerful signal, a triangle can be erected and solved, using simple geometry, to pinpoint the source at the vertex of the triangle opposite the baseline.

Denzin (1978) has suggested that four different modes of triangulation exist: the use of multiple and different *sources, methods, investigators,* and *theories.* The first of these, sources, is what people seem to mean most often when they speak of triangulation. One often encounters phrases such as, "No report was credited unless it could be verified by another person," or "The information forthcoming in interviews was discounted unless it could be checked in the available documents." These expressions suggest that "multiple sources" may imply *multiple* copies of one *type* of source (such as interview respondents) or *different* sources of the *same* information (for example, verifying an interview respondent's recollections about what happened at a board meeting by consulting the official minutes of that meeting [but note that if the minutes do not support the recollections, all one can infer is that *one* of the sources is probably in error]). Diesing (1972, pp. 147-148) supplies yet another possible meaning with respect to sources in his discussion of *contextual validation:*

Contextual validation takes two main forms. First, the validity of a piece of evidence can be assessed by comparing it with other kinds of evidence on the same point. Each kind . . . has its own characteristic

> ambiguities and shortcomings, which are unlikely to coincide with those of another kind. . . .
>
> The second kind of contextual validation is to evaluate a source of evidence by collecting other kinds of evidence about the source . . . to locate the characteristic pattern of distortion in a source.

The first kind of contextual validation seems to be similar to Denzin's use, the second seems to be a new form in which the source itself is called into question. The presumption seems to be that if one can establish a particular *pattern* of distortion (false or biased premises, for example), then one is in a position to *correct* the information forthcoming from that source, including that which cannot be verified elsewhere.

The use of different *methods* for triangulation also has a distinguished history. Webb et al. (1966, p. 3) conclude that while triangulation by methods may be difficult, it is very much worth doing, because it makes data believable:

> Once a proposition has been confirmed by two or more measurement processes, the uncertainty of its interpretation is greatly reduced. The most persuasive evidence comes through a triangulation of measurement processes. If a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence should be placed in it.

The unobtrusive measures proposed in their classic volume are, among other things, intended to provide for such triangulation. They also make the point that different quasi-designs, while each subject to one or more of the Campbell-Stanley "threats," may be used in tandem—a kind of triangulation—so that the imperfections of one are cancelled out by the strengths of another. It is as though a fisherman were to use multiple nets, each of which had a complement of holes, but placed together so that the holes in one net were covered by intact portions of other nets.

The concept of triangulation by different methods thus can imply either different data collection modes (interview, questionnaire, observation, testing) or different designs. The latter concept makes sense only within the conventional paradigm, however, for if the design is emergent, as in a naturalistic study, it would not be possible in advance to patch together multiple designs that had the property of warding off threats to which they might individually be exposed. The naturalist thus falls back on different modes of data collection, using

any that come logically to hand but depending most on qualitative methods.

The use of different *investigators,* a concept perfectly feasible for the conventionalist, runs into some problems in the naturalistic context. If the design is emergent, and its form depends ultimately on the particular interaction that the investigator has with the phenomena (Axiom 2), then one could not expect corroboration of one investigator by another. The problem is identical to that of expecting replicability for the sake of establishing reliability. However, the naturalitst sees it as perfectly possible to use multiple investigators as part of a team, with provisions being made for sufficient intrateam communication to keep all members moving together. The fact that any one team member is kept more or less "honest" by other team members adds to the probability that findings will be found to be credible.

Finally, the use of multiple *theories* for the sake of triangulation is a formulation that the naturalist cannot accept. What can it mean that certain facts can be consistent with two or more theories? In what sense can it be the case that facts can be given more weight if they are consistent with multiple theories? We have noted repeatedly the likelihood that facts are, in the first instance, theory-determined; they do not have an existence independent of the theory within whose framework they achieve coherence. If a given fact is "confirmable" within two theories, that finding may be more a function of the similarity of the theories than of the empirical meaningfulness of the fact. Further, theories can be interrelated; many "facts" within Newtonian theory are also facts within relativity theory, for example, because, in one sense, Newtonian theory can be taken as a "special case" of relativity theory. But the fact is no more believable because it has meaning within both these theories than if it had meaning in only one of them. The use of multiple theories as a triangulation technique seems to us to be both epistemologically unsound and empirically empty.

In summary, we believe it to be the case that the probability that findings (and interpretations based upon them) will be found to be more credible if the inquirer is able to demonstrate a prolonged period of engagement (to learn the context, to minimize distortions, and to build trust), to provide evidence of persistent observation (for the sake of identifying and assessing salient factors and crucial atypical happenings), and to triangulate, by using different sources, different methods, and sometimes multiple investigators, the data that are collected. At the same time the naturalist must guard against overrapport (going native) and premature closure, and take care that modes of triangulation inconsistent with naturalist axioms are not employed.

(2) Peer debriefing. This is the second of the techniques useful in establishing credibility. It is a process of exposing oneself to a disinterested peer in a manner paralleling an analytic session and for the purpose of exploring aspects of the inquiry that might otherwise remain only implicit within the inquirer's mind.

Multiple purposes are served by such a debriefing. First, and from the point of view of credibility, foremost, the process helps keep the inquirer "honest," exposing him or her to searching questions by an experienced protagonist doing his or her best to play the devil's advocate. The inquirer's biases are probed, meanings explored, the basis for interpretations clarified. All questions are in order during a debriefing, whether they pertain to substantive, methodological, legal, ethical, or any other relevant matters. The task of the debriefer is to be sure that the investigator is as fully aware of his or her posture and process as possible (remembering that while it is not possible to divest oneself of values, it is at least possible to be aware of the role they play).

Second, the debriefing provides an initial and searching opportunity to test working hypotheses that may be emerging in the inquirer's mind. Hypotheses that may seem perfectly reasonable to an isolated investigator desperate for *some* kind of closure may appear otherwise in the view of a disinterested debriefer. If the inquirer cannot defend the direction in which his or her mind is taking him or her to a questioner, he or she may very well wish to reconsider that position.

Third, the debriefing provides the opportunity to develop and initially test next steps in the emerging methodological design. Indeed, it is a function of the debriefer to push the inquirer on such steps, perhaps even suggesting some or asking whether certain ones have been considered.

Finally, debriefing sessions provide the inquirer an opportunity for catharsis, thereby clearing the mind of emotions and feelings that may be clouding good judgment or preventing emergence of sensible next steps. Naturalistic inquiry is a lonely business, as the literature well attests (see, for example, Reinharz, 1979; Wax, 1971; Zigarmi & Zigarmi, 1978). The debriefer who listens sympathetically to these feelings, defuses as many as possible, and assists the inquirer to devise coping strategies makes an important contribution to the quality of the study.

There is no formula to prescribe how a debriefing session should be conducted, any more than one can give a prescription for a psychoanalytic interview. It is clear that the debriefer must be someone who is in every sense the inquirer's peer, someone who knows a great deal about both the substantive area of the inquiry and the methodological issues. The debriefer should be neither junior—lest his or her inputs are disregarded—nor senior—lest his or her inputs be considered as mandates, or lest the inquirer "hold back" for fear of being judged incompetent. The debriefer should not be someone in an authority relationship to the inquirer (a matter of particular note in the case of a doctoral study, which should avoid using members of the research committee as debriefers). The debriefer should be someone prepared to take the role seriously, playing the devil's advocate even when it becomes apparent that to do so produces pain for the inquirer. Both inquirer and debriefer should keep written records of each encounter, partly for the sake of the audit trail (see below), and partly for reference by the inquirer as he or she later seeks to establish just why the inquiry emerged as it did.

Debriefing has several dangers. The inquirer may come to feel that his or her progress, or judgments, or insights, are not what they should be, and therefore may suffer diminished enthusiasm and energy. A careful and empathic debriefer can do much to avoid giving that impression. There is the distinct possibility that the inquirer may be influenced by the debriefer to a greater extent than should be the case—a tendency especially likely if the debriefer operates too directly from a conventional framework and is too demanding in terms of conventional criteria. Naturalists are, it should be recalled, the methodological out-group; it is they and not the conventionalists who must prove the utility of their approach. Too much criticism can be damaging in the extreme. Yet, despite these dangers, debriefing is a useful—if sobering—experience to which to subject oneself; its utility, when properly engaged, is unquestionable.

(3) Negative case analysis. A most useful discussion of this technique has been provided recently by Kidder (1981), who sees it as analogous, for qualitative data, to statistical tests for quantitative data. The reader should be forewarned, however, that Kidder takes an avowedly conventional posture; one might regard her work as one of those attempts at striking a compromise between the "qualitative and quantitative paradigms." Nevertheless her treatment is instructive, and we shall draw heavily upon it.

Negative case analysis may be regarded as a "process of revising hypotheses with hindsight." The object of the game is continuously to refine a hypothesis until it *accounts for all known cases without exception.* Hypotheses take the form, "All members of Class X have characteristics A, B, and C." So, for example, the hypothesis might be, "All learning disabled children will exhibit poor performance in

school, a 'spiked profile' of intellectual competencies (high in reading and social studies, say, but low in mathematics and science), and poor personal/social adjustment." Or, "All bureaucratic organizations exhibit subunit agreement on a common overall goal, perform complementary subunit functions (the output of one becomes the input of the next, and so on; commonly called "tight coupling"), and shared reward systems." These hypotheses are tested and refined so that, ultimately, the pattern exhibited in Table 11.1 is obtained, that is, all members of the class do share the characteristics named in the final version of the hypothesis.

Kidder cites as an example a study reported by Cressey (1953) on embezzlement. Five different versions of a hypothesis about the characteristics of embezzlers were formulated at various stages of the study, with each revision coming after certain findings inconsistent with earlier versions were obtained. Kidder (1981, p. 241) observes:

> Cressey formulated and revised his hypothesis five times before he arrived at his conclusion about the cases of embezzlement. Each time he formulated a new hypothesis, he checked it against not only new interviews but also all of his previously recorded interviews and observations. This ex post facto procedure is a necessary practice. . . . [it] forms the basis for analytic induction and negative case analysis. Negative case analysis requires that the researcher look for disconfirming data in both past and future observations. A single negative case is enough to require the investigator to revise a hypothesis. When there are no more negative cases, the researcher stops revising the hypothesis and says with confidence, "This caused that."

Leaving aside the quarrel we may have with the causal interpretation that Kidder implies is possible with negative case analysis, we may nevertheless find the Cressey example instructive. The first of the five

TABLE 11.1 Ideal Configuration After Negative Case Analysis

	Hypothetical Class	
Characteristics	*Member* %	*Nonmember* %
Present	100	0
Absent	0	100

hypotheses devised by him—and the data that required revision—took roughly the following form:

> Embezzlement occurs when someone "has learned in connection with the business or profession in which he is employed that some forms of trust violations are merely technical violations and are not really 'illegal' or 'wrong.' " (Conversely, if this definition has not been learned, violations do not occur.)

Cressey had to abandon this formulation as soon as interviews with incarcerated embezzlers made it plain that they knew all along that embezzling was illegal. Thus the second formulation:

> Embezzlement occurs when the incumbent of a position of trust "defines a need for extra funds or extended use of property as an 'emergency' which cannot be met by legal means."

This formulation had to be rejected when some interview respondents indicated that they had taken money without being confronted by an emergency; others said that they had at other times been confronted by emergencies and had *not* taken money. Hence the third formulation:

> Embezzlement occurs when persons in positions of trust "conceive of themselves as having incurred financial obligations which are . . . nonsocially sanctionable and which . . . must be satisfied by private means."

Cressey checked this formulation against both previous and subsequent interviews and found instances in which nothing existed that could be considered a financial obligation—a past debt for which the person felt responsible—and he found other instances in which nonsanctionable obligations had existed without embezzlement. Thus the fourth version:

> Embezzlement occurs not only for the reasons cited in the third hypothesis, but also "because of present discordance between the embezzler's income and expenditures as well."

This revision did account for some previously unaccountable types, but again negative instances were formed in which the conditions existed but embezzlement had not occurred. Finally, the fifth version:

> "Trusted persons become trust violators when they conceive of themselves as having a financial problem which is nonshareable, are aware that this problem can be secretly resolved by violation of the posi-

> tion of financial trust, and are able to apply to their own conduct in that situation verbalizations which enable them to adjust their conception of themselves as users of the entrusted funds or property."

Kidder (1981, p. 243) comments:

> Cressey tested this hypothesis against all the data he had gathered, against two hundred cases of embezzlement collected by another researcher, and against additional interviews that he conducted in another penitentiary. He found no negative cases.

Thus negative case analysis eliminates all "outliers" and all exceptions by continually revising the hypothesis at issue until the "fit" is perfect. Kidder (1981, p. 244) suggests that negative case analysis is to qualitative research as statistical analysis is to quantitative:

> Both are means to handle error variance. Qualitative research uses "errors" to revise the hypothesis; quantitative analysis uses error variance to test the hypothesis, demonstrating how large the treatment effects are compared to the error variance.

Of course, as Kidder also notes, proponents of the conventional statistical approach take exception to negative case analysis because it seems to build upon chance variations in the data at hand. But she rejects this criticism and endeavors to show the parallelism that exists between statistical analysis and negative case analysis. Whether she succeeds in that attempt is not a particular issue here; what is important to note is that the technique of negative case analysis does provide a useful means to make data more credible by reducing the number of exceptional cases to zero.

But perhaps the insistence on *zero* exceptions may be too rigid a criterion. Indeed, on its face it seems almost impossible to satisfy in actual studies (the contention that Cressey found *no* exceptions in all his own data, not to mention hundreds of other cases developed by a colleague, is a little hard to believe). In situations where one might expect lies, fronts, and other deliberate or unconscious deceptions (as in the case of self-delusions), some of the cases ought to *appear* to be exceptions even when the hypothesis is valid simply because the false elements cannot always be fully penetrated. Yet, if a hypothesis could be formulated that fit some reasonable number of cases—even as low, say, as 60 percent—there would seem to be substantial evidence of its

acceptability. After all, has anyone ever produced a perfect statistical finding, significant at the .000 level? The naturalistic inquirer who would cite such evidence would have piled up a convincing argument in favor of credibility.

(4) Referential adequacy. The concept of referential adequacy was first proposed by Eisner (1975), who suggested it as a means for establishing the adequacy of critiques written for evaluation purposes under the connoisseurship model. Videotape recordings and cinematography, he asserted, provide the means for "capturing and holding episodes of classroom life" that could later be examined at leisure and compared to the critiques that had been developed from all of the data collected. The recorded materials provide a kind of benchmark against which later data analyses and interpretations (the critiques) could be tested for adequacy.

But there is no need to confine such referential tests solely to electronically recorded data segments. Indeed, it seems likely that many investigators will lack the resources if not the expertise to utilize such high-tech devices as video recorders or movie cameras. Further, the collection of information by such means is highly obtrusive. But the concept can still be utilized if the investigator will earmark a portion of the data to be archived—not included in whatever data analysis may be planned—and then recalled when tentative findings have been reached. Aside from the obvious value of such materials for demonstrating that different analysts can reach similar conclusions given whatever data categories have emerged—a matter of reliability—they can also be used to test the validity of the conclusions. Skeptics not associated with the inquiry can use such materials to satisfy themselves that the findings and interpretations are meaningful by testing them directly and personally against the archived and still "raw" data. A more compelling demonstration can hardly be imagined.

Of course, there are drawbacks to the referential adequacy approach. First and foremost, the investigator will have to surrender some of his or her hard-won raw data to the archives, agreeing not to use those materials to further the purposes of the inquiry per se but reserving them exclusively for this adequacy test. Inquirers may be reluctant to give up appreciable portions of data for what may seem to them at best a tangential purpose. Further, it is likely that conventional critics will not accept these materials unless they can be shown to be representative—in the classical sense of the term. Since naturalists do not sample with representativeness in mind, they may be hard put to meet such a criterion, and may feel (rightly) that it is not an appropriate

requirement to lay on them. Naturalists using the referential materials are likely to want to "peel the onion" to a different layer, demonstrating less interest in the original analyst's findings than in developing their own. For all these reasons the referential adequacy approach does not recommend itself well to the more practical-minded or resource poor. Nevertheless, when resources and inclinations permit, the storage of some portion of the raw data in archives for later recall and comparison provides a rare opportunity for demonstrating the credibility of naturalistic data.

(5) Member checks. The member check, whereby data, analytic categories, interpretations, and conclusions are tested with members of those stakeholding groups from whom the data were originally collected, is the most crucial technique for establishing credibility. If the investigator is to be able to purport that his or her reconstructions are recognizable to audience members as adequate representations of their own (and multiple) realities, it is essential that they be given the opportunity to react to them.

Member checking is both informal and formal, and it occurs continuously. Many opportunities for member checks arise daily in the course of the investigation. A summary of an interview can be "played back" to the person who provided it for reaction; the output of one interview can be "played" for another respondent who can be asked to comment; insights gleaned from one group can be tested with another. Such immediate and informal checking serves a number of purposes:

- It provides the opportunity to assess intentionality—what it is that the respondent *intended* by acting in a certain way or providing certain information.
- It gives the respondent an immediate opportunity to correct errors of fact and challenge what are perceived to be wrong interpretations.
- It provides the respondent the opportunity to volunteer additional information; indeed, the act of "playing back" may stimulate the respondent to recall additional things that were not mentioned the first time around.
- It puts the respondent on record as having said certain things and having agreed to the correctness of the investigator's recording of them, thereby making it more difficult later for the respondent to claim misunderstanding or investigator error.
- It provides an opportunity to summarize—the first step along the way to data analysis.
- It provides the respondent an opportunity to give an assessment of overall adequacy in addition to confirming individual data points.

However, more formal checking is necessary if a claim to credibility is to be entertained meaningfully. For this purpose the investigator may wish to arrange a session, perhaps lasting an entire day or even several days, to which are invited knowledgeable individuals from each of the several interested source groups. Copies of the inquiry report may be furnished to such a member-check panel in advance for study and written commentary, while at the session itself, representatives of different groups may wish to air their disagreements with the investigator, or with one another. Clearly the investigator is not bound to honor all of the criticisms that are mounted, but he or she is bound to hear them and weigh their meaningfulness.

Of course problems emerge with the member-checking process. Most obviously, the groups brought together for the review may be in an adversarial position. The issue may turn out to be less one of the adequacy of the reconstructions than of their fairness. Checkers may be able to agree that reconstructions are fair even if they are not in total agreement with them. Care must be exercised that in an attempt to be fair the investigator does not simply reconstruct an "average" or "typical" position, which is not only in conflict with the naturalistic position on generalizability but which at bottom represents *no one's* reality.

Moreover, member checks can be misleading if all of the members share some common myth or front, or conspire to mislead or cover up. We have already noted that the naive investigator may be taken in through conspiratorial agreements about what he or she should or should not "discover" (Douglas's 1976 treatment about the several levels of fallback fronts utilized by massage parlor girls is instructive). Should he or she be so taken in, it is an easy next step for the member checks to affirm the validity of what has been "found." Unless one has reason to doubt the integrity of informants, however, the member check is probably a reasonably valid way to establish the meaningfulness of the findings and interpretations. The investigator who has received the agreement of the respondent groups on the credibility of his or her work has established a strong beachhead toward convincing readers and critics of the authenticity of the work.

The reader should be careful not to confuse the concept of member checking with that of triangulation. Superficially these two techniques appear identical, but there is a crucial difference. Triangulation is a process carried out with respect to *data*—a datum or item of information derived from one source (or by one method or by one investigator) should be checked against other sources (or by other methods or investigators). Member checking is a process carried out with respect to

constructions. Of course, constructions may be found to be noncredible because they are based on erroneous data, but the careful investigator will have precluded that possibility by virtue of assiduous earlier triangulation. Member checking is directed at a judgment of overall credibility, while triangulation is directed at a judgment of the accuracy of specific data items.

Transferability

The establishment of transferability by the naturalist is very different from the establishment of external validity by the conventionalist. Indeed, the former is, in a strict sense, impossible. For while the conventionalist expects (and is expected) to make relatively precise statements about external validity (expressed, for example, in the form of statistical confidence limits), the naturalist can only set out working hypotheses together with a description of the time and context in which they were found to hold. Whether they hold in some other context, or even in the same context at some other time, is an empirical issue, the resolution of which depends upon the degree of similarity between sending and receiving (or earlier and later) contexts. Thus the naturalist cannot specify the external validity of an inquiry; he or she can provide only the thick description necessary to enable someone interested in making a transfer to reach a conclusion about whether transfer can be contemplated as a possibility.

The question of what constitutes "proper" thick description is, at this stage in the development of naturalist theory, still not completely resolved. Clearly, not just any descriptive data will do, but the criteria that separate relevant from irrelevant descriptors are still largely undefined. One primitive attempt to define them is detailed in Chapter 13. The reader may regard that statement as a specification of the *minimum* elements needed. The naturalist inquirer is also responsible for providing the widest possible range of information for inclusion in the thick description; for that reason (among others) he or she will wish to engage in purposeful sampling (described in Chapter 9).

It is, in summary, *not* the naturalist's task to provide an *index* of transferability; it *is* his or her responsibility to provide the *data base* that makes transferability judgments possible on the part of potential appliers.

Dependability

In an earlier paper (Guba, 1981a) one of the authors made a number of arguments useful in shoring up dependability claims:

(1) Since there can be no validity without reliability (and thus no credibility without dependability), a demonstration of the former is sufficient to establish the latter. If it is possible using the techniques outlined in relation to credibility to show that a study has that quality, it ought not to be necessary to demonstrate dependability separately. But while this argument has merit, it is also very weak. It may serve to establish dependability in practice, but does not deal with it in principle. A strong solution must deal with dependability directly.

(2) A more direct technique might be characterized as "overlap methods." In effect, overlap methods represent the kind of triangulation urged by Webb et al. (1966) and reviewed in relation to credibility. But, as noted by Guba, triangulation is typically undertaken to establish validity, not reliability, although, by Argument 1 above, demonstration of the former is equivalent to demonstration of the latter. The "overlap methods" are simply one way of carrying out Argument 1 and not a separate approach.

(3) A third technique suggested by Guba is the method of "stepwise replication," a process that builds on the classic notion of replication in the conventional literature as *the* means of establishing reliability. The approach, somewhat analogous to the "split-half" mode of determining test reliability, requires an inquiry team of at least two persons, and preferably multiple persons, who can be divided into two inquiry teams. These teams deal with data sources separately and, in effect, conduct their inquiries independently. But there is the rub. Such an approach is quite possible within the conventional paradigm, in which a detailed research design that both teams could follow independently with no difficulty is laid out in advance. But the naturalistic design is emergent; it is precisely because the two teams could, for reasons independent of the instability problem, diverge onto two quite different lines of inquiry that stepwise replication is a dubious procedure. Guba recognized this problem and proposed to deal with it by making extraordinary provision for communication: on a daily basis, at milestone points, and whenever either of the teams saw a need for deviating from an originally chosen path (that is, a need to change the design). While such an approach may be feasible (although no doubt many conventionalists would argue that such arrangements utterly destroy the condition of independent inquiry), it is very cumbersome. Since other modes exist for establishing dependability, there seems to be little point in pursuing such a problematic alternative. It is therefore not recommended by us at this time.

(4) A fourth technique proposed by Guba is that of the inquiry audit, based metaphorically on the fiscal audit. Essentially, an auditor called in to authenticate the accounts of a business or industry is expected to perform two tasks. First, he or she examines the *process* by which the accounts were kept, to satisfy stakeholders that they are not the victims of what is sometimes called "creative accounting." The

concern here is not with the possibility of error or fraud, but with the fairness of the representation of the company's fiscal position. Accounting modes that would, for example, make the company appear to be more successful than it was, perhaps in the hope of attracting additional investors, are fair game for the auditor, who is expected to "blow the whistle" should such practices be detected.

The second task of the auditor is to examine the product—the records—from the point of view of their accuracy. Two steps are involved. First, the auditor needs to satisfy him- or herself that every entry in the account ledgers can be justified. So, for example, the auditor may send a letter to various involved parties asking them to confirm that the status of their account is thus and so, or that they did bill the company so many dollars for certain services on such-and-such date. In addition, the auditor may sample entries in the journal to ascertain whether they are supported by corroborative documents. So, for example, if an entry shows that a certain sum was paid to a salesman to reimburse expenses, the auditor may wish to see the voucher and its attached airline, hotel, car rental, meals, and other receipts. Second, the auditor reviews the amounts so as to be able to "verify the bottom line."

When the auditor has performed both these tasks to the standards required, he or she provides an attestation, for example, "Price, Waterhouse and Company have examined the books of the General Electric Company and find them to be in good order . . . ." In providing such an attestation the auditor certifies that both the process of accounting and the product—the account ledgers—fall within acceptable professional, legal, and ethical limits.

The two tasks of the inquiry auditor may be taken metaphorically as very similar to the tasks of a fiscal auditor. The former is *also* expected to examine the *process* of the inquiry, and in determining its acceptability the auditor attests to the *dependability* of the inquiry. The inquiry auditor also examines the *product*—the data, findings, interpretations, and recommendations—and attests that it is supported by data and is internally coherent so that the "bottom line" may be accepted. This latter process establishes the *confirmability* of the inquiry. Thus a single audit, properly managed, can be used to determine dependability and confirmability simultaneously. A fuller explication of the audit process is undertaken below.

Confirmability

The major technique for establishing confirmability is, as indicated above, the confirmability audit. Two other techniques (triangulation and the keeping of a reflexive journal) suggested by Guba (1981) for confirmability will be seen to dovetail with the audit process and hence are no longer discussed independently.

The major credit for the operationalization of the auditing concept must go to Edward S. Halpern, who in 1983 completed his dissertation at Indiana University on that topic. The major useful residues of that study are twofold: (1) a specification of the items that should be included in the audit trail—the trail of materials assembled for the use of the auditor, metaphorically analogous to fiscal accounts; and (2) an algorithm for the audit process itself. These two documents are included as Appendices A and B; they will be explicated here briefly.

(1) The audit trail. An inquiry audit cannot be conducted without a residue of records stemming from the inquiry, just as a fiscal audit cannot be conducted without a residue of records from the business transactions involved. Halpern suggests six classes of such raw records, which are outlined briefly below (see Appendix A for a fuller description). It may be noted in passing that the inquirer who keeps such records, suitably coded according to Halpern's notational system, will have greatly eased his or her own reporting problem. The inquirers who engaged Halpern to audit their studies were uniform in reporting that the discipline imposed on them by the need to provide an audit trail had innumerable payoffs in helping to systematize, relate, cross-reference, and attach priorities to data that might otherwise have remained undifferentiated until the writing task was undertaken. Thus there is utility in collecting information in accordance with audit requirements irrespective of whether an audit is intended and irrespective of which inquiry paradigm is being followed.

The six Halpern audit trail categories are these:

(1) *raw data,* including electronically recorded materials such as videotapes and stenomask recordings; written field notes, unobtrusive measures such as documents and records and physical traces; and survey results
(2) *data reduction and analysis products,* including write-ups of field notes, summaries such as condensed notes, unitized information (as on 3 × 5 cards), and quantitative summaries; and theoretical notes, including working hypotheses, concepts, and hunches
(3) *data reconstruction and synthesis products,* including structure of categories (themes, definitions, and relationships); findings and conclusions (interpretations and inferences); and a final report, with connections to the existing literature and an integration of concepts, relationships, and interpretations
(4) *process notes,* including methodological notes (procedures, designs, strategies, rationale); trustworthiness notes (relating to credibility, dependability, and confirmability); and audit trail notes

(5) *materials relating to intentions and dispositions,* including the inquiry proposal; personal notes (reflexive notes and motivations); and expectations (predictions and intentions)
(6) *instrument development information,* including pilot forms and preliminary schedules; observation formats; and surveys

Each of these categories is further subdivided by Halpern to provide illustrations of the kinds of evidence that might be useful for each category. Halpern's table is intended to be inclusive of all forms of inquiry and of the full range of information that might be available. Thus not all of this information would be placed before the auditor in any one situation. It is unlikely, for example, that a naturalistic study would produce much audit trail material in Category 6 (instrument development information). Probably no study would produce extensive files of *both* electronically recorded data and field notes; the inquirer relying on field notes would not be inclined also to audio- or video-record. Thus the actual task confronting the auditor may be much more manageable in practice than a casual inspection of Appendix A might suggest.

(2) The audit process. The Halpern algorithm is divided into five stages: preentry; determination of auditability; formal agreement; determination of trustworthiness (dependability and confirmability, and a secondary check on credibility); and closure. The reader should note that Appendix B provides, for each stage and its substages, a listing of tasks that should be carried out by the auditee and the auditor, guiding questions to help the auditor reach conclusions, and cross-references for the audit trail categories that must be consulted at each point.

Two considerations should be borne in mind in perusing Appendix B and in reading the following audit process description. First, the algorithm should be understood as a *reconstructed* logic, not a *logic-in-use* (Kaplan, 1964). While the stages and substages are described in a rational order, it is not the case that the sequence is inviolable; in an actual situation some of the steps may be interchanged and others may be omitted entirely. Further, there may be reiterations if circumstances require. Thus it is not order but the scope of coverage that is important. Second, the reader should note that the algorithm is based on the assumption that the auditor is called in at the very beginning of the study and thus can prescribe the nature of the audit trail as well as other helpful details. But just as evaluators are often not called in until the program they are to evaluate is well along in its development and implementation (*the* most common complaint of the evaluator is,

"If only they had called me in sooner . . . "), so auditors may not be consulted until the study is virtually complete. Indeed, there may be some utility in waiting until the end to avoid the possibility that the auditor might be coopted. After all, fiscal auditors are not consulted until after the accounts are closed; would one believe Price Waterhouse if they had been working with the General Electric accountants all year, advising them on what to do? Thus the reader should understand that (probably major) adjustments will need to be made in the algorithm depending on just when the auditor is initially contacted. If the auditor is not brought in until after the study is completed, it simply means that many of the steps of Appendix B will have to be carried out retrospectively. The danger in retrospective auditing is, of course, that deficiencies cannot be repaired; if, for example, the auditee has kept an inadequate audit trail, it may not be possible to carry out an audit at all. Problems of that sort ought to occur infrequently, however, particularly as auditees become more sophisticated about auditing requirements. There is no danger, for example, that a fiscal accountant will ever fail to keep the records that an auditor will require, for the fiscal auditor's needs are well understood and codified. One may confidently expect that an equivalent status will be reached in inquiry auditing before too long.

We turn now to a description of Halpern's five stages:

(1) Preentry. This phase is characterized by a series of interactions between auditor and auditee that result in a decision to continue, continue conditionally, or discontinue the proposed audit. Having determined that an audit might be desirable and useful, the auditee selects a potential auditor (the nature of persons suitable to be auditors is discussed below). An agreement is reached to have further conversation, in preparation for which the auditee prepares an outline indicating the kinds of audit trail materials that he or she will be able to collect and the format in which they will be made available. In their initial conversation, the auditee explains this record-keeping system to the proposed auditor, and describes the nature of the proposed study (as well as can be done in prospect). Finally, auditor and auditee discuss the three alternatives and decide to continue, continue conditionally, or discontinue their relationship. If the decision is to continue conditionally, the conditions are spelled out for the record, and the proposed audit trail is revised as necessary.

(2) Determination of auditability. This stage begins at whatever point the auditor and auditee have previously agreed should be the entry point; this may be after some specified time period or at some milestone event (if the auditor is to be involved during the course of the study),

or at the end of the inquiry (if the auditor is to perform ex post facto). The auditor's first task is to become thoroughly familar with the study: the problem (or evaluand or policy option) investigated (and how it may have changed with time), the paradigmatic and methodological approaches taken, the nature of the guiding substantive theory (and whether it is grounded or given a priori), and the findings and conclusions. The auditee's task is to arrange relevant materials in some convenient and easily accessible form, and to remain available for consultation as needed.

Next, the auditor must familiarize him- or herself with the audit trail as it has actually materialized. Presumably the trail will follow the structure and format previously agreed upon. The auditor in particular must become familiar with the linkage system that ties audit trail materials to actual events and outcomes. So, for example, if a datum is reported in a case study, the auditor must know how to trace that datum back to its original sources in interview and observation records, documents, videotapes, or whatever.

Finally, the auditor must make a determination of the study's auditability; in effect, this determination signals continuation or termination of the process. The auditor must be satisfied that the audit trail is *complete* (that is, that all of the elements in Appendix A are available or otherwise accounted for); that the trail is *comprehensible* (that is, that it can be understood and followed); that it is *useful* (that is, it is arranged in ways that make cross-referencing, indexing, organization, and the like evident); and that it is *linked* (that is, that the audit trail is systematically related to the methodological approaches, both in their initial and unfolded form). Following this determination the auditor and auditee engage in further negotiation, which may result, as in the preentry stage, in a decision to continue, continue conditionally, or discontinue the process. A decision to continue conditionally implies, of course, the auditee's ability to fulfill the conditions. A decision to continue if revisions are made in the audit trail may not be feasible, for example, if the auditor has not been consulted until after the study's completion, at which time it may not be possible to reconstruct missing items (it should be noted that even if reconstructions are possible, those reconstructions cannot be accorded the same weight as constructions made at the original time and place).

(3) Formal agreement. Assuming that a decision has been made in Stage 2 above to continue in some form, it is now appropriate to reach formal written agreement on what is to be accomplished by the audit. The agreement "locks in" the auditor; beyond this point there cannot (ethically or legally) be a withdrawal. The contract reached should do the following: establishing the *time limit* for the audit; determine the audit's *goals* (dependability, or confirmability, or both, with possibly a secondary check on credibility); specify the *roles* to be played by both auditor and auditee (along the lines of the tasks specified in the algorithm); arrange the *logistics* of the audit (time, place, support facilities, and so on); determine the product *outcomes* (reports, presentations, and the like); determine the *format* (a possible format for an auditor's report is discussed below); and identify *renegotiation criteria* (what to do in the event that the auditee finds the auditor's report faulty or erroneous, or if either party is impelled to alter the terms of the formal agreement in some way).

(4) Determination of trustworthiness. This stage is concerned with reaching assessments of confirmability, dependability, and, as an optional feature, providing an external check on steps taken in relation to credibility. The reader will note that the algorithm as displayed in Appendix B calls for the confirmability check to precede the dependability check, an order that reverses that which has characterized the discussion so far. The order is not, however, critical.

The assessment of *confirmability* itself involves several substeps. The auditor's first concern will be to ascertain whether the findings are grounded in the data, a matter easily determined if appropriate audit trail linkages have been established. A sampling of findings (it is suggested that findings that appear, on their face, to be most bizarre or unusual be among those sampled) is traced back, via the audit trail, to the raw data—interview notes, document entries, and the like—upon which they are based. Next, the auditor will wish to reach a judgment about whether inferences based on the data are logical, looking carefully at analytic techniques used, appropriateness of category labels, quality of interpretations, and the possibility of equally attractive alternatives. The auditor should then turn his or her attention to the utility of the category structure: its clarity, explanatory power, and fit to the data. The auditor will wish to make an assessment of the degree and incidence of inquirer bias (a clear judgment call), taking into account preponderance of inquirer terminology (as contrasted to grounded terminology), overimposition of a priori theoretical concepts (believing is seeing), and presence or absence of introspections. Finally, the auditor will assess the auditee's "accommodation strategies": the efforts made by the auditee during the inquiry to ensure confirmability (for example, triangulation), the extent to which negative evidence was taken into account, and the accommodation of negative examples (which should have been mostly eliminated through negative case

analysis). Upon successful completion of these steps the auditor will be able to reach an overall decision about the study's confirmability—the extent to which the data and interpretations of the study are grounded in events rather than the inquirer's personal constructions.

The assessment of *dependability* likewise involves a number of steps. First, the auditor is concerned with the appropriateness of inquiry decisions and methodological shifts: Are these identified, explicated, and supported? Inquirer bias is again reviewed to determine the extent to which the inquirer resisted early closure (early closure suggests too much dependence on the inquirer's own a priori constructs), the extent to which all data have been accounted for and all reasonable areas explored, the extent to which decisions about the conduct of the inquiry may have been overly influenced by practical matters such as arbitrary sponsor deadlines or client interests, and the extent to which the inquirer endeavored to find negative as well as positive data. Instances that suggest the inquirer may have been coopted are noted, as well as those in which premature judgments may have been reached. The possibility that the study may have been influenced by Pygmalion and Hawthorne effects is assessed, and the level of sophistication of the inquirer is taken into account. Sampling decisions and triangulation processes are again briefly reviewed. Finally, the overall design (as it emerged) is evaluated, and possible intrusion of instabilities noted. These several steps lead the auditor to a final overall assessment of dependability.

While it was not contemplated in early formulations of the audit process, Halpern found the auditor to have considerable leverage on the question of whether *credibility* had been appropriately dealt with in a study. Thus the algorithm contains an optional section (Step 10) in which the auditor can pursue that question. Essentially, this step requires the auditor to review the study from the point of view of techniques for credibility that have already been discussed—such as triangulation, peer debriefing, and member checks. To Halpern's list we would also add collection of referential adequacy materials and the application of negative case analysis.

(5) Closure. When the auditor has completed all of the tasks outlined in the Halpern algorithm, two steps remain: feedback and renegotiation, and the writing of a final report, which might more appropriately be called a "letter of attestation." In respect to the former, the auditor is obliged to review his or her findings with the auditee, for several purposes. The auditee has the right to know that all steps have been concluded in accordance with the previously negotiated agreement. If there have been errors of omission those can be called to the attention of the auditor, who should move to carry them out. Further, the auditee has the right to hear the findings and to register concurrence or exceptions. If exceptions are noted, there may be further negotiations between auditor and auditee to resolve them, for example, by carrying out some additional checks, reviewing work process steps, and the like. In the final analysis, if the auditor and auditee disagree, the auditor has the right to present the findings as he or she sees them, and the auditee has the right to append an exception report for the record.

In all events, the auditor must prepare a letter of attestation. While each case probably should be treated on its own merits, it seems likely that such a letter might be prepared according to the following outline:

(1) The charge: to determine (dependability) (confirmability) (both dependability and confirmability) (dependability, confirmability, and to review credibility measures).
(2) Theoretical basis for the audit (on the assumption that the typical reader may not be familiar with the concept).
  (a) Brief discussion of the methaphor of fiscal auditor.
  (b) Referencing of selected references (e.g., Guba & Lincoln, 1981; Halpern, 1983).
(3) Specification of particular goals of this audit: What are the particular questions that were agreed upon in the formal contract?
(4) Discussion of procedures used. Brief review of the Halpern algorithm (if used; if not, the actual procedures should be outlined). Additional steps or omitted steps should be described.
(5) Findings. Steps 8, 9, and 10 of the algorithm should be used as a guide for this presentation, as appropriate. Exceptions should be clearly explicated, together with the evidence in their support.
(6) Overall attestation, in conformity with 1 (the charge) above.
(7) Signature of auditor, together with typed name and professional affiliation (for identification only).
(8) A brief vita for the auditor (one or two paragraphs) that establishes the auditor's credentials to carry out audits.

It would not be surprising if the reader were to be overwhelmed by the apparent complexity of the auditing task, as imaged either by the preceding brief description or by the more detailed Halpern algorithm in Appendix B. In a real case, however, the steps are not so difficult to carry out as might be imagined. The question frequently comes up about the length of time it takes to do an audit; the way the question is asked suggests that is must be an overly long period. Related to that question is that of the resources (usually the fee involved) for having an audit carried out. It does not seem unreasonable to suggest that even for a complex project, a week to ten days will

be sufficient, including a day or so to browse through some initial orientational materials, three to five days to carry out the audit itself, and several additional days to prepare the report (much of which can be already available in the form of "boilerplate," once one or two audits have been done). The required resources may be no more than a typical fee for that amount of time, plus travel expenses to the site at which the audit is to be done. Some of our students have arranged "round-robin audits" for their dissertations, forming a pool from which each individual may draw someone to perform his or her audit (of course audits are not exchanged one-on-one; the possibilities for bias would be too great), and in return performing an audit for someone else.

The auditor should see him- or herself as acting on behalf of the general readership of the inquiry report, a readership that may not have the time or inclination (or the accessibility to the data) to undertake a detailed assessment of trustworthiness. If, as Cronbach and Suppes (1969) suggest, *disciplined* inquiry is inquiry that is open to inspection and verification, the role of the auditor is to make the inspection and verification on behalf of the reader and to attest to having done so. The role of inquiry auditors is thus exactly parallel to that of fiscal auditors, who, on behalf of a stakeholding group that may not be sufficiently sophisticated to read account statements themselves or may not be able to travel to the place at which such statements and their supporting documents are kept, examines the statements and attests to their accuracy and fairness.

The auditor must possess some rather special characteristics. Clearly he or she must be sufficiently sophisticated to act in such a role. Probably sophistication is most needed in the methodological arena, but knowledge of the substantive arena should not be minimized. The auditor must be someone who has sufficient experience to be trustworthy, whose judgments can be accepted as valid, and who is a disinterested party. At the same time, the auditor must be sufficiently close in peer status to the auditee that one does not dominate the other; the auditor can easily be overwhelmed by a more senior, widely published well-known auditee if he or she does not hold similar credentials and, conversely, the auditee may be overly responsive to criticisms and findings from someone who is clearly senior to him or her. The hope for an appropriate exchange and negotiation rests on roughly similar bases of power.

Finally, in the event that an auditor is involved early in the study, he or she must take great care not to be coopted. Early entry may imply a *formative* role, analogous to the role of formative evaluator. The

latter's task is to produce information that will help to refine or improve whatever is being evaluated, but if the formative evaluator's recommendations are accepted, he or she will, on the next data gathering round, be collecting data on something that is partly the product of his or her own interventions. Disinterestedness is thus immediately called into question. Evaluators have not produced a solution to this conflict, and there is little reason to suppose that auditors will fare any better. But the auditor must be aware of this possibility, and professional ethics demands that he or she assess the likelihood of cooptation before agreeing to produce a final attestation. If that likelihood is more than trivial, a second previously uninvolved auditor should be employed.

* * *

The techniques discussed in the preceding pages apply specifically to the establishment of credibility, transferability, dependability, and confirmability. One final technique should be mentioned that has broad-ranging application to all four areas and provides a base for a number of judgment calls the auditor must make, for example, extent to which the inquirer's biases influenced the outcomes. That technique is the reflexive journal, a kind of diary in which the investigator on a daily basis, or as needed, records a variety of information about *self* (hence the term "reflexive") and *method*. With respect to the self, the reflexive journal might be thought of as providing the same kind of data about the *human* instrument that is often provided about the paper-and-pencil or brass instruments used in conventional studies. With respect to method, the journal provides information about methodological decisions made and the reasons for making them—information also of great import to the auditor. While much thought remains to be given to the nature of such a journal, it would appear reasonable to suggest that it consist of separate parts that include the following: (1) the *daily schedule and logistics* of the study; (2) a *personal diary* that provides the opportunity for catharsis, for reflection upon what is happening in terms of one's own values and interests, and for speculation about growing insights; and (3) a *methodological log* in which methodological decisions and accompanying rationales are recorded. Entries should be made on a daily basis in the daily schedule and personal diary, and as needed in the methodological log. Useful suggestions for how to develop and manage such a journal are found in Lincoln (1981), Reinharz (1979), and Spradley (1979).